Index

Acemoglu, D. 68, 72
Africa
  labor markets 333
Africa, sub-Saharan 239
  growth of food production 460–62
  agglomeration economies 426–7, 526
  aggregate demand
    and economic growth 196–9, 201–2
    relevance in less-developed countries 196
    and underdevelopment 117–19
Aghion, P. 214
agricultural surplus 100
agriculture
  education and technological change in 348
  effect on growth 240–41, 497
  and the environment 408–9
  in India 20–21
  interaction with industry 239, 241–50
  technological change in 453–68
Ahlburg, D. 316–27
AIDS 77, 318
Alderman, J. 356–70
America, Central 73–6
America, Latin see Latin America
Amiti, M. 510
Anand, S. 9
Annales School 27
apprenticeship 378
Arestis, P. 290–300
Asia, growth of food production in 461–2
Asia, East 21–2, 177–8
  development path of 474–8
  labor markets in 335–6
Asia, South
  labor markets in 335–6
Atkinson, G. 406
Auty, R. 388–403
Ayres, C. 164–5, 171
Bacon, F. 19
Bagchi, A. K. 16–31
balance-of-payments, data 91
balanced growth 240, 243, 247–8
banks 290–98, 372
Baran, P. 27, 127–8, 151–3, 191
Bardhan, P. K. 68, 172
Barrett, C. 408–9
basic needs 8
Basu, S. 290–300
Bauer, P. T. 182
Baumol, W. J. 505, 512, 513
Becker, C. 516–31
Becker, G. 358–9
Behrman, J. 356–70
Berry, A. 328–40
Bhagwati, J. 499
big push 114, 120, 165–6, 470
birth rate 456, 464
Botswana 76–8
brain drain 524
Bukharin, N. 239, 240, 246
bureaucracy 76, 182–3
business groups 385
business process outsourcing (BPO) see outsourcing
capabilities 3
capacity utilization 218–20
capital accumulation 104–6
  see also investment
capital flows see international capital flows
capitalism 144–5, 147, 153
Cardoso, F. H. 126–7
Carnegie, A. 380
Cassen, R. 316–27
caste 381
centre-periphery model 136–7, 141, 252–3
Chang, H.-J. 171
Chenery, H. 183
China 16–22
  labor markets in 336
civil conflict 465
class struggle 23
classical development theory 111–22
classical economics 97–109, 207–8
classical growth theory 190–91
clusters
and growth 426–7
Coase, R., 182, 371
Colombia 73
common property, and the
environment 411–12
Commons, J. R. 164, 171
comparative advantage 178, 244
models of 51, 105–6, 473
computable general equilibrium (CGE)
models 86–8, 93, 172
contract enforcement 216
convergence 122
clubs 42–3
conditional 34
time series approaches to 36–7
unconditional 34
β– 34–5
σ– 35–6
corruption 201
Costa Rica 73–4
cost-benefit analysis 351, 363–4,
462–3
‘creative destruction’ 59–62, 224
credit 290–300
credit markets 376–7
culture 164, 184, 216
cumulative causation 167
Daly, H. 406
Dasgupta, P. 11, 405
data problems 88–92, 501–2
De Soto, H. 377, 489, 492
debt, corporate 296–7
deindustrialization 20, 55–6
‘demographic bonus’ 319–20
dependency theory 108, 125–34,
152–4, 167
devaluation
contractionary effects of 259–60
development
historical perspectives on 16–31
interdisciplinary approaches to
68–81
meaning of 3–11
measurement of 11–14
developmental state 394
division of labor 102–4, 145
international 137, 140, 142
social 100
disease 68
disguised unemployment see
unemployment, disguised
distribution, income 6, 12
and economic growth 198, 203–4
effect of education on 343–5
and health 361, 363
Dobb, M. 25
dual economy 241
dual economy model 50, 191, 518 see
also Lewis, W .A.
Durlauf, S. N. 32–47
Dutch disease 260, 262, 389–91
Dutt, A. 189–206, 504–6, 512
Dutta, D. 497–515
East Asia see Asia, East
ECLAC 126–7, 130–31, 136–42
econometric models 84–5
economies of scale
in services 507
see also increasing returns
education 8–9, 210–11, 341–55
effect on growth 343–4
effect of health on 367
and entrepreneurship 377–8
primary 344–5
quality of 349
return to 341–3
secondary 344–5
university 343
and urbanization 520
education policy 350–51
effective demand see aggregate demand
efficiency
dynamic 328
El Salvador 73–4
elasticity of factor substitution 418,
423
employment
in services 501–3
endogenous growth theory
see growth models, endogenous; new
growth theory
Engel’s law 54, 57
Engels, F. 144–6
England 18–19
entrepreneurs 202–3, 371–3
entrepreneurship 371–87
gender and 381–2
policy for promoting 385–6
socio-economic factors and 379–82
environment 404–16
and agriculture 409–8
and common property 411–12
and the Green Revolution 467
and growth 404–8
and population growth 321–4
and poverty 407–8
and services 500
urban 409–10
valuation of 410–11
environmental Kuznets curve 406–7
ethnic fractionalization 393–4
ethnicity 216
and entrepreneurship 380
Europe, Central and Eastern 165–6
Europe, Western 16–20
Evans, P. 68–81
Evenson, R. E. 453–69
evolutionary economics 220–21
exchange rate 258–61
policy 229–31
regime 230, 263
export base theory see staple theory
export promotion 421 see also outward orientation
and total factor productivity growth 426
exports 374–5
external disequilibrium 138
externalities 210, 359, 361
from education 346–9
and the environment 413
from health and nutrition 368
information 474
pecuniary 112
technological 112, 114
extreme bounds analysis 40
factor accumulation 116–17
contribution to growth 421
factor market
imperfections 419, 435–50
family, and entrepreneurship 379
farm size
and agricultural yield 443
Fay, M. 301–15
Feldstein-Horioka puzzle 282
fertility 316–17, 319, 345
fertilizers 463, 467
finance 290–300
financial institutions 290–91 see also banks
financial liberalization 21, 232, 290–91, 295
financial markets 295
financial repression 290
fiscal policy 199, 225–6
foreign direct investment (FDI) 374–6, 512
foreign direct investment in services 503–4
foreign exchange constraint 255–7 see also gap models
free trade 23
freedom 4, 13–14, 21–2
freedom index 13
functionings 3
Galton's fallacy 36
game theory 179
gap models, 87, 255–7
Gemell, N. 505
gender 8
and entrepreneurship 381–2
General Agreement of Trade in Services (GATS) 508–9, 511
General Agreement on Trade and Tariffs (GATT) 508
geography 68–9, 216
Gerschenkron, A. 26, 384
Ghosh, P. 435–52
Gibson, B. 83–94
global climate change 412
global production networks 374–6, 385
globalization 52, 282–4, 428, 500, 527
Goodwin, R. 24
governance 377
government, and services 512
government failure 471–2
government intervention, in financial sector 291–8
Grameen Bank 382
Green Revolution 408–9, 453–69
Green Revolution Modern Varieties (GRMV) 453–60
Gresham’s law 98
Griliches, Z. 562
Gross Domestic Product (GDP) 64–5
data on 90–91
per capita 11
growth 5–6, 189–206, 207–21
accounting 38–9, 421–4
collapses 390–91, 396–400
contribution of education to
343–4
contribution of services to 501–2, 504–6
in East Asia 177
effect of health and nutrition on
364–5
empirics of 32–43
and the environment 404–8
of food production 460–62
fundamental and proximate
determinants of 39–41, 204
stylized facts of 32–7
growth models
endogenous 34, 115–16, 120–21 see
also new growth theory
environment in 405
neoclassical 34, 112, 115–16, 121, 191–4 see also Solow, R.
Von Neumann model of 50
see also Keynesian growth theory,
neoclassical growth theory
Guatemala 73–4, 367
Gunder Frank, A. 108, 126, 128–9, 141, 152
Gutierrez, M. 269–89
Hagen, E. 379
Hamilton, A. 26
Hamilton, B. 378
happiness 4
Harris-Todaro model 519, 525
Harrod, R. 219
Harrod-Domar model 191, 212
Heal, G. 405
health 8, 201, 345–6, 356–70
health policy 361–4, 369
Hilferding, R. 148–9
Hirschman, A. O. 162–3, 166–8, 170, 173
Hirschmeier, J. 378
HIV 77, 318
Hoddinott, J. 356–70
Hodgson, G. 171–2
Howitt, P. 214
human capital 210–12, 341–55, 358–9, 497
human development 5–11, 14
Human Development Index 9, 11–14, 65
Hume, D. 100
ILO see International Labour
Organization
imperialism 20, 25, 27–8, 125, 132–3, 146, 148–51, 153
import substitution 182
as export promotion 254
industrialization 131, 140, 257, 417–18, 420, 471–2, 477
‘impossible trinity’ 227, 231, 261–2
income, national 6
increasing returns 112, 115–18, 195, 201, 238, 241, 245, 500
India 16–21, 376–7, 379, 381
India
financial sector in 292–9
industrial policy 470–78
rationale for 470–71
Industrial Revolution 17–19
industrialization 53, 117, 138–42, 257
in resource poor economies 394–6
rural 522
industry see industrialization
interaction with agriculture 241–50
inequality
and entrepreneurship 381
and the Green Revolution 467
and infrastructure 305–6
international 37
see also distribution
infant industry protection 18, 107
inflation
effects on growth 225
targeting 227–8, 263
informal sector 89, 330–31, 334, 483–94, 497, 501, 519, 524, 527
alternative conceptualizations of
484–91
policies towards 491–4
and state regulations 487–90
information and communications
technology (ICT) 375, 507–8, 513
information externalities 474
Index

information technology services (ITS) 509, 513
information technology see Information and communications technology
infrastructure 301–13
  barriers to progress 308–12
  and income inequality 305–6
  and poverty 305
  private investment in 310–12
  progress in delivery of 306–8
  regulation 303, 312–13
  and total factor productivity 306
Ingram, G. K. 301–15
innovation 213–14, 372
institutional approach 68–80
institutional change 80
institutional economics 162–75
institutional economics, new 169–70
institutions 68–80, 215–16, 377
intermediate products 213
intermediate technologies 485
international capital flows 254, 261, 282–4, 500
  controls on 231, 235, 262–4
  liberalization of 231, 262
  ‘sudden stops’ of 234
International Labour Organization (ILO) 484, 490
international migration 523
international production networks 427
international trade 54, 374–5
  liberalization of see trade liberalization
  in services 503, 508–10
  theory of 252 see also under specific theories
internet 507–8
investment 219–20, 270–72
  determinants of 276–80
  and growth 280–82
Jameson, K. P. 162–75
Japan 26, 374, 378–9, 384–5
Johnson, S. 72
Kaldor, N. 24, 219
Kalecki, M. 24, 25, 116–17, 196, 219, 270–71, 275
Kautsky, K. 149
Kenya 367
Keynes, J. M. 118–19, 196, 276
  Keynesian growth theory 196–9, 218–20
Knight, F. 371
knowledge 210, 213–15
  codified 500, 512
  tacit 500, 512
Korea, South 384–5
Korea, South, credit and financial sector in 291–3, 295–9
Kotwal, A. 435–52
Kremer, M. 367–8
Krugman, P. 245
Kurz, H. 207–22
Kuznets, S. 48, 50–51, 58–9, 64–5, 183
Kourtellos, A. 32–47
Lal, D. 182–3
labor
  productive and unproductive 103
  market 328–38
    imperfections 330–31, 333, 337
    relations 335
  land reform 181, 443, 445, 450
  late industrialization 26, 384, 385
Latin America
  development path of 130–31, 142, 474–8
  growth of food production in 461–2
  labor market in 333–5
  learning by doing 244, 246, 248
Lee, D. R. 409
Lee, K.-Y. 504–6, 512
Leibenstein, H. 371–2
leisure 6
Lenin, V. 130, 147–9, 151, 153
Lewis model 191
life expectancy 11, 357
limited liability 440–41, 447
linkages
  backward 137, 391–2
  diffuse 393
final demand 391, 393
fiscal 391–2
forward 137, 391–2
point source 393
Amitava Krishna Dutt and Jaime Ros - 9781848442818
Downloaded from Elgar Online at 07/19/2019 02:48:07AM via free access
Index

Philippines, The 178–9
Physiocracy 99–102
planning 180–82
policy, for entrepreneurial development 385–6
population 22, 216–17, 316–24
and environment 321–4
growth of 190, 200, 453
macroeconomic role of 318–20
and poverty 316, 320–21
poverty 366
and agricultural production 450
and the environment 407–8
and the informal sector 484
and infrastructure 305
and population growth 316, 320–21
poverty trap 114, 119, 199–201, 241
PREALC 486–7
Prebisch, R. 136, 252–3, 486
Prebisch-Singer thesis 23, 182
Preobrazhensky, E. 239, 240, 243
privatization 365
of banks 297
production function
econometric estimation of 424
productivity growth 244–50, 505–6 see also technological change
property rights 69–72, 201, 215–16, 377
Protestantism, and entrepreneurship 380
‘proto-industrialization’ 18
Psacharopoulos, G. 341–55
public goods 80, 302
public sector
and agriculture 458–9
and services 512
see also government, state
quality of life 5
Quesnay, F. 100–102 see also physiocracy
randomized experiments 360, 368–9
Rao, R.S. 376–7
Rashid, S. 176–86
Rattso, J. 238–51
religion, and entrepreneurship 380
remittances, migrant 521, 523
rent contract, fixed 438
rent seeking 362, 376–7, 474
research and development 210, 213
in agriculture 457–8
resource allocation
patterns of 56–7
and productivity growth 58–9
resource curse 57, 388 see also Dutch disease
Ricardo, D. 22–3, 105–6
risk, political 71
Robinson, J. 72–3, 197
Rodrik, D. 68
Romer, P. 195, 210, 212–14
Ros, J. 111–24
Rosenstein-Rodan, P. 23, 111–12, 114, 116, 165–6, 180
Rostow, W. 25–6, 181
rural industrialization 522
rural–urban migration 516–28
Sachs, J. 68–9
Salvadori, N. 207–22
savings 270–72
determinants of 272–6
Say’s law
in underdeveloped countries 118
Schultz, T. P. 341
Schumpeter, J. 202, 214, 220, 371–3, 386
Schumpeterian growth model 214, 244
sectoral interaction 238–51
Self employment, and
entrepreneurship 378
Sen, A. K. 3, 9, 168
services 497–515
definition of 498–9
types of 499
Shapiro, H. 470–82
sharecropping 438
simulation models 85–8
Smith, A. 22, 102–5, 162, 176–7, 207, 214, 341, 505
social accounting matrix (SAM) 87, 91–2, 172
social capital 396–7
social factors, entrepreneurship and 379–82
socialism
transition to 125–6, 128, 150
Solimano, A. 269–89
Solow growth model 191–4, 209–10, 217–18
Solow, R. M. 271, 389, 405
South Asia see Asia, South
Soviet Union 238–40
specialization, pattern of
and coordination failures 120
and external disequilibrium 138
and growth 473
stage theory of economic growth 25–6, 181
staples theory 391–4
state 75–6, 78–80, 165–6
as entrepreneur 384–5
role in agriculture 457–8, 65
Steindl, J. 219
Stiglitz, J. 405, 519
stock markets 290
Streeten, P. 3–15
structural adjustment programs 21
structural change 48–65, 505–6
stylized facts of 52–6
structural heterogeneity 137–42
structuralism 126, 136–42, 167, 252–3, 255
Subsistence sector 113, 115, 118
Sunkel, O. 167
surplus labor 58, 112–13, 118, 191, 196, 249, 329–30, 396, 444
sustainability, environmental 405–6
sustainable development 404–5
Sutcliffe, B. 144–61
Swidler, N. 77–8
Syrquin, M. 48–67, 183
Tan, C. M. 32–47
Taylor, L. 196, 204
Taylor rule 228–9
technical choice 418–19
technological change 116–17, 139, 199, 202–3, 210–15, 244–9, 258, 348, 373–6, 421–9
see also innovation, learning by doing, total factor productivity growth
in agriculture 453–68
in ICT 507–8, 510
and services 499–500, 505
technology 164
technology transfer 203, 374–6, 385
terms of trade, agriculture-industry
239, 243
terms of trade, international 23
deterioration of 137–42
shocks 233–4
Toman, M. 405
Tornqvist index 422–4
total factor productivity 38–9
growth 55, 58, 61, 280–81, 373, 417–18, 421–8
growth, in agriculture 455, 463, 465
and infrastructure 306
trade, international see international trade
trade liberalization 249, 477–8, 509
tragedy of the commons 449
transactions costs 445
transnational corporations 375, 476, 504
Trotsky, L. 147–8
unbalanced growth 240
underdevelopment 111, 127–9
unemployment 137–8, 202, 329–30, 336
data on 89–90
unemployment, disguised 166, 191
uneven development 23
unorganized sector see informal sector
urban bias 527
urbanization 516–28
and the environment 409–10
United States of America 380–81
van Wijnbergen, S. 245
Veblen, T. 164–5
Verdoon’s law 512
Viner, J. 181–2
volatility
and long run growth 224–5
wage differentials
by education 332
formal-informal sector 331, 334, 336
by gender 331
public-private 331
rural-urban 333
wage flexibility 332
war 27–8 see also civil conflict
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warren, B.</td>
<td>153–4</td>
</tr>
<tr>
<td>Washington Consensus</td>
<td>169, 176</td>
</tr>
<tr>
<td>water, access to safe</td>
<td>365</td>
</tr>
<tr>
<td>Weber, M.</td>
<td>182, 380</td>
</tr>
<tr>
<td>Wei, S.-J.,</td>
<td>514</td>
</tr>
<tr>
<td>Weitzman, M.</td>
<td>214–15</td>
</tr>
<tr>
<td>well-being</td>
<td>4</td>
</tr>
<tr>
<td>Williamson, O.</td>
<td>169, 171</td>
</tr>
<tr>
<td>Winter, S.</td>
<td>221</td>
</tr>
<tr>
<td>women, education of</td>
<td>345–6</td>
</tr>
<tr>
<td>world-system theory</td>
<td>152</td>
</tr>
<tr>
<td>World Trade Organization (WTO)</td>
<td>508–9</td>
</tr>
<tr>
<td>Wunsch-Vincent, S.</td>
<td>513</td>
</tr>
<tr>
<td>Yamamoto, K.</td>
<td>378–9</td>
</tr>
</tbody>
</table>
International Handbook of Development Economics
Volume Two

Edited by
Amitava Krishna Dutt
University of Notre Dame, USA

and

Jaime Ros
University of Notre Dame, USA

Edward Elgar
Cheltenham, UK • Northampton, MA, USA
Contents

PART VI INTERNATIONAL ISSUES

36 International trade and development 3
   
   Francisco Rodríguez

37 Terms of trade and economic development 16
   
   David Sapsford

38 Trade policy and development 30
   
   Henry J. Bruton

39 Foreign direct investment 48
   
   V.N. Balasubramanyam

40 Private capital flows and development 59
   
   Stephany Griffith-Jones

41 International capital flows to emerging economies: short- and long-run effects 71
   
   Graciela L. Kaminsky

42 Foreign aid 98
   
   Steven Radelet

43 International migration and the brain drain 119
   
   Francisco L. Rivera-Batiz

44 International technology transfer: the role of foreign direct investment 137
   
   Amy Jocelyn Glass and Kamal Saggi

45 International institutions and development 150
   
   Kunibert Raffer

46 North–South issues 163
   
   Amitava Krishna Dutt

PART VII DISTRIBUTION AND POVERTY

47 Measures of the distribution of income and their interpretation in developing countries 179
   
   Samuel A. Morley

48 Income distribution: effects on growth and development 193
   
   Nancy Birdsall

49 Measuring progress in reducing poverty 213
   
   Lyn Squire
50 Gender issues in development
   Diane Elson
51 Children and development
   Paul Glewwe and Amy Damon
52 Ethnicity and economic development
   William A. Darity Jr and Russell E. Triplett

PART VIII  THE STATE, INSTITUTIONS AND DEVELOPMENT

53 The role of the state and markets in development
   Louis Putterman
54 Monetary policy
   Paul R. Masson
55 Fiscal policy
   Raghabendra Jha
56 Stabilization policy and structural adjustment
   Valpy FitzGerald
57 Economic planning in developing economies
   Bill Gibson
58 State-owned enterprises and privatization
   Anindya Sen
59 Corruption and development
   Cheryl W. Gray
60 Law and development
   Pranab Bardhan
61 Institutions, property rights and development
   Jean-Philippe Platteau and Jean-Marie Baland
62 Culture and development
   Roland Hoksbergen and Charles K. Wilber
63 The economics of war: causes and consequences
   Frances Stewart and Graham Brown

PART IX  ECONOMIC DEVELOPMENT EXPERIENCE

64 Economic development in the Third World: an international perspective
   Richard Kozul-Wright and Paul Rayment
65 Latin America and the Caribbean
   Albert Fishlow
66 Sub-Saharan Africa
   Benno J. Ndulu and Stephen A. O’Connell
## Contents

<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
</tr>
</thead>
</table>
| 67   | The Middle East and North Africa
      | Omar S. Dahi and Firat Demir |
| 68   | China |
      | Yasheng Huang |
| 69   | South Asia |
      | Prabhat Patnaik and Jayati Ghosh |
| 70   | The East Asian newly industrializing countries |
      | Jomo K.S. |
| 71   | Transition economies: lessons for development |
      | Ruud Knaack and Henk Jager |
|      | Index |
PART VI

INTERNATIONAL ISSUES
In July of 2007 more than 1000 economists – including four Nobel prize winners – signed a letter to Congress petitioning it not to impose tariffs on China. The phrasing of the petition is revealing of the extent to which many economists believe in the beneficial effects of free trade:

As economists, we understand the vital and beneficial role that free trade plays in the world economy. Conversely, we believe that barriers to free trade destroy wealth and benefit no one in the long run . . . There is no foundation in economics that supports punitive tariffs.\(^1\)

This view of trade is widely shared by the overwhelming majority of economists with a neoclassical training. In a recent survey of graduate students at the six top economics departments in the USA, only 7 percent of respondents disagreed with the statement ‘tariffs and quotas reduce general economic welfare’ (Colander, 2005, p. 189) Of all the policy statements presented in the survey, trade policy was the one that commanded the broadest consensus among students. As Alan Blinder recently commented, in the economics profession ‘anyone who says anything even obliquely that sounds hostile to free trade is treated as an apostate’.\(^2\)

This consensus is not due to the absence of theoretical results showing the existence of situations in which greater trade can decrease welfare. Students of international trade are commonly exposed to examples of optimal tariffs, infant industries and strategic trade policy, all of which can serve to justify intervention in international trade. Rather, the consensus in the profession emerges from the vision that these examples are not relevant in most real-world circumstances and that the dangers from attempting to carry out activist trade policy far outweigh its potential benefits. As one of the architects of strategic trade theory has himself written:

The gains from intervention are limited by uncertainty about appropriate policies, by entry that dissipates the gains, and by the general equilibrium effects that insure that promoting one sector diverts resources from others . . . It is possible, then, both to believe that comparative advantage is an incomplete model of trade and to believe that free trade is nevertheless the right policy. In fact, this is the position taken by most of the new trade theorists themselves. (Krugman, 1987, p. 143)
Deciding whether to accept this conventional wisdom or to question it is a key step in the framing of a country’s development strategy. The purpose of this chapter is to analyze critically the basis for the pro-trade consensus, and to lay out clearly the arguments of supporters and detractors of activist trade policy. Such an exercise requires a dual discussion of the theoretical literature and the empirical evidence, which I present in the next two sections. As I will argue, a careful examination of these two literatures provides grounds for the raising of serious questions about the desirability of outward-oriented trade strategies for development.

What the theory says
In its simplest version, the gains-from-trade theorem (Kemp, 1962; Samuelson, 1962) states that in the absence of distortions and when lump-sum transfers are feasible, all individuals in an economy can be made better off from an appropriate combination of full liberalization and compensating transfers. This result follows from the basic Ricardian insight that it will be more efficient for countries to shift production from goods for which their opportunity cost of production is high to goods for which it is low.

The static intuition can be readily extended to an intertemporal model of growth, as shown by Smith (1979). In such models, the efficiency effects of trade will show up in higher steady-state levels of consumption and welfare. This extension can generate considerable confusion as openness will not be associated with higher long-run growth rates. This is more than anything a consequence of the fact that the long-run growth rate in a Ramsey-style model of economic growth is determined by the rate of technological progress and unaffected by any other variables. However, openness does generate higher steady-state levels of income and higher growth rates on the transition to these new steady states.

The intuition readily carries forward to endogenous growth models. In an AK model of growth, static efficiency losses generated by distortions turn up as declines in the level of efficiency captured by the productivity term in the production function. Trade policy thus has an unambiguously negative effect on the growth rate. In more complex endogenous growth models that seek to endogenize productivity as a consequence of decisions to innovate, the public goods nature of knowledge introduces a distortion. The existence of this distortion opens up the possibility of second-best effects in which trade policy can potentially enhance welfare, a possibility that we discuss in more detail below. A full discussion of alternative models of this type is given by Grossman and Helpman (1991).

In all of these renderings, the gains-from-trade theorem is little more than an extension of the first and second fundamental theorems of welfare economics. The possibility of trading shifts the consumption possibilities
frontier outwards, and the competitive equilibrium allows us to attain any point on that consumption possibilities frontier. The possibility of lump-sum transfers allows us to redistribute the gains from trade in such a way that all individuals can be made better off.

The theorem is thus open to the same caveats as the first and second welfare theorems. In the first place, it is vulnerable to the existence of distortions. One of the earliest recognized distortions in the trade literature is the possibility of the home country being able to affect international prices. In this case the world economy becomes a monopolist in the world market and perfectly competitive levels of production are no longer optimal. In the second place, the result requires the existence of non-distortionary taxes and subsidies that allow the redistribution of the gains from trade. In the absence of these instruments, some individuals may – and generally will – be made worse off by greater openness.

Obviously, in the real world distortions do exist and lump-sum taxes do not. Most of the academic literature’s defense of the gains-from-trade principle as a useful guide to policy action comes from the interpretation of results that appear to indicate that it will generally be suboptimal to deal with distortions through trade policy, and that reasonable approximations to lump-sum transfers exist.

Let us first discuss the issue of distortions. In a set of classic articles, Bhagwati (1971) and Bhagwati and Srinivasan (1969) showed that in the presence of distortions, trade tariffs or taxes will generally not be the optimal way to address the distortions or objectives that are commonly taken as rationales for activist trade policy. The exception is the case in which the distortion is actually generated by the existence of monopoly power in trade, in which the first-best policy is a tariff. But in the case in which there is an externality that leads to underproduction of a given good, the optimal policy is to subsidize the production of that good. A tariff is suboptimal because it is actually a combination of a production subsidy and a consumption tax, and there is generally no reason to expect that one would simultaneously want to increase production and reduce consumption of any particular good. These seminal contributions are thus generally taken as the demonstration that trade policy is a misguided way to address most of the problems generated by the existence of distortions.

The Bhagwati–Srinivasan contributions reflect a vision of policy-making in which the first-best equilibrium can be attained and thus serves as a useful policy benchmark. The ideal policy will be the one that accurately identifies all existing distortions and introduces an optimal intervention to eliminate the effects of each distortion. The existence of a government with the capabilities to carry out this complex exercise is presumed. One way of understanding this presumption is as a reflection of the view that the
majority of distortions are policy-induced. In this case, the policy prescription is clear: a laissez-faire elimination of all government-induced distortions could lead the economy sufficiently close to its first-best optimum to take advantage of the full gains from trade.3

While such a vision of the world is certainly coherent, it is also quite reasonable to hold to an alternative vision of the world as completely ridden with distortions. In this world, interacting complex processes such as innovation, knowledge networks, geographical clusters, public goods and international market power combine to create an economy in which distortions are a fact of everyday life. The task of identifying all such distortions and crafting interventions to address each one of them is beyond the realm of rationally effective policy-making. Even if one could identify all existing distortions, the design of interventions to eliminate some of them may be out of the sphere of possible policy actions by the government because of institutional or political constraints.

In such a world, policy-making should not try to replicate the first-best equilibrium. Rather, it makes sense to think of policy reforms as taking place in a setting of radical Knightian uncertainty, where the expected effects of removing a policy distortion can only be deduced from local experimentation. Hausmann et al. (2004) have recently proposed such an approach to policy reforms. Rather than attempting to eliminate all distortions at once, they suggest that reformers should concentrate on the reforms that have the greatest expected pay-off, given that other distortions are in place. They suggest an empirical method to infer whether certain distortions are in effect binding constraints on growth, and thus to identify whether altering them will lead to greater growth. The resulting method of policy-making is akin to the use of non-linear programming algorithms to search for local maxima, which do so through the search for incremental improvements rather than by the explicit calculation of a global solution.

One of the consequences of taking seriously such a world is that it turns a common free trade argument on its head. Advocates of trade often argue that even though some level of intervention in trade may be optimal theoretically to address trade-induced externalities, identifying such interventions in a context of considerable uncertainty would be beyond the capacities of most governments (Krugman, 1987). However, if we view the real world as ridden by trade and non-trade induced distortions, the same dose of realism that leads us to conclude that the government cannot address all of these distortions also leads us to recognize that we are unavoidably in a second-best world, in which the incremental effects of trade policy on welfare could well be positive.

Let us now turn to the issue of lump-sum transfers. Non-distortionary taxes require conditioning on characteristics that individuals cannot
change. It is doubtful whether such taxes actually exist or not in real life, and even if they do it would be extremely difficult to design them to redistribute the gains from trade – as the observable characteristics that are out of the control of individuals are unlikely to correlate perfectly or even reasonably well with the benefit or loss from trade that these individuals experience.

The existence of compensation to losers is particularly relevant precisely because trade theory predicts stark effects on income distribution from trade openness. In the absence of compensation, the Stolper–Samuelson theorem predicts that the real return to a country’s scarce factor will decrease with greater openness. In this case developing-country unskilled workers would actually benefit from greater openness, so that trade may be distributively beneficial for poor countries with an abundance of unskilled labor. However, the factor endowments model of trade appears to have little empirical support (see Feenstra, 2004, Chapter 2), so that this may not be the most appropriate theory with which to think about the distributive effects of trade. Alternative theoretical frameworks can produce different predictions concerning income distribution. For example, a set of recent theoretical and empirical contributions (Rodrik, 1997; Reddy and Dube, 2000; Ortega and Rodríguez, 2006) have argued that trade can diminish the bargaining power of unions and thus lead to a decline in labor shares.

In the case of compensation, the literature’s optimism comes from the belief that while lump-sum taxes do not exist, reasonable approximations can be constructed in practice to carry out the necessary compensations to losers. Indeed, it has been shown that lump-sum taxes are not even theoretically necessary in this respect: factor and income taxes will suffice to enact the desired redistribution (Dixit and Norman, 1980, pp. 79–80). This result appears intuitive even if the real-world setting is much more complex than that of our models: it is generally possible to identify – at least ex post – the key groups that gain and lose from trade openness and to design transfer programs to redistribute the gains.

The key question regarding compensation is not whether it is feasible in a technical and operational sense to design and implement such compensation, as it almost certainly is, but whether this compensation is likely to take place in practice. There are a number of political economy reasons why one may expect that such compensation is unlikely to take place. One is that while it may be optimal for the gainers to promise to compensate the losers before the reforms are carried out, such promises are likely to be time-inconsistent, particularly if their gains are protected by some degree of irreversibility in trade reforms. While the manipulation of activist trade policies by interest groups has received considerable attention in the literature (for example, Grossman and Helpman, 1994) and has often been adduced as a reason for
the superiority of simple rules such as free trade, the political economy of compensation arrangements has received much less attention. An early exception can be found in the work of political scientists such as Ronald Rogowski (1987), who argued that increased trade alters the political landscape, making owners of abundant resources much more powerful and assertive, and thus much less likely to accept demands for compensation.4

In sum, whether one considers the free trade case a reasonable one or not on theoretical grounds depends on whether one considers two basic assumptions of the gains from trade theorem reasonable. The first one is that the extent of distortions is sufficiently restricted so that the government can identify them and deal with them through policies designed on the basis of the theory of optimal intervention. The second one is that redistributive policies can and will be implemented to compensate the losers from trade liberalization, particularly when these are the most disadvantaged groups in society.

The decision of whether these two assumptions are reasonable or not on purely theoretical grounds is far from clear-cut. What should be clear is that a critical vision, which is based on skepticism about the appropriateness of the first-best model as a guide for policy and the likelihood of implementation of appropriate redistributive strategies, cannot be deemed insensible on a purely theoretical basis. The belief in the optimality of free trade should thus be based on the belief that the empirical evidence decisively points in favor of a beneficial effect of trade on growth. We turn now to that issue.

What the data say
Broadly speaking, the empirical literature that has studied the effect of openness on growth has taken one of two vantage points. The first one is to analyze the correlation between openness and growth in data sets that cover a large section of developing and developed countries, in the tradition of cross-country growth empirics initiated by Robert Barro (1991). The second one is to concentrate on country or region-level analytical case studies of economic growth. Both literatures have been appealed to by proponents and detractors of trade-oriented development strategies. In what follows, I will attempt to shed some light on the reasons behind these differences in interpretation.

The empirical literature on openness and growth is voluminous indeed. I will not attempt to provide a full survey of the main contributions (the interested reader may consult Rodriguez, 2008). Broadly speaking, however, a number of findings appear to emerge from this literature.

First, there is no strong unconditional or conditional correlation between economic growth and a number of direct measures of trade policy,
such as weighted or unweighted tariffs, import quotas or other non-tariff barriers. This point was first made by Rodríguez and Rodrik (2001) and generated some surprise in the literature. It has since been confirmed by, among others, DeJong and Ripoll (2006), who argue that there may be a non-linear relationship where the effect of tariffs on growth depends on the initial level of a country’s income and may be positive or negative. In Rodriguez (2007), I have also shown evidence in favor of a non-linear effect, although I have argued that the precise form of that effect may be difficult to discern.

Second, there appears to be a reasonably strong correlation between growth or productivity and the ratio of trade to gross domestic product (GDP), especially when the latter is measured in prices of a constant base year (Dollar and Kraay, 2002; Alcalá and Ciccone, 2003). Some attempts have been made to discern whether this correlation actually embodies a causal relationship. The most well-known attempt, formulated by Frankel and Romer (1999), consists in using instrumental variables estimates of the effect of trade volumes on growth where the latter is instrumented with its geographic determinants as derived from the estimation of gravity equations. These results are controversial – as Rodríguez and Rodrik (2001) and Irwin and Tervio (2000) have shown, they are not robust to controlling for the direct effect of geographical variables on income or productivity. Other attempts to discern causality using alternative methods to instrumental variables do not confirm the existence of causal effect (Rigobón and Rodrik, 2005).

A drawback of using the trade to GDP ratio as an indicator of openness is that it may capture many non-policy-induced changes in trade openness which are largely irrelevant if one is preoccupied with designing a developing country’s trade strategy. Natural resource booms, the emergence of new export sectors, changes in other countries’ trade policies, and changes in foreign aid can all have an effect on the trade to GDP ratio without necessarily having an obvious link to trade policy. In sum, the key problem of the trade–GDP ratio is that it is an indicator of results and not of policy actions. To take just one example, if the infant-industry argument for protection were correct, initial levels of trade protection would lead to the development of productive, competitive domestic industries that would later on be capable of competing internationally. Tariffs would be associated with higher growth, but so would exports. A correlation between trade volumes and growth may thus not be very informative about the desirability of activist trade policies.

Some authors have tried to produce compound measures of trade policy that capture the different ways in which an economy can be closed to international trade. According to these authors, one would not expect to observe
a simple correlation between simple measures of trade policy such as tariffs and economic growth because countries can use many policy devices to impose trade protection, of which only one is import tariffs. The most famous of these measures was provided by Sachs and Warner (1995) and recently updated by Wacziarg and Welch (2003). What these indices actually measure is very controversial. Rodríguez and Rodrik (2001) argue that the Sachs and Warner variable’s effect on growth was purely driven by two subcomponents of the index – black market premia and export marketing boards – which are not obviously linked to trade policy. For example, they argue that the effect of export marketing boards on growth in the Sachs–Warner study comes from the fact that the variable was taken from a 1994 World Bank study called Adjustment in Africa that covered only 29 African economies undergoing adjustment programs during the 1980s, leading to the exclusion of non-African or African non-adjusting economies from the sample and strongly biasing the results in favor of a trade–growth correlation. Rodríguez (2008) levies similar criticisms at the Wacziarg and Welch (2003) exercise.

In recent years, there has been growing skepticism of the possibility of establishing strong conclusions regarding causal growth effects using the cross-country regression framework. A growing consensus appears to have emerged around the belief that the problems of causality, robustness and specification are simply too pervasive and difficult to solve in the context of highly aggregated cross-national empirical data. This skepticism has led authors such as Bhagwati and Srinivasan (2001) to discount the aggregate growth evidence altogether, and to call for concentrating exclusively on the evidence from case studies. While these criticisms should be taken seriously, it is important to note that even if one takes the cross-country evidence at face value, accepting the framework without questioning, it does not appear to lend the strongest of supports to the pro-trade view. As in the case of the theoretical literature, it appears to be open to multiple interpretations, some of which are consistent with the view that protection is not unequivocally harmful for growth.

Country-level studies of openness and growth are also open to multiple interpretations. Bhagwati and Srinivasan (2001) cite the Organisation for Economic Co-operation and Development (OECD) and National Bureau of Economic Research (NBER) studies of more than a dozen major developing countries carried out in the 1960s and 1970s, which uncovered key differences between the constraints on economic performance in countries that pursued import substitution strategies and those that pursued export promotion. A revised interpretation of this view was given by the World Bank’s 1993 study The East Asian Miracle. Broadly speaking, the key argument of this study was that the openness to trade and reliance on market
forces of East Asian economies played a fundamental role in making possible their sustained growth acceleration.

The World Bank’s characterization of the high-growing East Asian tigers as economies that followed a strategy of free trade has, however, been strongly questioned by several authors. Some of these criticisms were collected in a 1994 volume published by the Overseas Development Council (Fishlow et al., 1994) in which Dani Rodrik, Robert Wade and Stephen Haggard disputed the key findings of the World Bank study. In Robert Wade’s words, ‘the [World Bank’s] report uses standards of inference so elastic that practically anything could be confirmed’ (2003, p. xix)

One of the key points of dispute concerns whether East Asia can adequately be characterized as a region that followed a non-activist trade policy. The World Bank study had concluded that East Asia’s relative prices were closer to international averages than those of other regions, supporting the contention that its international trade was relatively undistorted. Wade pointed out that this is only true when one uses an unweighted average that includes the island economies of Hong Kong and Singapore, where price distortions were necessarily negligible. In contrast, during the 1976–85 period, relative prices in Japan, South Korea and Taiwan deviated more from international prices than those of countries which are generally perceived to have had strong records of intervention, such as India, Pakistan, Brazil, Mexico and Venezuela in the period 1976–85. Similarly, Alice Amsden’s (1992) in-depth study of South Korea’s industrialization contends that the success of its industrial policies was largely due to an active intervention in the determination of relative prices, a strategy that she labels ‘getting relative prices wrong’.

During the 1990s, the set of liberalization experiences that could be the subject of in-depth studies expanded dramatically. Between 1990 and 2002, the average tariff rate in the world went down from 10.5 percent to 6 percent, and the ratio of imports plus exports to GDP rose from 75.2 percent to 86.8 percent (World Bank, 2005a). In 1990, the General Agreement on Tariffs and Trade (GATT) had been signed by 96 countries: between 1990 and 2005, 65 countries joined it either as the GATT or in its most recent incarnation as the WTO.

While the result of these liberalization experiences has not yet been fully analyzed, what is clear is that aggressive trade liberalization proved to be very far from a necessary condition for a growth take-off. Some of the most aggressive liberalizers of this period were former communist economies such as Mongolia, Ukraine and Moldova, which suffered some of the deepest growth collapses in post-World War II history. But openness did not only fail to pay off in the former Soviet Bloc. With the exception of Cuba, the evidence suggests that virtually all Latin American
economies moved in a direction of greater trade liberalization during the 1990s. Yet the region’s growth performance during the post-reform period has been disappointing to say the least, with per-worker GDP and total factor productivity growing respectively at annual rates of only 0.1 percent and 0.2 percent between 1990 and 2002 (Ocampo, 2004). The region is said to have entered an era of ‘reform fatigue’ (Lora et al., 2004) in which voters are increasingly willing to vote for political platforms to roll back reforms.

In sum, neither cross-national empirical studies nor country-level case studies seem to give strong support to the idea that openness is unequivocally good for growth. A reading of the evidence in support of activist trade strategies is certainly possible and indeed has been carried out by reputable mainstream economists. These conclusions mirror our interpretation of the theoretical literature, which can also be interpreted as supporting a case for intervention in trade policy.

Concluding comments
One way to explain the apparent divorce between the favorable view that the majority of economists have about free trade and the lessons given by the empirical and theoretical literature is by thinking about free trade as one of the components of our discipline’s ‘hard core’ (in the sense of Lakatos, 1976), a set of beliefs and methodological assumptions that are not considered the appropriate subject of empirical tests. Since these core beliefs are never tested without auxiliary assumptions, any failure to explain the evidence can be handled by altering the assumptions but not the core belief. As a senior faculty member once quipped after seeing a presentation of my work, ‘if the data does not say that trade is good for growth, then the data must be wrong’.

It is not easy for a discipline to abandon or even begin to question a hard-core belief, but neither is it impossible. To take one example, during the 1990s the assumption of rationality has made the transition from a hard-core belief to an auxiliary hypothesis that is not even taken very seriously most of the time. This change has opened up a burgeoning new area of research in behavioral economics which has transformed our understanding of individual economic behavior.

Signs that this may be starting to happen in the study of the relationship between trade and development are beginning to appear. In 2005, the World Bank published a comprehensive assessment of the experience of the 1990s with economic reforms (World Bank, 2005b). The sobering assessment of this disappointing period recognizes that the results of economic reforms were far below what its proponents had expected and rejects the one-size-fits-all approach to reform that the institution espoused during
the greater part of the period in question. On the concrete matter of trade policy, the report concludes that: ‘[w]hile trade reforms can help accelerate integration in the world economy and strengthen an effective growth strategy, they cannot ensure its success’, and ‘the distributive effects of trade liberalization are diverse, and not always pro-poor’ (pp. 131–2). On the fairness of the world trading system, it states that ‘global markets are the most hostile to the products produced by the world’s poor’. As Dani Rodrik wrote in his review of this volume: ‘occasionally, the reader has to remind himself that the book he is holding in his hands is not some radical manifesto, but a report prepared by the seat of orthodoxy in the universe of development policy’ (Rodrik, 2006, pp. 974–5).

A reconsideration of the role of openness in countries’ development strategies would fundamentally alter the nature of the debate on generating and sustaining growth. Whether this occurs will probably depend not only on the internal dynamics of academia, but also on the extent to which outside reality exerts pressure for such a change. Political discontent with the experience of the 1990s is undoubtedly a key reason for the World Bank’s reappraisal of the reform experience. In the same way, the results of the current reassertion of state involvement in much of the developing world are likely to influence deeply the direction that development research will take in the future. Perhaps, to turn Keynes on his head, economists are nothing more than the slaves of long-defunct practical men.

Notes
1. At the moment of writing (August 2007), the process of signature collection is not yet finished and thus the letter has not been published. The text of the letter was obtained through personal communication with Andy Roth of the Club for Growth.
3. Obviously, even if only a few distortions remain there is no theoretical presumption that the resulting equilibrium will be better than the pre-laissez-faire equilibrium. It can be argued that at this point it becomes feasible to target the remaining distortions through optimal interventions. Alternatively, it can be argued that theoretical models are always approximations of the real world, and that a distortion-free model should be a reasonable approximation to a reality in which there is a reduced number of distortions.
4. For a recent exploration of this issue, see Davidson et al. (2006).

References
Amsden, Alice (1992), Asia’s next Giant: South Korea and Late Industrialization, Oxford: Oxford University Press.


Introduction
Why do countries trade with one another? What determines the terms on which trade between countries is conducted in the world marketplace? These two questions are perhaps the most fundamental to be considered in any discussion of international economic relations, be it trade between developed and developing countries or trade amongst countries in either the developing or the developed world. These questions are of especial importance in the context of economic development since if there are ‘gains from trade’ to be earned, the distribution of such gains between trading partners (especially when we are thinking of trade between developed or industrialized countries on the one hand, and developing nations on the other) carries important implications not only for living standards and economic welfare within the participating countries, but also for the continued willingness of developing and newly emerging economies to engage in processes, such as those initiated under the auspices of the World Trade Organization, designed to bring about further reductions in barriers which impede the process of international trade.

The classical economists, most notably Adam Smith (1723–90) and David Ricardo (1772–1823) (see in particular, Smith, 1776 and Ricardo, 1817), initially addressed the first of these questions and their, respective, analyses of absolute and comparative advantage as the basis for trade are widely recognized. Indeed, to this day the Ricardian model, despite its various (over)simplifying assumptions (including that of a two-country, two-good, one-factor world) still occupies centre stage in the economic theory of international trade. Despite its undoubted logical elegance, a major limitation of the Ricardian analysis is encountered when we move to the second of these two questions, because the model leaves us analytically with a range of indeterminacy, somewhere within which the terms of trade (defined in the usual way as the ratio of the price of a country’s exports to the price of its imports) must lie if trade is to offer benefits (in terms of increased output or consumption with unchanged resource endowments) to at least one of the trading partners. This indeterminacy is, in practical terms, more than a mere theoretical loose end, especially when we move
outside of the static world of Ricardo’s model and begin to seek answers to important real-world questions such as who gains most from trade and to what, if any, extent the pattern of winners and losers has changed over time with the evolution that has occurred in both the structure and performance of the world economy. Such questions are of particular importance in the context of the changes that have occurred over the long run in the terms on which trade between developed and developing nations has been conducted, especially since this carries important implications for the pace and nature of economic development in the Third World.

Who gains from trade?
While the elegance of Ricardo’s analysis and its correctness within the confines of its own assumptions can not be faulted it does, as noted, beg a question that is of central importance in the context of the trade that takes place between countries of the developed or industrialized world, on the one hand, and those of the developing or Third World on the other. While the analysis demonstrates quite clearly the potential benefits to trading partners from engaging in international trade in the world marketplace, it has nothing whatsoever to say about the division of these potential gains between them. Suppose that the two countries comprising the Ricardian world are, say, the USA and China, then it follows from the model that if relative prices in the world marketplace (the so called net barter terms of trade) were equal to US internal relative prices then China would effectively appropriate all of the gains from trade for herself, whereas at the opposite end of the spectrum, the USA would scoop all of the gains if Chinese relative prices prevailed.

In order to focus ideas let us consider trade between the countries of the developed or industrialized world and those of the developing world and, for simplicity, assume that the former produce manufactured goods while the latter produce primary commodities. The fact that Ricardo’s analysis did not shed any light on the issue of how the potential gains from trade are shared out in practice did not seem to constitute a problem in the minds of classical economists since in a related context Ricardo, like Smith before him, had argued that as an inevitable consequence of the twin forces of diminishing returns in the production of primary commodities from a fixed stock of land (including mineral resources) as population increased, and the downward pressures on production costs in manufacturing generated by the moderating influences of surplus population and urbanization upon wages, the price of primary products would rise over the long run in relation to the price of manufactured goods, thereby giving rise to an upward drift in the net barter terms of trade between primary commodities and manufactured goods. On the above assumptions this movement
will translate into an improvement in the terms of trade of developing countries vis-à-vis the developed countries. On the basis of this argument there was little, if any, reason to be concerned about the plight of developing countries in the context of their trading relations with the industrialized world since it predicted that over the long run the terms of trade would shift, as a consequence of the workings of the invisible hand of the market, steadily in their favour, with the result that they would in due course enjoy an increasing share of the potential gains from trade. Indeed, casual analysis of the views expressed by many politicians and international policy-makers over at least the last century would seem to suggest an underlying acceptance of this particular prediction of classical economics (Sapsford and Singer, 1998).

The evidence
In the early 1950s, the classical prediction of a secular improvement in the terms of trade of primary commodity-dependent developing countries was challenged by both Prebisch (1950 [1962]) and Singer (1950).3 Both argued forcefully that in direct contravention of the then still prevailing classical orthodoxy, the terms of trade faced by primary producers had actually as a matter of statistical fact been historically subject to, and could be expected to continue to be subject to, a declining as distinct from an improving trend. Both analyses therefore implied that contrary to the classical view, developing countries were actually obtaining a falling proportion of the potential gains from their trade with the countries of the developed world (Sapsford et al., 1992). The statistical techniques available for revealing and estimating trends in the middle of the twentieth century were by today’s standards extremely rudimentary. However, it is relevant to notice that some 50-plus years later the conclusion of a downward secular movement remains strongly intact, despite a rapid evolution in the sophistication of the time-series methods to which terms-of-trade data have been subjected.4 Figure 37.1 provides a time plot covering the period 1900 to 1999 of the log of most commonly analysed terms-of-trade series, together with the estimated least-squares trend line which provides clear evidence of the presence of a steady downward trend over the last century.5

From Singer I to Singer II
When the Prebisch–Singer thesis was launched in 1950 the world was, perhaps, a simpler place economically than it is today. In particular, the fact that the industrialized countries of the North specialized heavily in the production and subsequent export of manufactured goods, while those in the South concentrated intensively in the production or extraction and export of primary commodities, meant that the Prebisch–Singer focus upon the
terms of trade between manufactured and primary goods carried direct implications for the terms of trade of less-developed countries (LDCs) \textit{vis-à-vis} developed countries. As the post-1950 decades unfolded, and many of the LDCs of the South moved increasingly down a path of industrialization, Singer (1975) argued that terms-of-trade issues continued to be of vital importance to LDCs despite their greater reliance on manufactured exports. In this context he distinguished between what he called ‘Singer I’ and ‘Singer II’; with Singer I referring to terms of trade between different commodities (manufactures versus primaries) and Singer II referring to terms of trade between different countries (less-developed economies, the ‘South’, versus developed economies, the ‘North’).

As we shall see in the following section there has been a range of alternative explanations put forward in the literature for the observed secular deterioration in the terms of trade. However, we will also see that these explanations relate at least as much to the characteristics and structures of the countries themselves as to the commodities they trade. The distinction between Singer I and Singer II is an important one because it alerts us to the possibility that potentially serious terms-of-trade problems for LDCs can continue to persist despite their increasing tendency to switch towards manufacturing exports. Various pieces of empirical evidence support this concern. For example, econometric evidence reported by Sarkar and Singer (1991) suggested that for a large sample of less-developed countries over the period 1979 to 1987 the terms of trade between the manufactures exported by them and those imported by them from the developed economies was subject to a declining trend. A similar finding has also been
reported by Lutz (1999). Some new evidence relating to the experiences of a sample of the world’s very poorest countries is reported later in this chapter. In an overview of the performance of the Prebisch–Singer thesis over its first half-century, Singer (1987) presented some statistical evidence to suggest that the terms-of-trade deterioration experienced by LDCs since the mid-1950s can be attributed, in approximately equal proportions, to three causes: first, the deterioration in the price of primary commodities in relation to manufactures (as emphasized in Singer I); second, a more rapid deterioration in the prices of the primary commodities produced by developing countries than in the prices of those produced by developed countries; and finally, a fall in the price of the manufactures exported by developing countries relative to those exported to them by developed countries (as emphasized by Singer II).

**Explaining the downward trend**

A number of theoretical explanations have been put forward in the literature to account for the observed downward trend in the terms of trade of developing relative to developed countries, and these can be conveniently summarized under the following four headings:

- Differing price elasticities of demand for primary commodities and manufactured goods (with the inelastic nature of the former resulting in a tendency for increases in the conditions of commodity supply to be felt more strongly in price decreases than in quantity increases).
- Differing rates of growth in the demands for primary commodities and manufactured goods (with the demand for primaries expanding less rapidly than the demand for manufactures due to their lower income elasticity of demand – especially so in the case of agricultural commodities due to the operation of Engel’s Law – plus the development of synthetic substitutes and the occurrence in manufacturing of technical progress of the raw material saving sort).
- Technological superiority (the argument being that the price of manufactured goods rise relative to those of primaries because they embody both a so-called Schumpeterian rent element for innovation, plus an element of monopolistic profit arising from the monopoly power of multinational producers).
- Asymmetries in market structure (the argument here is that differences in market structure – with primary commodities typically being produced and sold under competitive conditions, while manufacturing in industrialized countries is often characterized by a high degree of monopoly by organized labour and monopoly producers – mean that while technical progress in the production of primary
commodities results in lower prices, technical progress in manufacturing leads to increased factor incomes as opposed to lower prices).

In his 1950 paper Singer placed particular emphasis on points listed under the first two bullet points, while Prebisch (1950 [1962]) specially emphasized those covered under the third and final bullets. A number of recent authors have developed what have become known as North–South models, which further develop the points grouped under bullets three and four. See, for example, Sarkar (1994) and Dutt (1998).

**Some policy implications**

Although space constraints do not allow us to discuss in any detail the policy implications of the observed worsening trend in the terms on which trade is conducted in the world marketplace between primary commodity-producing countries and manufacturing countries, it is nevertheless important to note that the Prebisch–Singer hypothesis is sometimes advanced as one argument in favour of development and trade policies of the import-substituting industrialization (IS) as opposed to export promotion (EP) variety. A number of early enthusiasts of the Prebisch–Singer thesis recommended the adoption of such a policy stance. However, the policy issues here are not clear-cut and the fact, already mentioned, that all four of the above explanations relate as much, if not more, to the characteristics of different types of countries as to the characteristics of different types of traded goods highlights the need to devise and implement policies that address differences and imbalances of the former as well as the latter sort (Singer, 1987). Although a number of recent analysts (for example Krueger, 1997) have drawn an association between below-average growth performance and the adoption of IS as distinct from EP policies, the real-world situation, as Singer frequently pointed out, was rather more complex when one looks beneath the surface, in that countries such as Korea (held up in some circles – including Washington, DC – as a beacon to demonstrate the superiority of outwardly oriented trade policies) appear in reality to have achieved their above-average rate of growth by adopting a subtle dynamically evolving policy mix involving a combination of IS policies in certain sectors of the economy simultaneously accompanied by EP policies in other sectors. As Singer was also quick to point out (for example Sapsford and Singer, 1998), Ricardian comparative advantage is seldom, if indeed ever, exogenously given in the manner assumed in the simple Ricardian model, still so central in trade analysis. Thus while it is logically correct that in the static Ricardian world primary commodity-exporting countries can still potentially gain from trade by specializing in those sectors in which they possess (static) comparative advantage, what is relevant in reality is the...
fact that many comparative advantages are actually either consciously engineered or acquired through learning by doing and increasing returns to scale. Accordingly, it has been argued that the Prebisch–Singer thesis provides one possible case for industrialization based on (a limited period of) infant industry protection. In development economics, as in prospecting, all that glitters is not gold.

Although outside of the scope of the current chapter, it is important to recognize that the terms-of-trade issue carries a range of important implications for issues other than the distribution of the gains from trade. These include questions relating to the level, and pattern, of Southern or LDC growth and the unevenness (or otherwise) of international development patterns. Some brief comments are, however, in order. As far as the implications of terms of trade for Southern growth are concerned, there now seems to be something of a consensus in the so-called ‘New’ (or Endogenous) Growth Theory literature that the terms of trade do matter when it comes to explaining inter-country differences in growth performance. In particular, in a study of the determinants of growth across countries Barro (1997) estimates the (partial) elasticity of real output with respect to the country-specific net barter terms of trade to be 0.137, with a standard error of 0.3 – implying an effect which is significantly different from zero. The long-run trends in the terms of trade, and the forces that drive these, also carry some important implications for our understanding of the observed patterns of international development. The literature here is extensive and is well covered by Dutt (1990, 1998). For the purpose of the current discussion it is relevant to notice that different causes of terms-of-trade deterioration carry with them different implications regarding the evenness of development. In cases where the predominant factor driving terms-of-trade deterioration is the income elasticity of demand, we would expect to observe uneven development. However, in cases where its roots lie in the pace of technological progress in the South then, especially in situations where there are inflows of foreign direct investment into the South accompanied by the possibility of significant spillovers of various sorts to the domestic economy, there is the potential for deteriorating terms of trade to go hand in hand with more even development (Balasubramanyam et al., 1996). In a nutshell: the terms of trade matter, in terms of both the speed of LDC growth and its evenness.

Some recent issues: country-specific evidence and terms-of-trade volatility
Terms-of-trade issues continue to attract the attentions of researchers with interests in a variety of trade-related fields and in this section we consider two of these. The first relates to the strength and periodicity of terms-of-trade volatility, while the second relates, in the spirit of Singer II, to the experiences of particular economies.
Trend versus volatility in the terms of trade

Inspection of the time plot in Figure 37.1 suggests that while the terms of trade declined secularly over the course of the last century they were, at the same time, subject to appreciable variability about that trend. In a statistical analysis of the time-series properties of the Grilli–Yang data (as plotted in Figure 37.1) Sapsford and Balasubramanyam (1999) produced evidence that suggested that the extent of volatility exhibited by the terms of trade about its trend increased over the twentieth century. This increased volatility was particularly marked after 1972 and appeared to have been accompanied by an acceleration in the trend rate of decline of the terms of trade. Taken together these results yielded a depressing picture for the primary commodity-dependent countries of the developing world for they indicated that since 1973 such countries have experienced a marked increase in the trend rate of deterioration in their terms of trade (from about 1.6 per cent per annum to around 4 per cent), accompanied by a marked increase in their volatility estimated to be to the order of 50 per cent. This clearly represents a doubly unpleasant state of affairs.

In a related study, Singer and Lutz (1994) report cross-country panel data estimates which seem to indicate that the magnitude of both the downward trend in the terms of trade and the extent of volatility about this trend exerted significant downward effects on gross national product (GNP) growth in a sample of some 79 non-oil-producing countries. In this study the authors argue that terms-of-trade volatility exerts a detrimental influence on growth performance via a number of channels, including: the increased risks which it imposes on investment (which they see as reducing the incentives to invest and thereby the realized volume of investment); its potential to disrupt development (including structural adjustment) plans, to increase price instability, to destabilize domestic incomes and to distort the structure of domestic prices, including the price of traded relative to non-traded goods.

The arguments developed and empirical results reported in these two studies would seem to imply that the trend and volatility in the terms of trade should not be treated as separable issues in the manner of classical time-series analysis. Instead these two issues are best seen from the analytical standpoint as twin pillars of the same underlying problem faced by LDCs: their heavy dependence on primary commodities (or more recently upon components and parts of manufactured goods) as a source of export revenue. The empirical work reported in these papers focused attention on both the long-run and cross-country evidence regarding the trend and volatility in the terms of trade seen as related dimensions of the same underlying problem. Amongst other things the results reported by Sapsford and Balasubramanyam (1999), as noted above, indicated the
occurrence after 1972 of a major increase in the trend rate of deterioration of the terms of trade and, moreover, that this was accompanied by a marked increase in volatility. This is the true ‘double whammy’ in which both of the twin pillars of the commodity problem seem to have turned simultaneously against these countries. However, the evidence reported by Singer and Lutz (1994) suggests that the bad news does not end there, in the sense that both the trend rate of terms-of-trade decline and the extent of terms-of-trade volatility would appear to exert a subsequent significant downward influence upon the rate of economic growth; a ‘triple whammy’.

Country-specific evidence
A number of studies in the recent literature have moved away from the continued analysis of the aggregate data, preferring instead to focus attention on country-specific evidence, especially that relating to low-income primary commodity-exporting countries. It is a regrettable fact of current economic life that it is the very poorest economies in the world who are the most dependent on either a single or a small number of primary commodities for the vast majority of their export earnings. In an UNCTAD study, Sapsford (2001) noted that according to World Bank 1996 data, all but two of the world’s poorest 20 economies were located in Africa and of these, some 13 were dependent on either a single primary commodity or a small number of commodities for in excess of 90 per cent of their export earnings. Two examples are as follows: Mali (with an estimated annual per capita income of US$260 and life expectancy at birth of 50 years) earned 99.8 per cent of its export revenue in 1996 from cotton, while Ethiopia (the then poorest economy in the world, with an estimated per capita income of only US$110 per annum, and life expectancy of only 43 years) depended on coffee for 99.8 per cent of its export earnings. Clearly, for these seriously poor economies both trends and volatilities in the terms on which they undertake their trade in these particular primary commodities with the outside world are of vital importance. One might say that they are quite literally a matter of life and death.

Sapsford (2001) provided a detailed statistical analysis of the terms-of-trade experiences of this sample of the world’s poorest economies over the period 1960 to 1998. Using a simplified version of the structural econometric model proposed in Bloch and Sapsford (2000), this study undertook a structural stability analysis of country-specific terms of trade. Notice that unlike cruder previous approaches to trend estimation, this approach controlled for the influence of fluctuations in the level of production in the industrialized world on country-specific terms of trade. Although a detailed discussion of the nature and implication of the results revealed by
this study is beyond the scope of the present short chapter it is relevant to notice that the main results can be summarized as follows:

- Of the world’s 15 poorest commodity-exporting countries, all but two experienced a significant change in the trend rate of growth of their terms of trade during the period 1960–93.
- In nine of these cases, the change in the trend occurred between 1972 and 1982.
- Nineteen out of 28 reported trend estimates are negative; only three of the reported trend estimates are positive.
- In nine out of the 13 countries where there is a trend shift, the pattern shows a worsening of the situation in respect of terms of trade.
- In six out of the 13 countries where there is a trend shift, the pattern shows an increase in the volatility of the terms of trade.

Taken together these results seem to indicate that over the post-1960 period many of the very poorest commodity-exporting LDCs in the world have indeed been subject to Prebisch–Singer effects on their terms of trade, effects which have exerted a continuous downward pressure on economic and export growth of a magnitude sufficient to more than offset the upward effects which they might have experienced as a result of the positive influence of expanding output in the industrialized countries.

Some concluding remarks
The terms-of-trade is a topic that continues to attract the attentions of researchers in the fields of international trade and development economics: scarcely a week passes without yet another addition to the literature. Within the confines of this short chapter it has only been possible to scratch the surface of the many and varied issues involved. Particular emphasis was placed in the first half of the chapter upon the famous Prebisch–Singer hypothesis, the empirical evidence relating to the hypothesis and its policy implications, especially as these relate to countries situated at the lower end of the per capita income distribution. In the second half of the chapter we focused attention on some recent analyses that have extended the earlier approaches to encompass the effects, from a country-specific perspective, of terms-of-trade volatility upon economic growth performance and the relationship between terms-of-trade trend and volatility.

As we have seen, the Prebisch–Singer hypothesis that there is a long-term decline in the price of primary commodities relative to the price of manufactures has historically been an object of controversy, although given the overwhelming weight of empirical evidence in its favour, it seems to have become widely accepted in the majority of circles since the 1990s. Most
tests of the hypothesis use time-series models to estimate trend growth rates in selected relative prices. The focus of concern has typically been the net barter terms of trade between producers of primary products (equated with developing countries) and producers of manufactures (equated with industrialized countries). A new approach which has been developed recently is to construct a structural model which seeks to identify the various different factors which impinge on the prices of manufactured goods and primary commodities (Bloch and Sapsford, 1997, 2000). Applying this approach, it has been found that the overall trend identified in the time-series models is the net effect of a number of separate divergent influences. On the one hand, there are Prebisch and Singer effects that exert a downward pressure on the commodity terms of trade. These effects arise because of differences in market structure (markets for primary products are more perfectly competitive) and differences in the factor bias of technical change (technical change in manufactures is assumed to save raw material inputs and labour). On the other hand, rising output in the industrialized countries can have an offsetting effect, as primary products used in manufacturing activity experience rising prices when the level of manufacturing activity increases.

The final section of this chapter discussed some country-specific results that were obtained by applying this basic approach. These results seemed to suggest that many of the very poorest economies in the world – which are also, as it happens, heavily (if not totally) dependent on either a single primary commodity or a very small number of them for their export revenue – have been subject over at least the last four decades of the twentieth century not only to the downward pressures of the sort emphasized by both Prebisch and Singer but also to the additional pressures generated by increasing terms-of-trade volatility.

Some two and a half centuries have elapsed since the classical economists first proffered their prediction that the tide would inevitably turn, over the long run, in favour of the LDCs. To offer what is perhaps logically the only admissible defence – namely, that some two and a half centuries is too short a time period for the classical mechanisms to fully work themselves out – seems implausible not only to trained economists but more importantly to those citizens of the world’s very poorest economies who seem condemned to remain forever at the very margins of survival.

Notes

1. This chapter is dedicated to the memory of Hans Singer (1910–2006), friend, collaborator and inspiration.
2. For brevity I refer hereafter to the net barter terms of trade between primary commodities and manufactured goods (that is, to the ratio of the price of primary commodities to the price of manufactured goods) as simply their terms of trade.
3. Although conventionally referred to in the literature as the Prebisch–Singer hypothesis, recent evidence reported by Toye and Toye (2003) suggests that it should correctly be termed the Singer hypothesis since their detailed archival research indicates that Prebisch’s interest in the topic stemmed directly from his receipt of Singer’s then unpublished UN manuscript on the subject.

4. Being in contradiction with the then prevailing orthodoxy, these papers attracted criticism on a number of (primarily statistical) grounds. However, detailed scrutiny of the literature seems to indicate that almost six decades after its initial launch the empirical validity of the declining-trend hypothesis first put forward by Prebisch and Singer has become pretty much universally accepted. See Sapsford and Chen (1998) for a review of the second wave of statistical studies that appeared during the late 1980s and the 1990s, and Spraos (1983) and Sapsford (1985) for reviews of the earlier criticisms. This longevity is all the more remarkable when one recognizes the wide array of statistical techniques to which the hypothesis has been subjected. There are few, if any, hypotheses in economics that can claim to have stood the test of time so well as this one. See Spraos (1980, 1983), Sapsford and Singer (1998), Sapsford and Chen (1999) and Bloch and Sapsford (2000) for detailed reviews.

5. The data series employed here were compiled by Grilli and Yang (1988). Dr Stephan Pfaffenzeller of the University of Liverpool compiled the post-1986 data according to the same definitions. This is the terms-of-trade series most commonly analysed in the literature, being compiled as the World Bank’s index of the prices of some 24 internationally traded non-fuel primary products deflated by the UN’s index of the unit values of manufactured exports from industrialized countries.


7. The early twenty-first-century experience of China may well turn out to offer an important lesson here regarding the distinction between what is logically correct and what is relevant.

8. Although outside of the scope of the present chapter it is important to notice that the existence and indeed persistence of volatility around the trend carries with it a potential policy implication regarding the possible merits of instituting some form of stabilization scheme (for example buffer stocks, compensatory financing mechanisms and the like) designed to at least smooth out, if not remove, fluctuations about the trend. The literature here is extensive and goes back in time beyond Keynes, although the latter’s views as expressed at the Bretton Woods Conference are of particular importance. For surveys of the major issues involved see Winters and Sapsford (1990), Maizels (1992) and Sapsford and Morgan (1994).

9. Or, perhaps more correctly, have chosen to move away from the continued reanalysis of the Grilli–Yang (1988) data set.

10. To put these life expectancy figures in perspective it should be noted that the corresponding figures for the USA, the UK and Japan are 76, 77 and 80 years respectively.

11. See UNCTAD (2002) for a summary. Notice that in this study (following Singer and Lutz, 1994, Sapsford and Balasubramanyam, 1999, and others) the standard error of estimate about the estimated regression plane was used as the sub-period measure of terms-of-trade volatility.

12. It is now the case that at least some of the international agencies involved in the world trading system have come to accept that primary commodity producers in developing countries do face real and significant uncertainties and risks regarding the prices that they will actually receive for their products when they come to the world market. At the time of writing in 2006, a task force set up under the auspices of the World Bank is investigating a range of possible ‘market-based’ approaches (including the formation of futures markets) for dealing with the price risks faced by primary commodity producers in developing countries. As pointed out by Morgan (2001) these approaches appear to represent an attempt to confront price risk by modifying the financial environment within which primary producers in less-developed countries operate. However, it remains to be seen whether such approaches will prove any more, or less, successful than the various policies which have preceded them.
References


Introduction
The countries of Western Europe, northern North America, and Australia and New Zealand (the North) began to achieve increasing per capita gross domestic product (GDP) in the first part of the nineteenth century. Growth, so measured, while not uninterrupted, became sustained enough that one may say that the routine functioning of these economies produced increasing per capita GDP. Growth became, in effect, built in. As a consequence of 150–200 years of this fairly routine growth, the countries of the North are now very rich, at least in terms of GDP per capita. The Great Question of Development Economics is: Why have not all countries been able to establish economies in which growth is built in to their routine operations? The broad policy question is therefore how to bring about modifications in these non-growing economies (the South) in such a way that growth becomes routine for them too. The theoretical question is to explain why and how one group of countries did grow while, another, much larger group, has not. In the best of all worlds, this explanatory story would be well grounded in fundamental economic principles, supported by convincing empirical evidence, and lead to policies that can be implemented and, when implemented, bring about the achievement of the objective stated above.

The purpose of this chapter is to study the way that foreign trade enters the development quest just defined. The approach is to tell a story of history and learning, of learning through time as evidence and theoretical insight accumulated. Such a historical, learning approach seems appropriate because of the changing views and continuing controversy, and because no widely held model or theory or policy position has become conventional wisdom.

The beginning
In the late 1940s and early 1950s when ‘development economics’ emerged as a distinct field of inquiry within the broad discipline of academic economics, empirical evidence and explanatory hypotheses about the process of development were relatively primitive. Yet it was widely accepted that the huge differences in per capita GDP between the North and South demanded action, demanded that the world ‘do something’. 

Trade policy
was, from the beginning, a major element both in terms of explanation and in terms of policy.

**Import substitution: argument and practice**

The earliest role of trade as an explanation of the failure of the South to grow rested on arguments and assumptions about the way that the ‘structure’ of production in a country affected its capacity to grow. Structure of production refers mainly to the composition of output of an economy, especially to the distinction between manufacturing on the one hand and agriculture and mining on the other. The economies of the South had, it was widely argued, become locked into agriculture and mining, largely due to foreign trade based (at least to some extent) on assumptions about comparative advantage. While this locked-in state may have resulted in the maximization of current production, the argument continued, it did prevent these economies from growing over time. There were several detailed arguments and assumptions on which these conclusions depended.

The earliest specific argument was that the terms of trade had consistently moved against the South countries. Several reasons were offered as to why this had been and was happening. Productivity growth in manufacturing was thought to be greater than in agriculture and mining, but this higher productivity growth was matched or more than matched by rising wages in the North. The rising wages, combined with the widespread monopoly power of producers in the North, prevented manufacturing prices from falling as productivity grew. While in the South, productivity grew more slowly and an abundant supply of labor prevented wage rates (and prices) from rising. Added to these considerations was the assumption that demand for food (and agricultural products in general) and minerals was inelastic with respect to income in world markets. As output of these commodities increased, their prices relative to those of manufactures tended to fall over time as per capita incomes rose in the North. The South could not shift out of agriculture and minerals into manufactures because its technological capacity was such that it could not compete with the producers of the North as long as free trade prevailed. This disadvantage with respect to technological capacity prevailed despite the fact that the South had access to the same array of equipment and ideas that most of the countries of the North had – mainly imported from Britain.

These arguments disputed any role that conventional trade theory – comparative advantage as formulated by Ricardo and by Heckscher–Ohlin – could play in a development context. Comparative advantage, it was often noted, is a static notion that assumes technical knowledge is the same in all countries, constant returns to scale everywhere, and that equilibrium states always prevail, that is, no second-best issues arise. It thus had nothing to
communicate with respect to formulating a policy aimed at producing marked changes in the structure of an economy. Added to these considerations were references to colonialism and to the manner and extent to which it had drained resources from the poor to the rich countries.

These explanations of the failure of the South to grow had obvious policy implications, especially for trade policy: abandon free trade, and behind protection change the structure of the economy, that is, industrialize. The idea was simple. Products – especially consumer durables – now imported would be subjected to prohibitive tariffs or banned completely and local firms would emerge to take advantage of the already existing demand. These arguments and their policy implications became identified as the import-substitution (IS) approach to development.

Growth theory available at the time was the Harrod–Domar model in which physical capital formation was the basic source of growth. The main (often only) requirement to achieve growth (and changed structure) was a rate of capital formation sufficiently high, given the productivity of capital, to achieve a target rate of growth of output. Since virtually all capital goods were imported from the North, many countries sought to subsidize capital formation by maintaining an overvalued exchange rate. This practice, along with low, often negative, real interest rates, did induce relatively satisfactory rates of investment. At the same time such policies made it particularly difficult for new firms to enter export markets, they penalized employment and resulted in a great deal of underutilized capacity in the newly created firms. Physical capital, deemed the major lack in the South, was frequently found to be unused and poorly maintained. The importation of capital goods along with the inability to export new products created balance-of-payments problems which were often met by further exchange controls or more widespread use of tariffs and quotas.

Distortions were added to by the common practice of many countries of establishing individual tariffs to whatever level was deemed necessary to enable a given activity to come into existence. The result was a wide range of tariff rates and an even wider range of effective rates of protection.

In these highly distorted economies, market prices were quite misleading as signals of social costs and social returns. Considerable attention was given to benefit–cost analysis of large-scale projects using shadow prices, rather than market prices. This was difficult to do with any confidence, and even where it was done, faced the problem that costs and prices paid and received were those prevailing in the market, so that profitability of firms in shadow prices did not assure profitability in market prices which somehow had to be achieved. The use of shadow prices did not, in any way, offset the distortions that the policies to implement the shift in structure imposed on the economies.
It also became clear that the distortions, especially those that affected access to foreign exchange and domestic investment permits, resulted in a substantial allocation of resources to rent-seeking, that is, to seeking permits and licenses to access underpriced resources, rather than seeking investment projects, new technical knowledge and focusing on their own business. There is little doubt that such activities were costly in many ways and dampened the search for ways to make the new firms increasingly productive.11

The general approach to development just described, despite violating many of the more established principles of textbook economics, was widely supported. The World Bank, for example, in its 1979 *World Development Report*, noted (pp. 67–8) that the policy had had important positive effects on entrepreneurial and technological capacity in many developing countries, and had induced the growth of a manufacturing sector behind high levels of protection. This bank report noted the inefficiencies described above, but with much more restraint than was the case a decade or so later. There were other favorable reports and very few explicit criticisms expressed in the literature of the day.

Some consequences
Tables 38.1 and 38.2 show the broad outlines of the post-World War II years when IS was widely followed. The 1950s were dominated by the recoveries from the dislocations of the war, but by the 1960s the effects of IS in many countries became paramount. During the 1960–73 period both labor productivity and total factor productivity growth were impressive over a wide range of countries. Most countries also experienced export growth that dispelled doubts about the ability of developing countries to export. An index from Little et al. (1970, p. 245) shows that manufactured exports from all developing countries increased from a base of 100 in 1953 to 283 in 1965. Capital formation also took place at impressive rates in most countries as the protection of domestic activities created many new opportunities. Even agriculture, despite being heavily penalized, grew at respectable rates in most – not all – countries pursuing IS.

In addition life expectancy increased almost everywhere, infrastructure was improved, and literacy increased. Clearly things got better in the 1950s and 1960s as a consequence of the IS strategy.12 By the end of the 1970s it appeared that IS was a great success and a sure guide to continued growth.

As the tables show however, difficulties appeared as the 1970s wore on. Total factor productivity growth (TFPG) slowed markedly, and many cases of negative TFPG appeared, a sure sign of increasing distortions and coordination failures. Falling growth rates also contributed to negative productivity growth. Export growth slowed and turned negative in many cases.
### Table 38.1 Rates of growth of productivity

<table>
<thead>
<tr>
<th>Country</th>
<th>1960/73 GDP/L</th>
<th>TFP</th>
<th>1973/84 GDP/L</th>
<th>TFP</th>
<th>1984/94 GDP/L</th>
<th>TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>East Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>2.2</td>
<td>1.4</td>
<td>4.3</td>
<td>0.5</td>
<td>4.4</td>
<td>1.6</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2.5</td>
<td>1.1</td>
<td>4.3</td>
<td>0.5</td>
<td>3.7</td>
<td>0.9</td>
</tr>
<tr>
<td>S. Korea</td>
<td>5.6</td>
<td>1.4</td>
<td>5.3</td>
<td>1.1</td>
<td>6.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Malaysia</td>
<td>4.0</td>
<td>1.0</td>
<td>3.6</td>
<td>0.4</td>
<td>3.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Philippines</td>
<td>2.5</td>
<td>0.7</td>
<td>1.2</td>
<td>1.2</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Thailand</td>
<td>4.8</td>
<td>1.4</td>
<td>3.6</td>
<td>1.1</td>
<td>6.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Taiwan</td>
<td>6.8</td>
<td>2.2</td>
<td>4.9</td>
<td>0.9</td>
<td>5.6</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>Latin America</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>2.6</td>
<td>0.2</td>
<td>0.4</td>
<td>1.0</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Brazil</td>
<td>4.4</td>
<td>2.9</td>
<td>1.0</td>
<td>0.8</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Chile</td>
<td>1.6</td>
<td>0.7</td>
<td>0.6</td>
<td>0.7</td>
<td>4.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Colombia</td>
<td>2.9</td>
<td>1.9</td>
<td>1.2</td>
<td>0.0</td>
<td>1.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Ecuador</td>
<td>4.4</td>
<td>3.3</td>
<td>1.7</td>
<td>0.5</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Mexico</td>
<td>3.8</td>
<td>1.6</td>
<td>0.7</td>
<td>0.8</td>
<td>1.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1.2</td>
<td>0.9</td>
<td>−3.1</td>
<td>4.3</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Middle East</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td>4.7</td>
<td>2.3</td>
<td>0.5</td>
<td>2.2</td>
<td>−1.1</td>
<td>1.5</td>
</tr>
<tr>
<td>Iran</td>
<td>6.1</td>
<td>2.4</td>
<td>−2.9</td>
<td>−5.7</td>
<td>−2.2</td>
<td>−2.2</td>
</tr>
<tr>
<td>Jordan</td>
<td>2.1</td>
<td>−0.9</td>
<td>6.7</td>
<td>2.3</td>
<td>1.2</td>
<td>2.9</td>
</tr>
<tr>
<td><strong>South Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>0.0</td>
<td>−0.6</td>
<td>2.5</td>
<td>1.8</td>
<td>1.1</td>
<td>0.7</td>
</tr>
<tr>
<td>India</td>
<td>1.8</td>
<td>0.1</td>
<td>2.4</td>
<td>1.0</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>2.1</td>
<td>1.0</td>
<td>3.2</td>
<td>0.7</td>
<td>2.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Pakistan</td>
<td>3.9</td>
<td>0.2</td>
<td>2.8</td>
<td>2.0</td>
<td>2.7</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2.2</td>
<td>0.2</td>
<td>0.0</td>
<td>−0.9</td>
<td>−0.2</td>
<td>−1.6</td>
</tr>
<tr>
<td>Ghana</td>
<td>0.9</td>
<td>−1.0</td>
<td>−3.2</td>
<td>−3.2</td>
<td>1.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Kenya</td>
<td>3.4</td>
<td>3.4</td>
<td>0.4</td>
<td>−0.1</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1.2</td>
<td>−0.9</td>
<td>−2.3</td>
<td>−4.6</td>
<td>1.3</td>
<td>2.0</td>
</tr>
<tr>
<td>Uganda</td>
<td>0.7</td>
<td>−0.3</td>
<td>−2.9</td>
<td>−3.0</td>
<td>1.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Tanzania</td>
<td>3.0</td>
<td>2.2</td>
<td>−1.1</td>
<td>−1.7</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>S. Africa</td>
<td>2.3</td>
<td>0.9</td>
<td>1.0</td>
<td>−0.3</td>
<td>−2.0</td>
<td>−1.8</td>
</tr>
<tr>
<td>Zambia</td>
<td>1.0</td>
<td>0.2</td>
<td>−2.3</td>
<td>−1.9</td>
<td>−2.5</td>
<td>−1.1</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>2.9</td>
<td>2.7</td>
<td>−0.8</td>
<td>−1.3</td>
<td>0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Capital formation in the poorer countries declined and, while remaining strong in a number of middle-income countries, declined in others as investment opportunities dried up. The slowing down of capital formation and productivity growth also dampened an already weak demand for labor.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low-Income countries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>6.6</td>
<td>−7.3</td>
<td>11.1</td>
<td>−7.8</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>3.7</td>
<td>−3.8</td>
<td>5.7</td>
<td>−0.9</td>
</tr>
<tr>
<td>Malawi</td>
<td>11.6</td>
<td>3.0</td>
<td>13.3</td>
<td>4.5</td>
</tr>
<tr>
<td>India</td>
<td>3.1</td>
<td>6.4</td>
<td>5.6</td>
<td>2.1</td>
</tr>
<tr>
<td>Pakistan</td>
<td>8.2</td>
<td>−3.0</td>
<td>6.9</td>
<td>−0.7</td>
</tr>
<tr>
<td>Tanzania</td>
<td>3.5</td>
<td>−7.2</td>
<td>9.8</td>
<td>2.7</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>4.6</td>
<td>−5.3</td>
<td>6.6</td>
<td>0.9</td>
</tr>
<tr>
<td>Kenya</td>
<td>7.2</td>
<td>1.2</td>
<td>7.0</td>
<td>−2.7</td>
</tr>
<tr>
<td>Uganda</td>
<td>5.0</td>
<td>−9.6</td>
<td>9.8</td>
<td>−11.5</td>
</tr>
<tr>
<td>Indonesia</td>
<td>3.5</td>
<td>7.5</td>
<td>4.8</td>
<td>16.6</td>
</tr>
<tr>
<td><strong>Middle-income countries</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td>3.2</td>
<td>−3.3</td>
<td>3.1</td>
<td>23.6</td>
</tr>
<tr>
<td>Ghana</td>
<td>0.1</td>
<td>−1.9</td>
<td>−3.2</td>
<td>−8.6</td>
</tr>
<tr>
<td>Nigeria</td>
<td>6.1</td>
<td>1.3</td>
<td>6.5</td>
<td>22.9</td>
</tr>
<tr>
<td>Thailand</td>
<td>5.2</td>
<td>12.1</td>
<td>15.4</td>
<td>6.3</td>
</tr>
<tr>
<td>Philippines</td>
<td>2.2</td>
<td>5.0</td>
<td>8.2</td>
<td>11.7</td>
</tr>
<tr>
<td>Zambia</td>
<td>2.2</td>
<td>−2.3</td>
<td>10.6</td>
<td>−5.9</td>
</tr>
<tr>
<td>Jordan</td>
<td>10.1</td>
<td>20.8</td>
<td>9.9</td>
<td></td>
</tr>
<tr>
<td>Colombia</td>
<td>2.2</td>
<td>1.2</td>
<td>4.5</td>
<td>3.6</td>
</tr>
<tr>
<td>Ecuador</td>
<td>3.7</td>
<td>9.0</td>
<td></td>
<td>12.1</td>
</tr>
<tr>
<td>South Korea</td>
<td>35.2</td>
<td>30.7</td>
<td>23.1</td>
<td>12.4</td>
</tr>
<tr>
<td>Peru</td>
<td>1.9</td>
<td>−4.4</td>
<td>2.4</td>
<td>7.1</td>
</tr>
<tr>
<td>Malaysia</td>
<td>6.1</td>
<td>5.2</td>
<td>7.2</td>
<td>10.0</td>
</tr>
<tr>
<td>Turkey</td>
<td>1.6</td>
<td>0.8</td>
<td>8.8</td>
<td>12.7</td>
</tr>
<tr>
<td>Mexico</td>
<td>3.3</td>
<td>1.9</td>
<td>9.5</td>
<td>8.1</td>
</tr>
<tr>
<td>Chile</td>
<td>0.6</td>
<td>7.7</td>
<td>4.2</td>
<td>−8.9</td>
</tr>
<tr>
<td>China</td>
<td>23.7</td>
<td>6.7</td>
<td>16.2</td>
<td>9.1</td>
</tr>
<tr>
<td>South Africa</td>
<td>5.5</td>
<td>6.7</td>
<td>9.0</td>
<td></td>
</tr>
<tr>
<td>Brazil</td>
<td>5.0</td>
<td>6.5</td>
<td>5.3</td>
<td>12.6</td>
</tr>
<tr>
<td>Argentina</td>
<td>3.3</td>
<td>5.5</td>
<td>4.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Iran</td>
<td>12.7</td>
<td>−0.2</td>
<td>12.2</td>
<td>22.6</td>
</tr>
<tr>
<td>Venezuela</td>
<td>2.0</td>
<td>−10.5</td>
<td>7.3</td>
<td>9.8</td>
</tr>
</tbody>
</table>

and unemployment became an increasingly evident issue. Similarly almost all countries had trouble maintaining both internal and external stability.\textsuperscript{13}

Two things were clear to most observers by the early 1970s: the first was that for well over a decade IS had been a genuine success, and the second was that, as practiced, it was not the strategy that could create an economy in which growth was built in to the routine operation of the economy.

Several aspects of the strategy became widely appreciated. That the assumption of fixed production coefficients of Harrod–Domar was the source of many difficulties especially became recognized.\textsuperscript{14} The assumption of fixed production coefficients was a principal reason for the failure of the high rates of investment to create the demand for labor that would match available supply as well as for the widespread underutilization. The assumption rested largely on the belief, widely held, that the technology of the North was so dominant that it made any other technology, any other factor combination, technologically inefficient, that is, inefficient at any set of factor prices. So in effect firms in the South had no choice of technique or machine if they were to compete in the world economy.

This last assumption rested on another, larger, more encompassing notion. The prevailing idea of development at this time was simply to duplicate the North. This idea is given full exposition by Arthur Lewis in a very influential article published in the early 1950s.\textsuperscript{15} Lewis divides the economy of a South country into a large, very poor, traditional sector and a small capitalist sector. Investment was to take place in the latter sector and was to be accomplished by importing physical capital and technology from the North. As investment continued, the capitalist sector would grow relative to the traditional sector until it encompassed the whole economy and the South would be as the North. Development was in effect imported, not indigenous, and capacity to import became a crucial constraint.\textsuperscript{16}

The dual economy model was thus a story of displacement, not an explanation of how a non-growing traditional economy changed itself into a growing economy. This pervasive idea, development as replication of the North, had several consequences. It detracted attention away from the development notion itself, that is, to achieve the metamorphosis of a traditional, non-changing economy into a growing one. A theory of development is concerned with explaining how this traditional economy can be turned into a growing one in such a way that those characteristics that, in effect, define the society – its history, its values, its ethos, its very meaning – are not violated. Fixed production coefficients, the idea that all technical and organizational knowledge from the North could be codified and immediately utilized as in the North, the achievement of the North’s product mix, all contributed to the dampening, even preventing, of any domestic efforts to adapt, to seek and to learn, and to recognize the importance of
building onto and from and within existing institutions and within the boundaries defined by the ideas of the good life of the population. It is evident, from this statement, that the role of trade (and other foreign activities) is vastly important, but subtle and complex.\textsuperscript{17}

Finally, the strategy as pursued not only neglected, but in general penalized agriculture. Agriculture was often taxed and more often price controls on foodstuffs were imposed to keep their prices (and therefore wages) low in urban areas as yet another means of subsidizing new manufactures. The neglect of agriculture also encouraged a more rapid exodus from rural to urban areas, the result of which was increased unemployment in the cities and the emergence of ‘informal sectors’ alongside the new manufactures. Agriculture in most countries of the South was the largest sector in terms of both production and employment. To penalize this sector meant that the major sector of the economy was being penalized. In a country where new exports were very slow in emerging, a weak agricultural sector meant that the growth of domestic demand for new, non-agricultural products was severely dampened, and this in turn impeded the learning-by-doing process in the new firms. That agriculture could be safely penalized was a by-product of the ‘structural change’ argument for import substitution: the other side of creating a new manufacturing sector was the killing off of agriculture. Where exports of the new activities were not possible and agriculture was penalized, then importing food grains often became necessary and was costly. The late 1950s in India were a prime example of this sort of phenomena.

That this set of policies did not result in the creation of a built-in growth process does not now appear surprising. Despite the burst of early growth in the 1950s and 1960s noted above, and the improvement in a number of welfare measures, problems began to emerge that convinced most observers that the import-substitution process was in no sense sustainable. The details of the story just summarized vary markedly among the countries of the South, but the broad picture seems generally applicable. For Africa less so, perhaps, than in most other places, and more so in India, Pakistan and most Latin American countries.\textsuperscript{18}

The fall of IS and the rise of openness
While recognition that the IS policies had created an unsustainable situation became widespread, alternative strategies did not convince many policy-makers in the South. That distortions were ubiquitous and were penalizing the economies was appreciated by most observers. Economists still had only rough ideas of how to make an economy grow, but of course distortions and optimal allocation of resources were textbook stuff. It is not surprising therefore that the new strategy proposed concentrated...
attention on eliminating the sources of these distortions: the ad hoc tariffs, internal controls of many kinds, overvalued exchange rate, soft budget constraint, public ownership of many firms, price controls and inflationary fiscal policies. The catchwords became ‘market friendly’, ‘privatizing’, ‘macro stability’ and ‘openness’. ‘Outward orientation’ and ‘Washington Consensus’ replaced IS as the summary term. This new approach originated largely in the North and the pressure to liberalize came from the North, the World Bank, the United States Agency for International Development (USAID) and Northern academic economists, hence the appellation Washington Consensus. Reluctance of South countries to move sharply toward liberalization was partly a matter of the immediate economic costs (especially the almost inevitable increased unemployment) to withdrawing protection from recently established activities that could not survive without protection, and partly a matter of genuine doubt that the newly pushed theory would in fact have the desired effect. In addition, there were numerous people who profited from IS, and who, therefore, opposed its abandonment.

There was no new ‘theory’ of development that led to this different strategy. The liberalization package as looked upon by the South as simply a return to pre-World War I international arrangements. In particular the idea that if property rights are clear and in place, if the price mechanism follows the textbook stories, if macrostability is maintained, and the investment rate is at least 15 percent, the economy will grow with full employment, was not a widely accepted view outside the Washington Consensus group. It was too easy to see that some countries – for example India, Korea, Vietnam, China and Botswana – were growing well by violating many of the conditions laid down as essential by the Washington Consensus. Similarly other countries – especially in Latin America – had followed the rule rather closely and performed much less well.19

The period from 1870 to 1914 is especially illuminating. During this period capital and trade moved easily to almost all parts of the world, government interventions in the economy were modest, migrations were large and widespread, foreign exchange markets were fairly stable due to the widespread commitment to the gold standard and to the British pound, the international capital markets were effective, price stability was common, and there were significant social and political changes throughout the world. In addition, there were marked improvements in transportation and communication around the world.20 These circumstances, however, did not produce industrialization in the South, and did not produce productivity growth, nor the growth of factory employment in the South.21 While there was some growth in the South during this interval, it is correct to say that the North grew faster and routine growth became even
more firmly established. This experience was rarely confronted by the advocates of openness.22

**Doubts about the Washington Consensus**

Distortions of course mean (by definition) that the economy operates within its production possibilities frontier with the given technology that is available. So eliminating the distortions would be expected to result in a once-over increase in output; but the real question was what produced growth after this one-shot effect has been absorbed.

The main evidence leading to the outward-oriented position was the success of Korea and Taiwan. Taiwan had begun to grow rapidly from the early 1950s and Korea from the early 1960s. The most obvious feature of these impressive performances was the high growth rate of exports, often non-traditional products and services that had come into existence behind the protection and the subsidies that IS had provided. Similarly it was widely believed initially that the two economies were largely market driven with very little government intervention. Both countries made marked policy changes in this direction in the 1960s. These changes did reduce distortions, but they also included other government policies that subsidized exports and capital formation. In particular it became clear that both countries maintained an undervalued exchange rate over much of the time after 1960. There were other policies and arrangements that impeded imports. Liberalization as practiced by these two countries did not mean ‘free’ trade. It also became understood that both countries had accumulated a great deal of collective learning during the Japanese occupancy of the first part of the twentieth century.23 Korea also learned greatly from the presence of United States armed forces engineers in their country. This collective learning had produced a labor force much more experienced and skilled than in most of the other countries trying the IS strategy. This last item is especially relevant in understanding how and why production in the two countries responded so well and so quickly to the incentives and subsidies offered by their governments.

There was one other characteristic of government policy-making that is highly relevant to this story. Both governments recognized that good policy-making required trial and error and hence willingness to change policies and try different approaches. The idea was not minimal government, but rather a searching, learning government.24 This searching for the right policies was necessary simply because it was recognized that there can be no right policy that can be arrived at in any other way.

The Washington Consensus became widely advocated and strongly pushed by international agencies and key academic figures. To repeat the point made above: that IS, as practiced, had run into a dead end was, in
general, appreciated, but the Washington Consensus seemed simply to go back to an approach to development that had also failed as evidenced in particular by the experience of the 1870–1914 period. Research in the 1970s and 1980s on Korea and Taiwan convinced almost all observers that their success was not due to a minimal government or a market-friendly approach, and that these economies were much less open than was thought to be the case in the 1960s and 1970s.25

There have been, from early on, efforts to examine empirically the effects of IS and outward-orientation policies on growth of GDP, on employment and productivity, and on macrostability. A survey of this literature is beyond the scope of this chapter, but a few remarks may give the flavor of the current status of this work.26 There are two general approaches: case studies of individual countries and cross-country regression analyses involving numerous countries. There are difficulties at all steps along the way in both approaches: definitions and measurement of inward- and outward-looking, quality of available data, the appropriate theoretical formulations, distinguishing the role of trade policy from that of macro stability, education and other possible factors, and many more. Cross-country regressions, once so widely used, are increasingly recognized as subject to so many difficulties that their results are essentially meaningless.27

Three issues of direct relevance to trade policy may be noted briefly. The first is the role of exports. Do countries grow well because they export or do they export because they grow well? At the aggregate level the evidence might well support the former position, but at the firm or sector level, the evidence generally supports the opposite view. The micro data appear somewhat more convincing, that is, firms must find their basic inducements to search and learn that lead to productivity growth within their own indigenous environment, and thereby become able to export. It does seem clear that once they begin to export, then they can gain additional knowledge from that activity. Secondly, evidence supports the view that exporting does not have much effect on productivity growth of traditional activities of the developing countries.28

Thirdly the empirical (and other) evidence supports the view that the exchange rate is a crucial policy variable. The overvalued exchange rate was a major, perhaps the major, reason for the failure of the IS approach. Recent empirical work and case studies show convincingly the strength of an undervalued exchange rate as an instrument that encourages import replacements and the search for foreign markets as well as the search for increased productivity and higher-quality output.29

While it is important that empirical and historical research continue, it now seems likely that for the foreseeable future dispute between the two approaches cannot be resolved by evidence of the conventional sort applied
in the usual way. The most illuminating approach is simply case studies of individual countries to ascertain what has gone on in these countries. Generalizations across countries and through time, supported by rigorous theoretical formulations and by equally rigorous econometric results, do not now appear possible, and certainly not possible to the extent that specific policy formulations emerge that are applicable to all countries. The last section addresses briefly the consequences of this position.

The upshot of it all
An important reason why the role of trade in development is difficult to isolate is that the development process itself is not yet fully understood. There are two main aspects of development that both IS and the Washington Consensus seemed not to have appreciated in their early formulation. The first was the assumption, implicit in most of the literature of both strategies, that technical and other knowledge are public goods, available to everyone and equally productive in all places. Research of the 1980s and 1990s has shown this assumption to be terribly misinformed.30 Productivity levels among firms in the same activity vary widely within a country and across countries, as do rates growth of productivity, profit rates, capital–labor ratios, product quality, and so on.31 These great and abiding differences are explained by the fact that much of knowledge cannot be codified and put into manuals, but is tacit and can be accumulated only by producing and is, therefore, unique to the firm where it is accumulated. That this is the case is due to the great differences in entrepreneurship, in the ingenuity of the shop floor people, and in the extent of commitment to searching and learning.32 This finding has had a fundamental impact on development theory and practice: since sustained growth requires sustained increases in productivity, which, in turn, requires continuing increases in knowledge, a growing economy must be one in which firms are, as a matter of routine, actively engaged in searching for new knowledge.33 Policy incentives must be geared to this objective, different from the incentives aimed at achieving an ‘optimal’ allocation of given resources and given technology.34 Trade policy, in particular, is greatly affected by the notion of tacit knowledge and its source.

The second important research finding has to do with institutions: institutions as norms and rules of the game. As such they have direct and significant effects on the response of economic agents to market incentives. Economic actors are rule-followers as well as profit-seekers, and the rules emerge from and in turn create institutions. So whether a policy will have the desired effect depends on the institutional environment within which it is applied, and any change is sure to be resisted to some extent. In particular it means that the usual market signals can be expected to
have diverse results in different countries and at different times in the same country.

Given the two preceding points, it is evident that the textbook notions of ‘optimal’ allocation of resources or some other ideal or maximum state are basically meaningless: technology is always changing, that is, its change is part of the firm’s activity, and institutions mean that the textbook ideal – for example perfect competition – cannot be defined independently of these institutions. To understand a given economy means to gain some insight into how it works and then to show that its workings can be improved upon according to some practical criteria. This is quite different from the more usual objective of trying to create an economy that follows a general strategy, for example perfect competition, Pareto optimality, and so on.

Incentives are to be aimed at creating inducements to search and learn. Import-substitution strategies created investment opportunities, but also dampened any inducements to search and learn, to create tacit knowledge and thereby continue to grow. The Washington Consensus assumed that simply eliminating distortions by virtue of an unfettered market and a ‘realistic’ exchange rate would produce growth. Both were and are inadequate, because they misunderstand tacit knowledge and the role of institutions.

The final question is what kind of a trade policy will induce firms to engage in the searching and learning process that does produce growth and improved quality of output. Foreign trade policy should aim at making it very profitable to export and to replace imports. Such an approach is different from that usually practiced of subsidizing inputs – capital, loanable funds, land, fertilizer, and so on – as noted earlier.

There are several trade policies available that reflect this point of view. The most evident is an undervalued exchange rate, that is, one that results in the accumulation of foreign exchange. Such an exchange rate policy has been pursued by China, Taiwan, South Korea, Brazil and Japan at times to considerable advantage. Foreign aid takes the form of guaranteeing favorable prices for a developing country’s exports, or prices that rise as the exporting firm increases its productivity or tax relief based on increased export earnings. A policy that rewards firms that increase employment over time with a given capital stock has had favorable effects on both employment and productivity growth.

The general conclusion as to the present state of play of the role of trade may be summed up in this way: we have learned over recent decades that simple notions such as IS and outward orientation lead nowhere. Trade policies – including exchange rate policy – must emerge from a clear vision of how development takes place. That vision must explicitly include knowledge accumulation and application and the recognition that much knowledge is necessarily tacit. It must also recognize that institutions are so
fundamental and so history-and culture-dependent that generalizations across countries are risky indeed. Thus a country, in determining its foreign trade policy, must study how its economy in fact works.\textsuperscript{35}

Notes

1. There is some evidence that inequality among nations increased between 1500 and 1800. See, for example, Lindert and Williamson (2003).
2. President Harry Truman, in his inaugural address in early 1949, committed the United States to supply technical and financial aid to the countries of Africa, Asia, and Latin America.
3. This argument was first developed in Raul Prebisch (1950) and by Hans Singer (1950). A later and slightly different version is Lewis (1978).
4. A formal model that demonstrates that free trade can be harmful to long-run growth in those countries that are locked into sectors where productivity is low and whose indigenous technological change is slow or completely absent, is Rodriguez and Rodrik (2001).
5. See the papers by Gregory Clark and Robert Feenstra (2003) and by Clark (1987).
6. The notion of an optimal rate of saving was often introduced. If this optimal rate of domestic saving was not high enough to produce the target rate of growth of output, then foreign aid or borrowing could supplement domestic saving to the extent necessary to reach the target.
7. Gordon C. Winston (1974) is one of the earliest analyses of the underutilization of capital issues.
8. Countries that had long relied heavily on one or two primary exports often experienced the ‘Dutch disease’ even before the beginning of import-substitution policies. The overvaluation pursued as a policy to encourage capital formation often exacerbated an already misleading exchange rate regime.
10. The study by Little and Mirrlees (1974) is a fine review of the state and role of cost-benefit analysis at the time.
11. Anne O. Krueger (1974) first drew the attention of the profession to this important cost of the distortions.
12. African countries shared less in this success than in countries in other parts of the world. This failure had little to do with IS as few African countries made a substantial commitment to the strategy.
13. Although the term ‘import substitution’ seems to imply that imports should be reduced relative to GDP, this was rarely the case. The strategy proved quite import-intensive. See Carlos Diaz-Alejandro (1965).
14. Robert Solow’s neoclassical growth model (Solow, 1956) was motivated largely by the recognition that the fixed-coefficient assumption severely penalized employment growth as well as impeding adjustment of production techniques to the factor supply situation in general.
15. See W. Arthur Lewis (1954) and, somewhat later, Fei and Ranis (1964) for elaborations of the labor surplus models of growth.
16. Chenery and Strout (1966) developed a model in which capacity to import was a constraint on growth along with the capacity to save.
17. The role of ‘tacit knowledge’, in contrast to codified knowledge, is a crucial part of the story and will be discussed later.
18. There were many studies that explored the import substitution experience in specific countries and across several countries. Little et al. (1970) and the six country studies that accompanied it were among the first that attracted attention. Two studies by Balassa (1971) and Balassa et al. (1982) were also important in spreading understanding about the problems of IS, as was Krueger (1978). There are many others.
20. Some calculations show that the correlation between domestic saving and domestic investment was lower in this period than in recent years when capital markets are assumed to work extra well. With ‘perfect’ international capital markets there should be no or very little relationship between domestic saving and investment.

21. Elaboration of this point is in Lewis (1978) and Dowrick and DeLong (2003). See also Maddison (1970).

22. The story is similar if one begins in 1820 except that from 1820 to 1870 there were more tariffs and other interferences with the flow of trade and other forms of market interventions.

23. See Kohli (1994) for a good discussion of the impact of the Japanese occupancy on the creation of human capital in the two Koreas and Japan.

24. Gustav Ranis (in Winrock International Institute for Agricultural Development and USAID, 1991, pp. 128–9) points out that: . . . ‘what happened in Taiwan was not Mandarins sitting around saying is this what we have to do now. There was a lot of bumbling and stumbling and going back and forth.’ See also Biggs et al. (1995).

25. The World Bank study (1993) acknowledged this picture in general, but argued strongly that Korea and Taiwan were distinctive in several ways and their experiences and policy packages could not be replicated in other countries. See also Stiglitz (2000).


27. Jesus Felipe (1999), Durlauf (2000) and Brock and Durlauf (2001) all have especially helpful discussions of cross-country regressions and growth accounting in general.

28. Sectoral and firm studies have become plentiful in recent years, and results can always be questioned. Michael Hobday (1995) is a thorough study of the relationship between exporting and learning. Roberts and Tybout (1996), Aw et al. (1998), Temple (1999), Tybout (2000) and Bernard and Jensen (2001) all have helpful discussions and much data. Westphal (1990) is a particularly good discussion of the role of exports.

29. See Bruton (1997, Chapter 8) and Bruton (1998) for a discussion of the undervalued exchange rate as an instrument of development policy.


31. There are many studies available that confirm this state of things. See the sources cited in notes 28 and 30.

32. Tacit knowledge has been discussed by many people. Nelson and Winter (1982) were perhaps the first to employ it in a strategic way in economic analysis. More recently good discussions are found in papers by Chandler et al. (1998) and Chandler (1992). See also Helleiner (1992).

33. John Williamson (2003, p. 324) lists ten ‘points’ that are included in the Washington Consensus and that will (presumably) produce growth. The list does not include any item that acts directly on growth, but essentially defines the conditions for a perfectly competitive economy with no inflation and a government role limited to conventional post office, defense, infrastructure, and so on. See also Krueger (2000).

34. The differences among firms means that it is a dubious business to speak of a country’s comparative advantage. Some firms in an industry export, some do not. For a country to seek to produce those goods and services in which it has a comparative advantage is to seek that which is not there. See Hausmann and Rodrik (2002) for an interesting empirical study of the determination of activities in a country.

35. This position is increasingly recognized even by economists who differ sharply in other ways. See almost any of the recent writings of Dani Rodrik, especially Rodrik (1999, 2000) and Stiglitz (2000) on the one hand and Srinivasan and Bhagwati (2001) on the other. This point is also discussed in Bruton (1998).

References


Hobday, Michael (1995), Innovation in East Asia, Aldershot, UK and Brookfield, USA: Edward Elgar.


Introduction
The vast literature on foreign direct investment (FDI) in developing countries may seem disproportionate to the volume of FDI they harbour. Developing countries as a group accounted for 25 per cent of the total stock of FDI of $8.9 trillion in the world economy at the end of the year 2004, most of which, around 70 per cent, was accounted for by 11 developing countries (Tables 39.1 and 39.2). One reason for the intense interest in FDI is the nature of the beast, which appears to evoke both admiration and opposition in equal measure; admiration because of the unrivalled ability of the multinational enterprises (MNEs), the main purveyors of FDI, to transfer technology and know-how across borders; opposition because the MNE is first and foremost a profit-maximizing entity. In the recent past attitudes towards FDI have turned from a mixture of suspicion and admiration towards one of unqualified admiration, shown by the eagerness of most developing countries to attract FDI. This change in attitude towards FDI on the part of host developing countries is influenced by a number of factors: a steep reduction in alternative sources of finance in the wake of the debt crisis, the collapse of the Soviet Union and with it a waning of ideological opposition to capitalism and its institutions, the demonstrable success of the East Asian countries based in part on FDI, and growth in knowledge and understanding of the nature and operations of FDI on the part of the host countries.

The issues that have surfaced in the recent literature on FDI reflect these changes. These have to do with the specific factors which figure prominently in the choice of locales for investment by foreign firms, and the factors which promote effective transfer of technology to the host countries and maximize the benefits they can expect from FDI. In addition, there is the suggestion mooted principally by the EU and Japan that FDI should be on the agenda of the World Trade Organization (WTO) on a par with trade in goods and services.

This chapter reviews these and other issues in the literature on FDI. The next section discusses the determinants of FDI, drawing on the theoretical and empirical literature. The subsequent section discusses the impact of FDI on growth and development in host developing countries. The final section draws some conclusions.
Determinants

Stephen Hymer (1976) initiated the discussion on the determinants of FDI with the thesis that firms go abroad to maximize the rents inherent in the advantages they own. These advantages range from the possession of a...
brand name to complex processes and product technologies. Hymer’s work set in train a number of studies on the foreign investment decision process of foreign firms: John Dunning (1973, 1981), Buckley and Casson (1991), and Markusen (2004) who encapsulated the key factors in the foreign investment decision process of firms as the ‘OLI paradigm’. Simply put, firms must possess ownership advantages (O), location advantages (L) and must be capable of internalizing operations (I) if they are to invest abroad. Internalization here refers to the ability of the firm to set up its own suppliers of materials and equipment (backward linkages) and marketing operations (forward linkages). Such internalization is necessary to overcome imperfections in the market which may result in a number of problems including the loss of ownership advantages through imitation by others, hold-up of operations by suppliers of materials and equipment, and the problems associated with decision-making with imperfect information (Williamson, 1981).

The concern of the host countries is with location advantages (L): how best to lure firms, which possess ownership advantages and are able to internalize, to their locale and benefit from the technology and know-how they bring along and the jobs they create. There is now much received wisdom on the sort of L factors which will attract FDI, the result of years of research and the experience of host countries with FDI. It is now well known that artificial incentives such as tax concessions and tax holidays offered to foreign firms do not always attract foreign investors; such incentives are not a substitute for the sort of economic environment foreign firms seek. Countries endowed with the sort of raw materials the foreign firms seek, those blessed with cheap but productive labour (low efficiency wages), and economies which enjoy macroeconomic stability and pursue stable economic policies tend to attract relatively large volumes of FDI. In the absence of these fundamentals, tax concessions and various sorts of subsidies are of little attraction to foreign firms.

In the same vein, it is suggested that policy-induced incentives such as tariffs on imports and subsidies to exports do not lure foreign firms. In fact a neutral trade policy which favours neither production for exports nor the production of import substitutes for the home market tends not only to attract large volumes of FDI but also promotes efficient utilization of FDI. The import-substituting domestic market-oriented strategy (IS strategy) is characterized by tariffs and quotas on imports, which on average outweigh the average level of subsidies given to exports. In the case of the export-oriented strategy (EP strategy) the opposite is the case. These sorts of incentives, which bias production in favour of either exports or the domestic markets, distort allocation of resources in the economy. They are also artificial and uncertain incentives. They are artificial in the sense that they do
not reflect the true market-dictated comparative advantage of various sectors in the economy; they are uncertain in the sense that they are subject to the vagaries of government policy. Foreign firms in general are not attracted by such artificial incentives and those that do respond and operate in the distorted environment may reap benefits but they tend to be transient. A neutral strategy favours neither the export markets nor the domestic markets; resource allocation in such an environment is dictated by market forces. Foreign firms seek such an environment, which allows for the full play of the comparative advantage they possess. There is robust statistical support for the proposition enunciated by Bhagwati (1978) that a neutral strategy attracts relatively large volumes of FDI and also promotes its efficacy (Balasubramanyam and Salisu, 1991; Balasubramanyam et al., 1996).

The recent experience of India and China with FDI, though, does not seem to conform to received wisdom on the determinants of FDI. They both possess most if not all of the location advantages foreign firms seek. Both countries are endowed with relatively cheap labour, both have liberalized their trade and FDI regimes to a considerable extent, they both possess large domestic markets and they have posted impressive growth rates in recent years. But they differ markedly in the volume of FDI they have attracted. China attracts ten times more FDI than India does: in recent years the annual average inflows into China have averaged around $50 billion compared with the $4 billion that India attracts (Figure 39.1 and

![Figure 39.1 Inflows of foreign direct investment in China and India: 1980–2004 ($ US million)](chart)

Source: UNCTAD

*Figure 39.1 Inflows of foreign direct investment in China and India: 1980–2004 ($ US million)*
Table 39.3. Indeed India’s FDI regime is reported to be much more liberal than that of China (Nagaraj, 2003). Even so, the volume of FDI in India is lower than that in China by a factor of ten or more, a fact frequently debated in the media.

A number of reasons are offered for the observed differences in the volume of FDI in the two countries including differences in the accounting procedures between the two countries, so-called round-tripping FDI in China, and the horrendous Indian bureaucracy which stiﬂes any sort of enterprise, be it foreign or domestic. There may be other explanations for the observed differences in the volume of FDI the two countries harbour. Yashang Huang (2003) is of the view that a considerable volume of FDI in

<table>
<thead>
<tr>
<th>Year</th>
<th>China</th>
<th>India</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>57</td>
<td>79</td>
</tr>
<tr>
<td>1981</td>
<td>265</td>
<td>92</td>
</tr>
<tr>
<td>1982</td>
<td>430</td>
<td>72</td>
</tr>
<tr>
<td>1983</td>
<td>916</td>
<td>6</td>
</tr>
<tr>
<td>1984</td>
<td>1,419</td>
<td>19</td>
</tr>
<tr>
<td>1985</td>
<td>1,956</td>
<td>106</td>
</tr>
<tr>
<td>1986</td>
<td>2,244</td>
<td>118</td>
</tr>
<tr>
<td>1987</td>
<td>2,314</td>
<td>212</td>
</tr>
<tr>
<td>1988</td>
<td>3,194</td>
<td>91</td>
</tr>
<tr>
<td>1989</td>
<td>3,393</td>
<td>252</td>
</tr>
<tr>
<td>1990</td>
<td>3,487</td>
<td>237</td>
</tr>
<tr>
<td>1991</td>
<td>4,366</td>
<td>75</td>
</tr>
<tr>
<td>1992</td>
<td>11,008</td>
<td>252</td>
</tr>
<tr>
<td>1993</td>
<td>27,515</td>
<td>532</td>
</tr>
<tr>
<td>1994</td>
<td>33,767</td>
<td>974</td>
</tr>
<tr>
<td>1995</td>
<td>37,521</td>
<td>2,151</td>
</tr>
<tr>
<td>1996</td>
<td>41,726</td>
<td>2,525</td>
</tr>
<tr>
<td>1997</td>
<td>45,257</td>
<td>3,619</td>
</tr>
<tr>
<td>1998</td>
<td>45,463</td>
<td>2,633</td>
</tr>
<tr>
<td>1999</td>
<td>40,319</td>
<td>2,168</td>
</tr>
<tr>
<td>2000</td>
<td>40,715</td>
<td>2,319</td>
</tr>
<tr>
<td>2001</td>
<td>46,878</td>
<td>3,403</td>
</tr>
<tr>
<td>2002</td>
<td>52,743</td>
<td>3,449</td>
</tr>
<tr>
<td>2003</td>
<td>53,505</td>
<td>4,269</td>
</tr>
<tr>
<td>2004</td>
<td>60,630</td>
<td>5,335</td>
</tr>
</tbody>
</table>

Source: UNCTAD (various issues).
China is a substitute for domestic investment. Despite the relatively high savings rate, domestic Chinese investors find it hard to obtain credit lines and borrow from the banks, and they turn to foreign investors for finance. Many of the state-owned enterprises on the verge of bankruptcy also seek FDI, and foreign firms invest in regions of China which local firms hesitate to enter. It may also be the case that the requirements of the Indian manufacturing and services sectors are relatively low compared with those of China, simply because India is better endowed than China with the sort of human skills which FDI provides. In sum the determinants of FDI are rooted in the endowments of the host countries including human skills, the infrastructure facilities they provide and, most importantly, a policy framework which provides distortion-free product and labour markets.

**Impact**

FDI is defined as ownership of facilities abroad with control over operations. The distinguishing feature of FDI which sets it apart from other forms of capital flows is the control over operations which the parent company exercises over its subsidiaries abroad. As stated earlier, such control over operations is essential to preserve the ownership over advantages the firm enjoys. Such control over operations is exercised by virtue of ownership of equity and possession and control of technology and know-how. The higher the degree of equity participation by the investor firm in an entity, the greater is its ability to exercise control over operations. It is thought possible for a firm to exercise control over operations if it owns say only 4 per cent of the total equity of an entity and the other 60 per cent is disbursed amongst a number of investors. In fact, the International Monetary Fund (IMF) defines any investment by a firm with an equity share of 10 per cent as FDI. The second attribute of FDI – ownership of technology and know-how – is a much more powerful tool for exercising control over operations than majority equity ownership.

The three attributes of FDI, equity ownership, control over operations and transfer of technology, are intertwined. Ownership of equity and technology enables the firm to exercise control over operations and preserve its monopoly over technology and know-how, which in turn enables it to transfer technology and know-how across frontiers. The essential point to note here is that ownership of equity is a means to an end, the end being control over operations and transfer of technology. If the market for technology were perfect and if technology and knowledge were not public goods, the multinational company – the purveyor of FDI – would prefer to enter international markets by contracting technology-licensing agreements. Such agreements, by definition, are bereft of equity participation: the firm transfers technology to the licenses in return for fixed technical fees.
and royalties tied to the profits of the licensee. But because technology can be easily imitated and it is difficult to arrive at a price for most technologies, multinationals engage in FDI, that is, they own equity in entities abroad.

This somewhat long-drawn-out discussion of FDI is to emphasize the fact that it is technology and know-how transfer which is the main attribute of FDI, and it is this attribute which is sought after by countries host to FDI. Admittedly the capital that accompanies technology and know-how is also a benefit to the host countries, but for reasons specified above, transfer of capital is not a major feature of FDI. In any case, the larger the equity share of the foreign firm in an entity, the higher would be the control over operations it exercises, something which most host countries would wish to minimize. Also, relatively poor countries, which have very few of the attributes discussed in the section on determinants, may not be able to attract FDI in large enough volumes to meet their demands for capital.

FDI is also sought for the foreign exchange it provides developing countries, in the form of both the capital that accompanies it and the exports it promotes. The contribution of FDI to China’s exports is well known (Wei, 2004), and also to those of other East Asian countries such as Malaysia and Singapore (Driffield et al., 2004). FDI can also save foreign exchange for the host countries with productive investments in import-substituting industries. All of this, of course, makes for a healthy balance of payments of the country. But here again it should be noted that any investment, be it export-oriented or domestic market-oriented, contributes to the balance of payments as long as it is socially productive. This is simply because the balance of payments, as Kindleberger (1969), commenting on the balance of payments effects of FDI reminded us, is a general equilibrium phenomenon. In other words, the balance of payments is an integral part of the total economy. An FDI project which is socially productive, in the sense that the private rate of return to the investment does not exceed the social rate of return, will contribute to the balance of payments. Host countries which offer various fiscal incentives such as tax holidays and the institution of export-processing zones to lure foreign firms may be giving away income to the foreigners if the private rates of returns exceed the social rates, a conceivable outcome in the presence of market distortions which the incentives are supposed to rectify.

In sum the most significant benefit from FDI to the host countries is the technology and know-how it transfers. It is such transfers which augment the skill endowments of host countries and promote employment for local labour. How does the technology which is transferred get transmitted to the local economy? There are several channels, including imitation, acquisition of skills, competition and various sorts of tie-ups between the
foreign-owned and locally owned firms. Imitation of the products produced by foreign affiliates through reverse engineering, an activity that enables local firms to copy the processes and design of new products, is a recognized channel for spillovers. The acquisition of skills occurs mainly through the movement of skilled labour employed by foreign affiliates to locally owned firms. Such internal migration of labour, trained by foreign affiliates, is a significant channel for spillovers. Labour employed in foreign affiliates may wish to set up their own establishments with the experience and skills gained from their sojourn in the foreign affiliates. Also, foreign affiliates may, either in response to performance requirements imposed by the host country or because of distinct cost advantages, train or establish local suppliers of components and parts. This too would be a channel for spillovers.

Another potent channel for spillovers – or, more to the point, growth of productive efficiency – is competition. The theory here is that the entry of foreign affiliates increases competition in the marketplace and locally owned firms are compelled to increase their productive efficiency. This is the sort of efficiency recognized in the literature as ‘X-efficiency’ rather than allocative efficiency. Finally, locally owned firms may learn marketing techniques and methods of penetrating export markets from export-oriented foreign affiliates. This would count as a specific sort of technology transfer.

These propositions have been extensively tested in the context of FDI in developed and developing countries (Haddad and Harrison, 1993; Blomstrom and Kokko, 1998; Borensztein et al., 1998; Gorg and Greenaway, 2001). These econometric studies have produced a mixed bag of results: some identify positive spillovers from the presence of foreign affiliates in manufacturing industries, and others find them to be either negligible or negative.

These studies identify a number of factors that are likely to promote spillovers of technology and know-how from foreign affiliates to locally owned firms. First, the magnitude of spillovers tends to be high in industry segments in which the gap in technological capabilities between foreign affiliates and locally owned firms tend to be narrow. Second, spillovers are likely to be high when the competition in the marketplace between locally owned firms and foreign affiliates tends to be intense. Third, the extent and magnitude of spillovers differ between industries and host countries. Fourth, several studies show that spillovers are proportional to the magnitude of foreign presence, measured by shares of foreign affiliates in total equity or sales of the relevant industry groups. Fifth, local capabilities (including research and development – R&D – and human skills) sustain high levels of spillovers. Finally, analogous to the last proposition, the liberalization of foreign trade, increased competition and development of local infrastructure all promote spillovers.
The message of all this is clear. Increased volumes of FDI alone are unlikely to generate widespread spillovers. In the absence of competition and cooperant factors such as local R&D and human skills, spillovers from FDI may be limited. Put another way, FDI is a catalyst of technical change and growth; it cannot be expected to be the prime mover. Indeed empirical research suggests that FDI is most effective as an agent of change in economies that possess a threshold level of human capital and skills and in those economies that have attained a threshold level of growth (Blomstrom et al., 1994; Balasubramanyam et al., 1999).

In sum, in the absence of the necessary ingredients and cooperant factors, large volumes of FDI alone may not be effective in promoting growth and may even be counterproductive. For these reasons the exuberance relating to the role of FDI in the growth process and exhortations that developing countries should adopt a wide open door policy towards FDI should be tempered by a recognition of the conditions necessary for the effective utilization of FDI.

Conclusions

Recent literature on FDI reflects the substantial change in attitudes towards FDI by the developing countries. In the past, attitudes towards FDI and its role in the development process ranged to extremes – from hostility to ardent advocacy. In the recent past, there is a growing appreciation of its role in the development process and most developing countries have sought to attract increasing volumes of FDI. Research on FDI reflects this change in attitude. Much of the literature now is centred on econometric testing of the determinants and impact of FDI rather than the polemical debates on the role of FDI. Detailed case studies which provide analytical insights into the nature of FDI and its impact on development, of the sort done by Sanjaya Lall (1983), are unfortunately few and far between.

The message of the recent literature, though, is clear. First, FDI is attracted to countries that can provide the sort of environment which allows foreign firms to establish a foothold and successfully exploit the rents in the advantages they possess. The sort of environment which is conducive to the operations of foreign firms is characterized by a stable macro-economic environment with stable FDI regimes and an assured supply of cooperant factors including human capital. Second, FDI is a superb catalyst of development but not an initiator. The new-found enthusiasm for FDI on the part of developing countries is based on the success stories of the East Asian countries with FDI and the drying-up of alternative sources of finance such as bank credit. It is, though, worth noting that the successful utilization of FDI is contingent upon a number of factors discussed in the foregoing. In this context it is worth recalling Paul Streeten’s observa-
tion written during the 1970s (Streeten, 1971) when controversy on the role of the multinational enterprise as the purveyor of FDI was at its height: ‘it is not sensible to transfer income by attempting to transform the MPE [multinational production enterprise] from what it is – a profit seeking animal – into something it is not – a public service’.

**Note**

1. The author is grateful to MS Jasleen Sindhu for research assistance with this chapter.

**References**


Standard economic theory traditionally argued that international private capital flows will make a major contribution to development to the extent that they will flow from capital-abundant industrialized countries to capital-scarce developing countries, and help to smooth spending throughout the business cycle in capital-recipient countries.

In recent years, reality has contradicted both aspects (WESS, 2005). Between 1997 and 2004, developing countries have transferred a large amount of resources to developed countries. In addition to this, private capital flows to developing countries are highly concentrated in a group of large middle-income countries and are particularly insufficient for low-income and small countries. Secondly, private capital flows to developing countries have been highly volatile and reversible; as a consequence, they have been a major factor in causing developmentally costly currency and financial crises. Rather than smooth domestic expenditure, private capital flows seem to have contributed to making it more volatile.

Boom–bust cycles of capital flows have been particularly damaging for developing countries, when they both directly increase macroeconomic instability and reduce the room for manoeuvre to adopt countercyclical macroeconomic policies, and indeed generate strong biases towards adopting procyclical macroeconomic policies (Kaminsky et al., 2004; Stiglitz and Levy, 2005). Furthermore, there is now overwhelming evidence – accepted by institutions like the International Monetary Fund – that procyclical financial markets and procyclical macroeconomic policies have not encouraged growth and, on the contrary, have increased growth volatility in those developing countries that have integrated to a larger extent into international financial markets (Prasad et al., 2003).

The costs of financial volatility for economic growth are high, as it can generate cumulative effects on capital accumulation (Easterly, 2001). Indeed, major reversals of private flows have led to many developmentally and financially costly crises, which have lowered output and consumption well below what they would have been if those crises had not occurred. Eichengreen (2004) estimated that income of developing countries had been 25 per cent lower since 1980 than it would have been had such crises not occurred, with the average annual cost of the crises being just over $100 billion. Griffith-Jones and Gottshalk (2006) have estimated a similar
though somewhat higher annual average cost of crises in the period 1995–2002, of $150 billion in terms of lost gross domestic product (GDP).

These features are by no means inevitable. An appropriate domestic and international environment can improve the capacity of developing countries to benefit from private capital flows. In what follows we examine both characteristics of private capital flows to developing countries and some policy options that would improve their development impact.

Main characteristics of private flows
The volatility and reversibility of capital flows to emerging countries and the marginalization of many of the poorer and smaller developing economies with respect to financial markets are rooted in the combination of financial market failures and basic asymmetries in the world economy (Ocampo, 2001).

Instability is inherent in the functioning of financial markets (Keynes, 1936; Minsky, 1982). Indeed, boom–bust patterns in financial markets have occurred for centuries (Kindleberger, 1978). The basic reason for the existence of these patterns is that finance deals with future information that, by its very nature, is not known in advance; therefore, opinions and expectations about the future rather than factual information dominate financial market decisions. This is compounded by asymmetries of information that characterize financial markets (Stiglitz, 2000). Owing to the non-existence or the large asymmetries of information, financial agents rely to a large extent on the ‘information’ provided by the actions of other market agents, leading to interdependence in their behaviour, that is to say, contagion and herding. At the macroeconomic level, the contagion of opinions and expectations about future macroeconomic conditions tends to generate alternating phases of euphoria and panic. At a microeconomic level, it can result in either permanent or cyclical rationing of lending to market agents that are perceived by the market as risky borrowers. In many cases it is the endogenous behaviour of international financial markets that conditions or strongly influences fundamentals in developing countries. A supply-led, large capital inflow affects the domestic economic situation (for example by generating an asset price bubble or an overvalued exchange rate) in a way that can increase inflows. This can lead to costly macroeconomic crises, which makes regulation and other state intervention in international financial markets essential.

Herding and volatility seem to be accentuated by some features of the functioning of modern markets. An important element in the increased volatility of international bank lending is the use of modern risk management models (such as Value at Risk). As Persaud (2003) points out, the intrinsic problem with market-sensitive risk management systems is that
they incorrectly assume that banks act independently when in fact their decisions are interconnected. When many banks try to sell the same asset at the same time, and there are few or no buyers, prices fall and volatility increases. As prices collapse, for liquidity reasons banks try to sell another asset, which may have been previously uncorrelated with the first. This increases the volatility of the second asset and also correlation. This prompts repeated rounds of selling among agents who use similar models, and generalized herding takes place. The adoption of banks’ own risk management models to determine their required levels of capital in the internal ratings approach, as proposed in the new Basel Capital Accord, could seriously increase banks’ tendency for procyclicality in lending, exacerbating both booms and crashes.

An additional source of concern is the evidence that the Value at Risk (VaR) models first developed by banks are being extensively adopted by fund managers and pension funds, leading to similar herding patterns and to procyclicality in their investment. Therefore, herding is not restricted to one class of actor, but is spreading among many actors. The increasing use of similar market-sensitive risk management techniques (Persaud, 2000) and the dominance of investment managers aiming for very short-term profits, evaluated and paid at very short-term intervals (Griffith-Jones, 1998; Williamson, 2003), seem to have increased the frequency and depth of boom–bust cycles. The downgrade by a rating agency or any other new information available to investors may lead them to sell bonds and stop banks from lending to specific markets; simultaneously, reduced liquidity — owing, for example, to margin calls associated with derivative contracts in these markets — or contagion of opinions about the behaviour of different market segments that are believed to be correlated with a market facing a sell-off, will lead market agents to sell other assets or to stop lending to other markets. Through these and other mechanisms, contagion spreads both across countries and across different flows.

Different types of capital flows are subject, however, to different volatility patterns. In particular, the higher volatility of short-term capital indicates that reliance on such financing is highly risky (Rodrik and Velasco, 1999), whereas the smaller volatility of FDI vis-à-vis all forms of financial flows is considered a source of strength. However, even FDI does have volatile components. A particularly recent concern is that multinational companies, especially those selling in domestic markets, hedge their foreign exchange rate risk. This is particularly problematic when such hedging is done far more when a major devaluation is likely, as this will put additional pressure on the exchange rate and on the reserves. Naturally, such risks tend to become less important as national financial development deepens.
Capital account cycles involve short-term fluctuations, such as the very intense movements of spreads and interruption (rationing) of financing. These phenomena were observed during the Asian and, particularly, during the Russian crisis. However, and perhaps more importantly, they also involve medium-term fluctuations, as the experience since 1980 indicates. During those decades, the developing world experienced two such medium-term cycles that left strong imprints on the growth rates of many countries: a boom of external financing (mostly in the form of syndicated bank loans) in the 1970s, followed by a debt crisis in a large part of the developing world in the 1980s, and a new boom in the 1990s (now mostly portfolio flows), followed by a sharp reduction in net flows since the Asian crisis.

Improved economic conditions in developing countries, as well as the higher global growth, drove a recovery of private capital flows to developing countries in 2003, 2004 and 2005, perhaps signalling the beginning of a new cycle.

More importantly, net transfers of financial resources from developing countries have not experienced a positive turnaround and, on the contrary, continued to deteriorate in 2004 for the seventh year in a row, reaching an estimated $350 billion in 2004 (see Table 40.1). Periods of negative net transfers of financial resources from developing countries (especially from Latin America) have been frequent throughout history; indeed, Kregel (2004) provides evidence that these negative net transfers have been the rule rather than the exception.

Recently, these large and increasing net transfers of financial resources are explained by the combination of relatively low net financial flows and accumulation of very large foreign exchange reserves. Indeed, the most significant aspect of the net outflows from developing countries in recent years has been the growth in official reserves, particularly in Asia (Table 40.1). Accumulation of reserves initially had a large component of ‘self-insurance’ against financial instability, a rational decision of individual countries in the face of the limited ‘collective insurance’, often accompanied by what countries see as undesirable conditionality, provided by the international financial system. However, reserve accumulation in Asia has now clearly exceeded the need in several countries for self-insurance, raising increasing questions about the balance of costs and benefits of additional accumulation, especially if such reserves are invested in low-yielding assets and particularly in a currency, the United States dollar, that may at some point fall quite sharply. At a more fundamental level, the fact that countries like China and India, with very low levels of income per capita and large numbers of poor people – even though they have such dynamic growth – are transferring significant resources to finance developed countries, and especially the US, contradicts theory and is ethically undesirable.
### Table 40.1  Net transfer of financial resources to developing countries and economies in transition, 1993–2004 (US$40 billion)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing countries</td>
<td>69.3</td>
<td>35.8</td>
<td>42.9</td>
<td>19.9</td>
<td>-5.2</td>
<td>-37.9</td>
<td>-127.4</td>
<td>-186.5</td>
<td>-153.7</td>
<td>-205.5</td>
<td>-274.8</td>
<td>-353.8</td>
</tr>
<tr>
<td>Africa</td>
<td>1.1</td>
<td>4.0</td>
<td>6.4</td>
<td>-5.8</td>
<td>-4.7</td>
<td>15.6</td>
<td>4.3</td>
<td>-26.2</td>
<td>-14.7</td>
<td>-5.6</td>
<td>-20.2</td>
<td>-32.8</td>
</tr>
<tr>
<td>Sub-Saharan</td>
<td>8.6</td>
<td>6.7</td>
<td>7.4</td>
<td>5.3</td>
<td>7.5</td>
<td>12.1</td>
<td>9.1</td>
<td>3.0</td>
<td>7.9</td>
<td>6.4</td>
<td>6.5</td>
<td>3.9</td>
</tr>
<tr>
<td>(excluding Nigeria and South Africa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern and Southern Asia</td>
<td>18.7</td>
<td>1.0</td>
<td>22.1</td>
<td>18.5</td>
<td>-31.1</td>
<td>-128.2</td>
<td>-142.7</td>
<td>-121.3</td>
<td>-113.1</td>
<td>-142.1</td>
<td>-147.5</td>
<td>-167.8</td>
</tr>
<tr>
<td>Western Asia</td>
<td>33.1</td>
<td>13.2</td>
<td>15.6</td>
<td>5.3</td>
<td>6.2</td>
<td>28.5</td>
<td>-0.9</td>
<td>39.1</td>
<td>32.0</td>
<td>26.7</td>
<td>47.6</td>
<td>79.9</td>
</tr>
<tr>
<td>Latin America</td>
<td>16.4</td>
<td>17.7</td>
<td>-1.2</td>
<td>1.8</td>
<td>24.5</td>
<td>46.2</td>
<td>11.8</td>
<td>-31.1</td>
<td>6.1</td>
<td>-31.1</td>
<td>59.5</td>
<td>73.4</td>
</tr>
<tr>
<td>Economies in transition</td>
<td>1.8</td>
<td>-3.9</td>
<td>-2.3</td>
<td>-6.2</td>
<td>2.7</td>
<td>3.0</td>
<td>-24.0</td>
<td>-48.8</td>
<td>-30.5</td>
<td>-27.0</td>
<td>-34.4</td>
<td>-57.6</td>
</tr>
<tr>
<td>Memorandum item:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavily indebted poor countries (HIPC)</td>
<td>8.5</td>
<td>7.1</td>
<td>6.3</td>
<td>6.8</td>
<td>7.1</td>
<td>8.6</td>
<td>10.1</td>
<td>8.8</td>
<td>8.8</td>
<td>9.9</td>
<td>10.6</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Sources: UN/DESA, WESS (2005).
As private flows recover, an important question for policy-makers in developing countries is whether they will be sufficient and particularly more stable and less reversible than in the past. In this regard, the dominant role of FDI and the fact that it has been relatively stable in times of crises, are positive. However, not all components of FDI are equally stable. Furthermore, multinational companies, especially those producing for the local market, increasingly hedge their short-term foreign exchange risks, particularly when devaluations seem likely. This can lead to major temporary outflows of capital and significant pressure on exchange rates (Ffrench-Davis and Griffith-Jones, 2003; Persaud, 2003). More generally, the increasing use of financial engineering and of derivatives (as well as the growing scale and complexity of derivatives discussed below) seems to make the hypothesis of a hierarchy of volatility, whereby some categories of flows are more stable than others, less clear-cut.

Another potentially positive effect is the greater interest shown by institutional investors (such as life insurers) in investing in emerging countries (European Central Bank, 2005). However, the large rise in ‘carry trade’ – that is to say, investment in high-yielding emerging-market instruments using debt raised at lower cost in mature markets – makes those flows vulnerable to narrowing of interest rate differentials. Furthermore, the large fall in emerging countries’ bond spreads during 2004–05 (while naturally positive in itself for borrowing countries) has raised concerns that this increases the vulnerability of developing countries to international changes.

Finally, there are two structural trends that may add stability. The first is attested by the greater importance of local currency bond markets in developing countries; the second by the fact that international banks have increasingly ‘crossed the border’, lending from their local branches in local currency, and usually fund themselves via domestic deposits. This makes countries less vulnerable to crises, although it also implies that foreign banks are contributing less – or no – foreign savings.

At the time of writing in 2005 there were thus mixed signs in respect of whether the new inflows will be more stable than in the past. Therefore, policy efforts must be made, both in source and in recipient countries, to encourage more stable flows and discourage large flows that are potentially more reversible.

**Measures to counter procyclicality of private flows**

To counter the boom–bust pattern that characterizes private capital flows, several options are available. We here consider two: (1) designing mechanisms to encourage more stable private flows (countercyclical guarantees) or that distribute better the risk faced by developing countries throughout the business cycle (indexed bonds and bonds denominated in the currency...
of developing countries); and (2) introducing prudential regulations on the capital account. We also consider the likely effect of the New Basel Capital Accord (Basel II) on patterns of capital flows to developing countries. The procyclical pattern of private capital flows gives a compensatory role also to official financing, in relation to official development financing and to emergency (balance-of-payments) financing, respectively (for a discussion of these, see for example Griffith-Jones and Ocampo, 2003 and WESS, 2005, Chapters III and VI).

**Countercyclical financing instruments**

One way of addressing the problems created by the inherent tendency of private flows to be procyclical is for public institutions to issue guarantees that have countercyclical elements (Griffith-Jones and Fuzzo de Lima, 2004). In this regard, multilateral development banks and export credit agencies could introduce explicit countercyclical elements in the risk evaluations they make for issuing guarantees for lending to developing countries. This would imply that when banks or other private lenders lowered their exposure to a country, multilateral development banks or export credit agencies would increase their level of guarantees, if they considered that the country's long-term fundamentals were basically sound. When private banks' willingness to lend increased, multilateral development banks or export credit agencies could reduce their exposure.

There have also been proposals to introduce GDP-indexed bonds. The coupon payments on these bonds would vary in part with the growth rate of the debtor's economy, being higher in years of rapid growth of GDP (measured in an international currency) and lower in years of below-trend growth. It has been argued that such instruments would improve the cushioning of emerging-market borrowers against adverse shocks by making debt payments more contingent on the borrower's ability to pay. GDP-indexed bonds would therefore restrict the range of variation of the debt-to-GDP ratio and hence reduce the likelihood of debt crises and defaults. At the same time, they would also reduce the likelihood of procyclical fiscal policy responses to adverse shocks (Griffith-Jones and Sharma, 2006).

Another alternative for better managing the risks faced by developing countries throughout the business cycle consists in the introduction of local currency-denominated bonds. These bonds offer, in particular, a cure against the currency mismatches that characterize the debt structure of developing countries. At the domestic level, the development of domestic capital markets, especially bond markets, also creates a more stable source of local funding for both the public and private sectors, thereby mitigating the funding difficulties created by sudden stops in cross-border capital. In addition to proposals for institutional measures to develop local capital
markets, there have also been innovative proposals to make local currency investments more attractive to international investors. Spiegel and Dodd (2004) have suggested raising capital in international markets by forming diversified portfolios of emerging-market local currency debt issued by sovereign governments. These portfolios of local currency government debt securities would employ risk management techniques of diversification to generate a return-to-risk that competed favourably with other major capital market security indices.

**Prudential capital account regulations**

The accumulation of risks that developing countries face during capital account booms depends not only on the magnitude of private and public sector debts but also on maturity and currency mismatches on the balance sheets of financial and non-financial agents. Thus, capital account regulations potentially have a dual role: as a macroeconomic policy tool with which to provide some room for countercyclical monetary policies that smooth out debt ratios and spending; and as a ‘liability policy’ designed to improve private sector external debt profiles (Ocampo, 2003).

Overall, the experiences with capital account regulations in the 1990s were useful for improving debt profiles, giving governments more latitude in pursuing stabilizing macroeconomic policies, and insulating countries from some of the vagaries of capital markets. There is much evidence that, if well implemented, the benefits far outweigh the costs (Stiglitz and Levy, 2005; Ocampo and Palma, 2005).

One type of capital account regulations are price-based regulations. The basic advantages of price-based instruments are their simplicity and their focus on averting the build-up of macroeconomic disequilibria and, ultimately, preventing crises. A highly significant innovation in this sphere during the 1990s was the establishment in Chile and Colombia of an unrenumerated reserve requirement for capital inflows.

It is noteworthy that institutions such as the International Monetary Fund and the Bank for International Settlements have increasingly concluded that these controls were effective in important aspects. There is broad agreement that they were effective in reducing short-term debt flows and thus in improving or maintaining good external debt profiles. There is greater controversy about their effectiveness as a macroeconomic policy tool. Nonetheless, it can be asserted that reserve requirements helped countries maintain higher domestic interest rates during periods of euphoria in international financial markets.

On the other hand, quantity-based capital account regulations might be preferable when the policy objective is to reduce significantly domestic macroeconomic sensitivity to international capital flows.
The experience of the Asian countries that maintained quantity-based restrictions throughout the 1990s suggests that those restrictions might indeed also be particularly effective in preventing crises. China, India, Taiwan Province of the Republic of China and Vietnam offer successful examples in this regard. Despite the slow and cautious liberalization that has taken place in several of these economies since the early 1990s, the use of such traditional regulations has helped them prevent contagion from the East Asian crisis (see for example, in relation to India, Reddy, 2000).

Malaysia offers an interesting example of the effective use of quantitative regulations during the 1990s. Kaplan and Rodrik (2001) and others provide evidence that Malaysian regulations during the Asian crisis gave the government space within which to enact expansionary monetary and fiscal policies that contributed to the speedy recovery of economic activity.

Although quantity-based restrictions can be effective if authorities wish to limit capital outflows during crises, crisis-driven quantitative controls generate serious credibility issues and may be ineffective in the absence of a strong administrative capacity. A tradition of regulation may be necessary, and the tightening or loosening of permanent regulatory regimes through the cycle may be superior to the alternation of different (even opposite) capital account regimes.

It should be emphasized that capital account regulations should always be seen as an instrument that provides an additional degree of freedom to the authorities with respect to their adopting sensible counter-cyclical macroeconomic policies, but never as a substitute for those policies.

**Basel II and developing countries**

The right regulatory and supervisory regime is essential for maintaining domestic financial stability. In a globalized economy, some common standards of regulation and supervision may be also essential to guarantee global financial stability. This has been the major motivation behind the principles adopted by the Basel Committee on Banking Supervision in recent decades. The second generation of these standards (Basel II), agreed to in June 2004, takes a further step in aligning regulatory capital with the risks in bank lending, and in adapting regulations to the complexities of risk management.

There are fears that Basel II creates the risk of a sharp reduction in bank lending to developing countries, and of an increase in the cost of a significant part of the remaining lending, particularly in the case of low-rated borrowing countries. An equal cause for concern is the danger that Basel II will accentuate the procyclicality of bank lending, which is damaging for all economies, but particularly so for fragile developing ones, which are more vulnerable to strong cyclical fluctuations of financing.
Indeed, the proposed internal ratings-based (IRB) approach of Basel II overestimates the risk of international bank lending to developing countries, primarily because it does not appropriately reflect the clear benefits of international diversification. However, there is a great deal of evidence that by failing to take account of the benefits of international diversification at the portfolio level, capital requirements for loans to developing countries will be significantly higher than is justified on the basis of the actual risks attached to this lending (see, for example, Griffith-Jones et al., 2003).

Therefore, one clear way in which Basle II could be improved so as to reduce the negative and technically incorrect effects on developing countries would be to introduce the benefits of diversification into the internal ratings-based approach. One of the major benefits of investing in developing and emerging economies is their relatively low correlation with mature markets. This hypothesis was tested empirically using a wide variety of financial, market and macro variables (Griffith-Jones et al., 2004a). Every statistical test performed showed that the correlation between developed markets only was higher, in every case, than that between developed and developing markets.

An additional positive effect of taking account of the benefits of diversification is that this makes capital requirements far less procyclical than they otherwise would be. Indeed, if the benefits of diversification are incorporated, simulations show that the variance over time of capital requirements will be significantly smaller than if these benefits are not incorporated. Therefore, introducing the benefits of geographical diversification significantly decreases, though it does not eliminate, the higher procyclicality that the internal ratings-based approach implies. This difference may well be significant enough to prevent a ‘credit crunch’.

However, even if the benefits of diversification are incorporated, the internal ratings-based approach will still be more procyclical than the standardized approach, which is closer to the principles of the first Basel Capital Accord (Basel I). Therefore, as well as introducing the benefits of diversification, it seems desirable to introduce countercyclical measures (for example, countercyclical provisioning against losses) at the same time as Basel II is implemented.

We can conclude that several measures can be taken to reduce boom–bust patterns of private flows, and thus enhance their contribution to development. However, given that the risk of costly crises will remain – even if such measures are introduced – maintaining and improving the supply of countercyclical official liquidity and development finance is essential.

Acknowledgement
The author would like to acknowledge the excellent assistance by Carmen Seekatz.
References


Ocampo, José Antonio (2003), *Capital-Account and Counter-Cyclical Prudential Regulations in Developing Countries*, United Nations Publications.


Prasad, E., K. Rogoff, S. Wei and M. Kose (2003), *Effects of Financial Globalization on Developing Countries*, International Monetary Fund, Washington, DC.


UN’s World Economic and Social Survey, 2005.

International capital flows to emerging economies: short- and long-run effects

Graciela L. Kaminsky

Introduction
The explosion of capital flows to emerging markets in the early and mid-1990s and their reversal following the crises in Asia, Latin America, and the transition economies have reignited a heated debate on the benefits and drawbacks of financial globalization. Many have argued that globalization has gone too far and that international capital markets have become extremely erratic, with ‘excessive’ booms and busts in capital flows triggering bubbles and financial crises and magnifying the business cycle. In contrast, the traditional view asserts that international capital markets enhance growth and productivity by allowing capital to flow to its most attractive destination.

Even if international capital flows do not trigger excess volatility in domestic financial markets, it is still true that large capital inflows can spark off inflation in the presence of a fixed exchange rate regime. Moreover, transitory capital inflows may distort relative prices, with the domestic economy losing competitiveness as a result of the appreciation of the real exchange rate. Therefore, it is no wonder that policy-makers have used a variety of tools to manage these flows, especially flows of the ‘hot money’ type.

This chapter re-examines the evidence on the characteristics of international capital flows to emerging economies, with particular attention to portfolio equity flows and bank lending around the time of the crises of the 1990s. The results suggest that episodes of surge in capital inflows do, in fact, end abruptly – whether owing to home-grown problems or contagion from abroad. This chapter also reviews the evidence on the short- and long-run effects of financial deregulation on financial and real cycles. Interestingly, the stylized evidence suggests that although financial liberalization may trigger excessive booms and busts in the short run, financial markets tend to stabilize and growth accelerates in the long run, in part because financial globalization seems to trigger institutional reform. The conclusion summarizes what we know about financial globalization and examines policy options.
The behavior of mutual funds

The booms and busts in international capital flows have brought international investors into the limelight. In this section, I examine the behavior of equity mutual funds in emerging markets. I use the information on portfolio allocations provided by Emerging Market Funds Research, Inc., which covers the positions of nearly 1400 international emerging market equity funds, with an average position of about US$120 billion in 1996. It includes United States registered and offshore funds as well as funds registered in Luxembourg, the United Kingdom of Great Britain and Northern Ireland, Ireland, the Cayman Islands, Canada and Switzerland. Both open-end and closed-end funds are also included in this data set, which starts at 1995.

Figure 41.1 shows the average quarterly net flows to these regions from 1995 to 1999. Mutual fund flows to emerging markets peaked in the second quarter of 1997, reaching about US$8 billion. Overall, booms in mutual fund flows were followed by reversals. Reversals were not persistent after the ‘Tequila crisis’. Outflows from Latin America reached about US$4 billion in 1995, but mutual funds increased their positions in Latin America by about US$2 billion in the first half of 1996. The Tequila crisis did not have any spillovers in Asia or in transition economies. In fact, flows to Asia ballooned to almost US$11 billion in 1996, while flows to transition economies remained stable throughout 1995–96. The picture changed after the Asian crisis. This time, mutual funds pulled out not only from Asia but also from Latin America, with net outflows in the latter region reaching about US$1 billion in the six months following the collapse of the Thai baht. Mutual fund withdrawals took a turn for the worse in 1998, reaching about US$4 billion in Asia and also in Latin America, with substantial outflows from transition economies after the Russian crisis.

Figure 41.2 assesses the problem of the sudden stops in times of financial turmoil. It reports the average quarterly flows (as a percentage of the mutual funds’ initial positions) to countries in Asia and Latin America, as well as to transition economies in the two quarters following three crises. The top panel looks at the aftermath of the Mexican devaluation in December 1994, the middle panel examines the aftermath of the collapse of the Thai baht in July 1997, and the bottom panel studies the aftermath of the Russian devaluation and moratorium in August 1998. To capture the magnitude of the sudden-stop syndrome, this figure reports total flows relative to average flows (also as percentages of their initial positions) during the whole sample (1995–99). Following the Mexican devaluation, for example, mutual funds sold about 5 percent of their Brazilian positions (relative to their average quarterly buying/selling from 1995 to 1999). Thus, as shown in the first panel in Figure 41.2, Brazil experienced unusual withdrawals of about 5 percent in the aftermath of the Mexican devaluation. As shown in the last
Notes:
Latin America includes Argentina, Brazil, Chile, Colombia, Mexico, Peru and Venezuela.
Asia includes China, Hong Kong, India, Indonesia, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Sri Lanka, Taiwan and Thailand.
Transition economies include Armenia, Azerbaijan, Belarus, the Czech Republic, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Moldova, Poland, Russia, Slovakia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

Source: Kaminsky et al. (2002).

Figure 41.1 Mutual funds: quarterly flows to emerging countries (billions of dollars)
Figure 41.2 Mutual fund flows: global spillovers

Source: Kaminsky et al. (2002).
panel, Malaysia was the country most affected in the aftermath of the Russian crisis, with abnormal outflows of approximately 30 percent.

The extent of the mutual fund sudden stop in the aftermath of the three crises was substantially different. The so-called Tequila crisis was circumscribed to Latin America. Moreover, ‘abnormal’ mutual fund withdrawals in the aftermath of the collapse of the Mexican peso were confined to a handful of Latin American countries, with only Brazil and the Bolivarian Republic of Venezuela – besides the crisis country, Mexico – suffering average withdrawals of 5 and 2 percent, respectively, in the two quarters following the devaluation. In contrast, mutual funds increased their exposure to Asian countries and transition economies, with (above-trend) flows oscillating around 4 percent for Asia and 11 percent for the transition economies.

The aftermath of the collapse of the Thai baht presents a different picture of the international mutual funds industry. It is in this episode that we first observe signs of a more general retrenchment of mutual funds in emerging markets. Mutual fund flows to Asian economies were well below trend in the two quarters following the collapse of the Thai baht. Only flows to China, Pakistan and Sri Lanka were above average. Interestingly, after the collapse of the Thai baht, we observe substantial withdrawals from Hong Kong Special Administrative Region, Singapore and Taiwan Province of the People’s Republic of China, with average quarterly withdrawals oscillating at about 12 percent above average in the case of Singapore and Taiwan and about 7 percent for Hong Kong. The retrenchment this time also affected Latin America and the transition economies, with withdrawals reaching about 6 percent for Colombia and 4 percent for the Czech Republic during the two quarters following the outbreak of the Thai crisis. Colombia, the Czech Republic, Chile, Hungary and Peru were the countries most affected in this episode, with sales averaging about 3 percent above average.

The flight away from emerging markets became more pronounced during the Russian crisis, with about half of the countries in the sample experiencing abnormal sales of about 10 percent or even larger. In some cases, withdrawals were massive. For example, average mutual funds sales (relative to trend) in Malaysia reached 30 percent while those in the Czech Republic were in the order of 16 percent. Some Latin American countries were also dramatically affected in the aftermath of the Russian collapse. Colombia and Venezuela, for example, suffered average quarterly outflows of about 8 percent. Mutual funds investments in Mexico and Peru were the only ones that did not suffer following the worldwide turmoil triggered by the Russian default. In fact, inflows to Mexico were 5 percent above the average observed in the 1995–99 period.

Table 41.1 examines in detail why some countries were severely affected by mutual fund withdrawals while others were left unscathed. Three factors
Table 41.1 The behavior of mutual funds during crises

<table>
<thead>
<tr>
<th>Region</th>
<th>Liquid financial markets</th>
<th>Fragility</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>0</td>
<td>42</td>
<td>25</td>
</tr>
<tr>
<td>Latin America</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>67</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>0</td>
<td>67</td>
<td>33</td>
</tr>
<tr>
<td>Transition economies</td>
<td>..</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>33</td>
<td>75</td>
<td>50</td>
</tr>
</tbody>
</table>

The Thai crisis

<table>
<thead>
<tr>
<th>Region</th>
<th>Liquid financial markets</th>
<th>Fragility</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>43</td>
<td>86</td>
<td>29</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>25</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Latin America</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>75</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Transition economies</td>
<td>100</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>0</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

The Russian crisis

<table>
<thead>
<tr>
<th>Region</th>
<th>Liquid financial markets</th>
<th>Fragility</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>40</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Latin America</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>50</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>20</td>
<td>60</td>
<td>0</td>
</tr>
</tbody>
</table>
are examined: economic fragility, liquidity of financial markets and economic and political risk. Economic fragility is captured using the probabilities of crises in Kaminsky (1998) that measure the likelihood of crises conditional on 18 indicators reflecting macroeconomic vulnerabilities in each country. These indicators provide information about fiscal and monetary imbalances, financial and real vulnerabilities, current account and capital account problems, and world factors. For Table 41.1, I classify an economy as fragile if the probability of a crisis is higher than 50 percent; otherwise it is considered healthy.

Liquidity is captured using four indicators. The first one – the volume traded in the stock market – provides an overall measure of the size and depth of the stock market. The second one – the share of the mutual funds portfolio in each country at the onset of the crisis – is related to mutual funds liquidity in each country, since investors cannot sell in countries in which they have basically no exposure. These first two indicators provide different pictures of liquidity of financial markets. The third indicator dates the time when firms in emerging markets start to trade in mature and more liquid financial markets. The fourth indicator captures the ability of investors to change their portfolio rapidly in a particular country. In particular, this last indicator evaluates the extent of restrictions to capital mobility in each country. Restrictions could add ‘sand in the wheels’ of capital markets and thus curtail liquidity.

Finally, the risk indicator captures both political and economic uncertainty. The political risk indicator captures uncertainty due to expected changes of authorities or future policy actions, and it also identifies

Table 41.1 (continued)

<table>
<thead>
<tr>
<th>Region</th>
<th>Fragility</th>
<th>Liquid financial markets</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition economies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>50</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: This table relates the mutual fund (MF) withdrawals (injections) of funds to the emerging markets shown in Figure 41.4 with indicators of fragility, liquidity of financial markets, and economic and political risk in those economies. Data not available.

Source: Kaminsky et al. (2002).
widespread social unrest. In particular, it includes major changes in the political arena or events of political instability that took place six months before and after the crisis. The risk indicator also captures economic risk, such as imposition of restrictions to capital mobility in response to crises. A country is classified as risky when there is at least either political or economic risk.

Table 41.1 shows the characteristics of countries that suffer abnormal withdrawals and injections in the aftermath of the three crises. The table groups the countries into three regions: Asia, Latin America and transition economies. As shown in the first column, countries with fragile economies constitute the bulk of the countries that suffer withdrawals. During the Mexican crisis, for example, Latin America was the only region that suffered withdrawals. Interestingly, 67 percent of the countries that suffered withdrawals in this episode were also countries with deteriorated fundamentals. Again, during the Thai crisis, at least 75 percent of the countries that suffered withdrawals in the transition economies group and Latin America were countries with economic vulnerabilities. Similarly, 43 percent of the Asian countries affected by abnormal withdrawals also had deteriorated economies. The Republic of Korea (South Korea), Colombia, the Czech Republic and Chile, for example, suffered huge withdrawals in the aftermath of the Thai crisis – the Czech Republic and South Korea were the two most vulnerable countries during the Asian crisis (Thailand ranked fourth) in the sample of 25 countries, while Colombia ranked sixth. In contrast, countries that did not experience mutual fund withdrawals were less fragile in general (see Goldstein et al., 2000).

Domestic fragilities were not the only explanation for the sudden-stop syndrome, however. China, for example, did not even suffer a mild hiccup in the midst of the Asian crisis, even when devaluation fears were widespread among investors and the vulnerability of its financial system was widely known. In contrast, Singapore, Taiwan and Hong Kong – countries with the most liquid financial markets in the region – suffered pronounced capital flow reversals even when their economies looked far healthier than that of China. Overall, 86 percent of the countries in the Asia-Pacific region that suffered withdrawals were countries with quite liquid financial markets. In contrast, all the countries in that region unaffected by the Thai crisis had illiquid financial markets.

Finally, risk also had an important role, with 40 percent of the countries most affected by withdrawals also experiencing political and economic risk. In 1994, for example, in the midst of the banking crisis, Venezuela abandoned convertibility. Far from discouraging capital outflows, the implementation of restrictions to capital mobility seems to have also contributed to the fire sales of Venezuelan assets. Similarly, Malaysia suffered...
substantial losses in the aftermath of the Russian crisis when it introduced outright controls on capital outflows. Interestingly, the withdrawals may have been triggered by the increased risk – perceived or real – associated with the country.

The behavior of banks
Bank-related lending has also been quite volatile since the late 1970s. This section examines the role of European, Japanese and United States banks in spreading the crises of the 1990s. The Bank for International Settlements (BIS) Consolidated Banking Statistics database is used to examine the role of the three international banking clusters. In particular, international claims of reporting BIS banks in emerging economies, including both total cross-border claims and local claims in foreign currency booked by foreign offices, are studied.

As shown in Figure 41.3, bank flows poured into Asia throughout most of the 1990s and accelerated following the Mexican crisis. Bank loans to emerging Asia expanded by 89 percent from June 1994 to June 1997. Part of the rise in lending was due to the European banks’ goal of achieving a higher profile in emerging markets, particularly in South Korea. Much of the lending boom, especially in the case of Thailand, Indonesia and South Korea, was due to a rapid expansion in credit from Japanese banks. Faced with a slumping economy and little domestic loan demand, Japanese banks increasingly looked overseas to the rapidly growing economies of South-East Asia as potential borrowers. United States bank lending to Asia was modest before the crisis. By June 1997, the United States banks’ positions in emerging Asia had only reached US$32 billion and only accounted for 20 percent of all United States bank lending to developing countries. In contrast, by the onset of the Thai crisis, Japanese banks had exposure to Asia four times as much as United States banks (US$124 billion). European bank lending to emerging Asia was also significant and, by the onset of the Thai crisis, the exposure of European banks to Asia surpassed that of Japanese banks, reaching US$161 billion. The exposure of European banks to emerging Asia accounted for about a half of all their lending to emerging markets; South Korea alone accounted for 40 percent of their lending to the developing world.

Japanese banks, heavily exposed to Thailand, were the first to pull out of emerging Asia. Between June and December of 1997, lending by Japanese banks fell by 8 percent. European banks, heavily exposed to South Korea, only began to pull out following the start of the crisis in that country in November 1997. In net terms, European bank lending to Asia continued to increase from June to December 1997. By June 1998, however, lending to emerging Asia was reduced across the board. Bank lending to Asia fell by
Notes:
Asia includes Afghanistan, Armenia, Azerbaijan, Bangladesh, Bhutan, British Overseas Territories, Brunei, Cambodia, China, Fiji, French Polynesia, Georgia, India, Indonesia, Kazakhstan, Kiribati, Kyrgyzstan, Laos, Macau, Malaysia, the Maldives, Mongolia, Myanmar, Nauru, Nepal, New Caledonia, North Korea, Pakistan, Papua New Guinea, the Philippines, the Solomon Islands, South Korea, Sri Lanka, Taiwan, Tajikistan, Thailand, Tonga, Turkmenistan, Tuvalu, US Pacific Islands, Uzbekistan, Vietnam, Wallis Futuna and Western Samoa.

Latin America includes Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, the Dominican Republic, Ecuador, El Salvador, the Falkland Islands, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Paraguay, Peru, St Lucia, St Vincent, Suriname, Trinidad and Tobago, Turks and Caicos, Uruguay and Venezuela.

Transition economies include Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, the Czech Republic, Czechoslovakia, Estonia, the German Democratic Republic, Hungary, Latvia, Lithuania, Macedonia, Malta, Moldova, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, the Soviet Union, Turkey and Ukraine.

Source:  Bank for International Settlements.

Figure 41.3  Bank lending: European banks, Japanese banks, US banks
US$46 billion, with European banks recalling US$12 billion, Japanese banks US$25 billion and United States banks US$9 billion, respectively.

Figure 41.3 also reports bank lending to Latin America and transition economies. Exposure to these regions increased sharply in the mid-1990s (in large part driven by the purchase of domestic banks by European banks), with claims on these regions increasing by about 50 percent from June 1994 to June 1998, immediately before the onset of the Russian crisis. During the 1990s, European banks had the largest exposure to these regions – accounting for 67 percent to Latin America and 84 percent to transition economies. The Russian crisis led to some withdrawals of Japanese and United States lending from both regions, but this was not the case with European banks that had acquired local banks. Total exposure to Latin America by European banks peaked in December 2000.

Figures 41.4 to 41.6 tally country-by-country bank flows originating in European, Japanese and United States banks in the aftermath of the Mexican, Thai and Russian crises. Each figure focuses on the year following the crisis. Figure 41.4 shows that, with the exception of Mexico and Venezuela (which had a banking crisis of its own making), Latin American countries did not suffer major reversals in bank lending following the Mexican crisis. Moreover, within a year of the crisis, lending to Latin America recovered and even surpassed the levels observed before the crisis. Brazil was the prime beneficiary of bank flows during 1995, with lending from European and United States banks reaching US$15 billion. Even in the case of Mexico and Venezuela, withdrawals were not made across the board. Only United States banks recalled loans from these countries. Figure 41.4 also shows that in Asia, the major recipients of capital flows in 1995 were South Korea, Thailand and Indonesia.

Figure 41.5 shows the behavior of bank lending in the aftermath of the Thai crisis. In contrast to the Tequila crisis, the Thai crisis triggered major reversals in bank flows from banks in Europe, Japan and the United States. Thailand, South Korea, Indonesia and Malaysia were the countries that suffered major withdrawals. Contagion was only regional in nature, with almost all of the Latin American countries, and to a lesser degree transition economies, continuing to have uninterrupted access to bank lending.

Figure 41.6 shows contagion from the Russian crisis. As was the case with mutual funds, the reversal in bank lending following the Russian default was not restricted to the Russian Federation or neighbouring countries. This time, the reversal was more widespread, and affected countries as far away as Brazil and South Africa. While Japanese banks continued to recall loans from Thailand, Indonesia and South Korea, reversals were not just restricted to these countries. Japanese banks, as well as United States banks, also recalled loans from Brazil, Mexico, India and South Africa.
**Figure 41.4** Bank flows: global spillovers – after the Mexican crisis: December 1994–December 1995

**Source:** Bank for International Settlements.
**Source:** Bank for International Settlements.

**Figure 41.5** Bank flows: global spillovers – after the Thai crisis: June 1997–June 1998
Source: Bank for International Settlements.

Figure 41.6 Bank flows: global spillovers – after the Russian crisis: June 1998–June 1999
More formal evidence suggests that international banks were at the centre of financial contagion in the late 1990s. For example, Kaminsky and Reinhart (2000) examine contagion during the debt crisis in 1982, the Mexican crisis in 1994 and the Asian crisis in 1997, and find that United States banks were at the core of the contagion during the debt crisis, while Japanese banks spread the Thai crisis to Indonesia, South Korea and Malaysia. Van Rijckeghem and Weder (2003) examine the Tequila, Asian and Russian crises and the flows to 31 emerging countries from 11 creditor countries using BIS banks. Their evidence supports the idea that the degree to which countries compete for funds from common bank lenders is a fairly robust predictor of the incidence of contagion. Finally, Caramazza et al. (2000) extend earlier work on indicators of vulnerability to currency crises by examining the role of financial linkages, while controlling for the roles of internal and external macroeconomic imbalances and trade spillovers. Their results indicate that financial links do matter while exchange rate regimes and controls on capital flows do not seem to.

Globalization and volatility
As discussed in the introduction, the views on the effects of financial globalization have been diverse; there are those who defend capital controls (Rodrik, 1998; Stiglitz, 1999) and those who maintain that capital should be allowed to move freely (Dornbusch, 1998). The rationale for restricting international capital flows is grounded in the belief that market failures and distortions pervade capital markets around the world. One of the most frequently cited distortions is that of asymmetric information, which is rampant in international capital markets due to geographical and cultural differences that complicate the task of obtaining information. In addition, imperfections in international markets are magnified by the difficulties in enforcing contracts across borders. With imperfect information, investors may overreact to shocks, withdrawing massively from countries at the first signs of economic problems, or become euphoric and pour in capital in quantities beyond those justified by ‘good’ fundamentals. On the other hand, those who consider international capital markets to be efficient favor unrestricted capital movements. Financial liberalization is believed to improve the functioning of financial systems, increasing the availability of funds and allowing cross-country risk diversification. Moreover, it is also claimed that financial integration tends to facilitate economic growth.

This section will summarize some of the findings in the literature on the effects of globalization, paying particular attention to the short- and long-run effects of financial integration on real and financial volatility.
Financial markets

The evidence from the crises of the 1990s suggests that crises are preceded by ‘excessive’ capital inflows that, in turn, fuel large expansions in domestic credit and bubbles in financial markets (see, for example, Sachs et al., 1996). There is also evidence that most episodes of banking crises are preceded by financial liberalization (see, for example, Kaminsky and Reinhart, 1999; Demirguc-Kunt and Detragiache, 1999). To reconcile the evidence that globalization is at the heart of financial crises with the hypothesis that international capital markets allow capital to move to its most attractive destination and promote more stable financial markets, I examine the possible time-varying effects of financial liberalization on stock market price cycles.

Figure 41.7 shows the average amplitude of booms and crashes in stock prices for 14 emerging markets during periods of repression, in the immediate aftermath of liberalization (the four years following liberalization), and in the long run. The evidence in this figure seems to point to excessive cycles, with larger booms followed by larger crashes in the immediate aftermath of financial liberalization. However, liberalization does not permanently bring about more volatile financial markets. If liberalization persists, stock markets in emerging countries become more stable. Kaminsky and Schmukler (2003) argue that these conflicting effects arise because during episodes of financial repression, banks are protected from outside competition and do not have the pressure to run efficiently. Liberalization in this

![Graph showing average amplitude of booms and crashes in 14 emerging markets](image)

**Note:** Gray: repression; Black: short-run liberalization; White: long-run liberalization.

**Source:** Kaminsky and Schmukler (2003).

**Figure 41.7** Average amplitude of booms and crashes in stock prices in 14 emerging markets (in percentage points)
scenario unveils a new problem, as protected domestic banks suddenly get access to new sources of funding, triggering protracted financial booms. But financial liberalization triggers reforms and better-functioning financial markets as domestic investors, now with access to international capital markets, demand better enforcement rules to continue to invest in domestic financial markets. Moreover, as pointed out by Stulz (1999), the liberalization and gradual integration of emerging markets into international financial markets may help strengthen the domestic financial sector, as foreign investors generally have better skills and more information and can thus monitor management in ways that local investors cannot. Liberalization also allows firms to access mature capital markets. Firms listed on foreign stock markets are in the jurisdiction of a superior legal system with higher disclosure standards that will promote more transparency in the management of the firm and can trigger improvements in corporate governance.

**Business cycles and growth**

The evidence in the previous section is suggestive of excessive booms and busts in financial markets in developing countries following globalization, but of more stable financial markets in the long run if globalization persists. This section will examine the relationship between globalization and business cycle fluctuations and growth.

Figure 41.8 shows international capital flows to emerging markets in Asia, Latin America and transition economies, as well as annual output growth rates. The panels suggest that capital flows have been procyclical, with large inflows in good times and outflows during recessions. For example, Latin America’s growth rates oscillated at around 4.5 percent in periods of capital inflows, while growth rates were about 1 percent in periods of sudden stops. Similarly, Asia’s economic activity collapsed to about 5.5 percent after the sudden stop in capital flows in the late 1990s, after growing at an average annual growth rate of 8.5 percent during the earlier period of large capital inflows. This evidence contrasts sharply with the prescription that international capital markets should allow countries to smooth out the effect of the business cycle. Countries seem to have lost access to international credit markets during recessions on a systematic basis.

This non-optimal behavior of international capital flows has also been studied by Calvo et al. (2004), who observe that sudden reversals in capital flows to emerging economies lead to large real depreciations and profound downturns. As reported in Kaminsky et al. (2004), macro-policies tend to be procyclical in developing countries while they are countercyclical or acyclical in industrialized countries. That is to say, macro-policies tend to
smooth out the business cycle in industrial countries but magnify it in developing countries, as shown in Table 41.2. The left panel in this table reports the correlation between the cyclical components of fiscal and monetary policy with the business cycle. The right panel shows the correlations of the cyclical components of fiscal and monetary policy with net capital inflows. Interestingly, the evidence suggests that international capital flows to developing countries may trigger procyclical macro-policies. Government expenditure (inflation tax) is positively (negatively) correlated with net capital inflows, indicating that periods of capital inflows are associated with expansionary fiscal policies, and periods of capital outflows with contractionary fiscal policies. While more research is needed, the stylized evidence suggests that international capital flows may trigger more volatile business cycles in emerging economies.

While this evidence points to links between financial integration and output instability over the business cycle, there is also evidence that financial integration promotes growth. A variety of authors have examined the effects of domestic and external deregulation of financial markets in emerging economies and found that they generally trigger sustainable growth in the long run. Bekaert et al. (2005), for example, examine the effects on growth of the opening of the stock market to foreign investors in a sample of about 90 developing countries and find that, overall, liberalization triggers an increase in growth by approximately one percentage point. Similarly, Galindo et al. (2002) study whether financial liberalization promotes economic growth by analyzing its effect on the cost of external financing to firms. They find that the liberalization of domestic and external financial markets reduces the cost of external funds faced by firms. In particular, they find that industries that depend on external finance grow almost 1 percent faster, relative to industries with low external financing dependence, in episodes of globalization compared to episodes of repression. The evidence on the links between financial liberalization and growth is not conclusive, however. Edison et al. (2002), for example, using data from 57 countries from 1980 to 2000, conclude that there is no robustly significant effect of financial integration on economic growth. Similarly, Kraay (1998), using a sample of 117 countries, finds no effect of financial liberalization on growth or, at best, mixed results.

Perhaps the inability of past research to agree on the effects of financial globalization on economic growth lies in the fact that liberalization has time-varying effects on growth. Loayza and Ranciere (2002) present some evidence that suggests this might be the case. These authors estimate transitory and trend effects of financial deepening on growth using a sample of about 80 countries and find that financial deepening, which in general is closely related to financial liberalization, harms growth in the short run but
Notes:
The countries comprising Asia are Bangladesh, China, Hong Kong, India, Indonesia, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Taiwan, Thailand and Vietnam.
The countries comprising the transition economies are Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Estonia, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Macedonia, Moldova, Mongolia, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.
The countries comprising Latin America are Argentina, Brazil, Chile, Colombia, the Dominican Republic, Ecuador, Guatemala, Mexico, Peru, Uruguay and Venezuela.

Source: World Economic Outlook.

Figure 41.8 Private capital flows to emerging markets and GDP annual growth rates
leads to higher growth in the long run. These latest results are closely linked to the evidence from stock market cycles discussed before and suggest that financial liberalization triggers growth in the long run because it fuels institutional reform.

Conclusions

The explosion of capital flows to emerging markets in the early and mid-1990s and the recent reversal following the crises around the globe have reignited a heated debate on how to manage international capital flows. Capital outflows worry policy-makers, but so do capital inflows, as they may trigger bubbles in asset markets and lead to an appreciation of the domestic currency and a loss of competitiveness. Policy-makers also worry that capital inflows are mostly of the ‘hot money’ type, which is why capital controls have mostly targeted short-term capital inflows. While capital controls may work, at least in the very short run, the introduction of restrictions to capital mobility may have undesirable long-run effects. In particular, capital controls protect inefficient domestic financial institutions and thus may trigger financial vulnerabilities.7 Capital controls may also delay improve-

Table 41.2 Correlations between the cyclical components of macropolicies, real GDP and net capital inflows

<table>
<thead>
<tr>
<th>Countries</th>
<th>Correlations with real GDP</th>
<th>Correlations with net capital inflows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fiscal policy</td>
<td>Monetary policy</td>
</tr>
<tr>
<td>OECD</td>
<td>Government expenditure</td>
<td>Inflation tax</td>
</tr>
<tr>
<td>OECD</td>
<td>−0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>Non-OECD</td>
<td>0.33</td>
<td>−0.15</td>
</tr>
</tbody>
</table>

Notes:
A positive (negative) correlation between government expenditure (inflation tax) and real GDP indicates pro-cyclical fiscal policy. A negative correlation between lending interest rates and real GDP indicates pro-cyclical monetary policy. A positive (negative) correlation between government expenditure (inflation tax and lending interest rates) and net capital inflows indicates that contractionary macropolicies are linked to episodes of low net capital inflows. The cyclical component of the various indicators was obtained using the HP filter.

Source: Kaminsky et al. (2004).
ments in corporate governance of non-financial firms because, as countries liberalize their capital accounts, domestic corporations start participating in international capital markets, mainly through cross-listing in major world stock exchanges, with higher disclosure standards and under the jurisdiction of a superior legal system. This certainly promotes more transparency in the management of the firm and can trigger improvements in corporate governance (see, for example, Stulz, 1999). Thus, regulation of capital flows may not only provoke financial vulnerabilities but also lower economic growth. Policy-makers have also resorted to sterilization of capital flows to regain control of monetary policy. While sterilization may provide some relief, it may also be quite costly to central banks. Moreover, the ability of governments to control international capital flows or to sterilize them diminishes with globalization.

In conclusion, there is no optimal policy to deal with the risks of volatile international capital flows, as policies that may work in the short run may have adverse effects in the long run. Since there is evidence that currency and banking crises tend to occur in economies with deteriorated fundamentals, conservative macroeconomic policies should be at the heart of dealing with volatile capital flows. Further research should examine whether countries can deregulate financial systems without becoming vulnerable to crises. Since the costs of crises have been quite large, this last question deserves much attention.

Notes
1. This chapter draws on previous research with Richard Lyons, Carmen Reinhart, Sergio Schmukler and Carlos Végh and is a shorter version of a paper entitled ‘International Capital Flows, Financial Stability and Growth’ (Kaminsky, 2006).
2. Liquidity may have an important effect on investors’ portfolio allocations since investors may want to avoid illiquid markets to minimize the price collapses always present when there is no ready market.
3. To identify liquid markets, countries are ranked by region according to their volume traded and according to their share in the mutual funds portfolio at the onset of the crisis. The dummy variable related to volume traded is given a value of one if the country ranks among the top 30 percent of most-liquid countries in the region in that category, and a value of zero otherwise. Similarly, countries are classified as liquid (that is to say, the dummy variable is given a value of one) if they rank among the 30 percent of the countries with the largest share in mutual fund portfolios for the region. A third dummy is created to capture whether emerging-market firms are trading in mature financial markets: the variable is given a value of one if they do, and zero if they do not. Finally, the variable capturing restrictions to entry and exit of foreigners in the stock markets of emerging economies is given a value of one if there are no restrictions, and zero if there are. All of this information is collapsed into a liquidity variable that is the average of the four univariate liquidity dummy variables. Thus, the general index of liquidity, the average of the four components, can have five values: 0, 1/4, 2/4, 3/4 and 1, with a value of one indicating a highly liquid market. I classify a country as having liquid financial markets when this dummy takes a value of 2/4 or higher.
4. See Kaminsky et al. (2002) for a country-by-country detail on fragility, liquidity, risk and mutual fund withdrawals.
5. The results discussed below are from Kaminsky and Schmukler (2003).
6. The 14 emerging economies are Argentina, Brazil, Chile, Colombia, Hong Kong, Indonesia, South Korea, Malaysia, Mexico, Peru, the Philippines, Taiwan, Thailand and Venezuela.
7. Claessens et al. (1998) present evidence that liberalization of the capital account and foreign bank entry lead to improvements in banking system efficiency.

References
Introduction
Controversies about aid effectiveness go back decades. Critics such as Milton Friedman (1958), Peter Bauer (1972) and William Easterly (2001) have leveled stinging critiques, charging that aid has enlarged government bureaucracies, perpetuated bad governments, enriched the elite in poor countries, or just been wasted. They cite widespread poverty in Africa and South Asia despite four decades of aid starting in the 1960s, and point to countries that have received substantial aid yet have had disastrous records such as the Democratic Republic of the Congo, Haiti, Papua New Guinea and Somalia. In their eyes, aid programs should be dramatically reformed, substantially curtailed or eliminated altogether.

Supporters counter that these arguments, while partially correct, are overstated. Nicholas Stern (2002), Joseph Stiglitz (2002), Jeffrey Sachs et al. (2004) and others have argued that although aid has sometimes failed, it has supported poverty reduction and growth in some countries and prevented worse performance in others. They believe that many of the weaknesses of aid have more to do with donors than recipients, and point to a range of successful countries that have received significant aid such as Botswana, Indonesia, Korea and, more recently, Tanzania and Mozambique, along with successful initiatives such as the Green Revolution, the campaign against river blindness, and the introduction of oral rehydration therapy.

This chapter explores trends in aid, the motivations for aid, its impacts, and debates about reforming aid. It begins by examining aid magnitudes and who gives and receives aid. It discusses the multiple motivations and objectives of aid, some of which conflict with each other. It then explores the empirical evidence on the relationship between aid and growth, most (but far from all) of which concludes there is a positive relationship (at least under certain circumstances). It examines some of the key challenges in making aid more effective, including the principal–agent problem and the related issue of conditionality, and concludes by examining some of the main proposals for improving aid effectiveness.
Donors and recipients

What is foreign aid?
The standard definition of foreign aid comes from the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD), which defines foreign aid (or the equivalent term, foreign assistance) as financial flows, technical assistance and commodities that are: (1) designed to promote economic development and welfare as their main objective (thus excluding aid for military or other non-development purposes); and (2) provided as either grants or subsidized loans.

Grants and subsidized loans are referred to as concessional financing, whereas loans that carry market or near-market terms (and therefore are not foreign aid) are non-concessional financing. According to the DAC, a loan counts as aid if it has a ‘grant element’ of 25 percent or more, meaning that the present value of the loan must be at least 25 percent below the present value of a comparable loan at market interest rates (usually assumed by the DAC – rather arbitrarily – to be 10 percent with no grace period). Thus, the grant element is zero for a loan carrying a 10 percent interest rate, 100 percent for an outright grant, and something in-between for other loans.

The DAC classifies aid flows into three broad categories. Official development assistance (ODA) is the largest, consisting of aid provided by donor governments to low- and middle-income countries. Official assistance (OA) is aid provided by governments to richer countries with per capita incomes higher than approximately $9000 (for example, the Bahamas, Cyprus, Israel and Singapore) and to countries that were formerly part of the Soviet Union or its satellites. Private voluntary assistance includes grants from non-governmental organizations, religious groups, charities, foundations and private companies.

When discussing foreign aid, most people have in mind ODA. Global ODA increased steadily from the 1960s until it reached a peak of $68 billion in 1992, just after the end of the Cold War (Figure 42.1), and then declined sharply to just under $55 billion in 1997. Aid flows began to rebound in the late 1990s following calls for greater debt relief and increased aid to new democracies, and accelerated very sharply after the attacks of 11 September 2001, reaching $120 billion in 2006 (all of these figures would be slightly higher if they included OA). In real terms, total ODA in 2002 was about the same as in 1992, and by 2006 was about 15 percent higher. Measured as a share of donor income ODA fell sharply during the 1990s, and has rebounded only slightly. Donors have pledged to continue to increase aid, most recently in July 2005 when the heads of state of the
Figure 42.1  Global ODA 1975–2006

Group of 8 industrialized countries promised to double aid to sub-Saharan Africa by 2010 and triple it by 2015, but growing budget tensions in donor countries may undermine these pledges.

Who gives aid, and who receives it?
Historically most aid has been given as bilateral assistance directly from one country to another. Donors also provide aid indirectly as multilateral assistance, which pools resources together from many donors. The major multilateral institutions include the World Bank, the International Monetary Fund (IMF), the African, Asian and Inter-American Development Banks, and various United Nations agencies such as the United Nations Development Programme.

In terms of total dollars, the United States has consistently been the world’s largest donor (except in the mid-1990s when Japan briefly topped the list). In 2006 the USA provided $22.9 billion in ODA, with Japan, France the United Kingdom, and Germany the next-largest donors. However, when aid is measured as a share of donor income, the most generous donors are Norway, Denmark, Luxembourg, the Netherlands, and Sweden, each of which provided between 0.81–1.02 percent of GNI in 2006. Saudi Arabia provided aid equivalent to about 0.57 percent of its income. The United States is one of the smallest donors by this measure, at about 0.18 percent of US income in 2006, just over half of the 1970 level of 0.32 percent and less than one-third of the US average during the 1960s. Donors have pledged since the 1960s to devote 0.7 percent of their income as aid, most recently at the Financing for Development Conference in Monterrey in March 2002, but only a handful of small donors have achieved this level of aid.

One hundred and forty-five countries and territories around the world received aid in 2006. Table 42.1 shows the ten largest recipients, each of which received more than $1.8 billion. Nigeria was at the top of the list in 2006, with measured aid of $11.4 billion. But this figure is misleading because it includes $9.5 billion for a one-time debt relief deal. Debt relief is accounted for differently than other components of ODA – the value of debt relief is the charge to the creditor country’s budget for writing off the debt in the year of the debt relief, and does not represent new funding to the recipient (although it does capture a future reduction in debt service obligations). Nigeria’s actual inflow of new finance in 2006 was $1.9 billion. Iraq and Afghanistan together received nearly $12 billion dollars, nearly unprecedented amounts, accounting for about 10 percent of the global total. Amounts to the other countries shown in Table 42.1 are more typical (by historical standards) for large recipients. Total dollar amounts are important, but they do not tell the entire story. On a per capita basis, the aid flows to some of these countries are fairly small. Vietnam received $1.8
<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Total ODA (millions US$)</th>
<th>Aid as % of recipient GNI</th>
<th>Aid per capita (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nigeria</td>
<td>11434</td>
<td>61</td>
<td>1866</td>
</tr>
<tr>
<td>2</td>
<td>Iraq</td>
<td>8661</td>
<td>58</td>
<td>1777</td>
</tr>
<tr>
<td>3</td>
<td>Afghanistan</td>
<td>3000</td>
<td>54</td>
<td>1740</td>
</tr>
<tr>
<td>4</td>
<td>Pakistan</td>
<td>2147</td>
<td>41</td>
<td>1614</td>
</tr>
<tr>
<td>5</td>
<td>Sudan</td>
<td>2058</td>
<td>36</td>
<td>1534</td>
</tr>
<tr>
<td>6</td>
<td>Congo, Dem. Rep.</td>
<td>2056</td>
<td>35</td>
<td>786</td>
</tr>
<tr>
<td>7</td>
<td>Ethiopia</td>
<td>1947</td>
<td>30</td>
<td>426</td>
</tr>
<tr>
<td>8</td>
<td>Vietnam</td>
<td>1846</td>
<td>29</td>
<td>409</td>
</tr>
<tr>
<td>9</td>
<td>Tanzania</td>
<td>1825</td>
<td>28</td>
<td>304</td>
</tr>
<tr>
<td>10</td>
<td>Cameroon</td>
<td>1684</td>
<td></td>
<td>282</td>
</tr>
</tbody>
</table>

billion in aid in 2004, but this was equivalent to just 3 percent of its gross national income (GNI) or about $22 per person. By contrast, Cameroon received a similar amount, $1.7 billion in 2006, but for its 16.6 million people this was equivalent to about $100 dollars per person. For small countries, a little bit goes a long way. Tiny Solomon Islands received just $205 million, but this translated into 61 percent of GNI and about $409 per person. Aid is typically measured as a share of GNI, but this can be misleading as a high ratio can just as easily be indicative of low GNI as of a large amount of aid.

On a regional basis, sub-Saharan African countries received aid flows of 5.8 percent of GNI in 2006, or $50.2 per person (Table 42.2), although close to one-third of this amount was due to several large one-time debt relief deals that are not new inflows and are not indicative of long-term trends. North Africa and the Middle East received more than $44 per person (largely on account of Iraq), and Europe and Central Asia received about $18 per person. For low-income countries around the world, donors provided aid averaging about $20.2 per recipient in 2006, although once again these figures are inflated by several one-time debt relief deals.

Generally speaking, aid is one of the largest components of foreign capital flows to low-income countries, but not to most middle-income countries, where private capital flows are more important. Aid flows averaged 3.1 percent of GNI in low-income countries in 2004, but just 0.2 percent of GNI in upper-middle-income countries. It is commonly claimed that the decline in aid flows to developing countries in the 1990s was more than offset by a rise in private capital. While this is true for developing countries in aggregate, the rise in private capital flows was heavily

Table 42.2 Official aid receipts by region, 2006

<table>
<thead>
<tr>
<th>Region</th>
<th>Billion US$</th>
<th>% of GNI</th>
<th>US$ per person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-Saharan Africa</td>
<td>38.2</td>
<td>5.8</td>
<td>50.2</td>
</tr>
<tr>
<td>South Asia</td>
<td>9.2</td>
<td>0.8</td>
<td>6.1</td>
</tr>
<tr>
<td>East Asia &amp; Pacific</td>
<td>7.4</td>
<td>0.2</td>
<td>3.9</td>
</tr>
<tr>
<td>Europe &amp; Central Asia</td>
<td>2.6</td>
<td>0.4</td>
<td>17.6</td>
</tr>
<tr>
<td>Middle East &amp; North Africa</td>
<td>14.6</td>
<td>2.0</td>
<td>44.0</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>6.0</td>
<td>0.2</td>
<td>11.0</td>
</tr>
<tr>
<td>Low-income</td>
<td>47.5</td>
<td>3.1</td>
<td>20.2</td>
</tr>
<tr>
<td>Lower-middle income</td>
<td>26.7</td>
<td>0.5</td>
<td>11.0</td>
</tr>
<tr>
<td>Upper-middle income</td>
<td>3.7</td>
<td>0.2</td>
<td>10.1</td>
</tr>
</tbody>
</table>

Source: Author’s calculations based on data from OECD 2007 Development Cooperation Report and World Development Indicators 2007.
concentrated in a handful of middle-income countries. In low-income countries, private capital rose much more slowly, and remained significantly smaller than aid.

Why do donors give aid?
Donors have a variety of motivations for providing aid, only some of which are directly related to economic development. There is little question that foreign policy and political relationships are the most important determinants of aid flows. During the Cold War, both the United States and the Soviet Union used aid to vie for the support of developing countries with little regard as to whether the aid was actually used to support development. The two largest recipients of US foreign aid (including both OA and ODA) from 1980 until very recently were Israel and Egypt, as the USA provided financial support to back the 1979 Camp David peace agreement. Beginning in 2002 Iraq became the largest aid recipient in the world, and its reconstruction is likely to become the largest single foreign aid program ever recorded. Taiwan and China have used aid (among other policy tools) to try to gain support and recognition for their governments from countries around the world. Many donors provide significant aid to their former colonies as a means of retaining some political influence (Alesina and Dollar, 2000).

Many people see the main rationale for aid as fighting poverty, and although this is less important than political considerations in donor allocation decisions, it still plays an important role. Donors generally provide their most concessional aid to the poorest countries, and some aid programs are designed explicitly with this objective in mind. For example, the World Bank’s concessional financing arm – the International Development Association (IDA) – has an income ceiling ($965 per capita in 2004). Once countries reach that ceiling, in most cases they ‘graduate’ from IDA to non-concessional International Bank for Reconstruction and Development (IBRD) loans. Other programs have less formal graduation rules, but still tend to provide less aid as incomes grow.

Country size matters as well. Large countries, such as Bangladesh, Indonesia, Nigeria and Pakistan receive relatively small amounts of aid on a per capita basis, even though hundreds of millions of people live in poverty in these countries. By contrast, some small countries receive very large amounts. For political reasons, donors generally want to influence as many countries as possible, which tends to lead to a disproportionate amount of aid going to small countries.

Bilateral aid is often designed at least partially to help support the economic interests of certain firms or sectors in the donor country. Multilateral aid is less prone to these pressures, although by no means immune. Many donors ‘tie’ portions of their aid by requiring that certain
goods and services be purchased from firms in the donor's home country, or that it be used for specific purposes that support groups in the donor countries (such as universities or business consulting firms). Automobiles, airline tickets and consulting services financed by USA foreign aid in most cases must be purchased from USA firms. Tying aid can give it more political support at home, but it can also make it more costly and less effective. If funds must be spent in the donor country, it reduces competition for services so that donors do not always use the least-cost provider. For example, the USA requires that food aid be purchased in the USA and shipped in US carriers to recipient countries, which can be much more expensive and take much longer than if food was purchased in a neighboring country. This means that recipients receive much less value for each dollar of aid allocated than they otherwise could. One study found that tying aid added 15–20 percent to its cost, thus significantly reducing its impact on recipient countries. Donors have begun to reduce the amount of aid that they tie, but the practice is still widespread among some donors. The USA no longer reports the share of its aid that is tied, but historically it has been around 75 percent. Greece ties about 70 percent of its aid, and Canada and Austria more than 40 percent. By contrast, Ireland, Norway, and the UK do not tie any of their aid.

Aid, growth and development
Most foreign aid is designed to meet one or more of four broad economic and development objectives: (1) to stimulate economic growth through building infrastructure, supporting productive sectors such as agriculture, or bringing new ideas and technologies; (2) to strengthen education, health, environmental or political institutions or systems; (3) to support subsistence consumption of food and other commodities, especially during relief operations or humanitarian crises; or (4) to help stabilize an economy following economic shocks.

Despite these broader objectives for aid, growth has always been the main yardstick used to judge aid's effectiveness. Debate has raged about the relationship between aid and growth for years, but there are some broad parameters of agreement. Even most aid pessimists agree that aid has been successful in some countries (such as in Botswana or Korea, or more recently in Mozambique and Tanzania), that aid has helped improve health by supplying essential medicines, and that aid is an important vehicle in providing emergency relief following natural disasters. Similarly, aid optimists concede that much aid has been wasted or stolen, such as by the Marcos regime in the Philippines and the Duvalier regime in Haiti, and that even under the best circumstances aid can create certain adverse economic incentives. Debate continues on the overall general trends, the conditions
under which aid works or does not work, and on what steps can be taken to make aid more effective. Although the majority of research since the mid-1990s has found a positive relationship between aid and growth, several studies have found no relationship. Three broad views have emerged on the relationship between aid and growth.4

_Aid has a positive relationship with growth on average across countries (although not in every country), but with diminishing returns as the volume of aid increases_

There are three key channels through which aid might spur growth:

- First, the classic view is that aid augments saving, finances investment and adds to the capital stock. In this view, poor countries are unable to generate sufficient amounts of saving on their own to finance the investment necessary to initiate growth, or at best only enough for very slow growth. In the strongest version of this view, the poorest countries may be stuck in a poverty trap in which their income is too low to generate the saving necessary to initiate the process of sustained growth (Sachs et al., 2004). A related argument is that aid might help relax a foreign exchange constraint in countries that earn relatively little foreign exchange, a view that was popularized through the early ‘two-gap’ models of economic growth.

- Second, aid might increase worker productivity through investments in health or education.

- Third, aid could provide a conduit for the transfer of technology or knowledge from rich countries to poor countries by paying for capital goods imports, through technical assistance, or through direct transfer of technologies such as the introduction of new seeds and fertilizers in the Green Revolution.

Several early studies found a positive relationship between aid and growth (for example, Papenek, 1973; Levy, 1988), but this strand of the literature took a significant turn in the mid-1990s when researchers began to investigate whether aid might support growth with diminishing returns. Oddly – given Solow’s response to the Harrod–Domar model in the 1950s – research until the mid-1990s only tested a linear relationship, a specification which persists in some studies today. A large group of studies that allow for diminishing returns have found a positive relationship although the direction of causality is a subject of ongoing debate.5 These studies do not conclude that aid has always worked in every country, but rather that on average and controlling for other factors, higher aid flows have been associated with more rapid growth. These studies have received much less public attention than
those that have found a zero or conditional relationship. The robustness of the conclusions of several of these studies has been the subject of on-going debate, as has the robustness of the conclusions of several studies that have reached the opposite conclusion, as discussed below. But since the mid-1990s the majority of published research on the topic has found a positive relationship either by allowing for diminishing returns, or by testing for conditional relationships as explored below.

Aid could also have a positive impact on development outcomes other than growth, such as health, education or the environment. Perhaps the best-documented area is health, where aid-supported programs have contributed to the eradication of smallpox, the near eradication of polio, control of river blindness and other diseases, the spread of oral rehydration tablets to combat diarrhea, and the dramatic increase in immunization rates in developing countries since 1970 (Levine et al., 2004). A recent cross-country study found a positive and significant relationship between health aid and infant mortality (Mishra and Newhouse, 2007). Undoubtedly, much aid aimed at health has also been squandered. But beyond the examples listed here, there is little systematic evidence on the relationship between aid and health, education, income distribution or other outcomes.

Aid has no affect on growth, and may actually undermine growth

Peter Bauer was perhaps the most outspoken proponent of this view (for example, Bauer, 1972), but he never provided systematic empirical evidence to support his argument. Many later empirical studies did reach this conclusion, although once again the robustness of these results is the subject of on-going debate. These researchers have suggested a variety of reasons why aid might not support growth:

- First, aid simply could be wasted, such as on limousines or presidential palaces, or it could encourage corruption, not just in aid programs but more broadly.
- Second, it can help keep bad governments in power, thus helping to perpetuate poor economic policies and postpone reform. Some argue that aid provided to countries in the midst of war might inadvertently help finance and perpetuate the conflict, and add to instability.
- Third, countries may have limited absorptive capacity to use aid flows effectively if they have relatively few skilled workers, weak infrastructure or constrained delivery systems. (Aid could help redress these weaknesses, but it may not be aimed to do so.)
- Fourth, aid flows can reduce domestic saving, both private saving (through its impact on interest rates) and government saving (through its impact on government revenue).
Fifth, aid flows could undermine private sector incentives for investment or to improve productivity. Aid can cause the currency to appreciate, undermining the profitability of the production of all tradable goods (known as the Dutch disease). Food aid, if not managed appropriately, can reduce farm prices and hurt farmer income.

The last two points merit further discussion. On aid and saving, while foreign aid adds to total saving (since aid is a form of foreign saving), some studies have shown that a dollar of aid adds less than a dollar to total saving and investment, since domestic savings may fall as aid increases. Some of these studies conclude that aid is ineffective because it 'leaks' to consumption. This approach is not particularly helpful in the aggregate since large portions of aid are in fact designed specifically to directly increase consumption and not investment, including food aid, immunization programs, purchases of textbooks, technical assistance, and the like. Nevertheless, even where aid is aimed at investment, the impact could be partially offset by a reduction in either private saving (through a decline in the rate of return on private investment) or government saving (through a fall in tax revenues). There is a wide range of estimates of the offset effect, but most find that $1 in aid translates to an increase in investment in the range of 33 to 67 cents. Much depends on the particular country, the type of aid, and other factors.

Aid also could undermine incentives for private sector activity. Aid can spur inflation and cause a real appreciation of the exchange rate, which reduces the profitability of production of all tradable goods, creating ‘Dutch disease’ effects. Aid flows can enlarge the size of the government and related services supporting aid projects, drawing workers and investment away from other productive activities such as agro-processing, garments or footwear exports. To the extent that these tradable activities are a key source of productivity gains, long-term growth may suffer. Similarly, food aid can sometimes undermine local food production if an influx of food drives down prices (it has less adverse impact on production when it displaces food imports).

The empirical studies that have found no relationship between aid and growth have been influential. However, very few published studies have reached that conclusion since the mid-1990s (a recent exception is Rajan and Subramanian, 2005a). Most of those that do use restrictive models that impose constraints such as a linear relationship between aid and growth, ruling out by assumption the possibility of diminishing returns. Most also only examine aggregate aid, imposing the restriction that all aid has a similar impact on growth, which is not particularly realistic, since famine...
relief, immunization programs and road projects are all likely to have very different impacts on growth.

Aid has a conditional relationship with growth, helping to accelerate growth under certain circumstances

This view holds that aid supports growth in some circumstances but not others, and searches for key characteristics associated with the difference. This ‘conditional’ strand of the literature has three subcategories, with the effectiveness of aid depending on the characteristics of the recipient country, the practices and procedures of the donors, or the type of activity that the aid supports.

Recipient-country characteristics Isham et al. (1995) found that World Bank projects had higher rates of returns in countries with stronger civil liberties. Burnside and Dollar (2000), in a very influential study, concluded that aid stimulated growth in countries with good policies, but not otherwise. Other researchers have proposed different country characteristics that might affect the aid–growth relationship, including export price shocks, climatic shocks, the terms of trade, macroeconomic and trade policies, institutional quality, warfare, type of government and location in the tropics. All of these studies rely on an interaction term between aid and the variable in question, and (not surprisingly) many of the interaction terms are fragile. Easterly et al. (2004) find that the original Burnside and Dollar results do not hold up to modest robustness checks. Roodman (2007) tests several other ‘conditional’ studies and find most of them to be relatively fragile, although the conclusions of Dalgaard and Tarp (2004) are more robust.

Nevertheless, the view that aid works better (or in a stronger version, aid works only) in countries with good policies and institutions has become the conventional wisdom among donors, partly based on this research and partly due to development practitioners that believe this to be the case based on their own experience. The appeal of this approach is that it can explain why aid seems to have supported growth in some ‘well-behaving’ countries but not others. These findings have had an enormous impact on donors (World Bank, 1998). The concept feeds directly into the World Bank’s Performance-Based Allocation (PBA) system for distributing concessional IDA funds, and was the foundation for the United States’ new Millennium Challenge Account (Radelet, 2003).

Donor practices Many analysts have argued that donor practices strongly influence aid effectiveness. For example, multilateral aid might be more effective than bilateral aid, and ‘untied’ aid is thought to have higher
returns than ‘tied’ aid, as discussed previously. Many observers argue that donors that have large bureaucracies, do not coordinate with other donors, or have poor monitoring and evaluation systems undermine the effectiveness of their own programs. Two influential and overlapping views argue that aid would be more effective if there were greater ‘country ownership’ or broader ‘participation’ among government and community groups in recipient countries in setting priorities and designing programs. There has been substantial debate about these issues, and in some cases these ideas have begun to change donor practices. But to date there has been very little systematic research connecting specific donor practices to aid effectiveness.

**Type of aid** Different kinds of aid might affect growth in different ways. Clemens et al. (2004) disaggregated aid into types most likely and least likely to affect growth within a few years, if at all. They separated aid into three categories: (1) emergency and humanitarian aid (likely to be negatively associated with growth, since aids tends to increase sharply at the same time growth falls following an economic shock); (2) aid that might only affect growth after a long period of time, if at all, and so the relationship may be difficult to detect (such as aid for health, education, the environment, and to support democracy); and (3) aid that is directly aimed at affecting growth (building roads, ports and electricity generators, or supporting agriculture). It found a strong positive relationship between the third type of aid (about half of all aid) and growth, a result which stood up to a wide variety of robustness checks. As expected, the relationship with the other types was less detectable.

To summarize the aid and growth research, it appears that aid has been successful in some countries but not others. The overall trend is a subject of debate, but most research has found a positive relationship although the direction of causality is not always clear. This research is only beginning to scratch beneath the surface and investigate what types of aid are most effective and the conditions under which aid has the largest impact on growth. Since disputes continue about the determinants of economic growth more broadly, perhaps it is not surprising that the aid–growth relationship continues to be a matter of sharp debate.

**Donor relationships with recipient countries**
The criticisms about aid have led to debates about how aid programs can be improved to support growth and development more effectively. But the challenge is not easy. Aid programs face some inherent difficulties in trying to achieve a wide range of objectives, provide financial oversight and ensure results.
The principal–agent problem

A key issue facing aid agencies is that there is only an indirect and distant relationship between the people actually providing the financing – taxpayers in donor countries – and the intended ultimate beneficiaries of aid projects – poor people living in low-income countries. In most aid programs, there is a long and complex chain of principal–agent relationships, starting with the taxpayers that delegate authority to elected officials, who in turn become principals that delegate authority to a new set of agents, the heads of aid agencies, which delegate to agency employees, contractors and consultants. In the recipient country, there are similar relationships between citizens, their government and those that actually implement programs. The objectives, incentives and information available to these agents are not always well aligned with the objectives of either the taxpayers or the beneficiaries.

All public sector agencies and many private companies are faced with these principal–agent problems, but the international dimension and physical separation between the original taxpayers and ultimate beneficiaries makes it an even greater challenge for aid. In domestic public programs (such as rubbish collection or local schools) the taxpayers and ultimate beneficiaries are the same people, so they have clearer information about success or failure and can reward or penalize their agents accordingly by re-electing them or voting them out of office. But this feedback loop is broken for aid agencies. Taxpayers cannot tell if their money is well spent, beneficiaries sometimes do not even know about local programs, and each have limited mechanisms for penalties and rewards.

The principal–agent problem affects nearly all aspects of aid delivery including program design, implementation, compensation, incentives, evaluation and allocation of funding. The problem can never be fully solved – private companies face similar issues between owners, managers and employees, as do private aid foundations and charities. The challenge is to design institutions and incentives that mitigate these problems as much as possible to clarify goals, objectives, incentives and rewards. In this regard, one of the key challenges for donors is if, when, and how to apply conditions to their aid, a subject to which we now turn.

Conditionality

Partly as a result of the principal–agent problem, donors often apply conditions on aid programs to encourage recipients to act more in accord with the donors’ (and possibly the ultimate beneficiaries’) interests. Donor conditions on recipient actions or policies are among the most controversial aspects of aid. Policy conditionality is most often associated with the IMF and World Bank, but all donors use conditions to some extent.
The rationale for economic policy conditions is straightforward: donors believe that certain policies and actions in different countries are important for growth and development, and that without them providing aid is futile. If government policies have led to high rates of inflation, massive inefficiencies and waste of public spending, and extensive corruption, then providing aid – whatever the specific purpose – without requiring fundamental change would provide no benefits and perhaps could perpetuate damage. Some even argue that the primary purpose of aid is not the money, but for aid to act as a lever for the policy reforms.

There are three key problems with conditionality. First, it is not always clear what policy conditions are the most appropriate to ensure sustained growth and development. Development doctrine has swung from a state-led approach in the 1950s and 1960s, to basic human needs in the 1970s, to a macroeconomic approach focused on open markets in the 1980s and 1990s, to a greater focus on institutions beginning in the mid-1990s. As a result, the list of conditions is constantly evolving. Debate has raged for decades about whether specific IMF and World Bank conditions are justifiable and whether they support or hurt stabilization, growth and development. And who should bear the costs if donor-imposed conditions make things worse?

Second, while donors are often criticized for imposing too many conditions, they are almost as often criticized for not imposing enough conditions. Some advocates that criticize the IMF for imposing too much fiscal austerity also insist that it should require governments to spend a minimum amount on health and education. The World Bank is often asked to add conditions to force governments to take specific actions, for example on projects that have potential adverse environmental consequences.

Third, conditionality does not seem to work. Most analysts agree that governments implement reforms only when it is in their interests to do so, and donor conditions have little if any impact on that decision. Many donors continue to disburse aid even when recipients fail to meet conditions, sometimes repeatedly so. Donors are faced with their own internal incentives to continue to disburse aid to support the contractors and recipients that depend on it. They also face a ‘Samaritan’s dilemma’ that withdrawing aid would create short-term pain for the very people it is aimed to help.\textsuperscript{10}

The nature of conditionality has changed over time as the most pressing issues have changed and as donors continue to wrestle with the best ways to apply conditions. During the 1980s, most conditions focused on macroeconomic issues, trade reforms and privatization, as reflected in IMF and World Bank-sponsored structural adjustment programs. During the 1990s as macroeconomic imbalances improved and following the end of the Cold
War, attention shifted to governance, corruption and institution-building. Debate has re-emerged as to whether aid should be conditioned on democratic reforms in recipient countries. Whether governance-focused conditionality is a good idea, or whether it will be more successful than structural and policy conditionalities, remains to be seen.

There are no clear-cut rules for conditionality. Striking the right balance between responsible oversight and accountability on the one hand, and ensuring against high bureaucratic obstacles and the imposition of unnecessary controls or unwarranted policy changes on the other, requires flexibility, judgment and the ability to balance multiple objectives – none of which are easy for aid agencies to achieve.

**Improving aid effectiveness**

The debates about the strengths and weaknesses of aid have led to specific ideas for change, some of which donors have begun to put into practice. Four stand out.

**Country selectivity** One influential idea is that donors should be more selective about the countries to which they provide aid, based on the view that aid works best in countries with good policies and institutions. In the strongest version, aid should be provided only to countries that meet these criteria. A more moderate view is that more aid should be allocated to countries with stronger policies and institutions, but some aid should be targeted to countries with weaker policies, especially post-conflict countries. This proposal turns the conditionality debate: instead of providing aid to encourage reforms, give it to countries that have already demonstrated a desire to implement key reforms. In the language of the principal–agent problem, donors should spend less time trying to write contracts that force an alignment of incentives and instead give more aid to countries that on their own demonstrate similar motivations and objectives. Some donors have begun to be more ‘selective’, including the World Bank in the allocation of its concessional IDA funds, some European donors in terms of providing budget support, and the USA with its new Millennium Challenge Account. But since so much aid is allocated for political, security and other foreign policy reasons, there are limits to how far donors are likely to go in this direction.

**Recipient participation** Many analysts argue that aid has been weakened by donor domination in setting priorities, designing programs and implementing projects, and push for either a more ‘country led’ approach in which recipient governments take a stronger role, or a ‘participatory’ approach in which various groups in recipient countries (government,
NGOs, charities, the private sector) play a more active role. The idea is to eliminate some of the problems in the long chain of principal–agent relationships, and more tightly integrate the ultimate beneficiaries in key aspects of the aid delivery process. The World Bank and IMF (by requiring Poverty Reduction Strategy Papers), the Global Fund to Fight AIDS, Tuberculosis and Malaria, and the Millennium Challenge Corporation have all moved towards greater local participation in designing and implementing the programs they finance. This approach is new, so there is no evidence yet on the extent to which (or the circumstances under which) it improves aid effectiveness. There is a clear and inescapable tension between country ownership on the one hand, and donor priorities and conditionality on the other. Donors are more likely to facilitate a participatory approach in countries in which governments show a strong commitment to sound development policies, and less so in countries with corrupt and dictatorial governments.

**Harmonization and coordination** Managing aid flows from many different donors is a huge challenge for recipient countries, since different donors usually insist on using their own unique processes for initiating, implementing and monitoring projects. Recipients can be overwhelmed by requirements for multiple project audits, environmental assessments, procurement reports, financial statements and project updates. According to the World Bank, developing countries typically work with 30 or more aid agencies across a wide variety of sectors, with each sending an average of five missions a year to oversee their projects. The donors all want to meet with the same top government officials, leaving them with much less time to deal with pressing matters. These concerns have led to numerous suggestions for donors to coordinate their activities more closely, harmonize their systems or ‘pool’ their funds (Kanbur and Sandler, 1999). But while there has been some progress, the pace of change amongst the donors seems glacial.

**Results-based management** The emphasis on demonstrating the effectiveness of aid has led to calls for improved monitoring and evaluation and results-based management. In this view, aid programs should aim to achieve very specific quantitative targets, and decisions about renewing or reallocating aid going forward should be based on those results. There are three basic objectives: (1) helping donors allocate funds towards programs that are working; (2) detecting problems at an early stage to help modify and strengthen existing programs; and (3) improving the design of future programs. Stronger monitoring and evaluation would help improve principal–agent relationships so that aid agencies have clearer incentives and
taxpayers have better information about the impact of aid on its intended beneficiaries.

**Summary and conclusions**

Aid flows fell in the 1990s after the end of the Cold War and aid was widely attacked for being ineffective in spurring growth and development. However, aid began to grow again in the late 1990s and indications are that it will continue to grow throughout the first decade of the twenty-first century, although probably less rapidly than donors have pledged.

Most empirical research on aid and growth conducted since the mid-1990s has found a positive relationship, in contrast to popular perceptions, particularly studies that have allowed for diminishing returns and have controlled for other factors that affect growth. Some studies have found that the aid–growth relationship is conditional on the policy or institutional environment, but many of those results have been fragile. Some studies have concluded that there is no relationship or even a negative one, but while influential, these studies are few in number and tend to use restrictive assumptions. Recent research that has explored how different types of aid might have different impacts on growth has suggested one key reason why earlier research has reached mixed conclusions.

Nevertheless, there is little doubt that aid has been less effective in spurring development than is often expected. Aid can keep bad governments in power for too long, and can undermine incentives for saving, tax collection and private sector production. Aid relationships are made much more difficult by a complex chain of principal–agent problems that weaken information flows, introduce myriad motivations for different actors, and make monitoring and accountability more difficult. Attempts to solve the principal–agent problem through conditionality have not been very successful. The newest wave of reform efforts aims to solve some of the weaknesses of aid and the principal–agent problem through greater donor selectivity in choosing aid recipients, increased recipient participation in setting priorities and designing programs, streamlining aid bureaucracies, increasing donor coordination, and establishing clearer goals for aid and stronger monitoring and evaluation of aid-financed activities. These ideas have been very influential in designing aid programs in recent years, but there is no systematic evidence at this point as to whether these changes will lead to greater aid effectiveness.

**Notes**

1. This chapter draws heavily from Chapter 14 of Perkins et al. *Economics of Development*, 6th edn, 2006 (New York: W.W. Norton & Co.), (used by permission), and from Radelet et al. (2006). I thank Bilal Siddiqi and Sami Bazzi for their research assistance, and Amitava Dutt and Jaime Ros for comments on an earlier draft. I also thank the William and Flora Hewlett Foundation for financial support.
2. Non-concessional loans from donor agencies are counted as part of official development finance, but not as official development assistance.

3. More precisely, assistance to countries with per capita incomes (for three consecutive years) above the World Bank’s ‘high income’ threshold, but the DAC makes some exceptions.

4. This summary draws heavily from the review in Clemens et al. (2004). For another recent review of the literature see Hansen and Tarp (2001).


References


for Research in Economic Development and International Trade, University of Nottingham.


Introduction
International migration flows have expanded by historical proportions in recent years. In 1960, there was a stock of slightly over 75 million people residing in countries other than their country of birth. By 2005, this number had grown to 190 million. For many developing countries, international labor flows have become a natural outcome of the globalization process, as much as trade and investment flows. The money the migrants send back home has become a major source of income for families in developing countries. In 2005, migrant remittances amounted to $600 billion. For some countries, the income received from the services of workers abroad is now a major item of the balance of payments.

But international migration flows are not without their costs. For many years, the issue of the brain drain has been studied by international economists. From software engineers in India to doctors in the Philippines and nurses in sub-Saharan Africa, the exodus of skilled migrants has been a policy concern for many developing countries. The impact on the skilled labor forces of some countries has been substantial, especially in sub-Saharan Africa. In Ghana, for example, over 40 percent of persons with a college degree or more have migrated to other countries. In Gambia, the corresponding proportion is close to 65 percent, and in Somalia it is 59 percent. What are the consequences of these labor outflows? What benefits or costs do they impose in the sending nations?

This chapter provides an analysis and survey of the key recent trends in international migration, their determinants and major consequences. The next section presents data on the growth of global migration flows since the mid-1950s. The section after that proceeds to examine the causes of these migration flows, focusing on the main economic factors involved. The subsequent section analyzes the consequences of international migration for source countries. The concluding section discusses implications for future research.

Trends in international migration flows
International migration refers to the movement of people across national borders. Data on migration are available directly from national immigration services.
authorities in recipient countries. However, the data available for many countries are sketchy and difficult to compare with data for other countries due to differences in migration policies, definitions of what constitutes immigrants, the presence of undocumented migration, and so on. Some international organizations gather cross-country data on migration and seek to provide more uniform, comparable statistics. The Organisation for Economic Co-operation and Development (OECD) has for many years collected information on the migration of OECD countries and has an extensive database for these countries (see, for example, OECD, 2006). The United Nations has the most comprehensive worldwide database on the number of migrants residing in different countries (United Nations, 2007). There are also data collection efforts supported by the World Bank (Docquier and Marfouk, 2006), the International Monetary Fund (Carrington and Detragiache, 1998), and other institutions such as the Development Research Centre on Migration, Globalisation and Poverty at the University of Sussex (see Parsons et al., 2007).

Table 43.1 shows the massive growth in the estimated number of international migrants between 1960 and 2005. In 2005, there were 190 million people residing in countries other than their country of birth, up from 75 million in 1960. Most of these migrants originated in developing countries: it is estimated that 123 million migrants, or 65 percent of the total, were born in developing countries. Mexico had the highest number of persons residing abroad, equal to over 10 million in 2000, largely in the United States. This was followed by India, which had close to 9 million persons residing abroad in 2000, with the country’s diaspora spread all over the world. Other countries with mass emigration include Bangladesh (6.6 million emigrants), the Philippines (3.4 million), Pakistan (3.4 million), Turkey (3.0 million), Afghanistan (2.7 million), Morocco (2.6 million), Egypt (2.5 million) and Algeria (2.1 million), among others.

The destination of emigrants from developing countries is evenly split between high-income and developing countries. In 2005, close to 62 million people born in developing countries were residing in high-income economies, but 61 million resided in other developing nations. Among high-income countries, those with the largest number of immigrants included the United States (38.4 million), Germany (10.1 million), France (6.4 million) and Canada (6.1 million). Developing countries with the highest immigrant populations included Russia (12.1 million, who moved mostly from elsewhere in the former Soviet Union), Ukraine (6.8 million), Saudi Arabia (6.4 million), India (3.3 million) and the United Arab Emirates (3.2 million).

The emigration of skilled workers is one of the major concerns relating to migration flows in developing countries. To measure the magnitude of
skilled emigration or brain drain, a country’s labor force is divided into those who are skilled – generally considered to be those who have some tertiary education, that is, 13 years of schooling or more – and those who are less skilled, who have achieved less than tertiary education. The emigration of the skilled has been rising, just as global migration flows have increased in general. In OECD countries, for example, there were 12.5 million skilled immigrants of working age in 1990, but by 2000 the number had risen to 20.4 million, equal to 34.6 percent of the total number of immigrants in these countries.

The skilled emigration rate is defined as the stock of skilled migrants from a country (all persons with tertiary education living abroad) calculated as a percentage of the total skilled labor force in the source country augmented by the skilled migrants themselves. This shows the percentage of workers with tertiary education who were born in a country but are residing outside its borders. Hence, it is a measure of the relative impact of the emigration on the sending country’s skilled labor market. The highest skilled emigration rates in the world prevail in the Caribbean, where in 2001

Table 43.1 Estimates of the Stock of International Migrants in the World, 1960–2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Stock of migrants</th>
<th>Change between years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>75463352</td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>78443933</td>
<td>2980581</td>
</tr>
<tr>
<td>1970</td>
<td>81335779</td>
<td>2891846</td>
</tr>
<tr>
<td>1975</td>
<td>86789304</td>
<td>5453525</td>
</tr>
<tr>
<td>1980</td>
<td>99275898</td>
<td>12486594</td>
</tr>
<tr>
<td>1985</td>
<td>111013230</td>
<td>11737332</td>
</tr>
<tr>
<td>1990</td>
<td>154945333</td>
<td>43932103</td>
</tr>
<tr>
<td>1995</td>
<td>165080235</td>
<td>10134902</td>
</tr>
<tr>
<td>2000</td>
<td>176735772</td>
<td>11655537</td>
</tr>
<tr>
<td>2005</td>
<td>190633564</td>
<td>13897792</td>
</tr>
</tbody>
</table>

as much as 42.8 percent of the region’s tertiary labor force resided outside its borders (Docquier and Marfouk, 2006). Skilled emigration has also been enormous for a number of countries in Africa. In East Africa, the skilled emigration rate was close to 20 percent in 2001.

The determinants of international migration flows

What has caused the mass migration documented in the previous section? There is a massive literature examining the determinants of migration flows. This section presents the main forces and conceptual approaches that seek to explain migration flows.

The economic approach to migration: costs and benefits

At the level of theory, the classical economic model of the decision to migrate was formalized by Sjaastad (1962) and has been extended in a number of directions (see, for example, Lucas, 1985; Borjas, 1999). In this approach, the decision to migrate is seen as an investment decision that depends on individual assessments of the net balance of the present and future costs and benefits of migration. For a worker \( i \), the net gain from migration, \( G_i \), is equal to the present discounted value of the benefits minus the costs of migrating:

\[
G_i = \sum_{t=1}^{T} \left[ \frac{B_{it}}{(1 + r)^t} - \frac{C_{it}}{(1 + r)^t} \right]
\]

where \( T \) is the lifespan of the worker, \( r \) is a discount rate, \( B_{it} \) are the benefits at any given time \( t \), generally in the form of higher wages or improved employment opportunities in the destination country, and the costs, \( C_{it} \), include the direct costs of the move (transportation costs), the foregone earnings when the individual migrates, and any utility losses associated with leaving the homeland.

Note that different individuals will face varying costs and benefits, and the probability of a person migrating from an origin to a destination area will vary. Older workers, for example, may suffer the greatest net losses in foregone earnings and may also face higher psychic costs of leaving the source country. As a result, the likelihood of migration declines with age. Location also matters. Indeed, most migration flows occur among geographically close, often neighboring, countries, where the costs of migration are lower. Empirically, the role of geographical and cultural proximity in determining migration movements has been emphasized by the so-called ‘gravity model of migration flows’, for which there is ample supporting evidence (see Karemera et al., 2000; Hatton and Williamson, 2005).
Labor markets and the returns to international migration

Due to the difficulties of estimating future costs and benefits, most studies examining the decision to migrate focus on the current labor market gains from migrating, $I_i$, given by:

$$I_i = \log W_{iD} - \log W_{iO}$$

(43.2)

where $W_{iD}$ is the wage rate that individual $i$ can obtain in the destination and $W_{iO}$ is the equivalent wage at home. These wage rates are influenced by the various characteristics of workers, including their schooling, $E_{Di}$, their on-the-job experience, $E_{Xi}$, their motivation, $M_i$, health, marital status, number of children, and so on. Suppose, for simplicity, that the following characterizes the wages of individual $i$ in the destination and origin regions:

$$\log W_{iD} = \alpha_D + \alpha_{1D}E_{Di} + \alpha_{2D}E_{Xi} + \alpha_{3D}M_i$$

(43.3)

$$\log W_{iO} = \alpha_O + \alpha_{1O}E_{Di} + \alpha_{2O}E_{Xi} + \alpha_{3O}M_i$$

(43.4)

where the $\alpha$’s represent how the various individual characteristics (education, experience, and so on) are rewarded in each country. The net gain from migration is thus:

$$I_i = \log W_{iD} - \log W_{iO} = (\alpha_D - \alpha_O) + (\alpha_{1D} - \alpha_{1O})E_{Di}$$

$$+ (\alpha_{2D} - \alpha_{2O})E_{Xi} + (\alpha_{3D} - \alpha_{3O})M_i$$

(43.5)

For any individual, the incentive to migrate will vary on the basis of the skills (education, experience, and so on) that the worker is endowed with, his or her motivation, and how skills as well as motivation are rewarded in relative terms in the source and destination countries.

A large body of research has now accumulated studying the labor market outcomes of developing-country emigrants in their destinations (see, for example, the collection of research in Zimmermann and Constant, 2004, as well as the surveys by Borjas, 1999 and Hanson, 2006). This literature suggests that the economic returns to migrating are substantial for most workers. But the economic progress of migrants varies according to the characteristics of the migrants themselves (age, schooling, immigration status, and so on), the timing of the migration, and the country of destination. In some European countries, for instance, immigrants have substantially lower rates of labor force participation as well as higher unemployment rates than the native-born population. The relatively poor labor market outcomes of some immigrants are partly related to lack of skills, but they also depend on labor market conditions in – and
time spent in – recipient countries (Fertig and Schmidt, 2002; Rivera-Batiz, 2007).

**Migrant selectivity**

One of the most discussed issues in the international migration literature is whether emigrants are the most qualified, skilled workers in the origin economy or not. If migrants are positively selected, then they will be more likely to succeed abroad, but their exit from the source country will drain the most capable, most skilled population from the nation, with a potentially negative impact on its economy and society.

What determines whether migrants are positively or negatively selected? There are forces that favor a positive selectivity and others that encourage a negative selectivity. The best-known hypothesis is that emigrants tend to be positively selected because in order to compensate for the substantial costs of migration, only those who have the strongest drive and motivation – and the expectations of great rewards – will actually undertake the migration process (see Chiswick, 1978, 1999). Of course, if costs of migration decline, then this aspect of selectivity will tend to become less significant. In addition, as equation (43.5) suggests, the impact of motivation on migration depends on whether motivational skills are more strongly rewarded in the destination region. Indeed, one suspects that holding constant the distribution of motivational skills in a population, if these skills are poorly rewarded at home but richly rewarded abroad, the incentives to migrate from source to destination regions will increase.

A second hypothesis (referred to in the literature as the ‘Roy model’) is that, if those with greater skills or abilities are rewarded more highly compared to the less-skilled in the origin area when compared to the destination region, this will generate less incentives for those at the top of the skills or ability distribution to emigrate compared to those at the bottom of the distribution, causing a negative selectivity of migrants. Therefore, the emigrant contingent will be positively or negatively selected depending on the relative inequality of the distribution of income at home and abroad. For instance, the more unequal the income distribution in the origin area as compared to the destination, the lower the incentives for the highly skilled to emigrate relative to the less skilled. This result is easily obtained from equation (43.5). Since \( \frac{\partial I_i}{\partial Ed_i} = (\alpha_{1D} - \alpha_{1O}) \), if greater inequality in the source country is associated with a rate of return to education in the origin area that exceeds the rate of return in the destination country, then \( (\alpha_{1D} - \alpha_{1O}) < 0 \). This implies that, as the education of the prospective migrant rises, the incentives to migrate tend to decline. Note also that under asymmetric information, employers in the destination region may not be able accurately to assess the skills of the migrants and they may offer lower
wages to the highly skilled migrants, when compared to employers in the source country, that are able to assess more accurately the potential migrants’ skills and pay them wages more consonant with their skills. This will again reduce the rate of return to education received by the emigrants in the destination region relative to the source country \((\alpha_{1D} - \alpha_{10})\) will decline), shrinking the incentive to migrate of the highly skilled relative to the unskilled (see Stark and Taylor, 1991).

Some evidence appears to support the implications of the Roy model (see Borjas, 1987, 2006). For instance, recent research finds a negative selectivity in the migrant contingent from Mexico to the United States (that is, the emigrant group tends to have lower average skills than the population left behind), as would be expected from the relatively more unequal distribution of income in Mexico relative to the United States (see Fernandez-Huertas Moraga, 2007, for this analysis and Chiquiar and Hanson, 2005, for alternative results).

*Income differences and international migration*

Overall, the evidence on the importance of economic factors in motivating migration flows, as presented in equations (43.1)–(43.4) is extensive. Both documented and undocumented migration flows have been found to be strongly correlated to the relative economic conditions in recipient and source countries (see, for example, Adams, 1993; Hanson and Spillimbergo, 1999; Drinkwater, 2003; Castaldo et al., 2005).

But despite the widespread support for the hypothesis that increased income differentials between recipient and source countries stimulates migration, there is also ample support for the view that this connection does not always work and may actually hold in reverse (see Hatton and Williamson, 1998, and the review by Waddington and Sabates-Wheeler, 2003). For instance, in recent research seeking to determine the impact of differences in income per capita on migration flows in the world, Hatton and Williamson (2005, p. 240) find that in sub-Saharan Africa, increases in income at home increase migration. Indeed, the evidence suggests that the relationship between income per capita in source countries and emigration has an inverted-U shape. For poor countries, as income rises, migration actually increases. But as the wealth of a country grows, at some point, further increases in income per capita actually reduce emigration.

One explanation for this behavior is the fact that, at low levels of income per capita, a large part of the population just simply cannot afford the monetary costs of migrating. However, as per capita income in a country rises, this allows some people to save enough to pay for the transport and other costs of migration, thus increasing migration flows (see Hatton and
Williamson, 2005). An additional explanation is that the massive structural changes occurring in the early industrialization of an economy (the shift from agriculture to industry and services, from urban to rural areas, and so on) leads to a dislocation of the population that fosters international migration. As the economic development process matures, however, these changes diminish and migration declines. A third hypothesis for the paradoxical rise of migration flows as income increases in many economies is based on ‘relative deprivation theory’. It suggests that, as inequality rises in the early years of economic development (a trend first noted by economist Simon Kuznets), those who become relatively poor will become increasingly dissatisfied with their relative standing in the community. This will stimulate them to emigrate in order to improve their standard of living (see Stark and Taylor, 1991).

Families and the decision to migrate
Despite its powerful role in explaining migration flows, another problem of using a simple economic approach based on income differences across countries is that it cannot explain temporary migration. If there is a significant and persistent wage and employment differential between origin and destination regions, why do so many migrants wish to stay in the destination only for short periods of time?

One explanation is offered by the so-called new economics of labor migration (see Stark and Bloom, 1985). In this approach, it is understood that migration decisions are often made by families and households, not individuals. In contrast to the analysis in equations (43.1)–(43.4), the decision-making is not in the hands of an individual who maximizes his or her utility. Rather, migration decisions are made by families or households that maximize household utility over time. This utility, \( U \), is derived from a stream of consumption by family members located in the home and/or destination regions, \( C_D \) and \( C_O \), added all the way up to the time horizon of the family, \( T \), and discounted to the present time, \( t \):

\[
U = \int_t^T \exp[-\rho(\tau - t)] \log U[C_D(\tau), C_O(\tau)] \, d\tau \tag{43.6}
\]

where \( \rho \) is a rate of discount. Equation (43.6) is maximized subject to a budget constraint establishing that the present discounted value of income earned in the source and destination areas is equal to or greater than the present discounted value of family consumption spending.

Visualizing migration as an intertemporal family or household decision can help understand that the migration of some household members may be part of the savings and investment decisions of a family. The idea is that, if low-income households encounter capital market imperfections at home,
which exclude them from access to the financing of investments in housing, durable goods or in self-employed businesses, migration abroad may lead to the accumulation of remittances that can then be used to finance those purchases and investments. Migration becomes a short-term activity needed by households to raise funds in the absence of local financing. Another application of this approach is that the migration of one household member may reduce the costs of migration for other household members. This is what ‘network or chain migration theory’ suggests (see Piore, 1979; Bauer et al., 2000).

The political economy of immigration policy

The discussion so far has described the wide array of forces that may generate a supply of potential migrants. However, in an international context, destination-country governments exert an enormous force in restricting immigration flows. As a result, the volume of migration between developing countries and high-income economies is often determined by the immigration policies imposed by the latter.

Research on the determinants of immigration policies follows a growing literature in political economy that seeks to answer the question of how the policies of a country are generated through the interaction of economic, political and social forces (Mayda and Patel, 2006). A wide array of economic forces may lie behind the setting of immigration policies. For instance, if immigration reduces wages and raises the profits of employers and owners of capital, then persons who own firms or farms or own relatively large amounts of capital will be in favor of immigration, but those who do not have wealth and only have their labor will be against immigration (see Benhabib, 1996). But if the distribution of capital in a country is highly concentrated, with a great part of the workforce laboring at low wages, then there may be very few persons supporting immigration and many opposing it. If immigration policy is determined by influence of voters, immigration policy restrictions may be high. This force is magnified if the immigrants are unskilled since in this case the immigrants may be perceived as competing with the poor, generating stronger cries for immigration restrictions (Hatton and Williamson, 2005).

The available evidence confirms the role that economic forces play in affecting the immigration attitudes of natives in recipient countries (see Gang et al., 1999, 2002; Scheve and Slaughter, 2001; Mayda, 2006). But economic factors are not the only ones affecting attitudes towards immigrants. Social phenomena, such as xenophobia and bias against foreigners, can have a major influence on immigration policies (see Gang and Rivera-Batiz, 1994; and Gang et al., 2002, for analyses of the determinants of attitudes towards immigrants).
International migration and economic development

What are the economic consequences of mass migration for the developing world? What changes in the source countries are generated by migrants? This section examines the existing theory and evidence on how emigration affects developing nations.

The distributional and welfare impact of emigration

The simplest theoretical framework that can be used to examine the effects of international migration is one that focuses on the aggregate economy, within a simplified setting where there are only two inputs: capital and labor (see Bhagwati and Rodriguez, 1975; Borjas, 1999; Mishra, 2006). Figure 43.1 shows the labor market in the source country before and after emigration. The demand for labor (the economy’s marginal value product curve) is \( L^D \) and the supply of labor before emigration is depicted by \( L^S \), where it is assumed that there is an inelastic supply of labor given by the country’s labor force. Under the assumption that the labor market is perfectly competitive and generates full-employment, the equilibrium before emigration is at point A, with a wage rate equal to \( W^* \) and employment \( E^* \). If the number of emigrant workers is given by \( M \), the labor force declines, shifting the aggregate supply of labor to the left, to \( L^S - M \). The equilibrium after migration occurs at point B. The shortage of labor induced by emigration reduces employment from \( E^* \) to \( E^{**} \), and raises wage rates from \( W^* \) to \( W^{**} \).

**Figure 43.1 The effects of emigration on the source country**
wages rates from $W^*$ to $W^{**}$ (this assumes that the domestic capital stock is unaffected by emigration).

The impact of emigration can be depicted in Figure 43.1 by noting first that national income is given by $Y = WL + rK$, with $K$ equal to the economy’s capital stock and $r$ the rate of return to capital. Diagrammatically, after emigration, the income received by the workers that remain in the economy is equal to $W^{**}E^{**}$, as represented by the rectangle $W^{**}BE^{**}O$. This rises compared to the situation before emigration, when the income of these workers was $W^*E^*$. The gain in income by labor is $(W^{**} - W)E^{**}$. But although labor gains by emigration, capital is hurt by it. The income received by capital is given by $Y - WL$. Diagrammatically, the value of national income, $Y$, is equal to the area below the demand for labor curve up to the level of employment. After emigration, the value of national income is $DBCE^{**}O$. As a result, after emigration, the income received by capital is $Y - WL = DBCE^{**}O - W^{**}E^{**} = DBW^{**}$. The income derived by capital before the emigration was $DBACW^*$. Hence, the income of capital declines by $DBACW^* - DBW^{**} = W^{**}BAW^*$.

In this context, emigration results in a redistribution of income from capital to labor. But the emigration has also a net, overall negative impact on the income of those left behind. Adding the loss of capital and the gain to labor leaves a net loss represented diagrammatically by the area $BAC$. Algebraically, this loss can be approximated by:

$$\frac{\Delta Y}{Y} = -(1/2)(\Delta W)M = -(1/2)S_L\eta_{LL}(M/L)^2$$

(43.7)

where $\Delta W$ is the change in wages ($\Delta W = W^{**} - W^*$), $M$ is the loss of labor ($M = -\Delta E = E^* - E^{**}$), and where $\eta_{LL} = -(\partial W/\partial E)(E/W)$ is the negative of the elasticity of the labor demand curve with respect to wages. The effects on the income of labor and capital are then: $(\Delta WE)/Y = S_L\eta_{LL}(M/L)(1-(M/L)) > 0$ and $(\Delta rK)/Y = -S_L\eta_{LL}(M/L)(1-(M/2L)) < 0$.

Although in this simplified setting emigration results in a gain to labor (the mobile factor), a loss to capital (the fixed factor) and a net loss to the overall economy, these conclusions can be reversed in more complex frameworks. First of all, the model is a partial-equilibrium, closed-economy model but most modern economies are both highly diversified and engage in substantial international trade. In a general equilibrium framework with trade, such as the Heckscher–Ohlin–Samuelson (HOS) model, emigration has no lasting impact on the source country, whether on income distribution or in terms of the net impact on economic welfare. The explanation is that the emigrants themselves induce a reduction in the demand for labor in the source country. As the migrants leave, they do place upward pressure on wages, as noted in the earlier model. But this makes employment in
capital-intensive sectors, such as heavy manufacturing, more profitable. As production shifts away from labor-intensive products and into capital-intensive sectors, the overall demand for labor in the economy shrinks. This puts downward pressure on wages, reversing the initial impact of the emigration. This result, where an outflow of labor leads to no change in wages, is based on the Rybczyński effect in the trade literature (see Bhagwati and Rodriguez, 1975; Rivera-Batiz, 1983).

But the assumptions of the HOS model under which this result is derived are stringent. Besides the assumptions of perfect competition and no distortions, the framework assumes the absence of externalities and increasing returns. In addition, the model does not distinguish between skilled and unskilled labor emigration. Once these assumptions are relaxed, the theoretical analysis may yield complex and ambiguous effects of emigration (see, for example, Dutt, 2005). Furthermore, when considering costs and benefits from migration one should consider as well the overall, global impact of migration flows. If there are net world gains, there are then strong reasons for governments in recipient and destination areas to coordinate their migration policies, so that both developing countries and high-income economies can profit from the migration flows (see Pritchett, 2006; Kapur and McHale, 2005).

The empirical evidence on the welfare and distributional effects of emigration on source countries is scant (the analysis of the impact of immigration has been more extensive). Recently, however, some studies have utilized Census data over time to examine the issue. Mishra (2006) examines emigration from Mexico to the USA, concluding that the outflow of Mexican workers to the United States between 1970 and 2000 has increased worker earnings in Mexico by 5.9 percent of GDP and has reduced the income of the owners of fixed factors by 6.4 percent of GDP, with a small, negative net impact on overall economic welfare of those left behind (see also Borjas, 2006). On the other hand, there are other potential impacts of emigration to consider that are not examined in this literature, as discussed next.

Remittances and the impact of emigration

One of the most visible impacts of the migrants on source countries is connected to the remittances that they send back home. These flows of resources have grown exponentially in recent decades. Measured in 2000 international purchasing power parity (PPP) dollars, developing countries received $50 billion in migrant remittances in 1980, but by 2005 this had multiplied to $605 billion (World Bank, 2007). The regions receiving the most remittances were East Asia and the Pacific, and South Asia (each receiving about $171 billion in 2005), followed by Latin America and the
Caribbean (with $86 billion), Europe and Central Asia ($78 billion), the Middle East and North Africa ($75 billion) and sub-Saharan Africa ($24 billion).

The significance of remittances in many developing countries can be seen by comparing the value of remittances with the value of the merchandise goods exported by the source countries. Table 43.2 shows how significant migrant remittances can be, rising in some countries to over 100 or 300 percent of exports.

Remittances clearly constitute an improvement in the standard of living for family members who are recipients of such income. And recent evidence suggests that remittances are connected to lower poverty levels (see Acosta et al., 2006; Adams, 2007). Some questions have been raised as to the extent to which the remittances simply raise current consumption instead of stimulating investment and future economic growth. Recently, however, a number of studies have documented that, first, a significant portion of so-called consumption spending consists of household investments in

<table>
<thead>
<tr>
<th>Country/region</th>
<th>Remittances in current $</th>
<th>Remittances in PPP-adjusted $</th>
<th>Remittances as % of merchandise exports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing countries</td>
<td>179,425</td>
<td>605,678</td>
<td>6.0</td>
</tr>
<tr>
<td>Mexico</td>
<td>21,772</td>
<td>32,222</td>
<td>10.3</td>
</tr>
<tr>
<td>India</td>
<td>19,843</td>
<td>105,564</td>
<td>26.2</td>
</tr>
<tr>
<td>Philippines</td>
<td>11,634</td>
<td>45,605</td>
<td>29.2</td>
</tr>
<tr>
<td>Lebanon</td>
<td>5,722</td>
<td>5,493</td>
<td>335.3</td>
</tr>
<tr>
<td>Morocco</td>
<td>4,221</td>
<td>12,325</td>
<td>42.5</td>
</tr>
<tr>
<td>Serbia/Montenegro</td>
<td>4,129</td>
<td>9,868</td>
<td>103.8</td>
</tr>
<tr>
<td>Pakistan</td>
<td>3,955</td>
<td>14,277</td>
<td>29.5</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>3,583</td>
<td>15,407</td>
<td>43.1</td>
</tr>
<tr>
<td>Brazil</td>
<td>3,540</td>
<td>7,682</td>
<td>3.0</td>
</tr>
<tr>
<td>Colombia</td>
<td>3,345</td>
<td>10,704</td>
<td>15.8</td>
</tr>
<tr>
<td>Egypt</td>
<td>3,341</td>
<td>8,018</td>
<td>43.5</td>
</tr>
<tr>
<td>Guatemala</td>
<td>2,592</td>
<td>5,962</td>
<td>88.2</td>
</tr>
<tr>
<td>El Salvador</td>
<td>2,564</td>
<td>5,718</td>
<td>77.8</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>2,471</td>
<td>6,671</td>
<td>43.0</td>
</tr>
<tr>
<td>Algeria</td>
<td>2,460</td>
<td>7,577</td>
<td>7.9</td>
</tr>
<tr>
<td>Jordan</td>
<td>2,288</td>
<td>5,010</td>
<td>58.9</td>
</tr>
</tbody>
</table>

Source: Data for remittances in current dollars are taken from World Bank (2007); other indicators are author's calculations using PPP adjustments and exports from World Bank (2007).
housing, automobiles and durable goods, whose long-term wealth-raising capacities are substantial; second, the use of remittances for community investment projects is not insignificant and also acts to stimulate local development; and, third, the multiplier effects of the increased consumption spending in generating local economic activity may be substantial (see, for example, de la Garza and Lowell, 2002; Adams, 2007).

On the other hand, remittances do tend to be spent largely on internationally non-tradable goods, which can result in rising prices of these goods compared to internationally traded goods, generating an appreciation in the value of the domestic currency, which adversely affects domestic export sectors (Rivera-Batiz, 1986). Evidence of this Dutch disease-type effect has emerged recently (see Amuedo-Orantes and Pozo, 2004; Acosta et al., 2007). A vicious cycle may develop, where emigration leads to remittances that then deteriorate domestic competitiveness and growth, leading to further emigration, and so on.

The impact of the brain drain

The contribution of migrant remittances for economic development must be weighted against any externalities generated by the migration flows. In countries where emigration leads to the loss of the most talented and skilled, the so-called brain drain, migration could result in significant negative externalities (see Bhagwati, 1979; Bhagwati and Rodriguez, 1975).

If the emigration of skilled labor is substantial and these workers are employed in local service sectors, the result can be acute shortages in the supply of essential services, from school teachers to professors and nurses. Note that the emigration of workers employed in sectors that produce exports and imports is not subject to these effects because local consumers can import these products from abroad when the laborers leave the country. But when the workers are employed in service sectors that produce internationally non-traded goods, the impact of emigration is more significant and potentially disastrous, because domestic consumers can only obtain those services locally (see Rivera-Batiz, 1982). If doctors and nurses emigrate, the supply of health services can collapse, resulting in higher prices and acute shortages. A brain drain can therefore reduce sharply the economic welfare of those left behind.

On the other hand, there may be positive externalities of skilled emigration on source countries. First of all, some researchers have recently suggested that a brain drain may actually raise the level of schooling of the population in the source country, at least in the long run. There are several reasons for this. One hypothesis is that the brain drain will raise the rate of return to education and, as a result, more young people in the country will decide to pursue higher education, thus raising educational attainment.
Indeed, the point that the emigration of skilled workers may lie behind the rising relative wages of skilled labor in many developing countries has frequently been made. However, evidence that these changes have stimulated local investments in human capital has not been produced yet (see Schiff, 2005).

Another possible positive externality associated with the brain drain relates to the fact that emigrants may generate international networks that could enhance the scientific and technological capacities at home. One example is the role that has been played by the Indian diaspora in Silicon Valley and elsewhere in the growth of India’s information technology sector. By generating greater flow of skills and information among countries, and by raising the reputation of a domestic sector internationally, this type of emigration can have significantly positive effects at home (see Kapur and McHale, 2005).

Implications for future research
Both the theory and empirical evidence on the determinants and consequences of international migration have grown enormously in recent years. This research matches the growing importance of migratory flows. But despite the progress, the literature so far has generally failed to consider the dynamic effects of emigration, focusing instead on analyzing – and estimating – static models. This is an especially relevant issue to discuss in relation to the brain drain since there are a number of possible mechanisms through which the emigration of the skilled can affect a country’s economic growth.

First of all, the mainstream analysis of economic growth, as developed by Solow and Swan, suggests that a drop in population growth should lead to an economic expansion for those left behind since the steady-state amount of capital per worker in the economy would rise, increasing per capita income. But there are a number of caveats to this story. First, the emigration of working-age people means that the dependency rate in the economy rises, which tends to absorb resources that would otherwise be dedicated to the accumulation of capital and economic growth. The working-age skilled emigrants are also more likely to have higher savings rates than the rest of the population. Indeed, the empirical evidence available is consistent with a negative impact of a brain drain due to these demographic effects (Bloom and Williamson, 1998).

Another issue is that the emigration of skilled labor may reduce the human capital available to those left behind. This can potentially have a devastating effect on economic growth. One of the main sources of economic growth is technological change, which depends to a large extent on human capital (Romer, 1990). But if human capital flees a country, then the
ability of those left behind to sustain innovation and technical change may be compromised, thus reducing an economy’s economic growth (Rivera-Batiz, 1996).

On the other hand, a growing literature suggests that in countries where the quality of the public sector governance is low, many educated workers tend to be employed in activities that are not necessarily high-productivity activities. In other words, in economies where the government is highly bureaucratic, where corruption is rampant, and the rule of law does not apply, many highly educated workers will not be able to be gainfully employed (see Rivera-Batiz, 2002). In this case, the potentially negative impact of the brain drain on technological change and, therefore, on economic growth diminishes. It is a matter for future research to examine theoretically and empirically in greater detail how emigration is connected to technical change and economic growth.

References


Amitava Krishna Dutt and Jaime Ros - 9781848442818
Downloaded from Elgar Online at 07/19/2019 02:48:07AM
via free access


International technology transfer: the role of foreign direct investment

Amy Jocelyn Glass and Kamal Saggi

Introduction

International technology transfer (ITT) refers to any process by which a party in one country gains access to technical information of a foreign party and successfully absorbs it into its production process. The importance of ITT for economic development is widely recognized, and it has been argued that barriers to technology adoption help explain the income gap between developed and developing countries (Parente and Prescott, 1994). Such barriers include regulatory and institutional constraints that entrepreneurs must overcome as well as low levels of human capital. Furthermore, the technology frontier is a moving target – new technologies are continually being introduced. To close the technology gap, developing countries must adopt new technologies, at a faster rate than they are being created. Both market forces and government policies have an important role to play in accomplishing this formidable task.

At the heart of ITT is the exchange of information and knowledge. Technology may be codified (for example, in blueprints) or uncodified (for example, know-how of engineers). It may be embodied in products or people, or disembodied in ideas or services. ITT often occurs between unrelated partners in market-based transactions. However, information also flows internationally between related parties on a non-market basis, within the boundaries of firms and joint ventures. Given the multifaceted nature of technology transfer, there exist numerous channels through which technology flows across international boundaries. One major channel is trade in goods and services. All exports bear some potential for transmitting technological information. Trade in capital goods and technological inputs can directly improve productivity by being integrated into production processes. Another major channel of ITT is direct trade in knowledge via technology licensing, which may occur within firms, among joint venture partners or between unrelated firms. The focus of this chapter is the channel of foreign direct investment (FDI).

All these channels may facilitate imitation and reverse engineering. Because imitation does not require compensating technology owners, it can be an attractive option for developing economies. As Hoekman et al. (2005)
note, the temporary migration of students, scientists and managerial and
technical personnel to universities, laboratories and conferences also plays
an important role in encouraging ITT. Furthermore, ITT can also result
from the temporary movement of professionals and other service suppliers
who enter a developed country to perform specific services and in the
process acquire additional knowledge and skills that are transferred back
to the home country upon completion of the contract. While such channels
of ITT are no doubt of crucial importance we do not discuss them here, in
order to limit the scope of this chapter.

Although use of the word ‘transfer’ in the phrase ‘international technol-
ogy transfer’ seems to suggest that the process of ITT is somehow smooth
and automatic, nothing could be further from the truth. The fact that devel-
oping countries lag behind the technology frontier merely creates the
potential for ITT. For ITT actually to occur, providers and acquirers of new
technologies have to undertake deliberate and often significantly large
investments.

Investment costs are not the only hurdle facing ITT. The market for tech-
nology is hampered by at least two significant market failures: the presence
of asymmetric information and/or market power. In fact, the superior
information possessed by sellers when protected by intellectual property
rights is often what creates the market power. By keeping transactions
within one firm, technology transfer via FDI can lessen some of the
difficulties that confront arm’s-length exchange of technology, yet many
problems remain. Even within the same firm, Teece (1976) finds the costs
of transferring technology to a foreign plant average 20 percent of the total
investment required for the plant.1

Fuller benefits for local technological capacity are realized if the technolo-
gies introduced from abroad diffuse locally. The first step – getting the tech-
nology into the country – is ITT, and the second step – getting the technology
into the possession of local firms – is what we call technology diffusion. While
the first step is typically a deliberate act, the second step can often be an unin-
tended consequence. What makes the role of FDI especially important is that
multinational companies dominate global research and development (R&D)
and therefore act as important conduits of ITT. Yet multinationals are in the
business of maximizing profit, not the economic development of host coun-
tries. However, available models and empirical studies argue that achieving a
balance between the objectives of multinationals and host countries is not as
difficult as it may appear on casual observation.

**Foreign direct investment as a channel of international technology transfer**

Today, intra-firm trade (that is, trade between subsidiaries and headquar-
ters of multinational firms) accounts for roughly one-third of total world
trade, and sales of subsidiaries of multinational firms exceed worldwide exports of goods and services. Thus, FDI is the dominant channel through which firms serve customers in foreign markets. While much of FDI occurs between industrial countries, developing countries are becoming increasingly important host countries for FDI. Approximately 33 percent of the global stock of FDI today is in developing countries (UNCTAD, 2003).

FDI is growing in importance as a channel of ITT. Multinational activity occurs primarily in industries that are characterized by a high ratio of R&D to sales and by large shares of professional, scientific and technical workers (Markusen, 1995). A basic tenet of the theory of the multinational firm is that such firms rely heavily on intangible assets, such as superior technology and well-established brand names, to offset the logistical and other disadvantages of operating in multiple countries as well as to compete successfully with local firms that are better acquainted with the host-country environment. In 1995, of all transactions in royalty and license fees, transactions within the same firm made up in excess of 80 percent, so most explicit trade in technology takes place within multinational firms (UNCTAD, 1997).

Virtually all empirical studies of FDI find that foreign-owned plants in developing countries are typically more productive than purely domestic ones. For example, a recent paper by Arnold and Javorcik (2005) provides direct evidence on the impact of FDI. Using data from Indonesia’s manufacturing sector during the period 1983–86, the authors focus on the effects of acquisitions of local firms by foreign ones. They find that foreign ownership leads to significant improvements in acquired plants: after three years, the acquired plants outperform the control group in terms of total factor productivity by 34 percent.

ITT through FDI can be either horizontal or vertical in form. When horizontal, FDI transfers the full technology needed to produce the good. When vertical, different stages of the production process are split across countries so only the technology for the stage (or stages) being produced in the host country is transferred. For developing countries, often the more labor-intensive stages are shifted abroad.

Of course, instead of opting for FDI, a firm may sometimes decide to contract with a local firm for production of components in a developing country rather than forming its own production subsidiary there. Due to the participation of local firms, licensing and joint ventures can yield both ITT and technology diffusion. As might be expected, it has been shown that costs of transferring technologies across countries work against FDI (and licensing) as mode choices. Norback (2001) confirms, using Swedish data, that high costs of technology transfer discourage production abroad in favor of exports.
Studies such as Mansfield and Romeo (1980) and Smarzynska (1999) have found that newer technologies are transferred through FDI, whereas older technologies are transferred though joint ventures and technology licensing. Mode choice may be one way that firms attempt to maintain their technological advantage by avoiding modes with high technology diffusion until technologies become somewhat dated. Or perhaps costs of transferring technologies via arm’s-length channels are larger for newer technologies due to greater information asymmetries. Moral hazard considerations can also be important in this context. For example, Ramachandran (1993) has shown that subsidiaries receive greater resources than partially licensor-owned or independent firms once the incentives for both sides to invest in transferring technology are considered. Strategic incentives can also reinforce moral hazard and asymmetric information considerations: Fosfuri (2000) constructs a model with the feature that firms strategically use the vintage of technology to deter imitation by licensees, so that more recent technology is transferred to affiliates than to outsiders (see also Saggi, 1999).

**Foreign direct investment and local technology diffusion: friends or foes?**

An important consequence of FDI is that shifting production to a developing country can reduce technology adoption costs for indigenous local firms. The degree to which imitation costs are lowered by FDI might be higher for process than for product technologies. For product technologies, reverse engineering may be the main way that imitation costs are reduced. Since better process technologies tend to be difficult to deduce from inspection of the final good, first-hand experience with the technology may be required. Multinational firms bring production to the host country, providing workers with experience using the new technology. Workers then often leave to work for rival local firms or to start their own firms. Either way, such worker turnover generates knowledge flows that may lead to local firms adopting some aspects of the ways in which the new technology is better than the old. Also, any degree to which multinational firms adapt technologies to the local economic environment reduces costs of technology adoption for local firms.

Since technology advantages are often needed to survive as a multinational firm, why do multinationals not do anything and everything possible to curtail diffusion of their technologies to rival firms? As argued in Glass and Saggi (2002a), when the gains to local firms are great, the costs of preventing leakage of technologies to rivals (the wage premiums required to keep workers from leaving) are apt to be great as well. Additionally, the presence of multiple multinational firms in an industry likely leads to positive externalities among them: costly efforts undertaken
by any one multinational to curtail spillovers to local competitors would benefit all multinationals. A multinational firm that seeks to protect its technology through litigation, for example, bears the full cost but not the full benefits of its action. As a result, multinationals might very well under-invest in activities that curtail the local diffusion of their technologies.

Focusing on vertical technology transfer from a multinational to its suppliers, Pack and Saggi (2001) have shown that technology diffusion among suppliers can benefit foreign firms sourcing components. Thus, fully integrated multinational firms would be expected to be more averse to technology diffusion than firms involved in arm’s-length production deals with local firms. Mexico’s maquiladoras appear to have benefited from the transfer of sophisticated production techniques and backward linkages, especially in the automobile industry. Goh (2005) finds, however, that diffusion of knowledge to other potential suppliers can either encourage or discourage technology transfer depending on the incumbent supplier’s cost of technological effort. Using firm-level data from Lithuania, Javorcik (2004a) finds evidence of spillovers from foreign affiliates to their local suppliers in upstream sectors, but only for projects with shared domestic and foreign ownership (not for fully foreign-owned investments).

Evidence regarding whether technologies transferred by multinational firms diffuse to competing local firms is mixed. Finding that sectors with more foreign involvement have higher productivity or faster productivity growth could stem from FDI being attracted to those sectors rather than FDI improving productivity or accelerating productivity growth. Plant-level studies are required to help alleviate any selection bias in industry-level studies. Haddad and Harrison (1993) find that foreign firms have higher levels of total factor productivity (TFP) but lower TFP growth than domestic firms in Morocco. A stronger positive effect of FDI in low-tech sectors than in high-tech sectors may indicate that local firms in high-tech sectors lack absorptive capacity. Or perhaps multinationals in high-tech sectors take more actions to preserve their technological advantages.

Aitken et al. (1996) provide an empirical assessment of the hypothesis that technology spillovers ought to increase the marginal product of labor, and this increased productivity should show up as higher wages. Their study employs data from manufacturing firms in Venezuela, Mexico and the United States. For both Mexico and Venezuela, a higher share of foreign employment is associated with higher overall wages for both skilled and unskilled workers. Furthermore, royalty payments to foreign firms from local firms are highly correlated with wages. Most importantly, the study finds no positive impact of FDI on the wages of workers employed by domestic firms. In fact, the authors report a small negative effect for domestic firms, whereas the overall effect for the entire industry is positive.
These findings differ from those for the United States, where a larger share of foreign firms in employment is associated with both a higher average wage as well as higher wages in domestic establishments. Putting Aitken et al.’s (1996) findings into the context of previous work, it is clear that wage spillovers (from foreign to domestic firms) are associated with higher productivity in domestic plants. Conversely, the absence of wage spillovers appears to accompany the existence of productivity differentials between domestic and foreign firms.

Using annual census data on more than 4000 Venezuelan firms, Aitken and Harrison (1999) provide a plant-level test of the spillover hypothesis. They find a positive relationship between foreign equity participation and plant performance, implying that foreign participation indeed benefits local plants that receive such participation. However, this own-plant effect is robust for only small plants, that is, those plants that employ fewer than 50 employees. For larger plants, foreign participation results in no significant improvement in productivity relative to domestic plants. More interestingly, they find that productivity in domestic plants declines with an increase in foreign investment – that is, they find evidence of negative spillovers from FDI. The authors suggest that these could result from a market-stealing effect: foreign competition may have forced domestic firms to lower output and thereby forgo economies of scale.

However, the results of Haskel et al. (2007) contrast with those of Aitken and Harrison (1999). Haskel et al. (2007) use plant-level panel data for all UK manufacturing from 1973 to 1992 to re-examine the issue of spillovers from FDI. As the authors note, there can be little doubt that local firms in the UK possess sufficient absorptive capacity to benefit from the introduction of newer technologies by multinationals. So if spillovers do not materialize, they cannot be attributed to the limitations of domestic firms. Across a wide range of specifications, the authors find that there are positive spillovers from FDI at the industry level. More precisely, they find that a 10 percent increase in foreign presence in a UK industry raises the total factor productivity of that industry’s domestic plants by about 0.5 percent. However, the authors also note that the large tax breaks and incentive packages given to multinationals seem out of proportion relative to the magnitude of spillovers they generate.

While some studies have cast doubt on the optimistic view that FDI generates positive spillovers for local firms, others have reached different conclusions. Regardless of one’s view of these findings, it is worth stressing that domestic firms should be expected to suffer from an increase in competition that often results from FDI; in fact, part of the benefit of FDI is that it can help weed out relatively inefficient domestic firms. Resources released in this process will be put to better use by foreign firms with superior
technologies, efficient new entrants (both domestic and foreign) or some other sectors of the economy. However, such resource reallocation does not occur instantaneously. Existing studies of spillovers do not cover a long enough period to be able to determine accurately how FDI affects turnover rates (entry and exit). Furthermore, horizontal studies miss spillovers that may result from FDI in industries other than the one in which FDI occurs.

In a critical discussion of the plant-level studies of horizontal spillovers from FDI, Moran (2004) argues that there is a substantial difference in operating characteristics between subsidiaries that are integrated into the international sourcing networks of the parent multinationals, and those that serve protected domestic markets and are prevented by policy restrictions (such as mandatory joint venture and domestic content requirements) from being so integrated. These different operating characteristics include size of plant, proximity of technology and quality control procedures to industry best practices, speed with which production processes are brought to the frontier, efficiency of operations and cost of output. He argues that while the former have a positive impact on the host country, often accompanied by vertical backward linkages and externalities, the latter may actually have a negative impact. Drawing upon a wealth of case studies and econometric evidence, Moran (2004) argues that this contrast in performance holds across different industries, countries and time periods. He astutely notes that the failure to differentiate between export-oriented FDI and import-substitution FDI, or between foreign investors free to source from wherever they wish and foreign investors operating with domestic content requirements, or between foreign investors obliged to operate as minority shareholders and those with whole- or majority-ownership, accounts for the inability of earlier studies to isolate the influence(s) of FDI on host-country welfare.

We noted earlier that arm’s-length technology transfer is usually of lower quality than its intra-firm counterpart. But is greater involvement of local firms, such as in the form of joint ventures, more likely to lead to diffusion? While this appears plausible, there is little empirical evidence in support of this idea. For example, Blomstöm and Sjoholm (1999) find that the degree of foreign ownership did not affect the productivity of local partners or spillovers to domestic firms in Indonesia for 1991. Yet having any foreign participation at all did matter: plants with no foreign participation were less productive. These findings could represent selection at the plant level – FDI is attracted to more productive plants; or a threshold effect – that foreign participation, not the degree of participation, is what matters most.

Although the extent of technology diffusion resulting from FDI is unresolved, that FDI stimulates economic growth in the host country enjoys strong empirical support. Balasubramanyam et al. (1996) find the
growth-stimulating effects of FDI are stronger for countries that pursue export promotion rather than import-substitution policies. So trade policy seems to affect the benefits of FDI, although trade orientation could proxy for other unmeasured differences across countries. For export-promoting countries, FDI stimulated growth more than domestic investment. Borensztein et al. (1998) find that FDI contributes more to economic growth than domestic investment for countries that have a sufficient stock of human capital. Countries with insufficient human capital presumably lack the ability to absorb technologies. Xu (2000) finds that countries need to achieve a minimum level of human capital in order for the technology transferred by US multinational firms to contribute to productivity growth, but most less-developed countries do not satisfy the required threshold.

**Policy options for acquiring and absorbing new technologies**

Separating the concept of ITT from that of local technology diffusion is especially important for analyzing policy choices. When assessing the impact of a policy option, the effects on both ITT and technology diffusion should be considered. Some policies might promote ITT but not technology diffusion. Others might promote technology diffusion but then adversely affect ITT through discouraging FDI. Multiple policy instruments are likely to be needed to achieve the ideal combination of ITT and technology diffusion. Too much technology diffusion, and there may not be much technology to diffuse since the potential for local spillovers may deter FDI. Too much ITT, and few of the advanced technologies may ever be used by indigenous firms.

Many countries such as Japan, South Korea and China have historically restricted FDI, often in favor of technology licensing or joint ventures. Foreign firms were often required to share technologies with local firms in order to conduct business in Japan. It is difficult to judge whether countries restricting FDI would have performed better or worse than if they had taken a more liberal approach, since the counterfactual is not observed. The findings that newer technologies are transferred through FDI rather than through joint ventures and technology licensing call into question the wisdom of policies that favor technology licensing or joint venture over FDI. Even if more technology diffusion results, the technology obtained may be far below the state of the art. It is possible that policy interventions could act to improve the terms of licensing contracts for local firms by removing alternatives (or making the alternatives far less desirable) for the foreign firm.

More recently, developing countries have become quite eager to attract FDI. Part of this eagerness may stem from enhanced awareness that FDI can serve as an important channel of ITT, although employment issues
surely also play a role. Many countries, both developed and developing, offer fiscal and financial incentives to attract FDI. Eliminating restrictions on FDI is likely to be beneficial, at least at the world level, since foreign firms would be freer to choose between modes without interference. However, when it comes to promoting FDI, a few words of caution may be in order. You can have too much of a good thing. Similar to free trade being best and export subsidies being distortionary, care is needed to avoid overstimulating FDI. Incentives could lead to FDI being attracted to the wrong countries—countries where costs will be higher than alternative locations. Excessive competition for FDI between locations could bid away all potential benefits.

As noted above, empirical evidence on technology spillovers from FDI is mixed, so benefits to local firms might not be realized. If governments use incentives to try to obtain the right kind of FDI, one has to question whether the government can indeed pick industries with the best potential for spillovers. Adequate human capital and infrastructure are needed for absorption, and any bureaucratic impediments to technology adoption should be minimized. So much focus on FDI risks overlooking opportunities to improve the diffusion of technologies.

Stronger protection of intellectual property (IP) has often been suggested as a means of attracting FDI. The thought is that firms will avoid FDI in favor of exports to countries with weak protection of IP (although they may also shift from licensing to FDI). Theoretical studies such as Glass and Saggi (2002b) and Glass and Wu (2007) based on the quality-ladder model of growth cast doubt on the idea that FDI rises with stronger IP protection, once the repeated nature of innovation is captured. Taylor (1993) has suggested that poor protection of IP may lead firms to mask their technologies in order to make them harder to imitate.7

However, in a recent paper Branstetter et al. (2006b) have argued that in a variety expansion North–South product cycle model with endogenous Northern innovation, Southern imitation and FDI, intellectual property rights (IPR) reform in the South leads to increased FDI from the North, as Northern firms shift production to Southern affiliates. Furthermore, this increased FDI drives an acceleration of Southern industrial development, as the South’s share of global manufacturing and the pace at which production of more recently invented goods shifts to the South both increase. In addition, their empirical results show that US-based multinational corporations (MNCs) expand the scale of their activities in reforming countries after IPR reform, and this effect is disproportionately strong for affiliates whose parents rely strongly on patented intellectual property as part of their global business strategy. Furthermore, they also provide evidence from highly disaggregated trade data that suggests that the expansion of multinational activity leads to a higher net level of production shifting
to developing countries, more than offsetting any possible decline in the imitative activity of indigenous firms.


Stronger IP protection may be more important for technology licensing than for FDI due to the risk of opportunistic behavior and difficulty enforcing contracts. Yang and Maskus (2001) consider a model in which stronger IP protection increases the licensor’s share of rents and reduces the costs of licensing contracts. Thus, better IP protection may indeed stimulate licensing and technology transfer.

**Concluding remarks**

International technology transfer is a complex, multifaceted phenomenon. In this short chapter, we have chosen to highlight the role foreign direct investment and multinational corporations play in the process of international technology transfer, paying little attention to international trade of goods and services, movement of workers and professionals, and other related phenomenon. While we do feel that FDI is the most important channel, we do not mean to imply that the other channels are not of considerable importance as well.

That multinational companies are pivotal in introducing new technologies to host countries is fairly well established. However, lagging countries have not only to obtain foreign technology but also to learn how to use it to its fullest potential. In this context, we have found it useful to make a distinction between initial international technology transfer and subsequent technology diffusion within host countries. This distinction can be important since some policies could promote technology transfer but deter technology diffusion, or promote technology diffusion but deter technology transfer. With respect to the contribution of FDI, there is some good news and bad news. First the bad news: multinationals will usually lose from further horizontal diffusion of their technologies and should be expected to take actions that thwart that process. The good news is that technology transfer to local suppliers is compatible with the motives of multinationals and a plethora of empirical evidence indicates that vertical linkages between multinationals and their local suppliers play a crucial role in the industrial development of host countries.
An important policy conclusion of this analysis is that host countries are better off facilitating processes that are compatible with the motives of multinationals. In other words, a developing country should perhaps be less concerned about being able to produce an automobile of its own and more concerned about developing a competitive network of suppliers that can serve (and gain from) well-established foreign firms. It is in this mutually beneficial exchange that the most productive policy intervention might lie. Of course, if both sides are indeed willing participants, policy intervention required would be ‘light’ as opposed to ‘heavy’. Furthermore, it would not be targeted in nature. Instead it would ensure that local businesses have access to adequate infrastructure and skilled workers, and that their expansion or downsizing decisions are not hampered by burdensome regulations. In our view, this is another plus of pursuing policies that take proper account of the incentives multinational firms have (and do not have) to encourage industrial development in host countries.

Notes
1. See also Mansfield and Romeo (1980) and Ramachandran (1993).
2. That FDI leads to ITT may seem obvious, yet Glass and Saggi (1999) have argued that whether FDI creates ITT in aggregate depends crucially on whether substitute channels of ITT, such as imitation, exist. FDI could merely displace imitation that otherwise would have occurred, leaving ITT essentially unchanged. See also Glass and Saggi (1998) for a model in which narrowing the technology gap induces multinationals to transfer state-of-the-art technologies.
3. A nice feature of their approach is that they control for the self-selection problem – that is, while it is true that multinationals typically acquire firms that are relatively more productive, they further contribute to the future productivity of acquired firms.
4. See Saggi (2002) for an extensive discussion of this literature.
5. Cheng et al. (2005) examine the impact of the ability of workers to absorb foreign technologies on the extent of production by multinational firms.
6. Nevertheless, on balance, Aitken and Harrison (1999) find that the effect of FDI on the productivity of the entire industry is weakly positive. They also note that similar results are obtained for Indonesia, except that the positive effect on own plants is stronger, whereas the negative effect on domestic plants is weaker, suggesting a stronger overall positive effect.
7. See also Taylor (1994) for the effects of IP protection on technology transfer.
8. In a related paper, Branstetter et al. (2006a) have shown that there is a significant increase in technology transfer following reforms among affiliates of firms that make extensive use of the US patent system.

References
Introduction

Those international financial institutions (IFIs) most important for development nowadays were not established for this purpose. Keynes drafted the Bretton Woods system as the allied counter-proposal to Nazi Germany’s ‘New Order’ of a European Economic Community, basically structured like the present EU, even including the prospect of a European currency union, though no European ‘Parliament’ without real parliamentary powers (Raffer and Singer, 2001 [2002, 2004], pp. 1–2). The OECD (1985, p. 140) describes the initial tasks: ‘The IBRD was there to guarantee European borrowing in international (North American) markets; the IMF was there to smooth the flow of repayments.’ Because Southern delegations demanded resources for development, ‘and Development’ was added to ‘International Bank for Reconstruction’ (IBRD).

When the communist threat made European economic recovery a very important issue, other programs, mainly the Marshall Plan, took over. Apparently, IBRD loans were considered inappropriate for successfully reconstructing Europe, which raises questions whether they can develop much poorer countries. The addition ‘and Development’ allowed the IBRD to turn South (Caufield 1998, p. 56). Somewhat later the IMF also shifted totally to the South in spite of its more general mandate.

The International Trade Organization, the last pillar of the Bretton Woods system, did not come into being. Only the ‘provisional’ General Agreement on Tariffs and Trade (GATT) regulating trade in manufactures was established. Keynes’s ideas of developmental interest, such as stabilizing commodity prices, were not implemented. The GATT’s successor, the World Trade Organization (WTO) is characterized by a sharply different view on development, based on the Washington Consensus.

The Bretton Woods institutions (BWIs) and the WTO reflect global power asymmetries. Unlike the Inter-American Development Bank, where regional developing members must have 50.005 percent of total votes, industrialized countries (ICs) control the BWIs by clear voting majorities. The WTO’s one-country-one-vote principle is in practice overturned by the consensus approach, bilateral pressure on developing countries (DCs), and
the so-called ‘Green Room’, the practice of backroom negotiations to which only a few countries are invited, whose results are then presented to the rest for ‘consensus’, usually under time and other pressure. At Seattle this triggered strong protests. Southern discontent has also surfaced during the Doha Round.

Shifting focus

The IBRD (2005) and its low-income-country window IDA (International Development Association) declare ‘global poverty reduction and the improvement of living standards’ as their ‘mission’. The bank is not ‘a bank in the common sense’ (ibid.). These statements differ fundamentally from its initial businesslike approach precluding financing social activities. It not only refused ‘messing around with education and health’ or a water treatment plant, but even forced Columbia not to accept a French loan for waterworks (Caufield, 1998, p. 64). McNamara was the first president focusing on poverty. McNamara had the merit of giving credibility to the idea that helping the poor is not wasting resources, but makes economic sense. Brought about by Euromarket lending, the debt problem shifted the orientation of the BWIs fundamentally. The IBRD and IDA moved from project financing towards program lending, which should be exceptional pursuant to their Articles of Agreement.

The IMF was not designed as a development organization, but to enable members of the Bretton Woods system with short-term balance-of-payments problems to stay within the agreed parity bands. The demise of Bretton Woods left the IMF with very few remaining tasks, such as the Compensatory Financing Facility (CFF). These could have been transferred to another institution. Arguably, the IMF should have been dissolved. The debt crisis 1982 provided a new role.

Until the Cologne Summit entrusted both BWIs with the Highly Indebted Poor Country Initiative II (HIPC II), which explicitly includes anti-poverty measures, the IMF had usually and rightly stated that this was not its mandate. Turning debt managers, both BWIs gained strong control over development policies, characterized by the term ‘conditionality’. Conditionality became part of the IMF’s statutes as late as 1969. Until its introduction the IMF fulfilled a highly useful role of emergency lending. Forced to get the BWI ‘seal of approval’ in order to get urgently needed new loans, DCs in distress have to fulfill a wide range of conditions, not all visibly connected to economic necessities. BWI loans might carry over 100 conditions, which raises questions whether all can be complied with. All debt relief measures have increased their leverage.

Introduced to compensate export earnings shortfalls beyond the member’s control, the CFF illustrates the strengthening of conditionality
over time. Initially, a statement sufficed to cooperate with the IMF where required. Eventually, the ‘Fund has increasingly come to the realization that even though a country’s export shortfall was both “temporary” and largely beyond its control the country might still have balance-of-payments difficulties attributable to inappropriate policies and that large amounts of unconditional credit might cause the country to delay adopting needed policy adjustments’ (Polak, 1991, p. 9). Polak (1991, p. 12), an influential IMF theoretician, is outspoken: ‘The purpose of the Fund’s conditionality is to make as sure as possible that a country drawing on the Fund’s resources pursues a set of policies that are, in the Fund’s view, appropriate to its economic situation in general and its payments situation in particular’ – even if the country’s economic policy is not at all the reason for temporary problems.

The IBRD has never made unconditional loans. Conditions requiring policy changes have even been attached to projects (cf. Mosley et al., 1991, p. 27). ‘Programme lending’ increased conditionality: ‘[T]he Bank felt that it needed a place at the top policy-making table’ (Mosley et al., 1991, p. 34) beyond what it could expect from project monitoring. Stern (1983, p. 91), the IBRD’s Senior Vice-President, praised structural adjustment lending as enabling ‘the Bank to address basic issues of economic management and of development strategy more directly and urgently’, as a ‘unique opportunity to achieve a comprehensive and timely approach to policy reform’ (Stern, 1983, p. 104), the response to a ‘feasible . . . call for increased sacrifices’ (Stern, 1983, p. 91).

The IMF started adjustment measures in sub-Saharan Africa after 1973. After decades of adjusting debtor countries and ‘appropriate’ development strategies, no country regained economic viability. Attempts to prove success econometrically were given up long ago. Often no statistically significant difference between program and non-program countries was found. Khan (1990), an IMF econometrician, found significantly reduced growth in program countries; as Polak (1991, p. 42) points out, a predicted reduction in the growth rate of at least 0.7 percent of GDP each year a country had an IMF program. After years of ‘Structural adjustment’ the IBRD (1989, p. 6) found a lack of ‘an integrated analytical framework to understand better the links between a program and its expected macroeconomic outcomes’. In spite of little success, official creditors have steadily increased the role of the BWIs.

Rodrik’s (1996) analysis of neoliberal reforms might explain this behavior better. He sees the debt crisis as an opportunity seized for a ‘wholesale reform of prevailing policies’, offering the chance ‘to wipe the slate clean and mount a frontal attack on the entire range of policies in use’ (ibid., p. 17). A crisis brought about by overspending, overlending and the sudden
change of economic policy in ICs, which sent interest rates skyrocketing, was declared to stem from disliked policies in DCs.

The BWIs have strongly pushed for policy change while refusing to subject themselves to market mechanisms and basic legal principles. This produced an economically inverted incentive system absolutely at odds with market incentives. Although IFIs (co-)determine their clients’ policies, they refuse to share the risks involved appropriately, insisting on full repayment, even if damages caused tortuously by their staff should have occurred (Raffer, 2004). Borrowers have to pay for such damages. IFIs may gain financially from tortuous behavior and errors by extending new loans necessary to repair damages done by prior loans. New and larger crises increase their importance. Grave negligence creating damages leads to new loans correcting such damages, increasing IFI incomes. The IMF’s proposal of a Sovereign Debt Restructuring Mechanism would have perpetuated this situation, granting IFIs de jure preferred creditor status and increasing the Fund’s role in debt management (cf. Raffer, 2005).

The WTO
The WTO (2005a) presents itself as a ‘negotiating forum’, ‘a set of rules’ and a help ‘to settle disputes’. It sees its mandate in preventing ‘self-defeating, destructive . . . protectionism’ (WTO, 2005b). Theoretically it rests its case solely on comparative advantages (ibid.), although this theorem only works in a two-countries-two-goods world and constant returns to scale are necessary to guarantee welfare gains (cf. Viner, 1937, pp. 470–79).

Mattoo and Subramanian (2005, p. 19) argue that the WTO seems to be the ‘best vehicle’ for advancing Northern corporate interests, seeking ‘the opening of markets in developing countries for manufactured goods’. While able to retain high protection where wanted, ICs have managed to restrict or outlaw protection where it could be in the interest of DCs. The principles of the WTO are in many important respects the very opposite of the ideas behind the creation of UNCTAD.

In spite of rhetoric, tariff escalation continues to exist: ‘OECD tariffs on finished industrial products are about eight times higher than on raw materials . . . These barriers delay entry into the export-oriented industries, which are most accessible to developing countries’ (OECD, 2000, pp. 31–2). Average tariffs on Southern manufactured exports continue to be a multiple of those on imports from other ICs. The Doha Development Round again pressures for more market access of ICs. Market access for non-agricultural products is, for example, hardly of overwhelming developmental interest. DC interests can hardly be identified except, arguably, in trace elements. Special and differentiated treatment practically abolished by the WTO is again discussed. The Doha Declaration calls for a review to
strengthen it, without practical noteworthy effects so far. The WTO’s record has triggered proposals for making it more development-friendly, even from within the BWIs (Hoekman, 2005).

Heavy subsidizing of domestic Northern agriculture and agro-exports conform to WTO obligations. US exports are priced 65 percent below production costs. The EU exports sugar and beef at less than half their production costs. WTO cotton and sugar panels legally established that ICs had failed to abide by the loose rules on subsidies they had crafted during the Uruguay Round. ‘Comparative access to subsidies, not comparative advantage’ (Oxfam, 2005, p. 9) shapes ‘world markets’. Institutions interlink. In the name of economic efficiency the IBRD pressured Mali to pay local cotton producers this subsidy-determined ‘world market price’ in 2004. The government ultimately refused to bankrupt domestic peasants.

‘Voluntary Export Restrictions’ have become legalized. The Trade-Related Investment Measures (TRIMS) treaty restricts developmental options to industrialize. Politics successfully pursued by Asia’s dragons are now outlawed. The Trade-Related Intellectual Property Rights (TRIPS) agreement does not strictly speaking protect intellectual property, because the host of tribal knowledge in many DCs remains unprotected. It ‘increased the monopoly power of patent holders and limits the ability of generic producers to compete’ (Mattoo and Subramanian, 2005, p. 20), enabling pharmaceutical companies to raise prices far above what many poor people can afford. DCs are pressured not to use those WTO safeguards to protect public health, which the USA or Canada have used. Over the years the USA has threatened trade sanctions against countries revising their legislation to incorporate TRIPS safeguards, such as Thailand and South Africa. Complaints were launched against Brazil and South Africa using their WTO rights to fight AIDS via affordable drugs. DC protests brought about change for the better. At Doha the right of WTO members to use, to the full, the provisions in the TRIPS agreement was ‘reaffirmed’. This would have been unnecessary if DC rights under TRIPS had been respected.

Dispute settlement was one of the main chocolates on the tray to convince smaller and weaker countries to sign the WTO treaties, promoted as a rule-based system protecting the rights of the weak. It had been argued that the WTO would substitute bilateral (and GATT-violating) measures such as the US Super 301, a unilateral measure implemented in breach of international treaties. Meanwhile, the WTO accepted Super 301 (WTO, 2000, pp. 67–8). A review process had been agreed at Marrakesh, but ICs blocked any change. Although Doha is called a ‘single undertaking’ encompassing all issues, dispute settlement, a main concern of DCs, is not part of it.
Dispute settlement has no mandate to protect members’ rights. Article 3.7 of the Understanding on Rules and Procedures Covering the Settlement of Disputes states: ‘Before bringing a case, a Member shall exercise its judgment as to whether action under these procedures would be fruitful. The aim of the dispute settlement mechanism is to secure a positive solution to the dispute.’ The probability of success, not the rule of law, is explicitly established as the guiding principle. This is a unique and unfortunate approach. Suing big guys is often fruitless, not least because of the way ‘relief’ is organized. There is no right to compensation for damages suffered by violations of contractual obligations. The winning party may be authorized to suspend WTO concessions subject to strict and constraining rules. After winning against the USA, Antigua, whose exports of Internet games were blocked in breach of contract, was authorized to sanction the USA. The positive side is that DCs can and do win cases and can often obtain relief.

Less agricultural export subsidies and reviewing dispute settlement are not the only unfulfilled promises made while the signature of DCs was coveted. Before Marrakesh, the countries of the Organisation for Economic Co-operation and Development (OECD) apparently perceived a need to assure DCs of relief measures against expected higher food prices. After ratification net importers were referred to existing BWI facilities. A list of net food-importing developing countries exists meanwhile, but being listed does not confer automatic benefits. Donors and international institutions want a role to play. Estimates of the WTO’s benefits to DCs were ‘exaggerated and its costs were underplayed’, ‘liberalization assumptions were disconnected from what the [Uruguay] Round actually achieved’ (Mattoo and Subramanian, 2005, p. 21). Calling Doha a ‘Development Round’ continues this tradition.

The WTO is used to cleanse trade relations from disliked historical obligations. WTO compatibility is presently the EU’s argument to remove those remnants of the Lomé framework that had been adopted in favor of and due to pressure by DC signatories. The Final Act of the Uruguay Round also increased IFI influence. IFIs are to cooperate more closely with the WTO to achieve greater coherence in economic policy. Mali’s example above illustrates how this is done. Structural adjustment lending forced DCs to open and liberalize their economies to the extent of making the ‘WTO process a “victim” of the success of the World Bank and the IMF’ (Mattoo and Subramanian, 2005, p. 20). The WTO treaties are now useful to lock in liberalization, to restrict the options of DCs. Simultaneously, bilateral treaties try to gain further concessions, which are likely to become enshrined into WTO treaties at a later stage.

Unlike in other cases where signing means accepting all obligations of treaties, accession to the WTO means prior bilateral agreements with all
members. These are used to extract further, individual concessions. Small countries have little choice, unlike big ones. The EU demanded that Russia increase its domestic oil price as a precondition for being allowed to join the WTO. Russia declined. Apart from China, few DCs could defend issues important for them as easily.

Massive protests, the events of Seattle, and the forming of the G20 have improved the position of DCs somewhat. Especially Brazil and India have acquired new importance. But it remains to be seen how much influence any DCs will finally have.

**Statutory rights and developing members**

Critical minds point at the considerable difference between statutory rights and the policy space DCs enjoy de facto, as well as at the costs resulting from this discrepancy. These rights, the fear that bilateral arrangements might be worse, promises (such as large cuts in agro-export subsidies), and pressure seem to have enticed DCs into signing the WTO treaties. Once members, they are safe from bilateral accession agreements.

The WTO does not fully protect the rights of weaker members, as pharmaceuticals illustrate. Big players have a choice and cannot be forced to honor contractual obligations. When the EU complained against the Helms–Burton Act, the US observed that this would not lead to resolving the dispute but pose serious risks for the WTO. After agreeing bilaterally not to apply Helms–Burton to EU corporations the complaint was dropped. In a dispute with Brazil, Canada simply refused to provide information it was obliged to disclose promptly and fully pursuant to Article 13.1 of the Dispute Settlement Understanding. Expressly mentioning this and its potential to undermine the dispute settlement system, the WTO (2000, p. 59) found against Brazil, as Canada’s WTO-violating behavior could not be proved because of Canada’s additional violation of WTO rules on providing information.

DCs exercising contractual rights may raise the WTO’s concern. Though ‘not extensively used’ in Asia after the 1997 crisis, selective tariff increases by some countries remaining ‘within the flexibility allowed by bindings under the WTO agreements’ gave ‘cause for concern to the extent they may distort the pattern of production and trade’ (WTO, 1998, p. 28). The WTO has never voiced similar concern on potential distortions regarding ICs, including agrarian subsidies or the long phasing-out of GATT-inconsistent restrictions.

Capital controls are a membership right pursuant to the IMF’s constitution explicitly restricting the use of Fund resources to finance ‘large and sustained’ outflows. Even current transfers can be restricted with the Fund’s approval. The IMF may, but is not obliged to, request controls.
statutes clearly show that it is not supposed to press for liberalization of capital movements. Asian countries had the right to control capital outflows – as the IMF admitted when Malaysia exercised it (cf. Raffer and Singer, 2001 [2002, 2004], p. 157). IMF programs financing large and sustained outflows by speculators violated the IMF’s constitution, causing damages to DCs while increasing IMF drawings and thus earnings. Pressure to liberalize capital accounts has made increased stocks of international reserves necessary. These have become ‘one of the widely-used targets of poverty reduction strategies in Africa’ (UNCTAD, 2002, p. 31). This money is lost for poverty eradication, debt service or financing the Millennium Development Goals (MDGs). As DCs pay higher interest than they receive on reserves (often US Treasury Bonds), they face substantial annual costs which they could avoid by exercising statutory rights.

The IBRD’s statutes demand debt relief in the case of default, without any conditions. Article IV.4.c confers a right onto members suffering from an acute exchange stringency to ‘apply to the Bank for a relaxation of the conditions of payment’. Article IV.7 contains the obligation to reduce claims in the case of default. The statutes of the Asian and Inter-American Development Banks are similar. The African Development Bank’s new statutes do not contain this clause. The European Bank for Reconstruction and Development writes off losses, proving that IFIs can survive while doing so. By simply refusing to acknowledge default, even if countries have not paid anything for six or seven years, the IBRD does not act according to its statutes. This creates damages by delaying solutions.

**Attempts at reforms by the South**

The preponderance of Northern interest in most international institutions triggered Southern attempts to create new global institutions more in line with developmental needs and less under Northern control. Success has remained limited.

Early attempts focused on mechanisms within the UN, the most prominent being SUNFED (Special UN Fund for Economic Development). Inspired by the generosity of the Marshall Plan, the ‘wild men’ at the UN advocated setting up a UN Fund to administer large-scale soft aid. Theoretically, this idea was based on the then generally accepted Keynesian consensus that capital availability determined growth, which in turn was needed to improve the lot of the poor. It was easy to argue that European countries had a moral obligation to help as they had been helped by the US. ICs opposed the idea of the UN administering large funds. Demanding harder terms, nearer to the market, the IBRD was particularly strongly opposed to the principle of soft financing on economic
grounds. Once it became clear that soft multilateral financing would be done by an institution administered by the IBRD, the Bank dropped all reservations. Helped by the Cuban revolution, which sparked a wave of US ‘generosity’, and US interests in disposing of embarrassingly high, practically useless holdings of inconvertible currencies, IDA was established in 1960, attached to the IBRD, which ICs control. ICs preferred IDA to any UN Fund, ‘because the structure of the World Bank ensured weighted voting in their favour’ (OECD, 1985, p. 141). The Marshall Plan’s participatory and recipient-friendly approach was not to be repeated. The OECD (1985, p. 146) observed a certain differentiation regarding multilateral aid: ‘by and large the largest donors have favoured the World Bank, while the smaller donors have favoured the United Nations’. The UN obtained two valuable consolation prizes: the UNDP and the World Food Programme.

The UN Economic Commission for Latin America (ECLA; Spanish: CEPAL) took the lead in defending Southern interests by publishing Raúl Prebisch’s (1949) findings on the effects of trade on DCs, one pillar of the Prebisch–Singer thesis. This publication angered the USA to the point of attempting to close ECLA (Toye and Toye, 2003, p. 463). ECLA’s economists, the ‘Cepalistas’, continued over decades to produce divergent views, a source of heterodox theories emanating from the South.

Prebisch was also the engine behind the establishment of the United Nations Conference on Trade and Development (UNCTAD) created in order to reduce or eliminate damaging effects of world trade on DCs. Its ideas on appropriate trade-related development policies differ pronouncedly from the WTO approach. Commodity agreements were established, all of which collapsed later on. In the 1970s the South demanded a Common Fund to stabilize commodity prices. Its ‘Second Window’ was to finance projects such as research and development (R&D) and economic diversification. After protracted bargaining the Common Fund was agreed on before the UNCTAD meeting of 1979. Voting shares were allocated to the North (42 percent), the Group of 77 (that is, the South, 47 percent), communist countries (8 percent) and China (3 percent). ICs (East and West) stipulated financial contributions higher than their shares in voting rights. It took roughly a decade until enough countries had ratified to bring the Fund into existence, although the Organization of Petroleum Exporting Countries (OPEC) bankrolled the contributions of poor DCs. Inadequate resources prevented the Common Fund from playing any role. Several of this Fund’s ideas were taken up by an eminent person’s report on commodities in 2003 (Khor, 2005).

A highly useful recent activity of UNCTAD is its Debt Management Financial Advisory Services (DMFAS) program helping DCs to improve
debt management. As effective debt management is part of good governance, and accounting errors by creditors to the detriment of DCs are documented, one might have expected the BWIs to help debtors in establishing appropriate controlling. They did not. Advice by UNCTAD is preferable because UNCTAD is, unlike the BWIs, not a creditor.

There exist other multilateral institutions, notably those established by OPEC or Arab countries, such as the OPEC Special Fund. Established in 1977 with strong political and financial support of OPEC to finance projects in favor of the rural poor, IFAD (International Fund for Agricultural Development) was a successful attempt to make multilateral structures more democratic (Maurizio, 1983). Active participation by the projects’ beneficiaries was sought. Votes were equally split among the North, OPEC countries and other DCs, separating votes and contributions. This did not make IFAD particularly popular with ICs. The first replenishment led to quarrels. Soon ICs demanded a restructuring of IFAD so as to align votes and contributions. Meanwhile over 56 percent of all votes are distributed in accordance with cumulative convertible currency contributions. IFAD, a child of the South’s drive for a New International Economic Order in the 1970s, was finally brought in line with present realities. At a time when the power of ICs has reached its apex since decolonialization, a multilateral institution where they have only one-third of the votes is simply unacceptable.

Before 1997, Asian countries wanted to establish an Asian Monetary Fund as a regional counterweight to the IMF. The Asian crisis put an end to this, at least for some time. It is interesting to note that the IBRD (1999, p. 2) knew years before that the liberalization policies it encouraged in Asia would lead to disaster (cf. Raffer and Singer, 2001 [2002, 2004], pp. 150–51). This helps us to understand why conspiracy theories abounded. Even as distinguished an economist as Bhagwati (1997) spoke of a ‘Wall Street–Treasury complex’ dictating the agenda.

**Conclusion**

Evaluating the record and policies of international institutions does not suggest that they fully support development. Attempts by DCs to establish more development-friendly international institutions corroborate this conclusion.

At present one sees an interlinking of important institutions. The BWIs force debtor DCs to liberalize quickly and strongly. The WTO then protects and perpetuates the results. DCs suffering from negative WTO effects may draw on IMF resources, thus increasing their dependence on the Fund. Bilateral treaties go beyond the WTO, using the greater leverage ICs have vis-à-vis most DCs, justifying DC fears that bilateralism might protect them even less. But bilateral treaties also prepare the ground for further
concessions by DCs that may eventually become WTO obligations. Typically, more recent treaties extend the definition of ‘investment’ to loans, thus increasing creditor leverage. The WTO process moves so quickly that small DCs simply have no chance to implement agreements before new changes are negotiated or even implemented. Unequal treatment of members depending on their political clout is another fact.

Chang (2002, p. 139) argues that ICs are ‘kicking away the ladder’ of development by ‘insisting that developing countries adopt policies and institutions that were not the ones that they had used in order to develop’. Chang feels this may be done ‘out of’ genuine (if misinformed) good will’, nevertheless with catastrophic results. He might as well have quoted List’s recommendation of North–South relations. Better known for his opposition to the ‘English philosophy’ of free trade as harmful to Germany in its early development stages, List (1920 [1841], p. 211) advocated joint exploitation of DCs as ‘promising much richer and more certain fruits than the mutual enmity of war and trade regulations’. International institutions seem to show such ‘neo-Listian’ (Raffer, 1987) tendencies.

Important international institutions are constructed in such a way that they cannot but implement the wishes of ICs. Within IFIs, ICs usually enjoy comfortable voting majorities. Within the WTO, unequal economic and political influence asserts itself in spite of ‘one-country-one-vote’. Even if they wanted to, international institutions could not act against their powerful members. ICs seem more interested in gaining economic and political advantage than in supporting development. As selective WTO liberalization proves, ICs do not wish to approximate free market conditions globally, but seek export and investment possibilities where this is in their interest. Thus, Northern FDI in the South is said to be good for anyone needing WTO protection. Chinese FDI in the USA, however, is not – and is treated differently.

Finally, international institutions have self-interests. They want to gain influence and enlarge their mandates. This seems to explain the BWIs’ role in debt management particularly well. Regarding the three most important institutions, these factors combine to produce effects that hinder rather than foster development.

References


Introduction
It has become established practice in development economics to view the world as being divided into two parts, one comprised of rich, developed countries and the other comprising poor, less-developed countries. Although other terms have also been used to describe this division – including Center (or Core or Metropole) and Periphery, the West and the Rest, the First World and the Third World – the North–South terminology, despite its less than complete geographical accuracy, has become the most popular one, as reflected by its usage in North–South (or Brandt) Commission, ‘North–South’ negotiations in international organizations and other arenas, and the literature on ‘North–South’ models.

The relevance of the division between the North and South has sometimes been questioned because of the lack of homogeneity within the North and especially the South (as reflected, for instance, in the emergence of the newly industrialized countries or NICs from the ranks of the South, with characteristics very different from the least-developed and often stagnating economies of Africa and elsewhere) and because of divergent interests within them which has reduced their ability to bargain as unified blocs of countries. Despite this, however, the North–South dichotomy has remained an enduring one – in both descriptive and analytic senses – for a number of reasons. First, there are great differences in levels of economic development – however measured – between rich and poor countries, and it is of enormous ethical importance to examine whether such differences are persistent or not, and if they are, what can be done to remove them and bring about greater international equality. Second, one of the central issues in all of economics is why poor countries remain poor and what can be done to raise their levels of income and production, and it is possible that much can be learnt by comparing the experiences of rich and poor countries. Third, it is widely perceived that there are major asymmetries in the economic characteristics of rich and poor countries, and that this may well imply that trade, factor movements and other interactions between them will have asymmetric impacts on them. As the interaction between countries seems to be growing due to the widely noted process of globalization, it is important to explore the effects of such interaction on rich and poor countries. Fourth, there are many issues, including those of trade
policy, the activities of transnational corporations (TNCs), international labor migration, intellectual property rights and the global environment, on which the interests of rich and poor countries appear to be opposed, and on which cooperation between them is arguably required for their mutual benefit. All these issues can be examined and analyzed most simply by dividing the world into a rich and a poor part. It is little wonder that large and growing theoretical, empirical and policy-oriented literatures have developed, analyzing the world economy in terms of the North–South dichotomy.

This chapter provides a flavor of the main issues raised in these discussions. First, it discusses the main findings on whether the gap in levels of development between the North and the South have been growing or falling. Second, it briefly reviews some contributions to the theoretical literature on North–South models. Following that, it examines in turn the implications of, and issues concerning, trade, capital flows, labor migration, technology transfers and environmental interactions between the North and the South. The final section mentions other North–South issues and some issues which are downplayed by dividing the world into a North and a South.

North–South inequality

A large body of empirical research has examined changes in the level of international inequality and the gap between rich and poor countries. This issue has been investigated in various ways, mostly using purchasing power parity-adjusted per capita real GDP figures. Many observers have employed standard indicators of inequality – such as the Lorenz curve, the Gini coefficient and the Theil index – to show that inequality across countries has increased; Sala-i-Martin (1996) shows that the standard deviation of the log of per capita real GDP for 110 countries increased more or less steadily between 1960 and 1990, implying what is called $\eta$-divergence. An alternative technique regresses growth rates of per capita GDP for the 1960–90 period on the logarithm of initial level of per capita GDP to find a positive coefficient, implying that richer countries on average grow faster, so that there is $\beta$-divergence (Sala-i-Martin, 1996). Quadratic regression equations involving the same variables find an inverse U-shaped relationship, implying a positive relationship between starting income level and per capita growth for most of the sample, and a negative one for a small group of high-income countries. Quah (1993), examining the distribution of per capita GDP levels (relative to the world average), finds that the distribution tends over time to one with a thinning middle and accumulation at the two tails (the so-called twin-peaks phenomenon), and that countries very seldom move from low to high ends of the distribution. In sum, these
figures suggest growing inequality among countries, with convergence among a group of rich countries. Overall growth figures for groups of countries tell a similar story: according to World Bank data, the richest one-third of countries on average grew by an annual rate of 1.9 percent between 1970 and 1995, whereas the middle third grew by only 0.7 percent and the bottom third showed hardly any growth at all (Scott, 2001, pp. 162–3). Pritchett (1997), using different plausible estimates for initial levels of income for poor countries (for which hard data are not available), finds that there has been ‘divergence, big time’ between rich and poor countries over the last 150 years.

These findings, however, have been disputed by some analysts, who point out that by treating each country as one observation, most studies do not give adequate weight to the large low-income countries, China (which is sometimes left out of the sample due to lack of data) and India, which have experienced relatively high rates of growth in recent years (Dollar and Kraay, 2002; Bhalla, 2003). They argue that correcting the problem yields the result that poor countries – as a whole – are growing faster than rich countries. However, there may be a case for giving each country an equal weight, since each country follows a particular set of policies and is thus a single observation. Thus, if we are interested in the relative performance of countries (rather than individuals) it is more appropriate to give equal weights to countries. If we are interested in inequality among people in the world, population-weighted measures are appropriate, but because such measures implicitly assume that there is no inequality within countries, they have to be supplemented by data on within-country inequality to measure inequality among people (see Milanovic, 2005).

Even if the North and South are experiencing divergent patterns of development, it does not imply that this divergence is due to the nature of their interaction. Divergence could be occurring due to factors internal to them, as explored in models of poverty traps which formalize the earlier literature on vicious and virtuous circles in closed economies (see Azariadis and Stachurski, 2005). Since the empirical exploration of the relationship between North–South interactions and divergence is still in its infancy, we may turn to theoretical North–South models to examine the implication of such interaction.

North–South models
North–South models are not new. The classical economists, including Adam Smith (who discussed the importance of increasing productivity due to the division of labor as economies grow) and David Ricardo (who examined the role of trade in postponing the arrival of the stationary state in rich countries by enabling cheap food imports and in having the opposite effect
for poor countries), discussed North–South issues. Subsequently, Marxist writers on imperialism and dependency theorists, and development economists more generally, stressed the role of the South in providing markets and investment outlets for the North, and examined the problems of surplus transfers from the South, the deterioration of the Southern terms of trade, and of uneven development. For instance, Lewis’s (1954), pioneering contribution on economic development with unlimited supplies of labor examined the determination of the terms of trade between rich and poor countries in a world with three goods: a Northern good, a Southern good and a non-traded good which both regions produce. Lewis showed that technological change in the Southern goods sector, low and stagnant productivity of the non-traded goods sector (food) in the South, and high and rising productivity in that sector for the North, have the effect of turning the adverse Southern terms of trade further against the South.

More mathematically explicit general equilibrium North–South models which endogenize capital accumulation in the two regions emerged from the early 1980s. Findlay (1980) examines capital accumulation in a global economy with the North growing with full employment as in Solow’s model and the South with unlimited supplies of labor at a fixed real wage as in the Lewis model. Taylor (1983) allows for unemployment in the North as well, assuming that effective demand determines Northern growth as in the Kaleck–Keynes approach. These and other models can be thought of as special cases of a general framework (see Dutt, 1990) in which the North and South are completely specialized in the production of their good, where the Northern good is a consumption-cum-investment good, and the Southern a pure consumption good. The models embody specific behavioral and institutional assumptions for the North and the South, thereby highlighting their structural differences, and assume that fixed fractions of consumption expenditure in each region are spent on the two products (that is, preferences are homothetic), and trade between the two regions is balanced. The framework examines short-run equilibria with given stocks of capital with markets for both goods clearing, and long-run equilibria in which the stocks of capital in the two regions grow at the same rate due to saving and investment. The models are used to examine the effects of changes in such things as technology, consumption expenditure patterns and savings rates. Of particular interest are results which demonstrate that Southern growth depends on Northern growth, which is determined independently of the South (which imply that attempts by the South to grow faster by saving more merely imply a deterioration in its terms of trade, revealing its dependent status), the relation between the Southern terms of trade and Southern growth, and the possibility of uneven development (reflected by a rise in the relative capital stock of the North to the South)
due, for instance, to technological change, changes in consumer preferences and industrial concentration in the North (see Dutt, 1990). These results confirm some of the informal ideas of earlier writers on uneven development, but they depend on some of the specific assumptions made about the structures of the Northern and Southern economies. The models have also been extended to incorporate issues such as international capital flows and technology transfers.

The models stressing structural asymmetries between the North and the South do not explain why such asymmetries arise. Implicitly, they assume that events in the past, such as the Atlantic slave trade (Darity, 1992) or colonial domination and consequent policy regimes (Dutt, 1992a), related to the balance of military power (Findlay, 1992) create and lock in these structural differences. Models which assume identical structures for the two regions have also been developed to show how ‘small’ historical events can make one region (the North) end up exporting goods exhibiting increasing returns to scale and productivity-enhancing learning effects, while the other region (the South) becomes more oriented towards goods exhibiting constant returns and relatively technologically stagnant sectors, so that there is uneven development (Krugman, 1981, 1990).

The models discussed so far can be seen as reactions to the dominant neoclassical Heckscher–Ohlin–Samuelson (HOS) trade models which contain optimizing agents, are usually static in nature, and assume that markets clear so that labor and other resources are fully employed everywhere. However, the neoclassical approach has also contributed to the development of North–South models. Most of the North–South models from the neoclassical perspective, however, have followed the contributions of new growth theory, which emphasize the role of increasing returns and externalities in the growth process (see Darity and Davis, 2005). Many new growth theory models imply economic divergence between rich and poor countries due to economies of scale along Smithian lines even without any interaction between the two. These results often carry over to models with North–South trade, as we shall see below.

North–South trade

According to the canonical HOS theory of trade, with its assumptions of constant returns to scale and perfect competition, countries export goods which use their abundant factor intensively. It is generally assumed that the North is capital-abundant and the South labor-abundant, and trade with the North exporting capital-intensive goods and the South labor-intensive ones results in gains from trade through the reallocation of resources according to comparative advantage. With the addition of the assumption of identical technology across trading partners and some other conditions,
the theory also implies – according to Samuelson’s factor price equalization theorem – that trade equalizes factor returns across countries even without any factor mobility between them, since in effect abundant factors move abroad embodied in traded goods. These implications are frequently invoked as proof of mutually beneficial North–South trade, and of convergence. Modifications of the approach, which focus on skilled and unskilled labor as the two factors and assume that the North is skilled-labor abundant, have also been used to examine how the North can suffer increasing inequality and (with rigid wages) the unemployment of unskilled labor (Wood, 1994).

The HOS approach, however, does not imply that countries necessarily gain from trade or from moving to freer trade. If economies are ‘distorted’, for example because of the existence of production externalities or factor market rigidities, the economy may well lose from trade (see Bhagwati and Srinivasan, 1983). If such ‘imperfections’ are more prevalent in the South, as commonly supposed, North–South trade may not benefit the South. Dynamic extensions to the approach which incorporate some of the insights of new growth theory even imply uneven development. For instance, when trade liberalization leads to a rise in the wages of skilled workers or the return to capital in the North and reductions in the South, it can speed up human and physical capital accumulation in the North and slow it down in the South, resulting in a growing gap between the two regions (Baldwin, 1992; Stokey, 1991).

More drastic departures from the HOS approach may also imply divergent growth patterns. North–South models in which the North is specialized in goods with higher income elasticities than Southern goods imply that balanced growth for the two regions creates an excess supply of the Southern goods and a deterioration of the Southern terms of trade (Thirlwall, 1979). Long-run equilibrium in models with such a property – which seems empirically plausible – leads to uneven development (see Dutt, 2003). Models such as Krugman’s (1981 [1990]), in which the North specializes in the production of relatively sophisticated goods which exhibit increasing returns and dynamic learning and spin-off effects, while the South specializes in goods which do not have these properties, also imply uneven development and possible losses from trade for the South, a result which is found in a range of similar models (see Ethier, 1982; Boldrin and Scheinkman, 1988).

If insights such as these have any validity, the policy implication for the South is to attempt to change its pattern of specialization with industrial and trade policies towards goods with favorable demand and technological characteristics. Whether individual Southern countries can do so effectively is, of course, another matter. But the experiences of late industrializers in
the past, such as the USA and Germany (where such policies were espoused by Alexander Hamilton and Friedrich List) and Japan, and more recent success stories of South Korea and Taiwan, certainly point to their importance (see Amsden, 2001).

**North–South capital movements**
According to standard neoclassical theory the mobility of factors of production leads to production convergence. The argument can be made in its simplest form for a one-good, two-factor – capital and labor – world with diminishing returns to factors of production, perfect competition, flexible prices and profit-maximizing behavior for the case of capital movements (see Bhagwati, 1979). With the North being capital-abundant, the marginal productivity of capital is low compared to that of the South, where capital is scarce (assuming superior Northern technology does not outweigh the strength of diminishing returns). With the rental rate equal to the marginal product of capital, capital will move from the North to the South, increasing per capita production in the South and reducing it in the North, increasing production in the world as a whole, and increasing per capita income (taking into account payments to capital) in both regions.

In fact, capital does not move as much from the North to the South as is suggested by this theory, as shown by the fact that for substantial periods of time there has been a reverse transfer of financial resources from the South to the North. To the extent that capital does move from the North to the South, it moves to a small group of countries. Modifications of the neoclassical model provide reasons why this may be so. If we replace the assumption of diminishing returns by increasing returns, introducing external economies, or internal scale economies with imperfection competition, it is no longer the case that the capital-abundant North has a lower return to capital than the South, and capital will move in the reverse direction (Lucas, 1990). If lenders do not know exactly what borrowers do with borrowed funds and can only observe outcomes of their activity, while borrowers know what they are doing, we have the problem of asymmetric information, and lenders will require collateral to ensure that borrowers do not willfully default. The implication of this is that borrowers in rich countries who have higher initial endowments of capital will be able to borrow more than those in poor countries, because they can put up collateral to overcome moral hazard problems, while borrowers in poor countries are less able to do so (Gertler and Rogoff, 1990; Matsuyama, 2004). This may imply that capital will flow from poor to rich countries, making rich countries even richer, resulting in a process of uneven development.

If capital does move from the North to the South, the effects may not be as implied by the simple neoclassical theory. If we depart from the
one-good assumption, and allow the North and South to trade in different products, capital flows from the North to the South can turn the terms of trade against the South (Singer, 1950), and lead to Southern immiserization of the type discussed by Bhagwati (1956). These insights have been incorporated into some North–South models. Burgstaller and Saavedra-Rivano (1984) introduce capital mobility into Findlay’s model and find that with capital mobility (compared to the case without it) Northern per capita income will be higher, but Southern income per worker will be lower (because of the payments that have to be made for foreign capital), and the terms of trade of the South deteriorates. Relative Southern employment will also fall if the Northern propensity to spend on the Southern good is less than that out of Southern profits, since there is a reduction in the demand of the Southern good due to the redistribution of income caused by capital mobility. In Blecker’s (1996) model in which Northern growth is driven by aggregate demand, greater capital mobility leads to uneven development in the sense of an increase in the stock of Northern capital to Southern capital, due to Southern terms-of-trade deterioration. Such outcomes, however, depend on the assumption that capital flows increase the production of the Southern good, thereby resulting in a deterioration of its relative price. If, as pointed out in some empirical studies, capital flows in the form of foreign direct investment feature deep integration, involving the production of typically Northern goods in the South, capital flows could lead to more even development patterns as the North loses markets to the South and experiences greater excess capacity and unemployment and lower growth, and lower profits encourage more capital flows to the South (see Dutt, 1996). TNCs bring in not only capital, but also technology and exporting capabilities, which make these effects more likely. Such outcomes are more supportive of those who fear Northern stagnation due to capital outflows.

However, North–South capital flows may not bring about Southern development because they may not be productively invested. Borrowing by the South in the past, often due to loan-pushing by Northern banks, led to the accumulation of Southern debt but often went right back to the North through capital flight by corrupt elites (Darity and Horn, 1988). Debt-ridden Southern countries are then forced to make interest payments and repay loans by running current account surpluses and reducing growth. Default can threaten financial stability in the North as well. Portfolio flows and bank loans often finance short-term investment in the South in stock markets and real estate, leading to speculative bubbles which, when they burst, bring about sharp capital outflows, currency crises and macroeconomic contraction (see Stiglitz, 2002).
Labor migration
The standard one-good neoclassical model with diminishing returns and perfect competition discussed earlier also implies convergence of per capita production due to the movement of people from the labor-abundant South to the labor-scarce North. Such an approach, which implies that migration leads to a reduction in the Northern wage, can be used to explain some of the opposition to immigration in the North. It can also be used to explain income losses to those who are left behind in the South, because of the disappearance of the surplus produced by the workers who migrate (see Bhagwati, 1979).

Such a simple framework does not capture the complexities of real-world migration, however. Countries of the North heavily restrict the migration of unskilled workers from the South, allowing mainly the legal immigration of skilled workers. If one distinguishes between skilled and unskilled workers, and introduces scale economies and imperfect competition, uneven development due to brain drain from the South may occur. For instance, if unskilled workers and intermediate services are used in the production of the final good under conditions of constant returns to scale and perfect competition, while non-traded intermediate services are differentiated products, each produced by a monopolistic competitor under conditions of increasing returns with skilled labor as the only factor of production, the migration of skilled labor from the South to the North reduces the number of intermediate goods, quantity of the final good produced, and per capita income in the South, and has an opposite effect in the North, implying uneven development (Dutt, 2005). The endogenization of skilled labor supply, due to tax-financed government educational spending or the decision by workers to accumulate human capital, into new growth theory models can also imply divergence growth patterns (Haque and Kim, 1995; Wong and Yip, 1999).

Technology transfers
If knowledge is something that all countries can share, it may be supposed that the South will eventually catch up with the North. Lucas (2000) develops a simulation model in which countries take off in sequence (depending on their internal conditions) and in which latecomers grow faster than leaders because they have access to technology and policy experiences of the latter. His model implies that although international inequality increases initially, ‘sooner or later everyone will join the industrial revolution . . . economies will grow at the rate common to the wealthiest economies, and . . . percent differences in income levels will disappear’ (Lucas, 2000, p. 166).

Such an outcome, however, does not occur even in standard neoclassical models of technology transfer. In Krugman’s (1979 [1990]) model of
product innovation which formalizes the product life cycle approach, and which assumes that Northern ‘new’ goods become ‘old’ goods produced by the South at a constant rate of ‘radioactive’ decay, the ratio of Northern to Southern income stabilizes in equilibrium, but does not become unity. In models of technology transfer involving process innovation, where the change in Southern productivity depends positively on the North–South productivity gap (reflecting more opportunities for transfer), a narrowing of the technology gap occurs if there is initially a large gap, but complete technological catch-up does not occur unless transfers occur at an infinite rate.

Models in which rates of technology transfer are either constant or monotonically related to the relative technology gap (a higher technology gap leading to a faster rate of technology transfer) may not capture the realities of the process of technology transfer. Since much technological knowledge is tacit, and requires constant modification and adaptation, the process of transferring technology is not so very different from that of innovation, and successful technology transfer requires the development of some amount of social and technological capability of the South (Abramovitz, 1986; Bell and Pavitt, 1993). This can have a number of consequences. First, if we measure (relative) technological capability by the South–North productivity gap, a large productivity gap may increase the potential for technology transfer, but reduce the ability of the South to do so successfully. In this case, even standard models will imply convergence of the technology gap if the gap is not too large, but divergence if the North and South are too far apart (Verspagen, 1991), which may explain the formation of convergence clubs of rich countries and the exclusion of many poor countries. Second, technology transfer may be linked to other aspects of North–South relations. To the extent that technological capability is enhanced by the presence of TNCs, FDI can speed up technology transfer. However, an adverse impact of indigenous technological development cannot be ruled out if competition between TNCs and domestic firms hurt the latter. Trade liberalization and its consequence on the composition of production may, as mentioned earlier, have adverse effects on learning by doing and slow down the accumulation of technological capability (see Van der Klundert and Smulders, 1996).

In addition to these issues, the protection of intellectual property rights (IPRs) can serve as a barrier to technology transfers. Indeed, the protection of IPRs has been a major source of conflict between the North and the South, as reflected in the Uruguay Round discussions of the WTO. Since most technological innovation occurs in the North, it is not surprising that the North, especially its innovating firms, has an interest in protecting IPRs internationally, while the South, which is more interested in the speedy
diffusion of technology from the North to the South, is against such protection. These insights are confirmed by simple partial equilibrium models of product and process innovations in terms of consumer and producer surpluses in the North and South (see Chin and Grossman, 1990; Deardorff, 1992).

Although the interests of the North and South may be in conflict, it is often argued that the international IPR protection may be good for the world as a whole, because in its absence there would be insufficient innovation, since innovating firms would receive smaller rewards for their research and development activity and therefore conduct less of it. However, the partial equilibrium models imply that the effect of international IPR protection on world welfare (measured by the total surplus accruing to the North and the South) is ambiguous: if the North is a large part of the economy and protects IPRs internally, there will be sufficient innovation in the North to make world welfare higher without international IPR protection than with it. This result can carry over to a dynamic general equilibrium setting: in Grossman and Helpman's (1991) model weaker protection of IPRs will not only speed up technology transfers, but also accelerate innovation in the North, because Northern resources are devoted more to innovation, rather than to production, which increasingly moves to the South.

While it is often the case that tighter international IPR protection slows down technology transfer, this may not be true under all circumstances. For instance, with IPR protection TNCs may more willing transfer better technology abroad, while without it they are likely to hold it back, fearing imitation (see Lai, 1998). However, Glass and Saggi (2002) find that although stronger IPR protection makes TNCs and Northern firms safer from imitation, the greater difficulty in imitation results in more resources being absorbed in imitative activity which reduces FDI, as well as Northern innovation.

Environmental issues
The literature on the environmental Kuznets curve suggests that the rich North is able to deal better with environmental problems than the South, where poverty, population growth and lax pollution control have an adverse environmental effect which can in turn have an adverse effect on growth and other development indicators, such as health. North–South interactions through trade and capital movements may exacerbate such problems, if they lead to the movement of the production of relatively dirty and resource-intensive goods from the North to the South, for instance, because of less-stringent environmental regulation; although, if capital-intensive Northern goods are more pollution-intensive, this may not necessarily occur (Copeland and Taylor, 2003).
Further North–South interactions regarding the environment occur because of the global public goods nature of several environmental issues, such as global warming, depletion of the ozone layer and loss of biodiversity (Sandler, 1997). The North blames the South for its increasing contribution to damaging the global environment, because of high population growth and pollution-intensive growth, while the South blames the North because of its high per capita contribution to this damage. The solution of global environmental problems requires such conflicts to be resolved by mutually beneficial international agreements between the North and the South which take into account the poverty of the South.

Conclusion
The North–South approach can be criticized for downplaying the differences within rich and poor countries, and the possibility that poor countries may grow and join the ranks of the rich. However, two-region models have been extended to include a third, consisting of newly industrialized countries, to explore the causes of its growth and to analyze whether its growth results in the end of uneven development or the exacerbation of the gap between the rest of the South and the North (see Dutt, 1992b). Moreover, multi-region models have been developed to show how the countries can endogenously sort themselves into Northern and Southern groups (see Matsuyama, 1996, 2004).

This brief review has confined attention to only a few of the major North–South issues, not examining other areas of interaction (for instance, due to cultural and political influences), conflict (such as those over labor standards) and cooperation (related to the existence of global public goods such as global health and security). Nevertheless, it suggests that the North–South framework remains a useful way of looking at many important and interesting development issues.

Notes
1. I am grateful to Jaime Ros for his helpful comments on an earlier draft of this chapter.
2. More details can be found in the other chapters in Part VI of this Handbook.

References


PART VII

DISTRIBUTION AND POVERTY
There is a rising level of concern about income inequality and poverty in developing countries. The development process has brought rising inequality within many countries and has also increased the gap between the poorest and richest countries. All this has led to an increasingly vocal demand for a more equitable style of development. Certainly no one would wish for more inequitable development, but it is still important to stop and ask oneself what distribution measures really measure, and what comparisons of these measures over time or across countries tell us. In particular, what do trends in inequality mean in developing countries? How are they affected by the development process? Are they something that governments in those countries can affect and should worry about?

I am going to address these questions by first considering at the individual or family level what should be included in a distribution measure. Once one has a satisfactory measure at the individual level, one needs a number or index which aggregates or summarizes all the individual distribution data across some reference group. The best-known is the Gini coefficient, but there are many others whose properties will be discussed further on. But in addition to this one has to decide the group over which to aggregate the individual distribution data. What group should that be? Should it be a country, a region within a country, the whole world, or should it be confined to a group such as an age cohort within a country? Responses to this question will have a big impact. My focus in this chapter is going to be on the distribution within countries rather than for the world as a whole, or between countries, because I believe that the national state is the basic economic and administrative unit which is able to influence the distributions that matter to its citizens.

Next I want to consider briefly some problems in the interpretation of the aggregate measures. What do changes over time or comparisons across countries tell us? Finally I will discuss several of the determinants of the distribution and the ways that they are likely to affect the distribution during the development process.
Measuring inequality at the individual or family level

Even though in some ultimate sense ‘fairness’ may imply an equal distribution of welfare across a population, I am going to limit my discussion to the distribution of goods or income produced by an economy. I am going to assume that a more equal distribution of goods or income is in the economic sense a fairer distribution, even though I have no way of knowing or of proving that a fairer economic distribution is also a more equitable distribution of welfare. I am bringing welfare into the discussion here because it will guide me in my choice of what to measure and, more importantly, what the reference group should be for my aggregate distribution statistics.

All of my statistics on distribution come from household surveys which are now available for a large number of countries. These measures can be based on income or on consumption. The World Bank favors consumption because it believes that it is both more accurately measured and a better proxy for lifetime income than observed income. If one thinks that people are concerned with their lifetime income rather than their income in any particular year, then the distribution measure should also be based on lifetime income. While I cannot measure lifetime income, it is likely that consumption is a better proxy than observed income in any single year. Families smooth out temporary fluctuations in income by either running down saving during recessions or adding to their assets when they have a temporary windfall. In some situations consumption is not a perfect measure, however. Suppose one is interested in the relationship between stabilization or recession and the distribution of welfare. Poor families may maintain their consumption by distress selling of assets or by borrowing at high interest rates. In either case the distribution of welfare is made less equal, even though the distribution of consumption may not show that, or may understate the rise in welfare inequality. If one is interested in short-run fluctuations and their distributional impact, it is probably better to use an income-based distribution measure. For longer-run questions, the consumption-based measure is better.

Every family receives income in kind. While the statisticians have learned how to include the imputed value of home-grown food or the rent on owned houses in the total value of family consumption, they do not generally include as consumption the value of government services such as public health care or education. These can have a fairly sizable effect on measures of the distribution.

For labor market questions, the distribution of earnings is relevant. However, since welfare is based on consumption, and since the family is the social unit which converts the earnings of some of its members into consumption for all its members, one should use the distribution of income among families. I can convert that into an individual distribution for
poverty analysis by dividing total family income or consumption by the number of family members.\textsuperscript{2}

**Aggregation problems**

I now want to address several aggregation questions.\textsuperscript{3} The first is what group the distribution should be based on. Using welfare as our criterion for answering this question, the group should be the one to which the individual belongs and across which he or she makes comparisons. That turns out to have important implications. It is customary to calculate distributions at the national level at a point in time. That is reasonable if national relative position is an important consideration to individuals in a society. But in large countries it may well be that a regional distribution is more consistent with how people see their own welfare. If inequality indexes are thought to be good proxies for social tension or the likelihood of populist demands for redistribution, it is important that the aggregation reflects the interpersonal comparisons made by people in a society.\textsuperscript{4}

Another possible sub-national group classification is by age or cohort. Suppose that people's primary concern is their relative position in their own age cohort. Young people presumably know that there is a positive age–earnings profile. Incomes rise with age and experience. It that is so, it would seem reasonable that purely age-based inequality would be less significant as a source of social tension than inequality within an age cohort. To take this to an extreme, suppose the individuals are concerned only with intra-cohort inequality. Now imagine an economy in which each age cohort has exactly the same expected lifetime income and age–income profile. In such an economy, there would be no income inequality provided that we were able to base the distribution measure on permanent income. That would not be true if the measure was based on consumption, if consumption, like earnings, has an inverted U-shape, rising through most of an individual's working life, and then falling in retirement. In that case any aggregate distribution measure based on observed income or consumption at a point in time will show a good deal of inequality even when there is no intra- or inter-cohort inequality. In any year the younger members of the society will be earning or consuming less than their elders. Observed inequality will be higher, the steeper the age–consumption profile, the greater the return to experience or the more credit-constrained the younger members of the society are. Apparent inequality will also be higher, the faster the rate of population growth, not because younger cohorts have lower expected lifetime incomes, but simply because there will be more of them at the bottom of the income pyramid at any point in time.

Statistically in this hypothetical economy there is inequality between the old and the young. But the significance or meaning attached to that
inequality depends on the attitudes or preferences of the members of that society. If the comparator group for each member is his or her own age cohort (or their own region of the country) then a national distribution of income or consumption will be a misleading indicator of what one might call ‘socially relevant’ inequality.

**Aggregate measures of inequality**

Once a choice has been made of what to measure, one is left with the problem of how to aggregate the information into a small number of understandable statistics. There are many ways of doing this, each of which implicitly assigns certain weights to each individual observation. The most straightforward statistic is the shares of total income going to different shares of the population such as the poor and the rich, or the bottom and top quintiles.

Perhaps the best-known measure is the Gini coefficient, now used around the world to compare distributions between countries or within countries over time. The Gini coefficient is the ratio of the area between a 45º line and a curve called the Lorenz curve showing the shares of total income accruing to different shares of the population. The Gini varies between zero and one. If there is perfect equality, each individual or family earns the same income, and the actual Lorenz curve overlays the 45º line. Here the Gini coefficient is zero since the gap between the lines is zero. Maximum inequality is when one individual or family owns all the income. In this case the Lorenz curve has a backward L-shape, and the Gini is equal to one since the gap between the Lorenz curve and the 45º line is equal to the entire area under the 45º line.

There are a number of alternative aggregate measures of the distribution found in the literature. One is the coefficient of variation (CV). It is defined as the square root of the variance of income divided by the mean to make it, like the Gini, independent of the level of income. The CV has the somewhat counterintuitive property that it gives equal weight to transfers between individuals regardless of their incomes. That is, inequality changes by the same amount if changes in income occur close to or further away from the mean, provided only that the changes are of the same size:

\[
CV = \frac{\text{Var}}{\mu}^{1/2}
\]

If one wishes to put more weight on what happens at the bottom of the distribution, one can transform income into the log of income and then take the standard deviation of this transformed income measure. Since the log of income falls rapidly for small values of income and since the standard deviation uses the square of the deviation from the mean, the
transformation will increase this measure of inequality when there are many people with levels of income far below the mean. The measure is defined as:

\[
SDY = \sqrt{\frac{\sum (\log \mu - \log X_i)^2}{n}}
\]

**Distribution comparisons among countries and over time**

Two of the main uses of distribution statistics are to compare distributions across countries or to measure changes in a single country over time. Deininger and Squire (1996) at the World Bank recently collected a set of comparable national Gini coefficients for 108 countries around the world. Many of these countries have estimates going back over 30 years. Table 47.1 gives their estimates of the median Gini by region and decade.

Other than the very large differences between Latin America and most other regions of the world, what is striking is that there is no convergence over time especially in Latin America. Latin inequality fell slightly in the 1960s, but then it reversed after 1970. Meanwhile there was some improvement in the Middle East, Africa and South Asia, and worsening in Eastern Europe in the 1990s after the fall of communism. In the developed countries inequality has remained roughly constant at a low level. In the other developing regions where inequality was once high, it has fallen quite sharply. Only in Latin America has inequality remained near its high initial level.

The differences between regional Gini coefficients translate into large differences in the amount or share of income going to the rich and poor. In the 1990s on average in Latin America the top 5 percent of the population received 25 percent of total income while the bottom 30 percent got only 7.5 percent. In South-East Asia the top 5 percent received only 16 percent of income while the bottom 30 percent got 12.2 percent. The comparable

**Table 47.1 Decadal median Gini coefficients by region**

<table>
<thead>
<tr>
<th>Region</th>
<th>1960s</th>
<th>1970s</th>
<th>1980s</th>
<th>1990s</th>
</tr>
</thead>
<tbody>
<tr>
<td>E Europe</td>
<td>25.1</td>
<td>24.6</td>
<td>25.0</td>
<td>28.9</td>
</tr>
<tr>
<td>S. Asia</td>
<td>36.2</td>
<td>33.9</td>
<td>35.0</td>
<td>31.9</td>
</tr>
<tr>
<td>OECD &amp; high-income</td>
<td>35.0</td>
<td>34.8</td>
<td>33.2</td>
<td>33.7</td>
</tr>
<tr>
<td>E. Asia &amp; Pacific</td>
<td>37.4</td>
<td>39.9</td>
<td>38.7</td>
<td>38.1</td>
</tr>
<tr>
<td>M. East &amp; N. Africa</td>
<td>41.4</td>
<td>41.9</td>
<td>40.5</td>
<td>38.0</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>49.9</td>
<td>48.2</td>
<td>43.5</td>
<td>46.9</td>
</tr>
<tr>
<td>Latin America</td>
<td>53.2</td>
<td>49.1</td>
<td>49.7</td>
<td>49.3</td>
</tr>
</tbody>
</table>

figures for Africa show the top 5 percent got 24 percent, the bottom 30 percent got 10.1 percent. In the developed countries the top 5 percent got only 13 percent of total income while the poorest 30 percent got 12.7 percent. These shares imply that in Latin America the average income of the richest 5 percent is 20 times that of the poorest 30 percent. In South-East Asia the richest 5 percent have average incomes only 7.9 times that of the poorest 30 percent.

Problems in comparisons across countries
One of the problems with interpreting differences in any of the aggregate measures between countries is that they force one to make value judgments about the weights that one gives to welfare or income at different income levels. One can see this most easily for the Gini coefficient. Lorenz curves for different countries may cross as may the curves for the same country over time. The Gini will say unambiguously which country has a more equal distribution. But even assuming that one thinks that more equal distributions are preferable, can one really be sure that one prefers the more equal distribution in all cases? Take the case of two countries with the same mean income. Suppose that in country A there are relatively few poor people because the middle class has been taxed to support a generous safety net. But there are a lot of rich people. In country B the poor have a much smaller share of income than in country A, but there is a big middle class and a relatively small group of the rich. Here the Lorenz curves of the two countries cross. The curve for country A lies above that of country B at the bottom of the distribution, but below it at the top of the distribution. If one is really concerned about the welfare of the poor, one may prefer the distribution of country A, even though its Gini is larger (more unequal) than the Gini of country B.

Large countries are likely to be more heterogeneous and have significant backward regions and significant differences in regional incomes. That will make their measured income distribution less equal than in smaller countries. Similarly, national distributions are generally less equal than purely urban distributions because on average, rural incomes are lower than urban. Countries with big indigenous populations are also likely to have more unequal distributions as will be seen below. In all these cases the reason for higher inequality is the existence of a large group within a national border which does not fully participate in the process which generates economic growth. Note that none of this would be particularly relevant to cross-country comparisons if comparator groups were truly national. As mentioned above, welfare statements about distributions are based on the position of individuals relative to others in their comparator group. If that group is local, or limited to members of an indigenous group
or the rural population, then comparisons of national distributions between countries may have misleading welfare implications.

Comparisons for the same country over time
Comparisons over time in single countries may have the same ambiguities and problems of interpretation that I noted above in the discussion of crossing Lorenz curves. They have additional interpretation problems in economies in which the population is growing. In that situation, by definition, over time the populations whose distributions are being measured are not the same and it is important to make a distinction between what I will call the base period population and the later group to which it will be compared. Apart from any questions about comparator groups, changes in the observed distribution will be affected by the earnings of the new entrants. Apparent earnings of groups such as the rich or the poor will appear to rise or fall even if the earnings of the base period population do not change. Distribution statistics are based on shares of total income received by different shares of the population. When there is an increase in the income-earning population, both income shares and population shares are affected. Total and average incomes are affected by where these new entrants come into the income pyramid.

When studies of the distribution report statistics on the incomes or income shares of the rich or the poor, they mean the group at the top or the bottom of the income pyramid at different moments. They do not mean the base period rich or poor. Where population growth is rapid or where comparisons are made over long periods of time, the distinction between the distribution or growth rate of income of the base period population and the observed population is significant (see Morley, 1981).

Determinants of the distribution
I turn now to a short discussion of the determinants of the distribution of income. One should distinguish between what I will call the primary or earnings distribution and the family distribution. This first is the distribution of income to the owners of the factors of production that produce it. The family distribution is derived from the primary distribution according to what factors of production each family owns and what each of those factors earns in economic activity. Economic factors mainly affect the primary distribution while demographic factors play a key role in the corresponding family distribution.

A useful abstraction here is the factor market where the demand and supply of each factor determines its earnings. For simplicity consider four factors, skilled and unskilled labor, capital and land. Like other economic markets, prices are determined in factor markets by the interaction of the
supply of each factor and the demand for factor services as well as institutional factors such as the minimum wage. Markets clear at a price at which someone is willing to use the available supply of each of the factors. That set of market-clearing factor prices determines the relative price of skilled and unskilled labor and the rate of return on capital and land. In the labor market, it may well be the case that demand is so low that many are forced to work part-time in the informal sector or are unemployed altogether because the legally prescribed minimum wage in the formal sector exceeds the marginal product of the fully employed labor force.

On the demand side, economic growth shifts out the demand for each of the factors, which tends to raise each of their prices. What happens to relative factor prices depends on the nature of the growth process. If it is skill-intensive, skill differentials widen. If it is led by sectors such as agriculture which use mainly unskilled labor, the reverse should occur. Structural reforms such as trade liberalization also affect factor demand by changing relative goods prices and the composition of output.

The supply side is critical to understanding the dynamics of the process. In the short run the supply of factors is fixed because each of the factors is a stock which produces a flow of services. That stock can change, but that takes time. Thus in the short run there is a fixed stock of factors determined by past investment decisions, and a set of factor demand curves. The two together produce the short-run set of factor prices. While this process sets the valuation of the factors of production, one needs to know the distribution of ownership of the factors of production in order to determine the primary earnings distribution, since it is the owner that receives the payment for the factor services. The primary distribution in the short run is determined jointly by the relative factor prices that come from the factor market and by the pattern of ownership of the factors of production.

The dynamics of the process come from investment. Investment increases the stock of physical capital, while the education system produces graduates who enter the labor force and change the supply of both skilled and unskilled labor. Training and experience also increase the stock of skilled labor or what I will call from here on, ‘human capital’. Migration also changes the position of factor supply curves, and that could be for either unskilled or skilled labor.

It is important to remember that investment takes time and that the amount that can be added to the stock of any of the factors is relatively small over any short-run time period. Thus, in any period as short as say a year, it is impossible to shift the factor supply curves out by more than 3–5 percent. That means that in the short run, it is changes that come from the demand side that are most likely to explain observed changes in the primary
distribution. Over time these stocks change because of additions to the physical capital stock, education, retirements and so on.

An important feature of the distribution process is the dynamic feedback between factor prices and changes in factor supplies through investment. Relative prices or returns on physical capital determined in the factor market affect investment and education decisions. If there is a rise in the rate of return on capital, investment in physical capital increases. Similarly if there is an increase in the wage differential between skilled and unskilled labor, or between university and high school graduates, that will tend to increase the demand for university education. Fewer people will enter the labor force after high school. More will continue on to earn a university degree. Over time, those decisions will increase the supply of educated labor in the labor force. If there were no changes on the demand side, this increase in the supply of skilled or educated labor should reduce the skill differential. One could thus think of the initial rise in the skill differential or the return to capital as a dynamic signal which sets in motion an investment process which eventually adjusts factor supplies and relative factor prices to a long-run equilibrium. By definition that equilibrium is one in which additions to supply are just sufficient to offset changes in the demand for factor services coming from economic growth and technological change.

From the point of view of the primary earnings distribution, a rising skill differential or profit rate which tends to increase inequality in the short run is also a signal which expands the supply of educated labor or physical capital in the long run. These long-run changes on the supply side may well reverse the short-run rise in inequality that induced them since they tend to drive down the rate of return to capital, both human and non-human.

There is thus an important distinction or ambiguity between the short-run and the long-run meaning of a rise in the skill differential or the rate of return to capital. In the short run an increase in either of these two is almost surely regressive. But in the long run, so long as the supply side reacts positively to these changes in the rate of return, the change could be progressive either because of upward mobility, or because the increase in physical capital drives down the rate of return and raises the productivity and the average wage of workers.

The ambiguity I am discussing here is a specific example of the dual function of income in a market system. On the one hand relative income determines the distribution at each point in time. Any relative increase in the income of the rich is regressive. But on the other hand income is also the signal or incentive by which economic agents are encouraged to change their behavior. A rise in the skill differential induces socially desirable education investment. Similarly a rise in profits induces investment and a shift of productive resources from less-desirable to more-desirable uses. One
makes a serious error of interpretation if one concentrates only on the short-run regressive effect of changing factor returns without taking into account progressive long-run supply responses.

**Inequality and economic growth**

In a classic article, Simon Kuznets empirically analysed the relationship between growth and equity, using the historical experience of England and the United States (Kuznets, 1955). He found that over the course of the nineteenth century, inequality rose as both countries grew, but somewhere around the beginning of the twentieth century the trend reversed and inequality began to decline. This non-linear, inverted U-shaped relationship between income and inequality has been labelled the Kuznets curve and has been the subject of a vast empirical literature looking for similar relationships in other countries and in international cross-sections between countries.

Kuznets's rationale for the relationship is based on the important idea that growth starts somewhere specific, in either a region, a sector or a city. From that starting point its effects spread through a variety of linkages to the rest of the economy. In the first stage of the process, inequality rises. Later, as the effect of the growth stimulus spreads out in the population, inequality begins to decline.

In the Kuznets study, the growth process was the Industrial Revolution. Industrialization started in the urban centers of agricultural economies. Incomes in the small industrial sector were much higher than those in the agricultural sector. As industry expanded, inequality at first increased because the higher urban wages benefitted only a small fraction of the population. This is the stage when growth and inequality are positively related or when the Kuznets curve has a positive slope. At some point, when the urban sector got big enough, inequality trends were determined by those left behind in low-wage agriculture. At that point, further industrial urban growth began to reduce inequality because it reduced the population share in low-wage agriculture.

The point here is that the growth process starts somewhere specific, after which its effects spread by a variety of linkages to the rest of the economy. In the first stages of this process inequality is almost certain to rise. Later, as the effect of the growth stimulus spreads in the population, inequality will begin to decline. How long this takes to happen, or indeed whether the spread effect is big enough to offset the initial increase in inequality, depends on the strength of linkages. Linkages are the connections between other sectors or economic actors and the sector in which growth is occurring.

Linkages are a key determinant of the relationship between growth and inequality. They determine the ‘spread effect’ of growth in the economy.
The stronger they are, the faster and further the benefits of growth will spread out, and the more equitable growth is likely to be. This notion can be applied in a variety of important ways. In some economies there are big backward regions or indigenous populations which are only weakly or marginally connected to the modern, dynamic sector where growth is occurring. The areas themselves have a significant proportion of total population, which means that their relative income levels will have a noticeable effect on inequality. But for reasons that are not entirely understood, growth in the dynamic sectors does not induce much forward or backward linkage activity. As a result, when these countries grow, there is not much of a spread effect to their backward or poor regions. Growth under these conditions tends to be inequitable. One could say that in these countries inequality is high because of growth, in the sense that if the entire country had remained in the same conditions as its backward regions, inequality would be lower. Here, growth leaves behind significant proportions of the population. Growth always leaves some people behind, if one accepts the idea that growth starts in a particular sector or area of the country. The key thing here is that if the linkages are weak and the areas left behind are large, then the interval in which growth is inequitable is likely to be a long one.

For the same reason, inequality is likely to rise with growth in countries with large indigenous populations. Typically the links between indigenous people and the rest of the economy are weak. When growth occurs, it provides little stimulus to incomes of the indigenous. Conversely one could say that inequality is likely to be lower the smaller and more homogenous the economy. Small countries with homogenous populations are unlikely to have backward regions or groups which are disconnected from the modern economy. Examples are countries like Uruguay or Hong Kong where the bulk of the population lives in a small number of interconnected urban areas. When this sort of country grows, a greater share of the population benefits because most people are linked either directly or indirectly to the sector where the growth stimulus began.

Two examples illustrate the point I am making here. In Brazil growth was very rapid between 1960 and 1980, but it was centered in the south-east, both because of rapid industrialization in the area around Sao Paulo and also because of the expansion of the agricultural frontier. The north-east, which in 1960 contained 33 percent of the population, also grew, but far more slowly than the south and south-east. Partly as a result, the Gini for Brazil rose from 0.53 in 1960 to around 0.58 in 1980, one of the fastest increases in inequality observed anywhere. Similarly China has had a period of explosive growth since the mid-1980s. But that growth was concentrated on the coast, leaving the interior of the country relatively untouched. Between 1990 and 1999 average income in the coastal provinces
and in urban centers grew far faster than in the inland provinces and the rural sector. This was a major contributor to rising national inequality indexes. For the country as a whole the Gini increased from around 0.26 in the mid-1980s to 0.37 in 2000 (Kanbur and Zhang, 2005). Over one-half of that increase was caused by the increase in the gap between rural and urban incomes, and about one-third by the rising inland–coastal gap (Kanbur and Zhang, 2005).5

Despite the predictions of Kuznets that at some point in all countries inequality should decline, there seem to be at least two features of growth under current conditions that may reverse this trend. First, modern growth has become increasingly skill- and capital-intensive. That has raised the return to both capital and skilled labor. Wage differentials for the educated have widened considerably and profits have gained relative to labor. Second, and this is more controversial, removing the barriers to capital mobility, an important part of the current generation of economic reforms, has increased the bargaining power of capital and its ability to extract wage concessions under the threat of moving to lower-cost developing countries. At the same time trade liberalization has lowered the prices of simple manufactures produced in developing countries. That has also put pressure on the wage differential in developed countries. All of this may make the world distribution more equal by narrowing the wage differential between developing and developed countries, but at the same time the within-country distributions that I have been discussing may well get less equal in both the advanced and the developing countries.

Governments can do a number of things to make the distribution more equal. The burden of taxes can be shifted toward the more affluent. Government can spend more or subsidize goods and services such as education and health care that benefit the poor. It can create a safety net financed out of general revenue to reduce or even eliminate extreme poverty. It can sponsor public works programs to absorb the unemployed and build useful infrastructure. But probably the two most important things that government can do are to maintain a sustainable and stable growth rate for the economy and to reduce the ranks of unskilled labor by effective programs of education and training.

Conclusions
In this chapter I have discussed how to measure the distribution, how distributions are likely to change as countries develop, and what government can do to make the distribution more equitable. I argued that the effect of growth on the distribution is determined jointly by the skill-intensity of growth, the structure of the economy and its factor markets and by how region- and factor-specific it is. The stronger the linkages between the
leading sector, factor or region and the rest of the economy, the more equitable growth will be.

I paid particular attention to several problems of interpretation of changes in the distribution as indicators of social welfare either within countries over time or between countries. To be a socially relevant measure, it is important that the distribution be defined over the appropriate comparator groups – those with whom people actually make welfare comparisons. The typical aggregation is by country. But distributions by regions or by age cohorts may be better indicators of welfare, even at the national level.

Comparisons of distributions over time are particularly troublesome. First, the populations are different and the lifetime distribution of any cohort is likely to be very different from the national distribution. Second, the distributions that are observed are short-run and any growth process that drives up skill differentials and profit rates will be judged regressive in the short run. But it is unclear how to evaluate this short-run rise in inequality if it generates a progressive factor supply response in the long run by increasing either the supply of skilled labor or employment-creating investment.

Notes
1. The author would like to thank the editors for comments on a previous version of this chapter.
2. This measure is called family income or consumption per capita.
3. For a fuller discussion of aggregation problems and measures see Sen (1973).
4. The distinction I am making here is closely related to the notion of horizontal equity discussed by Ravallion (2004). National distribution statistics implicitly assume that welfare is invariant with respect to where income is earned, or in other words that welfare is unaffected by equivalent transfers among groups with the same income, such as the rural and urban poor. A trade reform which lowers the price of food may leave aggregate national distribution statistics unchanged, by helping the urban poor and harming the rural poor. That could exacerbate social tensions even if the national distribution of consumption became more equal.
5. One could cite a third example. Bourguignon and Morrisson (2002) showed that there was a very large increase in world inequality between 1820 and 1950, almost all of which was caused by an increase in the gap between the fast-growing industrial economies in the USA and Europe and those in the rest of the world. Weak linkages between these economies and those on what many have called the periphery were presumably responsible for this result.

Bibliography


Morley, Samuel A. (2001), The Income Distribution Problem in Latin America and the Caribbean, Santiago: CEPAL.


Sen, Amartya (1973), On Economic Inequality, Oxford: Oxford University Press.
Introduction
Until the end of the Cold War, most development economists were not particularly concerned with the distribution of income, but instead with understanding growth and reducing absolute poverty in the developing world. For one thing, Kuznets (1955) had suggested that a deterioration in the distribution of income might be the natural outcome of the early stages of development, as people begin the shift from low-productivity subsistence agriculture to high-productivity sectors. And mainstream economists’ starting assumption, rooted in the Smithian trade-off between efficiency and equity, was that in the other direction of causation, inequality resulting for example from increased security of property rights would enhance growth by encouraging investment and savings and creating a necessary incentive for individuals to work hard.2

But beginning in the 1990s, as panel data on changes in the distribution of income in developing countries became available, as mainstream development economists became more concerned with political economy analysis, and – perhaps – once the fall of the Berlin Wall liberated the mainstream from the taboo of Marxian analysis, economists became more interested in assessing the effects of income distribution on growth. In the last 15 years a major focus of new theoretical and empirical work has been the effects of income inequality on growth and other indicators of inequality in the developing world. Much of that work has been ably reviewed in major reports of the United Nations Development Programme (UNDP), the Inter-American Development Bank and the World Bank.3

Still there is no consensus among economists that income inequality matters, and little attention among development practitioners to policies to address inequality as opposed to growth and poverty reduction.4

Obviously if people care about their relative income status then ipso facto inequality matters. That they do, to some extent, has long been remarked; consider Adam Smith, who noted that a man to retain his dignity may in one society need enough income to buy a linen shirt, and Veblen (1970) who noted that the absolutely well-off worry about their status relative to the more absolutely well-off.5 Hirschman (1973) observed that people stuck in
a tunnel in a stopped lane of traffic eventually become deeply frustrated if the other lane, but not theirs, has been inexplicably (and presumably unfairly) moving – quite independent, to extend the metaphor, of the kind of car they drive. Easterlin (1995) noted that happiness (or subjective well-being, or utility, to use the economists’ term) varies directly with one’s own income and inversely with the income of others, that is, that relative as well as absolute income matters. He comes to that conclusion in part based on surveys of happiness within countries over time; the average level of happiness has not increased even where average incomes have increased substantially. It is possible in fact that inequality of income reduces the utility or happiness not only of the relatively poor but of the better-off, who may enjoy their own affluence less if others are visibly worse off.

In this review I focus, however, on the instrumental reasons why a highly unequal distribution of income matters in developing countries. I review a large body of work, primarily of economists, indicating that beyond some level inequality in developing countries matters because: (1) where markets are underdeveloped, high income inequality is likely to inhibit growth; (2) high income inequality can discourage the evolution of the economic and political institutions associated with accountable government (which in turn enable a market environment conducive to investment and growth); and (3) high income inequality can undermine the civic and social life that sustains effective collective decision-making, especially in multi-ethnic settings.

Theory and some empirical work suggest that inequality does not undermine growth directly. Instead it is the interaction of inequality with imperfect markets or with unaccountable or incompetent governments (increasingly labeled weak ‘institutions’ in the latest literature on growth – see for example Acemoglu et al., 2000) that harms growth. In addition, and conceptually different, inequality (that is high enough) may directly create conditions that lead to or exacerbate poor governance and thus poor economic policy, and/or weak social and economic institutions and thus ineffective implementation of stable and sound policies – reducing growth through the effect on economic, political and social institutions. Weak institutions broadly defined are increasingly viewed as the key cause of low growth in developing countries. Since weak markets, poor governance and underdeveloped institutions might be said to be the very characteristics that define a country as ‘developing’, it follows that inequality is a key factor in understanding the dynamics of growth and institutional development in the developing world.

The reader will note that I do not discuss the effect of growth on inequality, the subject of the Kuznets hypothesis, nor the evidence that inequality and growth may each be simultaneously affected (Lundberg and Squire, 2003), either similarly or differently, by still other economic and non-economic
variables such as inflation and increased access to education. Once panels of household data enabled analysis of changes in the distribution of income over time within countries, the existence of a stylized Kuznets effect was not supported by the evidence (for example, Deininger and Squire, 1996), almost certainly because so many other country-specific factors compound any fundamental relationship there might be.

**Effect of inequality on economic growth and poverty: theory and evidence**

Two stylized facts emerge from the growing literature on the effects of inequality on growth. First, the evidence suggests that inequality above some level is more likely to reduce growth. Second, theory and empirical work suggest that high levels of inequality are more likely to harm growth in developing than in developed countries.

Barro (2000), in a study of the determinants of growth, was among the first to report a structurally different relationship of inequality to growth in developing compared to developed countries. Across developed and developing countries combined, he found no clear effect of inequality on growth. However, dividing the sample into the two groups he found the relationship is structurally different. In higher-income developed countries inequality may indeed be associated with higher growth (as often referred to in contrasting the USA and countries of Western Europe). Below a certain income level (about $2000 US 1985 dollars – equivalent to about US $3200 in 2000 dollars), higher income inequality is associated with lower growth. (The simple relationship is illustrated for developed and developing countries in Figure 48.1.) Cornia et al. (2004), using data from a more comprehensive set of household surveys, tested the relationship between changes in inequality and growth over almost four decades for 25 countries. They report a positive effect on growth as the Gini coefficient increases from very low levels (from the .15 typical say of subsistence economies and of the former socialist economies to .30) and a negative effect as the Gini coefficient rises from .45 (typical in Latin America and sub-Saharan Africa) to higher levels.8

The specific thresholds should not be taken too seriously, given poor measurement particularly of the distribution of income. However, they allow for a rough assessment of how widespread across people and countries in the developing world the resulting vulnerability might be. The critical thresholds of a Gini at or above .45 and income per capita at or below $3200 affect a significant number of countries and people in the developing world. Virtually no developing or transitional economies have income Gini coefficients below .30, though India and China did at about that level for much of the post-World War II period until the 1990s. About 15 percent of the population of the developing world currently lives in countries (33
Figure 48.1 Inequality and per capita income growth in developing and rich countries, 1970–2000
countries) with reported Gini coefficients of .45 or higher and per capita income below $3200 (in 2000 dollars), mostly in Latin America and sub-Saharan Africa. But that percentage mounts to 40 percent if China, whose reported 2003 Gini coefficient was 44.9, is included, and rises further to 44 percent if Brazil, whose per capita income now just exceeds $3200, is included. Other countries with per capita income below the Barro threshold where the income Gini has risen in the last 15 years and is now above .40 are Bangladesh and Pakistan. In India and Vietnam, inequality has also risen rapidly since the 1990s but reported Gini coefficients are still below .40.9

These findings are broadly consistent with theory. Why might some level of inequality enhance growth? First, inequality can be too low, as when it was imposed in state-managed economies where planning and controls replaced price and other market signals, encouraging ‘shirking’ and free-riding. A certain degree of inequality may be necessary to permit the incentives that induce individuals to work hard, innovate and undertake risky but productive investment projects, resulting in higher output and productivity, and therefore higher average incomes and growth rates. (For economists, these incentive effects are the backbone of the moral hazard argument against tax-financed distribution; Okun, 1975). Second, some concentration of income could encourage growth if high rates of saving enable more investment, and if savings rates are greater where income is concentrated in the hands of the rich whose marginal propensity to save is higher than that of the poor (Galenson and Leibenstein, 1955; Kaldor, 1961). A related idea is that investments in infrastructure and industry critical to development are large and indivisible; in the absence of well-functioning capital markets, wealth and income need to be highly concentrated to generate the minimum required resources to undertake new investment projects.10 (Recent ‘endogenous’ models of growth, however, rely much more heavily on the incentive effects of institutions and policy than on high savings and investment as the keys to sustained growth.)

The incentive effects of inequality can be thought of as the outcome of ‘constructive’ inequality, that is, income inequality that reflects solely differences in individuals’ responses to equal incentives or opportunities, and is thus consistent with efficient resource allocation.11 In contrast would be ‘destructive’ inequality, reflecting inefficient privileges for the rich, social and economic discrimination which reduces incentives for effort, investment and innovation by some groups, and in general reduced potential for productive contributions of the already poor. In a kind of tautology, destructive inequality can be defined as that inequality which results in lower, rather than higher economic growth (Birdsall, 2001).

The idea of destructive effects of inequality is consistent with the empirical evidence noted above of lower growth at very high measured levels of
inequality. Theory suggests that inequality is also more likely to be destructive in developing countries (as Figure 48.1 suggests). The remainder of this section sets out why that is likely to be the case. In summary it is because inequality tends to undermine growth when it combines with or interacts with weak markets and poor government policy. In general in developing compared to developed countries, financial and other markets are less complete and public policy is less effective in addressing market failures and imperfections.

**Imperfect credit and other markets**

Benabou (1996) and Aghion et al. (1999) develop models in which inequality exacerbates the effect of capital and other market failures on growth. When creditworthy borrowers cannot borrow because they lack collateral to comfort lenders (given imperfect information, a market failure in itself), then their lack of income or wealth limits their ability to invest. In addition, given limited liability (the borrower cannot repay more than his or her net worth), borrowers with less wealth have less incentive to exert effort to ensure success of an investment since they must pay lenders a higher portion of their returns (a moral hazard effect). In this case redistributing wealth has no adverse incentive effects – on the contrary it creates a positive incentive – and will be growth-enhancing. Weak or non-existent insurance markets will also force those without assets to bypass high-return projects. Galor and Zeira (1993) and earlier Louy (1981) suggest that the distribution of wealth affects output due to the indivisibility of investments in human capital. When it is difficult to borrow, lack of liquidity limits investments in human capital despite prospective high returns; this obviously affects the poor but may also affect the large majority of middle-income people in developing countries with a high concentration of income at the top of the income distribution. Birdsall et al. (1998) note that even where the poor are credit-constrained, they can exploit an increase in the return to potential new investment (in education or their own farm or business) by increasing their work effort. They will do so as long as the returns to their labor are adequate – as was the case in Korea and Taiwan in much of the post-war twentieth century. If labor markets are functioning well, and returns to education or other investments are rising, the credit market may not matter as much. Or in those countries, lower overall inequality of wealth, income and land (well below the Gini of 0.45 on the distribution of income), may have minimized the negative effect on growth of an interaction between inequality and weak markets.12

Obviously weaknesses in capital markets are greater in developing countries, as are compensatory policies such as enforcement of creditor rights. They are also more likely the lower average income is and the higher the
proportion of poor people, making it difficult to distinguish empirically between the negative effect of inequality per se (whether of income, wealth, education or land) interacting with weak markets from the negative effects of high rates of poverty. In any event, whether or not because capital markets are weaker and more people are poorer, it is not surprising that inequality undermines growth in developing countries though not necessarily in developed countries.

In these models, it is not actually income inequality but inequality of financial wealth or other assets that interacts with weak capital markets to reduce growth. (Only recently has household level data on financial wealth of reasonable comparability across countries become available.) But increasing evidence suggests that other assets – land and education – tell the same story. Latin America still appears to bear the costs of its historic land inequality. Carter and Coles (1998) show that concentration of land ownership is associated over long subsequent periods with concentration of income, even in countries where the economic relevance of agriculture has declined. Birdsall and Londono (1997) show that across countries inequality in the distribution of education reduces growth, and that once inequality of land and education are accounted for, inequality of income washes out as a factor affecting growth. In that respect, market economies in Latin America compared to East Asia, discussed below, do not operate differently – it is just that they operate in a context of high concentration of land and education.

Ineffective or corrupt institutions of the state and resultant poor public policy

As with weak markets, weak governments and poor public policy are likely to exacerbate the effects of inequality (of income, assets, education and so on) on growth. Behrman et al. (2000) show that differences across countries in social mobility, measured by differences in the effect of parents’ income and education on children’s education, are robustly and systematically affected by differences in two factors: public spending on primary education and the depth of financial markets. Repressed interest rates and directed credit programs that end up limiting access to credit except for privileged insiders worsen the effect of inherently imperfect capital markets on growth. Lack of adequate public spending on basic health and education means that public policy is not correcting for the inherent inability of markets alone to compensate for differences across households in endowments of all kinds. Growth is then lower than it could be since aggregate accumulation of human capital is reduced.

If income inequality interacting with poor policy reduces growth, then it is implicated in reduced poverty reduction – given that empirically, growth
has seemed necessary if not sufficient for reducing poverty, and since whatever growth occurs will help the poor less in an accounting sense the less equal the distribution of income (Ravallion, 1997, 2001). There may also be a more substantive link of inequality to the persistence of poverty where state institutions and government policy fail to ensure equal opportunities for the poor, even when there is income growth on average. Birdsall and Londono (1997) report that across countries in the period 1960 to 1990 greater land and education inequality reduced the income growth of the poorest quintile about twice as much as they reduced average income growth for all quintiles. In the extreme, unequal distribution of land may cut off altogether the usual effect of growth in agriculture on reduction of rural poverty. Some evidence suggests that agricultural growth in Latin America in the 1970s and 1980s failed to reduce poverty at all (De Janvry and Sadoulet, 2000), as large landowners captured most of the benefits. In contrast, in Indonesia, where small farmers provide the bulk of agricultural production, growth was good for the rural poor even in the days of Sukarno, and still better in the days of Suharto (Timmer, 2006a, 2006b).

Political instability and social conflict
Initial theorizing put any negative effect of inequality on growth not on its interaction with weak markets or poor public policy, but through a direct effect in the political sphere, as Benabou put it: ‘where asset markets are complete and distributional effects arise solely from the balance of power in the political system’. Economists suggested that higher inequality causes lower growth because the median voter, who is relatively poorer where inequality is high, votes for inefficient redistribution financed by growth-reducing higher taxes (Persson and Tabellini, 1994; Alesina and Rodrik, 1994). Their cross-country tests were not, however, convincing. Moreover, the median voter theory did not square with anecdotal evidence that policies in unequal countries are often shaped not by the relatively poor median voter (even where there is democracy) but by a more politically influential elite, and with lack of any evidence that redistributive policies, measured in terms of the marginal tax rate, are associated with lower growth (Easterly and Rebelo, 1993).

An alternative political explanation blames political instability on ‘social discontent’ (associated with inequality among other things) (Alesina and Perotti, 1996). Socio-political differences that reduce the security of property rights and the expected return on investment thus reduce investment and subsequently growth. In a test of the determinants of growth collapses after 1975, Rodrik (1999) found that high inequality and the quality of institutions that manage conflict were key underlying factors – not the size nor the intensity of external shocks. He argued that with high inequality,
distributional conflicts would be more difficult to resolve, delaying fiscal and monetary adjustment and diverting productive resources to bargaining over distributional changes. Benabou (1996) notes that if the rich understand the implication for growth of rent-seeking in unequal societies and of populist revolts, it may be in their collective interests collectively to transfer wealth to the poor through land reform, education subsidies or trade protection.\textsuperscript{18} The problem may be (as experience in Latin America and Africa suggests) that for such transfers to be efficient and growth-enhancing requires effective institutions of the state.

**Effect of inequality on political and economic institutions**

A large literature is concerned with the importance of effective institutions for growth (for example North, 1990; Acemoglu et al., 2004). Does inequality (in some ‘initial’ state) contribute to the failure of effective institutions to emerge in some societies? Engerman and Sokoloff (1997, 2002) suggest that differences in the factor endowments of colonial North and South America contributed to differences in the concentration of income which in turn affected the evolution of different economic and political institutions. Abundant slave or indigenous labor, and soil and climate conducive to large plantation agriculture in the south, and the opportunities for extraction of mineral wealth, were conducive to the high concentration of income, human capital and political power. The elite in the south then tended to create and sustain institutional arrangements that protected their interests but did not encourage broad-based investment, for example in education or productive economic activity. In contrast were the smallholder farms of the north, where the soils and climate were conducive to wheat, for example, and cheap labor was not available. In these settings, more democratic institutions evolved, property rights were broad-based, and a thriving smallholder class supported public financing of education and in general created local governments that were accountable to most citizens.

Public-choice models similarly attribute poor public policy to government regimes in which bureaucrats and insiders face no real checks on the pursuit of their own interests (Buchanan and Tollison, 1984). If the rich favor public policy that preserves privileges independent of their economic efficiency, inequality may not only inhibit growth by interacting with government failure and poor public policy, as set out above, but may contribute to poor institutions and government failures in the first place. The problem seems especially great when concentration of income at the top is combined with substantial poverty at the bottom, and there is not a large middle class to demand accountability from government. Easterly (2001) and Easterly et al. (2006), use country-level data on size of the middle class (instrumented
by differences in commodities produced, recalling Engerman and Sokoloff),
to study the determinants of good ‘institutions’ (measured in terms of
survey results on accountability, corruption, property rights, and so on).
They conclude that a small middle class is implicated in weak institutions,
and through weak institutions in low growth.

An example is the apparent relationship between a high concentration of
income in a society and differences across countries in the policy and insti-
tutional capacity that ensure access to education – as in the difference
between East Asia and Latin America in educational opportunities for the
poor (Birdsall et al., 1997). Supply of publicly subsidized education is likely
to be limited where the rich resist a large tax burden to finance services
which they can purchase privately. Targeting social services to the poor can
help reduce the fiscal burden of greater public spending, but easily leads to
loss of political support from the working and middle class. Without
middle-class interest and pressure, the quality of public services deterio-
rates (and the middle class resorts to private services). Thus it is possible
for income inequality to contribute to poor public policy and institutions
even where there is little or no absolute poverty – for example in US cities.

It is also likely that high income inequality encourages rent-seeking by
the rich through bribes and extortion in the political sphere, and populist
and protectionist policies when those who feel disadvantaged acquire a
that income inequality is associated with weakening of the protection of
property rights.

In short, not only does theory and some evidence suggest inequality
harms growth in interaction with poor public policy, but it is plausible that
high inequality more directly undermines good public policy by delaying or
stalling the emergence of the political and economic institutions (property
rights, an independent judiciary, accountability to voters and checks on
abuse of privileges and power) – institutions that are increasingly viewed as
fundamental to sustaining growth.

Effect of inequality on social institutions, social capital and collective
decision-making

Amartya Sen places considerable emphasis on individuals’ ‘capability’ to
participate in the life of the community as an aspect of development inde-
pendent of any implications for economic growth (Sen, 1992, 1999). Par-
ticipation in the life of the community suggests there are assets that are
held not individually but only in relation to others; Putnam (1993) defines
the asset of social capital in terms of trusts, norms and networks that can
improve the efficiency of society, ‘facilitating coordinating actions’. Social
capital may also have economic value to the extent that it reduces the cost
of transactions and of contract enforcement, and as Rodrik (1993, 1999) argues, reduces resistance of losing groups to political compromises.

There is good evidence from microeconomic analyses that income inequality adversely affects some of the inputs or correlates of social capital. In Tanzania, informal insurance is higher in communities where income inequality is lower (La Ferrara, 2000). Among sugar cooperatives in India, where land ownership is more unequal, cooperatives are less productive (Banerjee et al., 2001). The literature on local public finance addresses the same issue indirectly, in assessments of the link between income levels and the formation of communities with different amounts of heterogeneity. A typical finding is that the quality of publicly provided education is inversely related to income inequality, controlling for average income (Fernandez and Rogerson, 2003).

Finally there is the evidence from studies of crime and violence. Fajnzylber et al. (2002) assessed the impact of inequality on homicide rates in a cross section of 39 countries over the period 1965–95. Income inequality measured by the Gini coefficient had a significant and positive effect on homicide rates, robust to a variety of specifications. Ratios of income of contiguous quintiles starting with the second quintile (that is, third to second, fourth to third, and fifth to fourth) exacerbate crime, and at an increasing rate. In other words, it was not poverty nor inequality at the bottom that explained crime, but the disparity between the middle strata and their richer counterparts. It was not absolute but relative income that mattered.

It is difficult to distinguish conceptually between the effects of inequality on political and economic institutions and on such ‘social’ institutions as social capital and shared civic customs and habits. To some extent that may be because across societies such ‘institutions’ as broad-based property rights, democracy with checks on abuse of power, and ‘trust’ among citizens, tend to be correlated with each other. In any event, evidence suggests that in each category, such institutions have evolved less successfully where income inequality has been high.

**Inequality and growth in East Asia versus Latin America**

In 1960, average real per capita income in Latin America was higher than in East Asia. Since then, average per capita income has risen almost tenfold in East Asia whereas in Latin America it has less than doubled (Table 48.1). In 1960, income and land inequality were significantly higher and income concentration much more extreme in Latin America compared to East Asia. (Table 48.2); Taiwan and Korea both benefited from externally imposed land reform after World War II. The contrast over four decades between fast-growing East Asia, with its relatively low inequality in 1960
### Table 48.1 Inequality, Income and Growth in Latin America and East Asia, 1960 and 2000

<table>
<thead>
<tr>
<th></th>
<th>Income share of poorest 10% of population (%)</th>
<th>Income share of richest 10% of population (%)</th>
<th>GDP per capita (constant 2000 US$)</th>
<th>Average real GDP per capita growth rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin America¹</td>
<td>0.51</td>
<td>0.53</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td>East Asia²</td>
<td>0.42</td>
<td>0.43</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>China</td>
<td>0.32</td>
<td>0.39</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.53</td>
<td>0.54</td>
<td>1.3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

**Notes:**
- All group averages are unweighted.
- n/a indicates data not available.
- Latin America includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. 1960 and 2000 income Gini data not available for Ecuador, Guatemala, Nicaragua and Paraguay.
- East Asia includes Hong Kong, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand, but excludes China.
- East Asia excludes Singapore, Latin America excludes Bolivia, Chile, Honduras and Peru.

**Sources:** WDI (2005) and WIDER (2005).
(compared to Latin America, and in particular in fast-growing Korea and Taiwan compared to Thailand and Indonesia) and slow-growing Latin America, with its very high inequality (Figure 48.2), is consistent with the theory and evidence reported above: that high inequality in developing countries, where it is likely to be combined with imperfect and weak markets and poor government policy, reduces an economy’s growth prospects; and that high ‘initial’ inequality puts at risk the development of the economic, political and social institutions that support deeper markets, better government and sustained growth.23

Rapid growth in East Asia is associated with the region’s early export push, supported by high savings and investment and healthy rates of total

---

Table 48.2  Income, Education and Land Inequality in Latin America and East Asia, 1960 and 2000

<table>
<thead>
<tr>
<th></th>
<th>Income Gini(^3)</th>
<th>Education Gini(^4)</th>
<th>Land Gini(^5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin America(^1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.51</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.06</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>East Asia(^2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.42</td>
<td>0.43</td>
<td>0.58</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.04</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>China</td>
<td>0.32</td>
<td>0.39</td>
<td>n/a</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.53</td>
<td>0.54</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Notes:
All group averages are unweighted.
n/a indicates data not available.
1. Latin America includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. 1960 and 2000 income Gini data not available for Ecuador, Guatemala, Nicaragua and Paraguay.
2. East Asia includes Hong Kong, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand.
4. Education Ginis for population aged 15 years and over.
5. East Asia average excludes Hong Kong, Singapore, Taiwan and Malaysia for which data are not available. Latin America average excludes Bolivia, Chile, Ecuador, El Salvador and Guatemala.

factor productivity growth in manufacturing (World Bank, 1993). Behind export success were other factors rooted in rapid changes in household decisions and behavior. Those other factors included unprecedented gains in small farmers’ agricultural productivity, high demand for schooling including of girls, and declines in fertility far more rapid than and at lower income levels than had occurred in the industrialized economies (Birdsall and Sabot, 2002). Governments generally ensured that exchange rates were competitive and that fiscal discipline kept inflation low. Governments also favored public investment in basic (primary and secondary) education.

In Latin America, inflation and overvalued exchange rates penalized agriculture, and were combined with tariff and other protection of industry and subsidies to capital that may have reduced the demand for labor. Spending on education was comparable to that in East Asia but was much more concentrated on highly subsidized university education for a select few, responding to the demands of richer households. In the 1960s, educational attainment of the adult population was at roughly the same levels in East Asia and Latin America, and inequality of education (measured in terms of number of years of schooling achieved) was actually higher in East Asia. Since then, educational attainment has risen more quickly in

Sources: WDI (2005), WIDER (2005), and authors’ calculations.

Figure 48.2 Income Inequality and GDP per capita growth in East Asia and Latin America, 1960–2000
East Asia and education inequality has fallen faster (Birdsall and Londono, 1998). Broad-based investment in basic education in East Asian countries led to substantial growth of labor productivity and enabled firms to acquire and adapt new technologies and move up the value chain as increasingly skilled cohorts of workers became available (Schultz, 1961; Romer, 1994). The export-push, labor-demanding strategy chosen by East Asian countries generated the conditions for a savings and investment boom in middle-income and poor households and farms (Birdsall et al., 1998).

It seems plausible that one region’s lower inequality compared to the other, among other things, affected the difference in the two regions’ subsequent trajectories of growth, inequality and investment in human capital. The story is not straightforward. Latin America has a longer history of democracy, for example. But the differences do suggest that the potential negative effect of inequality – of income, land and other assets – on growth and on the evolution of institutions that support the development process, deserves continuing attention.

Notes
1. I am grateful to my former and current research assistants: Gunilla Pettersson, Christine Park, and Karelle Samuda; and to Lyn Squire, John Williamson and Amitava Krishna Dutt for their thoughtful comments on an earlier draft. This chapter was meant to be co-authored with Richard Sabot, with whom I enjoyed spirited discussions about its content and emphasis before his untimely death in 2005; this chapter reflects inadequately those discussions and the contribution he would have made.
2. For example, Finis Welch entitled his 1999 address to the American Economics Association ‘In Defense of Inequality’. The reference to the Smithian trade-off is to Smith’s Theory of Moral Sentiments published in 1759 (Smith, 1982 [1759]). Kaldor (1961) noted that a higher profit share would encourage savings on the assumption that capitalists have a higher propensity to save, from which it follows that when income is more concentrated, savings and investment and thus the equilibrium rate of growth will be higher.
4. Lyn Squire (personal correspondence; and see Lundberg and Squire, 2003) makes the point that policy recommendations for addressing inequality may not be much different from those meant to address poverty in a country with an egalitarian distribution of income. (An exception might be tax policy, which ideally might be more progressive in the former setting, if only to sustain politically open markets. In addition greater priority in the face of political and administrative constraints might go to anti-trust and anti-monopoly programs in high-inequality settings.) I do not try to address this point in this chapter since it is not focused on policy per se but on a review of the implications of inequality for the dynamics of growth in the developing world.
5. Graham and Felton (2006) provide a survey of recent studies linking measures of ‘well-being’ (or reported ‘happiness’ in surveys of individuals) to prevailing levels of inequality.
Results depend on setting, definition of reference group, and the particular measure of well-being. In Europe and the USA inequality has generally negative effects on reported measures of well-being.

6. Graham and Pettinato (2002) make the point that what is important is people’s perceptions about their current and future income relative to others. Graham and Felton (2006) report based on happiness surveys that people in Nigeria are as happy as people in France despite the huge discrepancy in per capita incomes.

7. Much of what I say about income inequality applies to consumption inequality, and much theory reviewed below applies better to wealth than to income and consumption inequality. In principle ‘income inequality’ as I use it refers to ‘permanent income’, though in fact empirical work on income inequality is almost always based on current income, and sometimes on wages and other pre-tax income. Elsewhere I have used the term ‘money inequality’ to distinguish income and consumption inequality from inequality of ‘opportunity’ (which is difficult if not impossible to measure) and of land, education and other non-monetary assets. See Birdsall (2001).

8. See Chapter 47 in this volume for a discussion of the Gini coefficient and other measures of distribution. The studies referred to all use panels of country observations and employ country fixed-effects estimations, so that they are assessing changes over time within countries, not differences across countries.

9. Reported Gini coefficients are from the WIDER (WIID2a) database; see http://www.wider.unu.edu/wiid/wiid.htm and WIDER (2005). Income per capita is from the World Bank World Development Indicators (http://www.worldbank.org/data). For the statements in this paragraph, I used Gini coefficients from as many countries as possible. For some countries only Gini coefficients of the distribution of consumption are available. The distribution of consumption will be more equal than the distribution of income so that the number of countries and people in the categories I defined may be higher than stated here.

10. With this in mind, many developing countries embraced the need for the state to assume the commanding heights of the economy and used tax and donor resources to finance state-led industrial investments throughout much of the post-World War II twentieth century. This approach almost certainly, and ironically, led to increased concentration of income. Worse, in some countries the later privatization of those investments further increased income concentration, though there is also good evidence that privatization of water, electricity and other utilities has improved access to these services by the poor (Nellis and Birdsall, 2005).

11. Rawls (1971) argued that unequal systems of incentives and rewards may be justified if they improve the position of the least advantaged. His fundamental point was that an increase in inequality can only be justified if the outcome is an improvement in the welfare of the worst-off.

12. A closely related more Keynesian point is that greater inequality may depress aggregate demand, and thus investment incentives and growth – even where markets are otherwise functioning well. See Chapter 14 in this Handbook.


14. Ravallion (2001) reported an average elasticity of poverty reduction with respect to growth of −2.5, implying that for every 1 percent increase in the growth rate in average income, the proportion of the population living below $1/day falls by an average of 2.5 percent.

15. See also Deininger and Squire (1996). These findings contrast with the conclusion of ‘Growth is Good for the Poor’ in which Dollar and Kraay (2002) find that average incomes of the poorest quintile rise proportionately with average incomes in a sample of 92 countries spanning the last four decades.


17. De Mello and Tiongson (2006) find no evidence that governments of highly unequal countries are more likely to attempt to redistribute income.

18. Similarly it is often in the collective interest of an ethnic or racial majority to support anti-discrimination and other policies and programs to reduce horizontal inequalities, that is, inequalities among groups in political, economic and social dimensions, as these
otherwise can provide the basis for inter-group animosity and fuel civil conflict (Stewart, 2001; Ostby, 2003). See Chapter 63 in this Handbook.

19. On the demand side, low public spending combined with pressure to maintain or expand enrollments has led to low-quality schools, reducing the economic returns to poor families of sending children to school who can otherwise help at home or work. In effect schooling could be analyzed in terms of a two-sector model, with poor families confined to one sector with low returns, and the rich going to the other sector where returns are high. The difference in returns between poor and rich would explain the high dropout rates throughout much of Latin America, even in the face of high returns on average to those who manage to complete secondary school (Behrman and Birdsall, 1983).

20. The importance of institutions in development is discussed further in Chapter 61 in this Handbook.

21. In the USA the percentage of households that participate in various membership organizations is higher in metropolitan areas with lower income inequality – controlling for racial and ethnic heterogeneity, income, education and other household characteristics. The effect is substantial. An increase in the Gini coefficient of inequality by one standard deviation leads to a reduction in the probability of participation of 24 percentage points – more than twice the effect on participation of an individual going from the status of high school dropout to high school graduate or higher (Alesina and La Ferrara, 1999).

22. Land inequality is still extremely high in Latin America.

23. Rapid growth in East Asia without accompanying increases in inequality also contravenes the pattern suggested by Kuznets. More recently in China, rapid growth has been accompanied by rising inequality.

References


Understanding poverty in all its dimensions

The eradication of poverty is a fundamental goal of development. Ameliorating the lot of the poor is central not only in its own right but because improvements in their health, education and access to credit, infrastructure and markets unleash the untapped potential of large sections of the population, thereby contributing to overall growth and development. Progress in reducing poverty is thus a key indicator of development. Webster’s Dictionary defines poverty as ‘the state of having little or no money and few or no material possessions’. Taking this definition as a starting point, one person out of every six worldwide struggles to survive each day on less than the price of a cup of coffee in the United States. This statement contains two immensely important messages: a large part of the world’s population is obliged to live on the merest pittance; and the disparities between the rich and the poor are truly staggering.

Prestigious though it undoubtedly is, Webster’s Dictionary may not, however, capture current thinking well about the term ‘poverty’ nor its current usage in much of the professional literature. As a result of intensive research on the subject, economists and other social scientists have developed new and richer perspectives on the manifestations of poverty that better capture the dire straits that constitute daily life for so many people. For example, Amartya Sen has emphasized the importance of an individual’s capabilities to determine their own lives (Sen, 1999), of which control over resources (income) is only one aspect. Others have explored the factors that empower the poor to lead full and productive lives (Stern et al., 2005). In addition to low incomes and inadequate consumption, poverty in these interpretations is understood to encompass lack of education, poor health, insecurity, violence, social and political exclusion, as well as absence of basic rights and human dignity. This broader conceptualization of poverty has pushed development experts to expand the conventional indicators of poverty to reflect today’s more comprehensive characterization of the phenomenon. With this in mind, this chapter undertakes three tasks.

It first describes the attempts to measure poverty as the definition of the term has expanded from an initial focus on subsistence to a broader appreciation of the many elements constituting a person’s well-being. As we shall see in the next section, even unidimensional measures of poverty based
exclusively on income or expenditure pose many conceptual problems as well as the ever-present difficulty of securing appropriate data. Moving to richer, multidimensional measures poses additional problems, among which the allocation of appropriate weights to the constituent elements is perhaps the most daunting.

The second task is the presentation of the best available evidence. Conceptual difficulties notwithstanding, statistics can be marshaled to demonstrate two important points. First and foremost, progress in reducing poverty on almost all fronts has been greater since 1980 to date than in any other similar period in the recorded history of mankind. And second, progress has been uneven in several respects: across countries; within countries; and across the various dimensions of poverty. Evidence on each of these points is presented in the third section. If progress has been uneven around the world – and it has – then it is of interest to explore why some countries or regions have done better than others and to share the lessons of their success with those that have fared less well. This, the third task, is undertaken in the fourth section. It concludes with some observations on what the available evidence suggests for future efforts to attack poverty.

**From unidimensional to multidimensional poverty**

_A unidimensional focus on income_

Consider first how to measure poverty when this is confined to meaning lack of income. For purposes of exposition assume that we have a universally accepted measure of income (as we shall see, matters are not so simple). Call this measure \( y \). Let us also assume that we know the income of each individual in a population so that they can be ranked according to income beginning from the individual with the least income and proceeding in order to the individual with the greatest income. Finally, let \( y^* \) be the poverty line, that is, the income threshold below which people will be said to be poor.

Even with all the elements apparently in place to measure poverty, the analyst still faces the task of deciding how to present the evidence on poverty in a single indicator. Here I present three well-known indices to illustrate the range of choice and the issues to be considered. Perhaps the simplest is to count the number of individuals for whom \( y < y^* \). Known as the ‘head count’, this indicator represents the number of individuals in poverty and is the most commonly used measure.

The head count, however, says nothing about the depth of poverty or the extent to which an individual’s income falls below the poverty line. To address this issue, the ‘poverty gap’ measures the average income needed to bring the incomes of all poor people up to the poverty line or \( y^* - \bar{y} \) where \( \bar{y} \)}
is the mean income of all individuals below the poverty line. The poverty gap however makes no allowance for the severity of poverty because it treats an extra dollar for the poorest person among the poor as equal in value to an extra dollar for the person just below the poverty line. The Foster–Greer–Thorbecke index overcomes this deficiency by appropriate weighting. Thus, the measure of severity for the \( i \)th individual is given by \( (y_i^* - y_i)^\alpha \) where \( y_i \) is the \( i \)th individual's income and \( \alpha \) is the weight. As \( \alpha \) increases beyond unity, the weight of those whose incomes are well below the poverty line is magnified in the summary index for all individuals (Foster et al., 1984). This feature, however, introduces a degree of arbitrariness because the appropriate value of the weight is inevitably a matter of judgment.

While the preceding points should be born in mind, much of the effort to track progress in reducing poverty worldwide has focused on the head count, a practice that we continue in the third section. Before looking at the numbers, however, I will first examine some of the difficulties in arriving at a measure of income and deciding on a poverty line, difficulties I had assumed away in the forgoing discussion.

**Measuring income and expenditure**

Researchers seeking to measure income in developing countries face many daunting challenges. Some of the main difficulties include: how to capture non-marketed and non-priced subsistence output such as crops grown by households for their own consumption; how to allow for free social services and pure public goods; whether to use data on income or expenditure; how to accommodate differences in the demographic composition of households; how to distinguish between temporary and permanent poverty; how to manage seasonal variations in income; and so on (Kanbur and Squire, 2001; Lok-Dessalien, 1999).

While there is no universally accepted procedure for dealing with each of these issues, in practice most analysts follow some conventions. Thus, expenditure is usually preferred to income since the former is thought to be a better indicator of permanent income. Since data are usually collected at the level of the household, most analysts present information on expenditure per household member by dividing household expenditure by the number of household members. Thus no allowance is usually made for the different consumption needs of children and adults although ‘equivalence scales’ are available and are used on occasion. Nor is any allowance made for distributional rules within the household based on gender, age or working status.

The researcher’s ability to deal with these issues depends crucially on the quality of the data, which are usually obtained from household surveys. In
addition to the difficulties noted above arising from the number of activities, products and services that are unrecorded, estimates of household expenditure are also affected by the limitations of sample surveys (Deaton, 2003). Recall errors, short reference periods and the exclusion from the sampling frame of people in remote areas and other marginal groups who are most likely to be poor are common problems. The quality of enumerators and the consistency of surveys over time are other factors complicating the task of the researcher. That said, surveys of reasonable quality are now available for 97 countries covering 93 percent of the population in the developing and transition worlds (Chen and Ravallion, 2004).

Setting the poverty line

Poverty lines are commonly distinguished according to whether they are absolute or relative. While arguments can be advanced that absolute measures best capture the notion of individual capabilities incorporated in the broader definitions of poverty, the issue is more open in the case of poverty confined to income or expenditure. In this case, ‘absolute poverty’ refers to subsistence below some minimum, socially acceptable norm, usually established on the basis of nutritional requirements plus other essential goods. ‘Relative poverty’ compares the proportion of the population below a poverty line defined relative to mean income or some measure of the overall well-being of the population. Relative poverty lines thus adjust to changing circumstances. Absolute poverty lines on the other hand remain unchanged and are therefore better able to track changes in poverty over time and, where they are based on equivalent real baskets of goods and services, can also be used to aggregate across countries to track global poverty. For this reason, the remainder of this chapter focuses on measures of absolute poverty.

The absolute poverty line currently used to track global poverty is the well-known figure of a $1 a day. Introduced first in the 1990 World Development Report (World Bank, 1990), this figure is based on actual poverty lines then prevailing in some of the poorest countries in the world (World Bank, 1990). These measures typically are based on the cost at local prices of a diet providing the minimum nutritional requirements for subsistence, plus some allowance for other basic needs, usually confined to clothing and shelter. To translate these figures into dollars, purchasing power parity (PPP) conversion factors are used. Although not designed to capture the purchasing power of the poor, they remain the best instrument for translating local currencies into dollar equivalents for purposes of international comparisons. This weakness notwithstanding, the comparison revealed a marked bunching of these national lines around the dollar-a-day mark and resulted in its selection as a reasonable indicator of absolute
poverty, at least as perceived by those in the countries suffering most acutely from low incomes. For the remainder of this chapter, we will use this poverty line to indicate extreme poverty but rely as well on a higher figure – roughly $2 a day – to provide a measure of poverty that is more than mere subsistence.

Multidimensional measures of poverty
Estimates of poverty based exclusively on income or expenditure fail to capture significant aspects of deprivation experienced by the poor. Hence there is a need to go beyond traditional methods of measuring poverty and address the broader multidimensional nature of poverty by including measures of health status, educational attainment, political voice and social inclusion as well as measures of control over material resources. While conceptually straightforward, this extension encounters several practical problems of implementation (Falkingham and Namazie, 2001; Boltvinik, 1999).

Data issues  A major difficulty is that many of the required measures are not readily available or, where they are available, suffer from problems of interpretation. Consider political voice or social inclusion. Neither has been measured on a routine basis or according to a widely endorsed procedure. While data collection efforts will undoubtedly continue to improve, it is currently difficult to track progress in all dimensions, especially for historical periods. Even where measures are available on a regular basis and are collected according to generally accepted conventions, quality remains an issue. For example, educational attainment is often represented by the net primary school enrollment rate but such measures are silent on the quality of education actually provided.

Moreover, many of the measures of interest are reported only as national aggregates with no distributional breakdown. Consequently, it is not possible to develop measures of who is poor or deprived and who is not for many of the non-income indicators. At best, one can report national averages. Increasingly, however, household surveys are incorporating questions about health status, educational attainment, and so on. A common result emerging from these surveys is that the poor in one dimension are, as one might expect, often the poor in other dimensions as well. For example, school attendance rates for 6- to 17-year-olds for the poorest decile of the population ranked by income were 31 percent in Pakistan, 32 percent in Nepal and 51 percent in Vietnam, compared with rates of 71 percent, 88 percent and 84 percent for the richest decile in the same countries (Appleton and Song, 1999). Evidence on health status tells a similar story. Stunting, as measured by height-for-age, affected 52 percent of preschoolers in the poorest decile in Pakistan, 55 percent in Nepal and 58
percent in Vietnam. In contrast, the corresponding rates for the richest segment were 32 percent, 27 percent and 23 percent (Appleton and Song, 1999).

While quantitative information on health status and educational attainment by households ranked according to income or expenditure is beginning to appear, more qualitative approaches are required to capture political voice, social exclusion, measures of self-respect, and so on. Interviews and focus-group discussions with villagers, urban slum dwellers and marginalized communities remain the best means of securing the fullest and most complete understanding of the plight of the poor (Narayan et al., 2000).

To aggregate or not Putting the data issues to one side, the focus on multiple dimensions of poverty raises the question of the most useful form of presentation. One possibility is to combine the various individual indicators into a single index. Thus, if there are now \( y_n \) indicators with corresponding weights represented by \( \alpha_n \), then \( \sum \alpha_n y_n \) provides an overall measure of well-being. In principle, households could then be ranked according to their value of this measure, and then those falling below some minimum could be identified as the poor, in the same manner as with the single indicator. In fact, this is never done because as noted above distributional data do not exist for many of the dimensions of interest. Nevertheless, attempts have been made to combine income and social indicators at the national level to provide a more complete measure of well-being.

The best-known such index is the Human Development Index introduced by the United Nations in 1990 (UNDP, 1990) and its subsequent extension, the Human Poverty Index. The latter aggregates three separate measures: longevity (the percentage of people expected to die before age 40); literacy (the percentage of adults who are illiterate); and deprivation in overall economic provisioning both public and private (the percentage of people without access to water and health services and the percentage of underweight children below the age of five). The basic problem with such aggregates is that there is no satisfactory way of arriving at the weights (Srinivasan, 1994). The United Nations Development Programme (UNDP) has assumed weights of one-third for each of the three measures without any real justification other than convenience. That said, composite indices are considered valuable at the global level for advocacy purposes, but less so for country-specific policy-making purposes since they effectively conceal knowledge on the individual measures through the process of aggregation. For this reason, I report an array of indicators in the remainder of this chapter to capture the main dimensions of poverty but stop short of using an aggregate index.
Reducing poverty: the record since 1980 to date

Progress in the aggregate: a success story

The previous section has introduced the basic proposition that poverty should be understood as a multidimensional phenomenon: it is the failure to meet minimally acceptable standards, not only with respect to income or expenditure, but also with respect to health, education, and social and political values. This richness carries a cost: it is difficult to arrive at simple statistics that fully capture all aspects of such a broad concept. We have also seen that attempts to aggregate individual indicators into composite measures pose insuperable problems and in fact conceal information. Accordingly, I present in Table 49.1 several indicators, each of which reflects an important dimension of poverty.

The table reveals two decades of genuine progress. Perhaps the most dramatic improvement is the virtual halving of the percentage of the population living in poverty in low-income and middle-income countries (following World Bank classifications). In Table 49.1, the poor, defined as those subsisting on less than $1 a day, accounted for 40 percent of the population in 1981 but only 21 percent in 2001. In terms of absolute numbers, however, the outcome is not so impressive. Because of population growth (roughly an increase of 1.5 billion people), the number of poor only fell by 390 million. On the other hand, had the percentage remained at its 1981 level, the number of poor would have increased to over 2 billion, implying that as many as 1 billion people, or one-fifth of the population, avoided poverty relative to what might have happened based on an extrapolation of 1981 circumstances.

Table 49.1 also presents the progress of low- and middle-income countries in reducing poverty in its non-income dimensions and in improving the

| Table 49.1 Measures of aggregate well-being in low-income and middle-income countries |
|-----------------------------------------------|---------|---------|---------|
| % of extreme poor                             | 40.4    | 27.9    | 21.1    |
| Number of extreme poor (billion)              | 1.48    | 1.22    | 1.09    |
| Life expectancy (years)                       | 60      | 63      | 65      |
| Under-5 mortality rate per ‘000               | 131     | 103     | 87      |
| Literacy (%)                                  | 61      | 68      | 78      |
| Net primary school enrollment (%)             | 78      | 95      | 97      |

Source: Chen and Ravallion (2004) and World Bank (various issues).
welfare of their people. It shows an increase of four years in life expectancy at birth for this group of countries during the two decades to 2001. To see an increase of one year in life expectancy every five years is remarkable progress. While child mortality rates also witnessed significant progress, they remain unacceptably high. The literacy rate also shows steady improvement in this group of countries and net primary school enrollment is now almost 100 percent.

**Failures and setbacks**

The progress in aggregate does not mean that everyone advanced and in all dimensions. Indeed, more disaggregated evidence reveals that some groups not only failed to participate in the overall progress, but they saw their situation worsen in one respect or another. At least four groups can be identified where progress in one or more dimension of poverty departs significantly from the overall progress revealed by Table 49.1: those in Africa; those in marginalized regions or social groups despite country-wide progress; those just above the poverty line; and those where indicators for health or education have improved while poverty as measured by expenditure has worsened, or vice versa. I briefly consider each in turn.

The overall progress revealed in Table 49.1 notwithstanding, one region of the world – Africa – has seen poverty worsen. Recall that the number of extremely poor in China fell by 420 million during the two decades of the 1980s and 1990s. This is more than the worldwide decline in the extremely poor – 380 million. Thus, in the rest of the developing and transition world the number of poor increased slightly during this 20-year period, and if some regions saw declines in poverty, as South Asia did, then other regions must have experienced substantial increases in extreme poverty. This happened in sub-Saharan Africa, and dramatically so: the number of extremely poor almost doubled from about 160 million in 1981 to 315 million in 2001 (Table 49.2). Sub-Saharan Africa is the only region in the developing world to see the head count index increase – from 41.6 percent in 1981 to 46.9 percent in 2001.3 As a result, Africa’s share of the world’s extremely poor increased from 11 to 29 percent during this period. Worse yet, not only has the number of Africans in extreme poverty increased, but the severity of their poverty has also increased. While the rest of the world’s poor saw their daily expenditure increase from $0.70 in 1981 to $0.77 in 2001, those in Africa saw theirs fall from $0.64 to $0.61. Poverty is worsening in Africa and becoming more concentrated in Africa (Chen and Ravallion, 2004).

Uneven progress is apparent among social indicators as well. Take for example life expectancy, perhaps the most fundamental measure of overall well-being. Table 49.1 reveals steady progress. This has been true of almost all countries, but 26 countries saw life expectancy fall during the 1990s and
in some cases precipitously so. For example, life expectancy fell by 17.8 years in Botswana, 16.3 years in Zimbabwe and 14.1 years in South Africa (World Bank, 2004). These figures depict a rapid undoing of steady progress over many years. Nineteen of the 26 countries are in Africa, and four are in the republics of the Former Soviet Union. While causes vary across countries, the advent of the AIDS crisis has played a major role, especially in Africa. In several countries the prevalence of HIV among the 15–49 age group in 2001 exceeds one-fifth: Botswana (38 percent), Lesotho (29.6), Namibia (21.3), South Africa (20.9), Swaziland (38.2) and Zimbabwe (24.9) (UNAIDS, 2004).

This phenomenon of uneven progress across regions of the world is also evident within countries. A second group therefore comprises those living in certain areas or belonging to certain ethnic groups who have not participated fully in their country’s overall progress. Consider the case of India. Using the official poverty line, the number of Indian poor fell from 320 million in 1993–94 to 260 million in 1999–2000. But not all states prospered equally. Indeed, two states – Madhya Pradesh and Orissa – saw the numbers in poverty increase, albeit only slightly (Government of India, 2002). Groups within states can also experience different outcomes. For example, the incidence of poverty in the coastal districts of Orissa declined from 46 percent to 32 percent during this period. The experience of the southern districts, however, was the reverse: the incidence of poverty increased from 43 percent to 50 percent. And the districts with the highest initial incidence of poverty – the northern districts – experienced a significant increase from 66 percent to 81 percent (de Haan, 2004). Isolation in economic terms, political terms, social terms and racial terms underlie many of these examples of lagging regions or lagging social groups.

A third, major group to witness a deterioration in their well-being is the less poor, or those just above the extreme poverty line. The dollar-a-day poverty line is usually characterized as depicting extreme poverty; it allows mere subsistence at best. Consequently, a higher poverty line – $2 a day – is also frequently employed. This provides important new information. Using this measure of poverty, the number of poor actually increased from

\[
\begin{array}{|c|c|c|}
\hline
& 1981 & 1990 & 2001 \\
\hline
% of extreme poor & 41.6 & 44.6 & 46.9 \\
Number of extreme poor (million) & 163 & 227 & 316 \\
\hline
\end{array}
\]

Source: Chen and Ravallion (2004).
2.4 billion in 1981 to 2.7 billion in 2001 (see Table 49.3). Given that the number below the dollar-a-day line fell, the net result is a significant increase in the number of people surviving on expenditure per day of between $1 and $2. In fact, the number increased from about 1 billion in 1981 to 1.7 billion in 2001. This is significant for at least two reasons. It suggests that the factors – be they growth, redistributive policies, or whatever – that produced the decline in extreme poverty have been less successful in helping those just above the extreme poverty line. And it suggests that a large number of people remain highly vulnerable to droughts, economic crises, conflicts or other calamities that could easily push them into the ranks of the extremely poor.

The final group comprises those who have experienced uneven progress among indicators. While Table 49.1 shows that in aggregate the various indicators all move in a positive direction, in some cases progress in reducing poverty as measured by increases in expenditure has co-existed with deteriorating outcomes for the social indicators. Thus, eight of the 19 African countries that saw life expectancy fall in the 1990s enjoyed increases in real gross domestic product (GDP) per capita. Since HIV was prevalent in all eight countries, it appears that the disease strikes life expectancy more or sooner than GDP. In still other cases, real GDP per capita has declined while life expectancy has continued to increase. Indeed, life expectancy increased in just over half of the 48 countries that saw real GDP per capita fall in the 1990s (World Bank, 2004). While many factors come into play, it is noteworthy that 22 out of the 26 countries that saw continued progress in life expectancy spent 2 or more percent of GDN on health. The often referred-to cases of Sri Lanka and the state of Kerala, India, are powerful illustrations of how sustained investment in nutrition and health can result in unusually long life expectancies at relatively low levels of income. Thus, life expectancy in Sri Lanka was 73 years in 2000 despite its relatively modest level of GDP per capita.

Table 49.3  ‘Vulnerable’ population

<table>
<thead>
<tr>
<th></th>
<th>1981</th>
<th>1990</th>
<th>2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of poor</td>
<td>66.7</td>
<td>60.8</td>
<td>52.9</td>
</tr>
<tr>
<td>Number of poor (billion)</td>
<td>2.45</td>
<td>2.65</td>
<td>2.74</td>
</tr>
<tr>
<td>% of ‘vulnerable’</td>
<td>26.3</td>
<td>32.9</td>
<td>31.8</td>
</tr>
<tr>
<td>Number of ‘vulnerable’ (billion)</td>
<td>0.97</td>
<td>1.43</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Source: Chen and Ravallion (2004).
Learning from the past

The review of progress in reducing poverty in all its dimensions presented in the previous section revealed substantial but uneven progress. Overall the well-being of the poor has improved, but many have been bypassed or marginalized and others have seen their situation worsen. This suggests two broad questions: What is driving the general progress? And what can be done to ensure that progress is more widespread and inclusive?

In answering these questions, the starting point for most analysts is the role of GDP growth. Table 49.4 illustrates the association between growth in GDP per capita and selected indicators of poverty. For the 13 countries that enjoyed growth rates of at least 2.3 percent a year in GDP per capita in the 1980s and 1990s, the head count index fell by seven percentage points, illiteracy by six percentage points, and life expectancy increased by more than three years. Thus, the people in these 13 countries saw average conditions improve in three important dimensions. At the other extreme, those living in the 39 countries that experienced low growth in GDP per capita saw the head count index increase, illiteracy fall by over seven percentage points, and life expectancy increase by 1.4 years.

The following interpretation of these aggregate trends draws on a wide range of other material including country studies and case studies as well as cross-country analyses that cannot be reported here because of space limitations. Table 49.4 plus the evidence of many other studies indicates a strong but imperfect relationship between growth in average incomes and the incomes of the poor. This reflects the tendency for inequality in national

<table>
<thead>
<tr>
<th>Change in indicator: comparing 1980s and 1990s</th>
<th>Unit</th>
<th>Period</th>
<th>High growth</th>
<th>Moderate or improved growth</th>
<th>Low growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Count Index % with less than US$1 a day</td>
<td>1990s</td>
<td>24.1</td>
<td>31.4</td>
<td>36.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1980s</td>
<td>31.0</td>
<td>32.1</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>Illiteracy %</td>
<td>1990s</td>
<td>17.2</td>
<td>31.2</td>
<td>31.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1980s</td>
<td>22.9</td>
<td>37.6</td>
<td>38.8</td>
<td></td>
</tr>
<tr>
<td>Life expectancy Years</td>
<td>1990s</td>
<td>70.0</td>
<td>62.9</td>
<td>59.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1980s</td>
<td>66.8</td>
<td>60.6</td>
<td>58.4</td>
<td></td>
</tr>
<tr>
<td>Number of countries</td>
<td>13</td>
<td>53</td>
<td>39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

income to change only slowly (see Li et al., 1998) implying that incomes of the poor increase more or less by the same percentage as the incomes of everyone else (Dollar and Kraay, 2002). The strong influence of growth is seen clearly in Table 49.4. Results of this sort, however, say nothing about policy or causality. Thus, the observed transformation of growth into a reduction in the head count index may or may not require a wide range of redistributive policies (Kanbur, 2003). The evidence presented thus far is silent on this issue. And with respect to causality, both growth and inequality are outcomes of the same economic system and all the factors that influence that system (Lundberg and Squire, 2003). Thus, growth and poverty as measured by the head count index are jointly determined.

Reductions in illiteracy, on the other hand, are not obviously correlated with growth and are therefore presumably driven by other factors. For countries with universal enrollment, reductions in illiteracy are largely determined by each country’s population dynamics and the share of school-age children in the population. And for those without universal enrollment, the capacity of the school system is the decisive factor. The key policy instrument in these circumstances is the provision of adequate support from the public budget to maintain schools, improve quality and expand capacity where necessary. While funding alone is not sufficient – many other factors influence the delivery of public services (World Bank, 2004) – it is nevertheless an essential ingredient. As long as countries can maintain budget outlays for schooling, we should not therefore expect to see a strong relationship between movements in the growth rate and reductions in illiteracy. In the extreme, however, a collapse in GDP as has occurred in failed states or countries in major transitions will inevitably undermine the educational system. For example in countries like Albania, Bosnia, Congo, the former Yugoslavia, Rwanda, and so on, various shocks and ethnic conflicts have resulted in a colossal destruction of human and social capital (World Bank, 2000).

Improvements in longevity, however, do appear to be linked with growth in GDP per capita (see Table 49.4). Nutritional intake, a key factor influencing morbidity and longevity, depends to a considerable degree on the incomes available to households to buy food of increasing variety and quality. At the same time, life expectancy, like literacy, reflects the quantity and quality of public spending on, in this case, health services. Extraordinary examples like Sri Lanka and the Indian state of Kerala reveal the power of intensive and sustained public support for health services. By the same token and as we have seen above, major new diseases like HIV/AIDS can quickly undermine years of progress in extending people’s lives.

The key question facing the policy-maker is, of course, the policies and programs required to achieve growth in GDP and to ensure that all, and
especially the poor, benefit. While it is not possible to go into detail, the broad elements of a strategy consistent with the existing evidence can be sketched. In short, the evidence from the success stories suggests a ‘two-part strategy’ that on the one hand harnesses market incentives, social and political institutions, infrastructure and technology to promote growth, and on the other hand, supports the poor through the provision of health, education and other basic services (World Bank, 1990). The first part of the strategy provides income-earning opportunities for the poor, while the second part strengthens the capabilities of the poor to take full advantage of those opportunities. In this sense, the two parts of the strategy are mutually reinforcing.

A recent assessment of development experience that takes a very broad definition of poverty as its starting point has elaborated and refined this ‘two-track’ approach (Stern et al., 2005). The first pillar of this strategy calls for creating a supportive investment climate to encourage firms and farms, small and large, to invest, create jobs and increase productivity. The investment climate is determined by several factors that can be classified under three broad groups: macroeconomic and trade policies, infrastructure, and governance and institutions. The second pillar involves empowering and investing in poor people by enhancing their health, education and security and by fostering mechanisms for them to participate in the growth process of the economy. This approach is broadly consistent with the evidence reported here.

In conclusion it is worth making two cautionary remarks. First, useful though it is to have some general view on strategy, country circumstances vary so much that careful and possibly major tailoring may be required to arrive at an appropriate national strategy. For example, many developing countries have suffered from ethnic strife. In Sierra Leone, the prolonged strife has exacted a heavy human and economic toll. Other countries have been ravaged by the AIDS epidemic. Any national poverty-reduction strategy would have to recognize and respond to such local circumstances. Second, special actions may be required for groups which may otherwise be excluded from the national strategy’s reach such as ethnic minorities, HIV/AIDS-positive individuals and those damaged by war, or which have suffered temporary setbacks arising from price fluctuations, unemployment or natural disasters.

Notes
1. The valuable research assistance of Partha P. Sahu, Intern, global development network (GDN), is acknowledged with much appreciation.
2. The exact figure is $1.08. The original figure of $1 was recalculate using the new PPP of 1993 (Chen and Ravallion, 2004).
3. Eastern Europe and Central Asia also saw the head count index rise, but it remains negligible; the paucity of survey data for this region in the 1980s should not be forgotten.
Thus the estimates are heavily based on interpolations, which do not allow for any changes in distribution (Chen and Ravallion, 2004).

4. The exact figure is $2.15. The original figure of $2 was recalculated using the new PPP of 1993 (Chen and Ravallion, 2004).

References


World Bank (various issues), World Development Report.
Introduction
Two questions are at the heart of economic analysis of gender equality issues in development:

1. Are the fruits of economic growth and development in developing countries fairly shared between women and men, girls and boys?
2. Does gender inequality promote or hinder economic growth and development in developing countries?

There is a vast literature addressing the first question, which will be considered in the next section; and a much smaller but growing literature addressing the second question, which will be considered subsequently. The challenges of designing macroeconomic policies to promote gender-equitable growth and development will be briefly considered in the final section.

Gender and the distribution of the fruits of growth and development
A gender analysis of the distribution of the ‘goods’ produced by development (not only income and wealth but also capabilities) goes beyond a focus on women as an isolated group and beyond mere disaggregation by sex. It situates distribution of these ‘goods’ in the context of the social construction of gender (that is, the social construction of norms of masculinity and femininity), which shape the choices made by people; and the consequences of those choices. It recognizes not only differences between the sexes, but also structural inequality between the sexes, embedded in institutions, including not only families but also markets, businesses and states (see for example, Sen, 1983; Folbre, 1986; Sen, 1996; Agarwal, 1997; Harriss-White, 1998; Elson, 1999).

Prior to the 1970s, there was little discussion of gender issues in economic development. If the topic was mentioned, there was a tendency to assume that women were an underutilized factor of production which could be mobilized for structural transformation and economic growth, and that economic growth and structural transformation would in turn be good for women. An example is the work of Arthur Lewis. In his famous model of ‘economic development with unlimited supplies of labour’, he identified ‘the wives and daughters of the household’ (in the subsistence
sector), as a source of labour for the modern ‘capitalist’ sector, arguing that this would lead to gains for women: ‘because most of the things which women otherwise do in the household can in fact be done much better or more cheaply outside, thanks to large scale economies of specialization, and also to the use of capital (grinding grain,fetching water from the river, making cloth, making clothes, cooking the midday meal, teaching children, nursing the sick etc.)’ (Lewis 1955, p. 404). In his book on the theory of economic growth, Lewis was in no doubt about the benefits to women: ‘Women benefit from growth even more than men . . . Woman gains freedom from drudgery, is emancipated from the seclusion of the household, and gains at last the chance to be a full human being, exercising her mind and her talents in the same way as men’ (Lewis, 1955, p. 422).

This optimistic view was challenged by Ester Boserup (1970), who argued that modernization of economies in Africa and Asia had marginalized women. In the agricultural sector, she argued, women had been deprived of access to training, land rights, education and technology, by both colonial and post-colonial administrators, who could not conceive of women being farmers in their own right, even though in much of sub-Saharan Africa and South-East Asia women enjoyed a significant autonomous role in traditional agricultural production. This lack of access to resources meant that while men’s productivity in farming increased, women’s productivity did not.

In the industrial sector, she argued, women accounted for a much lower percentage of the industrial labour force in large-scale modern factories than they did in home-based handicraft manufacturing. She pointed to obstacles on the demand side, including labour market regulations, and employers’ prejudiced perceptions of women’s capacities and work commitment; and on the supply side, she suggested that women had difficulties combining work in the modern sector with their family responsibilities, and were hindered by the view that work outside the home was not proper for women. Above all, women were hampered by their lack of appropriate skills, stemming from their lack of formal education. As a result of all these factors, women had been left marginalized and excluded from development. Boserup’s remedy for this was investment in more and better education and training for women – planners must change their view that women were primarily housewives, and train women to compete equally with men in the marketplace, so that women could be included in economic modernization.

As Naila Kabeer (1994) points out, Boserup’s book laid the foundations for a large body of ‘Women-in-Development’ literature, and a large number of policy initiatives aimed at ‘integrating women into development’. Irene Tinker (1990), in describing the making of the field of ‘Women-in-
Development’, calls Boserup’s book ‘the fundamental text for the UN
Decade for Women’ (1975–85) (Boserup, 1970). Boserup’s ‘marginaliza-
tion’ thesis found support from other authors, such as Saffiotti (1978), who
examined the implications for women’s employment of import-substitution
industrialization in Brazil and found that during the 1950s and 1960s, while
women’s industrial employment increased overall, their share relative to
men in the industrial labour force declined.

Addressing women’s marginalization through education has been a con-
stant theme in the ‘Women-in Development’ literature. There have been
significant increases in girls’ enrolment in primary school since the 1970s,
and by 2000 almost all girls (and boys) were enrolled in primary school in
developing countries in East Asia and the Pacific, Europe and Central Asia,
and Latin America and the Caribbean (UNESCO, 2004). However, in the
Middle East and North Africa, South Asia and sub-Saharan Africa,
significant numbers of children remained out of school, of which 54 per
cent were girls. In South Asia, the gender gap was particularly wide, with
Even in regions with parity in enrolment, girls drop out of school more
than boys, so that, with the exception of Latin America and the Caribbean,
boys’ completion rates remained higher than those of girls.

At the secondary level, by 2000, no region had succeeded in enrolling all
children, and girls’ enrolment lagged that of boys in all regions, with the
exception of Latin America and the Caribbean, where the reverse was true.
Moreover, girls’ secondary enrolment rates remained substantially lower
than their primary enrolment rates. In sub-Saharan Africa and India, girls’
secondary enrolment was particularly low, at 29.7 per cent and 47.1 per cent
respectively. The comparable figures for boys were 35.6 per cent and 53.7
per cent (UNESCO, 2004).

Boserup’s concern about the lack of visibility of women farmers and
their lack of resources, including training, has been another continuing
theme in the ‘Women-in-Development’ literature. The importance of agri-
culture as a source of livelihood has declined in most developing countries,
though it remains high in some countries and regions. The proportion of
the economically active population in agriculture in 2000 was, on average,
21 per cent in Latin America and the Caribbean, 59 per cent in Africa and
36 per cent in Asia. In the latter region there was wide variation, with only
10 per cent of the economically active population in agriculture in South
Korea, while the figure for China was 67 per cent and for India, 60 per cent
(UNRISD, 2005, p. 91).

There is still a lack of reliable data on women’s share of the economically
active population in agriculture, since women’s work is still undercounted
in censuses and labour force surveys, despite some improvements (Beneria,
1992, 2003). In particular, women's 'own-account' or 'self-employed' farming is undercounted, and women are more likely to be enumerated as 'unpaid family workers' contributing to farms managed by their husbands, and 'agricultural wage workers', working for larger-scale commercial farms, than as farmers in their own right.

In much of sub-Saharan Africa, women and men in the same household farm and manage separate plots, while at the same time supplying labour inputs to each other's plots, so that many women are both 'own-account farmers' and 'unpaid family workers'. Moreover, migration, war and HIV/AIDS have reduced rural male populations in sub-Saharan Africa, and about one-third of all rural households are headed by women, leading to the 'feminization of agriculture', according to the UN Food and Agriculture Organization (FAO, 2005). In many Latin American countries, there has been a feminization of agriculture, as men have migrated from rural areas in search of better incomes, and women have taken over the management of family farms, and do the bulk of farm labour (UNRISD, 2005, p. 96). In Asia, similar factors have led to more women in China, India and South-East Asia taking on the management of farming activities, though there is not full agreement on whether farm management could be said to be 'feminized' (UNRISD, 2005, p. 97). In India, for instance, 48 per cent of self-employed farmers are women; and in dairying and animal husbandry, women farmers far outnumber men (Ministry of Agriculture, Government of India, 2005).

Farming women throughout the developing world have less rights over land than do men, as a result of a variety of factors, such as unequal inheritance practices, registration of land titles only in the name of male household heads, and land reforms that are biased against women (Grown et al., 2005, p. 75). Comprehensive data on the size of the gender gap in land ownership are not available. A study covering five Latin American countries found that women account for between 11 and 27 per cent of landowners (Deere and Leon, 2003). Women farmers in sub-Saharan Africa lack secure rights to the land they farm, and frequently lose their land when they are widowed or divorced (World Bank, 2001, pp. 121–2). In South Asia, few women own agricultural land; and of those that do, few exercise full control over it (Agarwal, 1994).

Since the early 1990s there have been some improvements in women's formal land rights in Latin America and sub-Saharan Africa. Joint titling of land to couples has been introduced in Brazil, Colombia, Costa Rica, Honduras and Nicaragua (Deere and Leon, 2003). In Uganda, Kenya, Tanzania, South Africa and Rwanda women's formal statutory land rights have been strengthened – but customary law is still a powerful obstacle, generally preventing women from owning or inheriting land in their own name.
(Grown et al., 2005, p. 81). Moreover, in other parts of sub-Saharan Africa, women's customary use rights over land have been weakened through the introduction of individual ownership of land (Lastarria-Cornhiel, 1997). Women's land rights were strengthened by new legislation in India in 2005 but many challenges of implementation remain (Agarwal, 2005).

Formal ownership rights do not necessarily ensure gender equality in livelihoods. For instance, land markets themselves are not gender-neutral: women buyers have lower bargaining power than men, and face discrimination (Deere and Leon, 2003). Moreover, women farmers still receive less technical support from agricultural extension workers, and only a tiny percentage of extension workers are women (World Bank, 2001, p. 52). In the case of women smallholders, many aspects of the disadvantages depicted by Boserup still persist. But efforts to promote gender equality are unlikely to yield substantial gains in terms of improved standards of living for women smallholders without more effective strategies for the agricultural sector as a whole (UNRISD, 2005, p. 104).

Some women have been moving into new areas of ‘modern’ commercial agricultural production. In the production of flowers, fruit and vegetables for export from Latin America and sub-Saharan Africa, women comprise between 45 and 90 per cent of contract workers (UNRISD, 2005, p. 98). However, women are mainly employed in temporary or seasonal low-paid jobs, with long hours of work, poor health and safety conditions and no social protection (UNRISD, 2005, p. 98). This illustrates a key shortcoming of the Women-in-Development (WID) ‘marginalization’ thesis: it identifies gender issues in development in terms of women's exclusion from development, whereas the problem is increasingly the terms and conditions of women's integration in development.

This type of criticism of the WID approach was initiated in the early 1980s, especially by women scholars from developing countries who argued that women were subordinated to development rather than marginalized from it (see, for example, Beneria and Sen, 1981). The disadvantage experienced by women in the process of development was not, they argued, primarily the result of the persistence of ‘traditional’ cultural practices and prejudices, but of the way in which development has created new job structures, in which gender inequality is embedded. For instance, women are concentrated at the lowest rungs, with lower pay, and less job security and social protection than men; and are expected to combine paid work with ongoing responsibility for the unpaid work of caring for families and communities, reproducing both the labour force and the social fabric. This approach argued for changing development strategies to make structural transformation more egalitarian (Sen and Grown, 1987).
This critique was reinforced by the emerging evidence that, unlike import-substitution industrialization, export-oriented industrialization did not marginalize women, but actually increased their share of manufacturing employment. Women’s share of employment in the growing service sector also increased. By 2003, women’s share of wage employment in non-agricultural sectors in Latin America and the Caribbean had reached 43.5 per cent, approaching that of the developed regions (46.4 per cent). In Eastern Asia, women’s share was 40 per cent, in South Eastern Asia 38.6 per cent and in sub-Saharan Africa 35.8 per cent. Lower shares prevailed in Southern Asia (18 per cent), Northern Africa (21.5 per cent) and Western Asia (aka the Middle East) (20.2 per cent) (data from UN Statistics Division, Millennium Indicators Database).

The rising trend in women’s share of employment in industrial and service sectors has been accompanied by a lively debate on the extent to which such employment has improved women’s lives and reduced gender gaps in well-being (see, for instance, for an early contribution, Elson and Pearson, 1981; and for a recent contribution, Kabeer, 2000; for a survey of the arguments, see Razavi, 1999). Gender inequalities persist in pay and conditions of employment, with most women segregated in a few occupations in which the vast majority of those employed are women (Anker, 1998). At the end of the twentieth century, on average, the hourly wages of women in developing countries were 73 per cent of those of men (compared to 77 per cent in developed countries). In both developed and developing regions, more than 80 per cent of the gender wage gap could not be explained by measurable differences in workers’ characteristics, and probably indicates discrimination in the labour market (World Bank, 2001, pp. 55–6). Women workers in developing countries are more concentrated than men in informal employment that lacks social protection; and within informal employment, in the more precarious types, with lower incomes (Chen et al., 2005). Informal employment has been growing as a share of total employment, as labour markets have become both more flexibilized and more feminized (Standing, 1999).

The conventional wisdom is that, despite problems of the quality of women’s paid employment, their earnings give them greater bargaining power within their households, and more influence over how household resources are allocated (see for example, Kakwani and Son, 2006). However, this is not universally the case (Elson, 1999). Women in Uganda report that when a woman starts to earn an income of her own, her husband is liable to withdraw his financial support, and shift to her the responsibility for paying school fees, medical bills, and buying food and clothing and other necessities for the whole household (Ellis et al., 2006, p. 24). Moreover, in most parts of the world, the division of unpaid domestic work between men
and women in the household is rarely renegotiated when women start earning (Kabeer, 2005). Case studies reveal that it matters where women obtain their earnings: for instance, in Ahmedabad, India, home-based paid work does not give women as much say in household decisions as employment outside the home (Kantor, 2003).

Women’s entry into the labour market was accelerated in the 1980s and early 1990s by stabilization and structural adjustment policies (Çagatay and Ozler, 1995). It has been suggested that this acceleration was widely linked to ‘distress sales’, as women are forced to try to make good shortfalls in household income following the loss of employment by male household members (for example, Moser, 1989; Gonzalez de la Rocha, 2000).

Elson (1991) argued that stabilization and structural adjustment polices implicitly assumed unlimited supplies of female labour, available to make good through unpaid work in families and communities any shortfalls in provision of public sector non-tradeable services (such as health, education, water and sanitation), and to increase production of exports, while at the same time maintaining household food security and the social fabric of family and community networks. Moreover the theory ignored the gender norms that structure the division of labour, and mean that men’s labour tends not to be reallocated to ‘women’s work’, where there is a decrease in what is considered to be ‘men’s work’ (for example construction) and an increase in what is considered to be ‘women’s work’ (for example garment-making, unpaid care work). Instead, a more likely outcome is unemployment and underemployment for men (who do less paid work but little or no more unpaid work), and overwork for women (who do additional paid work as well as unpaid work). Failure to take this into account in designing adjustment policies, argued Elson, results in extra burdens for women, and risks deterioration in health, nutrition and education.

A number of case studies provide empirical backing for Elson’s argument (for example Moser, 1989; Lim, 2000; Tanski, 1994), but lack of data and methodological problems hinder definitive conclusions about whether women and girls have, on average, borne a greater share of the costs of adjustment (Haddad et al., 1995). Nevertheless, there is widespread empirical support for the conclusion that women are disadvantaged in the adjustment process, unless specific measures are taken to address pre-existing gender inequalities (Haddad et al., 1995).

One of the important gaps in data is comprehensive information on the extent of men’s and women’s unpaid work. Time-use surveys have been used for more than two decades in developed countries to remedy this. In the late 1990s such surveys began to be introduced in a growing number of developing countries. They confirmed what small-scale case studies had suggested: women and girls spend more time on unpaid work than men and
boys; and when both paid and unpaid work is taken into account, women and girls have a longer total working day than men and boys. For example, in South Korea, in 1999, the total working time for males was almost six hours a day on average, while for females it was almost seven hours a day. Males spent on average only 50 minutes per day on unpaid work, while females spent almost four hours (Tae-Hong, 2001, p. 8). A similar picture was revealed by a time-use survey for six states of India in 1998–99: on average female total working time was just over 7.5 hours per day, while for males it was 6.5 hours per day. Females spent almost five hours a day in unpaid work, and males spent only about 30 minutes (calculated from Chakraborty, 2005, Table 3). Charmes (2006, Table 3.2) shows a similar picture for sub-Saharan African countries. For example, in Benin in 1998 the total working time for females was on average almost 7.5 hours a day, while for males it was about five hours a day; females spent almost 3.5 hours per day on unpaid work, while males spent just over one hour. In Madagascar in 2001, the total working time for females was on average almost 6.5 hours per day, while for males it was almost 5.5 hours per day; females spent just over 3.5 hours on unpaid work, while males spent almost 50 minutes. In Mexico, in 1995, the total working time for females was on average just over 8.5 hours a day, while for males it was almost 7.5 hours a day. Females spent four hours in unpaid work, while males spent just over 1.5 hours (calculated from Elson, 2000, p. 102). Very few developing countries have conducted time-use surveys at regular intervals, so it is not possible to examine trends. Cross-country analysis that includes both developed and developing countries shows that higher gross domestic product (GDP) per capita is associated with a decline in time spent on unpaid work, and smaller gender gaps in total hours worked and time spent in unpaid work (World Bank, 2001, p. 185). A key factor in this is investment in infrastructure and public services. Nevertheless, development does not eliminate unpaid work, as people value time to care for their family and friends (Folbre, 2001).

By the beginning of the twenty-first century, the context for discussion of gender issues in development had become the acceleration of globalization, the growth of income inequality between countries and between households within countries, and the growing differences in the trajectories of developing countries, with some, such as China and India, experiencing rapid structural transformation and growth, others, such as many countries in Latin America and the Caribbean, growing only slowly, and yet others, especially in sub-Saharan Africa, suffering deteriorating conditions.

Millions of women in China and India were not, however, benefiting from the rapid increase in national income because they were ‘missing’, in the sense of either having died prematurely, or not having been born at all.
This phenomenon was first brought to public attention by Amartya Sen (Sen, 1990a). More recent data (China 2000 and India 2001, see Klasen and Wink, 2003) confirms its persistence, and also its prevalence in middle-income developing countries such as Taiwan and Korea (Klasen and Wink, 2003). Preference for sons has not been reduced by rapid development. Indeed, in India the sex ratio imbalance is higher in the higher-income states than in the lower-income states.

Despite rapid growth in some developing countries, millions of people throughout the developing world continue to live in poverty. There has been considerable debate on whether poverty is ‘feminized’ in the sense of females being disproportionately represented among the poor (Çağatay, 1998; World Bank, 2001, p. 63). Measurement of poverty is dominated by a focus on consumption poverty, calculated on a household basis. This has led to a preoccupation with comparing the poverty rates of male- and female-headed households. The evidence is mixed: in some countries female-headed households are disproportionately represented in households below the poverty line; in others they are not (World Bank, 2001, p. 64). An alternative comparison is between proportions of the male and female populations that are in households below the poverty line. Again the evidence is mixed. In some countries a higher proportion of the female population than of the male population lives in households below the poverty line, but not in others (see for instance, ECLAC, 2002, Tables 6a and 6b). However, there is general support for the view that women are more vulnerable to poverty in old age than are men, as many more women are widows than men are widowers (World Bank, 2001, p. 67).

Moreover, women are certainly over-represented among the adult population who have no income of their own, because their participation in paid work remains lower than that of men. This limits their bargaining power within households and their capability to live a life of dignity, even if the household in which they live has an average income above the poverty line (Sen, 1990b). There is also plenty of evidence that poverty is differently experienced by males and females, and that gender gaps in education, health and work burdens tend to be larger in households in poverty (World Bank, 2001, pp. 61, 66).

Some studies have found that economic growth narrows gender gaps. For instance, Dollar and Gatti (1999) found that a higher level of per capita GDP was associated with greater gender equality in secondary schooling, in life expectancy and in representation in parliaments. However, when the labour market is brought into the picture the results are different for different groups of countries. Seguino used a composite gender equality indicator that includes relative labour force participation rates and female share of technical, professional and administrative positions, as well as the
indicators used by Dollar and Gatti, and clustered countries into four groups, ranging from poorest to richest in per capita terms. She found that in the highest and third-highest groups, there was a positive relation between growth and gender equality, but there was a negative relation in the lowest and second-highest income groups. Her conclusion was that economic growth is not sufficient by itself to achieve gender equality (UNRISD, 2005, p. 58).

These conclusions are reinforced by two regional studies of the impact of growth on gender equality. In Asia, in the period 1970–90, gender equality (assessed using a composite quality-of-life indicator that includes allowance for ‘missing women’) was highest in those countries that grew slowest (Seguino, 2002). In Latin America and the Caribbean, for the period 1970–2000, economic growth was found to have little beneficial effect on closing gender gaps in well-being, though the share of the manufacturing sector in GDP and the share of government expenditure in GDP were positively related to the reduction of gender gaps (Seguino, 2007).

Attention is now beginning to be focused on inequalities between women in developing countries. If inter-household inequality is rising, then it seems likely that inequality between women is also rising. There is as yet no comprehensive study on this topic. A recent study of maternal mortality and poverty in ten developing countries, based on demographic and health surveys, found that the proportion of women dying from maternal causes increases with the poverty of the households in which they live. For instance, in Indonesia, the probability of maternal death was three to four times greater in the poorest than in the richest quintile. In the Philippines and Tanzania the probability was two to three times greater (Graham, 2004). More research needs to be done on inequalities between women.

Gender and the determinants of economic growth and development
In this section, we examine whether gender inequality hampers or promotes economic growth and development. The 1990s saw the emergence of a growing literature on this topic, though it is still very small compared to the literature on the distribution of the fruits of growth and development. A number of cross-country econometric studies relating gender inequality to economic growth in developing countries is now available. They have produced a variety of results, depending on the dimensions of inequality considered, the selection of countries and the specification of the equations (World Bank, 2001).

Barro and Lee (1994) found a negative relationship between female secondary schooling and economic growth, though the relation between male secondary schooling and growth was positive. However, subsequent studies have found a positive relation between economic growth and gender
equality in education (for example Hill and King, 1995; Dollar and Gatti, 1999; Esteve-Volart, 2000; Klasen, 1999, 2002). Dollar and Gatti (1999) used data from over 100 countries covering three decades, and found that an increase in one percentage point in the proportion of adult women who have secondary education is linked to an increase in growth rates of per capita income of 0.3 percentage points per year. Klasen (2002) finds that the higher gender gaps in education in sub-Saharan Africa, compared to East Asia, and their slower reduction, accounted for 0.6 percentage points in the 3.5 percentage points difference in the growth rates in the two regions in the period 1960–92. Closing the gender gap in education enrolment by 2005 is a Millennium Development target, adopted by the UN General Assembly at the Millennium Summit in 2000. An estimate of the impact on the economic growth of countries that were not on track to meet this target found that they would have grown faster by about 0.1 to 0.3 percentage points if they had been on track to close the gap (Abu-Ghaida and Klasen, 2004).

Increasing the level of education of girls may increase growth in per capita incomes directly, by increasing the participation of women in the labour market and the productivity of women’s labour, and indirectly, by facilitating a transition from a high to a low rate of fertility (since educated women have fewer children). During the fertility transition, the working-age population grows at a higher rate than the dependent-age population. Bloom and Williamson (1998) refer to this as a ‘demographic gift’ and argue that it raises the rate of growth of per capita income during the transition (provided policies are in place to employ the working-age population productively). They estimate that it accounts for between 1.4 and 1.8 percentage points of growth in per capita income in East Asia, in the period 1965–90. Sub-Saharan Africa has yet to experience the transition. Its fertility remains high and its working-age population has not grown faster than its total population. A recent study of Uganda estimates that a period of sustained fertility decline could boost medium-term per capita growth rates by 0.5 to 0.6 percentage points a year (Klasen, 2005).

Nevertheless, questions remain about the direction of causation between the education of girls and economic growth. For instance, Robbins (1999) argued, in a study of six Latin American countries, that causation goes from increases in growth to increases in education of girls, rather than vice versa. He found that economic growth leads to rising educational attainment by drawing more women into the labour force, increasing the opportunity cost of women’s time, and thus reducing fertility and leading families to invest more in the education of their (fewer) children, girls as well as boys.

There is a general agreement that increasing the level of girls’ education improves outcomes for their children. Studies based on household data
show that the more educated are mothers, the lower their children’s mortality, controlling for household income and other aspects of socio-economic status; and child immunization rates rise with mother’s education (World Bank, 2001, pp. 79, 80). Cross-country regression analysis indicates that higher levels of female enrolment in school, and lower gender education gaps, are associated with lower rates of infant mortality (Hill and King, 1995). Increases in women’s education accounted for 43 per cent of declines in child malnutrition in 1970–95 (Smith and Haddad, 2000). A recent estimate of the costs of failing to achieve gender parity in educational enrolment by 2005 found that by 2015, such countries would have on average 15 per 1000 higher rates of under-five mortality and 2.5 percentage points higher prevalence of underweight children under five (Abu-Ghaida and Klasen, 2004).

Focusing on labour markets, rather than education, produces mixed results: greater gender equality in participation in the labour market seems to promote faster growth, but greater equality in wages does not. Recent studies on the Middle East and North Africa (Klasen and Lamanna, 2003) and India (Esteve-Volart, 2004) suggest that growth would be higher if the gender gap in labour market participation were reduced (through more women entering the market). However, cross-country regression analysis of growth and the gender wage gap in 20 semi-industrialized economies (such as those in East Asia) in the period 1975–95 found that, controlling for gender differences in educational attainment, gender wage inequality was a stimulus to growth. A 0.1 percentage point increase in the gap between female and male returns per year of secondary education is associated with a 0.1 percentage point increase in the growth of per capita GDP (Seguino, 2000). There is a strong inverse relation between the gender wage gap and the educational attainment gap in semi-industrialized countries (Seguino, 2005, Table 2). This implies that although narrowing gender gaps in education (through increasing the education of girls) tends to raise productivity, women’s lack of bargaining power in the labour market holds their wages down. Seguino argues that this stimulates growth through higher profits in female labour-intensive manufactured exports, higher investment and higher foreign exchange earnings. Of course, if high gender wage gaps held back household investment in girls’ education, the dynamic would be different, but this does not seem to have happened in semi-industrializing countries, especially in East Asia (Seguino, 2005, p. 22). Instead the combination of educated but cheap female labour has been critical to the expansion of exports of manufactures. Seguino has extended her analysis to compare the impact of different kinds of inequality on growth in 37 semi-industrialized countries in the period 1975–99. She finds that whereas income inequality between households is negatively associated with
growth, the gender wage gap is positively associated with growth (Seguino, 2005, p. 23).

However, in agrarian economies, in which self-employment and family labour are more important than wage labour, it seems likely that gender inequality in production can hamper economic growth. Lack of aggregate data means that cross-country regression analysis has not been used to investigate this relationship, but micro-level studies suggest that gender inequality in access to and control of assets and income is likely to hamper growth. For instance several studies of smallholder agriculture found that agricultural productivity could be increased if female farmers had the same levels of inputs (such as fertilizer, land and labour) and education and training as male farmers (World Bank, 2001, pp. 85–6). The following examples are highlighted by Blackden and Bhanu (1999). In Burkino Faso, output could be increased by 10–20 per cent by shifting resources from men’s plots to women’s plots within the same household. In Kenya, increasing the education and input levels of female farmers to those of male farmers could increase yields obtained by women farmers by as much as 22 per cent. In Tanzania, reducing the amount of unpaid work that women have to do could increase household cash incomes of smallholder coffee and banana growers by 10 per cent, labour productivity by 15 per cent and capital productivity by 44 per cent. In Zambia, if women farmers enjoyed the same level of investment in agricultural inputs as men farmers, agricultural output could increase by up to 15 per cent.

Gender inequality in control of resources and division of responsibilities within households can hamper the expansion of agricultural exports, including non-traditional agricultural exports, by smallholders. In sub-Saharan Africa, the production of export crops often requires inputs of women’s labour on plots controlled by their husbands, while the cash income from this production is controlled by their husbands. The Poverty and Social Impact Assessment of Uganda’s Strategic Exports Initiative (Booth et al., 2003) found that such intra-household inequalities limit the export supply response in Uganda because women prefer to put more of their labour into producing crops on their own plots for household use and sale in local markets, rather than into producing export crops controlled by their husbands. This is corroborated by other studies on Uganda (for example Muhereza, 2001), and there is evidence of similar effects in other countries, such as Zambia (Wold, 1997) and Burkino Faso (Smith and Chavas, 1999).

It is important not to overstate the role of household inequalities in constraining export expansion and productivity in smallholder economies (Whitehead, 2005). There are many extra-household gender inequalities that are also important, including unequal access to extension services,
markets, transport and credit, and unequal laws and social norms. Such inequalities also constrain the contributions to growth made by self-employed women in non-agricultural activities. This is important for future development, since the share of agriculture in employment and income generation has been declining and is expected to decline further. Self-employment is particularly important in sub-Saharan Africa. In this region (excluding South Africa), informal employment comprises 78 per cent of non-agricultural employment and self-employment represents 70 per cent of informal employment (ILO, 2002).

Many hopes have been invested in new forms of microfinance as a means of enabling self-employed women to both increase their own incomes and contribute to economic growth. During the 1990s there was a large expansion in poor women’s access to small loans from microfinance institutions, many of them based on the social collateral of group liability, rather than the economic collateral of individual assets, such as land. Optimism about the impact of microfinance reached a high in the 1997 Microcredit Summit, which issued a Declaration stating that:

> empirical evidence has shown that women, as a group, are consistently better in promptness and reliability of repayment. Targeting women as clients of microcredit programs has also been a very effective method of ensuring that the benefits of increased income accrue to the general welfare of the family, and particularly the children. At the same time, women themselves benefit from the higher status they achieve when they are able to provide new income (quoted in Mayoux, 2000, p. 3)

However, detailed evaluations of a range of programmes in Africa and Asia have shown that there are a number of limitations. Mayoux (2000, pp. 12–13) summarizes them as follows: women may not have control over the loans they get, but act as de facto intermediaries between male family members and microfinance institutions; even if women control the loans, they may not generate significant increases in income, because women are crowded into a narrow range of traditionally female low-return activities; even if there is an increase in women’s income, men may control its use, or men may reduce their own contributions to meeting household expenses. Kabeer (2005) finds that microfinance institutions that combine financial services with other forms of support, and which build the organizational capacity of poor women, are more likely to be empowering. Poor women need more than just loans and savings accounts.

To make a significant impact on their own incomes and on the national economy, women need to be able to move beyond informal micro-enterprises and develop small businesses that are registered and eligible for loans from the formal financial sector, and for support from government...
agencies. There remain many legal and social barriers which prevent women, more than men, from formalizing and growing their businesses, as shown in detail in the case of Uganda by Ellis et al. (2006).

Facilitating women’s increased participation in the market economy, either in self-employment or in wage employment, will only promote economic growth if there is sufficient demand for their products and their labour. This raises the question of appropriate macroeconomic policies, to which we turn briefly in the next section.

**Macroeconomic policy challenges for gender-equitable development**

At the sectoral level, there is a great deal of knowledge about economic policies that work to improve the position of poor women, to reduce gender inequality and to promote growth and development (Grown et al., 2005). Less attention has been paid until recently to appropriate macroeconomic policies. The importance of examining macroeconomic policies from a gender perspective was emphasized in contributions to two special issues of *World Development* (Çağatay et al., 1995; Grown et al., 2000). However, the World Bank paid little attention to macroeconomics in its 2001 policy research report, *Engendering Development*.

In this final section we briefly consider some aspects of monetary and fiscal policy and trade and investment policy that need to be addressed if the relations between growth, development and gender equality are to become mutually reinforcing.

Removing deflationary bias in monetary and fiscal policy is one important challenge. Employment has ceased to be a goal of macroeconomic policy. Instead, the focus is solely on financial variables, such as inflation, the fiscal deficit and debt-to-GDP ratios. Rates of inflation have been brought down to much lower levels than in the 1980s, but in many regions, this has been at a huge sacrifice in public investment, economic growth and decent jobs (for evidence, see, for example, UNRISD, 2005, p. 30). Deflationary bias in macroeconomic policy was identified as an important issue for women in the UN *World Survey on the Role of Women in Development* (UN, 1999), which argued that women disproportionately bear the costs of this bias. Much of the research and policy development on gender equality in employment has been focused on measures to enable women to compete with men on an equal basis. These measures are important, but they are not sufficient. To the extent that they are successful, they will simply redistribute some jobs from men to women. This will reduce gender gaps, but not in a way that provides ‘full and productive employment and decent work for all’, as called for by the UN Secretary-General (UN, 2006, p. 6). In order for gender equality to be realized in ways that ‘equalize up’, rather than ‘equalize down’, there needs to be an expansion
of the total number of decent jobs, as well as an improvement of women’s access to them.

Women are particularly likely to be disadvantaged by deflationary bias because it interacts with, and reinforces, other policy biases, such as male breadwinner bias, the assumption that men are more deserving of decent jobs because they are assumed to be the principal economic support of families, while women’s incomes are wrongly perceived to be merely supplementary, and not essential to family well-being (Elson and Çağatay, 2000, pp. 1354–56). Seguino (2003) finds this to be important in explaining why women are much more likely to be unemployed than men in Barbados, Jamaica and Trinidad and Tobago, using data from the period 1980–99. This inequality cannot be explained in terms of women being less educated than men, since women have a higher unemployment rate than men with the same education; nor in terms of the different sectoral distributions of men’s and women’s employment. While both male and female unemployment rates fell in economic upturns, male rates fell more than female rates. Male workers were the first to be hired in economic upturns, even in the female-intensive service sector.

Global unemployment rose from 5.6 per cent in 1993 to 6.2 per cent in 2003, and the female unemployment rate was slightly higher than the male rate; among young people, the gender gap was bigger (ILO, 2004). There were some regional differences: in the Middle East and North Africa, and Latin America and the Caribbean, the female unemployment rate was higher than the male, while in sub-Saharan Africa and East Asia the reverse was true (ILO, 2004, p. 2). However, in many countries, female unemployment rates are likely to underestimate the true extent of women’s unemployment because women are more likely to be ‘discouraged workers’ who respond to their failure to find jobs by ceasing to search for one actively, although they would like to have a job if one were available. The discouragement and underemployment of women appear to have been significant in the aftermath of the Asian financial crisis in 1997–98. For instance, in South Korea, the rate of job loss for women was higher than for men, but subsequently, male unemployment rates appeared to be higher than female rates, while a higher proportion of women than before were employed in various types of informal employment (UNRISD, 2005, p. 42).

In a pioneering study, Braunstein and Heintz (2006) investigated the link between monetary policies and gender equality in employment in 17 low- and middle-income countries in the period 1970–2003. They identified episodes when monetary policy led to contractionary inflation reduction, when the growth of employment fell below its long-run trend; and episodes when it led to expansionary inflation reduction episodes, when the growth of employment was faster than its long-run trend. (The contractionary
episodes were associated with real interest rates being maintained above their long-run trend – an indicator of deflationary bias.) The study found that in 67 per cent of the contractionary inflation reduction episodes, the female-to-male employment ratio fell below its long-run trend, indicating that women were disproportionately affected by the slowdown in employment. However, in expansionary inflation reduction episodes, there was no clear disproportionate effect on either women or men. The female-to-male employment ratio increased faster than the trend in 53 per cent of cases, and at or below the trend in 47 per cent of cases.

Braunstein and Heintz concluded that a policy of responding to inflationary pressures by raising positive real interest rates above their long-run trend, and reducing real money supply below its long-run trend, tended to be associated with a greater loss in female than in male employment (relative to long-run trends in both). They noted that in 33 per cent of contractionary inflation reduction episodes, women’s employment was not disproportionately affected by deflationary policies. They found that in these episodes, the real exchange rate either depreciated or remained at its long-run trend. They concluded that ‘maintaining a competitive exchange rate may offset some of the gender bias observed during contractionary inflation-reduction’ (Braunstein and Heintz, 2006, p. 12).

Inadequate levels of employment and decent work are also related to an emphasis on managing government budgets by cutting expenditure rather than raising tax revenues. Globalization has made it harder for governments to raise tax revenue. Çağatay (2003) summarizes the key aspects of this fiscal squeeze. Trade liberalization cuts import duties and export taxes, key sources of revenue in many poor countries. Competition to attract multinational corporations and their highly paid executives leads to cuts in corporation and capital gains taxes, and tax holidays and other exemptions and to cuts in top rates of income tax. Development cooperation grants have fallen as trade is supposed to replace aid. Governments have been encouraged or pressured into turning to sales taxes like value-added tax (VAT) to raise revenue, but such taxes fall most heavily on poor people and worsen the distribution of income. Revenue has also been raised through the sale of public enterprises and other public assets, but this only gives a one-time boost to revenue, and may result in costs for services, like water, that poor people cannot afford.

With revenue limited and debt burdens rising, the pressure has been on governments to make their budgets sustainable by cutting back on expenditure. This pressure has come from the public international financial institutions like the International Monetary Fund (IMF) and the World Bank, and also from private investors, who have seen budget deficits as harbingers of inflation, signals that the value of their assets would be eroded. In order
to build a reputation for ‘sound finance’ in financial markets, many governments have enacted legislation (such as balanced budget laws) that severely limits the fiscal space (Bakker, 2002).

More room for a variety of fiscal policies would not by itself ensure that fiscal policy is used to promote gender equality. Recognizing this, since the late 1990s, a series of gender budget initiatives (GBIs), in all parts of the world, have sought to improve the distribution, adequacy and impact of government budgets at national, regional and local levels; and to secure greater transparency in the use of public money; and to secure greater accountability to women as citizens. A number of tools have been developed for analyzing the gender dimensions of government budgets (Budlender and Sharp, 1998; Elson, 1998). By 2002, up to 50 countries in all parts of the world had hosted some kind of gender budget initiative (Budlender and Hewitt, 2002, p. 8). There is no one template: GBIs have taken place at all levels of government, involving regional and local government budgets as well as national budgets. Moreover, a multiplicity of actors have been involved: government ministers and officials (especially women’s ministries, sometimes Ministries of Finance), parliamentarians, women’s organizations and academic researchers (Budlender et al., 2002; Budlender and Hewitt, 2002). An important area of concern has been whether budgetary policies are reducing or increasing the amount of unpaid domestic work that poor women have to do; and whether they make it easier or more difficult for people (both men and women) to combine paid work and caring for their families without enduring excessive hours of work.

The most effective GBIs have produced some institutionalization of gender equality concerns in one or more stages of the budget cycle in one or more ministries, or have resulted in an ongoing public scrutiny of the budget from a gender equality perspective. Their aims have included: raising awareness and understanding of gender impacts of budgets and the polices they fund; making governments accountable for their budgetary and policy commitments; and changing and refining government budgets and policies to promote gender equality. Many examples of success in achieving the first two goals have been identified, though fewer in achieving the third (Sharp, 2002).

As well as the challenge of securing more gender-equitable public finance, there is also the challenge of securing polices on industry, trade and finance that enable increases in women’s wages without jeopardizing economic growth. This challenge is analysed by Seguino and Grown (2006), who identify the need for an approach they label ‘industrial policy under conditions of strategic openness’. This entails an industrial policy that promotes a shift of female employment to high-quality exports with a low price elasticity (so that higher wages for women will be less likely to have a
negative effect on exports). This should be complemented by policies that slow down the mobility of foreign direct investment, providing incentives for firms to respond to higher wages by investment in technological improvements, rather than by relocation. In addition, the maintenance of exchange rates should be maintained at competitive levels. Such a strategy would require a greater latitude for special and differential treatment of developing countries in international trade agreements.

It is also important to consider the impact of trade policies on unpaid domestic work. Pioneering research has been done on this topic by Fontana (2003), who used a computable general equilibrium model to simulate the impacts of trade liberalization on the paid and unpaid work of women and men in Bangladesh and Zambia. This technique is able to show the quantitative implications of the expansion of paid employment of women for the time they spend in unpaid domestic work and leisure, and the conditions under which there may be some redistribution of unpaid work between women and men.

The policy challenges identified above were considered by the United Nations Research Institute for Social Development in its comprehensive report on gender equality (UNRISD, 2005). The report concluded that to achieve greater gender equality, a new package of macroeconomic policies is required that puts more emphasis on redistributive taxation, gender-responsive public spending, the creation of decent work, universal social protection schemes, and policies to enable people more easily to reconcile their paid and unpaid work responsibilities, all in the context of a more just and equitable system of international economic relations (UNRISD, 2005). The challenge for the future is to ensure that not only micro and sectoral polices take account of gender issues, but also macroeconomic policies. This will require a more systematic incorporation of the unpaid work of the household sector into macroeconomic analysis, alongside the paid work of the public and private sectors.

References


51 Children and development
Paul Glewwe and Amy Damon

Introduction
In developing countries, 30 percent of the population is less than 15 years old. Thus any discussion of the well-being of the population in any developing country must examine the welfare of children. In general, the relationship between children and economic and social development is primarily one of the impact of development on children, rather than the impact of children on development. Yet today’s children will become adults in one or two decades, and events during their childhood will have a strong impact on their lives as adults, including their contribution to economic and social development.

This chapter summarizes recent research by economists on the status of children in developing countries. It begins by examining the status of children in those countries, and proceeds by assessing the impact of economic growth on children. It then presents policies that are likely to improve children’s health, education and employment status. The final section presents some concluding remarks.

Child welfare in developing countries
By virtually any measure, the welfare of children in developing countries has improved markedly over the past half-century, not only for developing countries as a whole but for all regions as well. This section reviews this progress and presents the current status of children with respect to their health, educational status and employment status.

Health
In almost every dimension the health status of children is improving in nearly all developing countries. Progress in reducing infant and child mortality rates is shown in Tables 51.1 and 51.2. These rates show how many children died, out of every 1000 born, before their first and fifth birthdays, respectively. For developing countries as a whole, the infant mortality rate has dropped dramatically, from 126 in 1960 to 53 in 2000, and the child mortality rate has declined sharply, from 197 to 78, over the same time period. This decline has occurred in all regions of the developing world.

Despite this progress, infant and child deaths are still common in developing countries. The situation in sub-Saharan African countries is especially
worrisome. In particular, progress in reducing infant and child mortality was much slower between 1980 and 2000 than it was from 1960 to 1980. While rates were similar in sub-Saharan Africa and South Asia in 1960 and 1980, and both made good progress in reducing rates in those two decades, South Asia was able to reduce infant and child deaths much more quickly between 1980 and 2000. The most obvious explanation is the advent of AIDS in sub-Saharan Africa, but slow economic growth is also likely to have played a role.

### Table 51.1 Infant mortality rate (per 1000 live births)

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>116</td>
<td>65</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>14</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>65</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>19</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>95</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>25</td>
<td>26</td>
<td>32</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>151</td>
<td>83</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>13</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>166</td>
<td>119</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>162</td>
<td>118</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>38</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>126</td>
<td>82</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>117</td>
<td>135</td>
<td>147</td>
</tr>
</tbody>
</table>


### Education

Schooling is another area in which child welfare has increased in the last half-century, as seen in Tables 51.3 and 51.4. These tables present gross enrollment rates, which are defined as the number of children enrolled as primary (or secondary) students divided by the number of children in the age range associated with that level of schooling. Note that it is possible for these enrollment rates to exceed 100 because it is common for ‘over-age’ children to be in a particular level because of delayed initial enrollment or grade repetition.

In developing countries as a whole, the primary school gross enrollment rate increased from 68 percent in 1960 to 99 percent in 2000. These
### Table 51.2  Child mortality rate (per 1000 live births)

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>175</td>
<td>94</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>14</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>88</td>
<td>57</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>19</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>139</td>
<td>77</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>25</td>
<td>26</td>
<td>32</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>237</td>
<td>117</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>13</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>265</td>
<td>182</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>268</td>
<td>191</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>38</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>197</td>
<td>124</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>117</td>
<td>135</td>
<td>147</td>
</tr>
</tbody>
</table>


### Table 51.3  Primary school gross enrollment rate

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>77</td>
<td>102</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>8</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>101</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>21</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>89</td>
<td>101</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>23</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>60</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>33</td>
<td>79</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>42</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>37</td>
<td>38</td>
<td>36</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>68</td>
<td>89</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>103</td>
<td>110</td>
<td>107</td>
</tr>
</tbody>
</table>
increases occurred in all regions of the developing world (except in Europe and Central Asia, where the enrollment rate was already 101 percent in 1960). Increases in the secondary school gross enrollment rate are even more striking, increasing from 19 percent in 1960 to 61 percent in 2000. These rates also increased in all regions.

Despite these gains, there are some troubling patterns. First, primary school enrollment rates are still below 100 percent in sub-Saharan Africa, South Asia and the Middle East and North Africa. Indeed, heavy grade repetition exaggerates the percentage of children who are actually in primary school. Sub-Saharan Africa and South Asia also lag behind in their secondary school enrollment rates, although they are increasing rapidly over time. Second, there is evidence that children in developing countries learn much less than children in developed countries for a similar amount of time in school. These issues are discussed in detail in Glewwe and Kremer (2006).

**Employment**

School-age children work in many poor countries, sometimes so much so that they leave school. Most children who work in developing countries work alongside of their parents in agricultural activities, but some work in factories or in other activities in urban areas. Concerns regarding child

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>15</td>
<td>41</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>52</td>
<td>91</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>21</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>18</td>
<td>47</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>23</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>12</td>
<td>37</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>10</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>4</td>
<td>17</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>36</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>19</td>
<td>42</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>102</td>
<td>108</td>
<td>91</td>
</tr>
</tbody>
</table>
labor have received increased attention in policy circles in the last one to two decades.

While child labor has received increased attention, child labor itself has decreased in almost all developing countries. The school enrollment trends discussed in the previous subsection suggest that this is the case, and the data in Table 51.5 confirm this. In developing countries as a whole, the labor force participation of children aged 10–14 has been reduced by almost half from 1960 to 2000, from 24.4 percent to 13.5 percent. Yet it is still high in two regions, South Asia and sub-Saharan Africa, which are the two regions with the lowest school enrollment rates. This may reflect lower economic growth in those two regions, an issue that is explored further in the next section.

The impact of economic growth on child welfare

Economic development has often been equated with income growth, but development includes not only income growth but also better health outcomes, higher levels of education, better housing (including potable water and hygienic sanitation conditions), and perhaps even democracy and respect for human rights. While it is very rare for income growth not to be accompanied by these other aspects of the quality of life, the nature of economic growth can determine how quickly economic growth leads to

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>28.4</td>
<td>21.3</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>3.8</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>14.5</td>
<td>9.8</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>17.2</td>
<td>10.9</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>40.0</td>
<td>31.5</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>40.6</td>
<td>35.2</td>
<td>27.7</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>24.4</td>
<td>19.3</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>134</td>
<td>134</td>
<td>134</td>
</tr>
</tbody>
</table>
improvements in other areas. This section presents evidence on the potential role that economic growth can play in leading to improvements in child welfare, first by showing that per capita income is positively correlated with indicators of child welfare and that income growth is correlated with improvements in child welfare, and then citing several studies that examine more carefully the causal impact of income on child welfare.

**Correlation between per capita income and child welfare**

Nations with higher per capita incomes have, on average, higher levels of child welfare, as seen in Table 51.6. More specifically, in the year 2000 in low-income countries (those with per capita income levels of $755 or less), 79 children out of every 1000 born died before their first birthday, and another 42 died before their fifth birthday. In contrast, in middle-income countries (those with per capita income levels above $755 but below $9266) only 27 out of 1000 children died before their first birthday, and only another seven died before their fifth birthday.

Middle-income countries also have much lower rates of child labor force participation than do low-income countries: 4 percent of children aged 10–14 work in middle-income countries, but 22 percent work in low-income countries. Primary and secondary school (gross) enrollment rates are also higher in middle-income countries; the rates for those countries are 110 percent and 77 percent, respectively; while the rates in low-income countries are much lower, namely 88 percent and 40 percent.

Not only is child welfare higher in countries with higher income, but improvements in child welfare are also positively correlated with the rate of

<table>
<thead>
<tr>
<th>Income level</th>
<th>Statistic</th>
<th>Infant mortality rate (per 1000 live births)</th>
<th>Child mortality rate (per 1000 live births, children &lt; 5 yrs)</th>
<th>Child labor (children 10–14)</th>
<th>Primary school gross enrollment rate</th>
<th>Secondary school gross enrollment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income</td>
<td>mean</td>
<td>79</td>
<td>121</td>
<td>22</td>
<td>88</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>72</td>
<td>72</td>
<td>69</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>Middle income</td>
<td>mean</td>
<td>27</td>
<td>34</td>
<td>4</td>
<td>110</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>72</td>
<td>72</td>
<td>59</td>
<td>56</td>
<td>51</td>
</tr>
<tr>
<td>Low and middle income</td>
<td>mean</td>
<td>53</td>
<td>77</td>
<td>13</td>
<td>100</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>144</td>
<td>144</td>
<td>128</td>
<td>104</td>
<td>89</td>
</tr>
</tbody>
</table>
income growth. This is seen in Table 51.7. Data from about 90 countries are used to divide countries into the 25 percent that had the slowest rates of growth from 1980 to 2000, the 25 percent that had the fastest economic growth, and the 50 percent with ‘intermediate’ rates of economic growth. Child labor and infant mortality drop faster in countries with higher economic growth. For example, from 1980 to 2000 labor force participation of children aged 10–14 dropped by 55 percent in countries with high economic growth but only by 43 percent in countries with low economic growth. Similarly, the infant and child mortality rates decreased by almost half in countries that had high economic growth while decreasing only by about one-quarter in countries with low economic growth.

The causal impact of income on child welfare

The correlations shown in the previous subsection are consistent with the hypothesis that income growth causes improvements in child welfare, but they do not constitute proof of a causal relationship. This subsection briefly reviews several recent studies that use microeconomic (household survey) data to provide more convincing evidence of a causal relationship.

Income growth can lead to improvements in child welfare, and in social welfare more generally, by two distinct pathways. The first, and most obvious, pathway is that households with higher incomes can purchase goods and services that improve children’s health and education outcomes, and improve child welfare in other ways. The theoretical literature has developed formal models that show how these income effects can occur (see, *inter alia*, Basu, 1999; Glewwe, 2002). Second, households with higher incomes, and more generally higher-income economies, generate more tax revenue (via either direct or indirect taxes) that governments can use to provide health, education and other services. Household survey data can be used to search for causal relationships that operate through one or both of these pathways.

### Table 51.7  Changes in child welfare by rate of per capita income growth

<table>
<thead>
<tr>
<th>Growth rate</th>
<th>Statistic</th>
<th>% Δ Child labor</th>
<th>% Δ IMR</th>
<th>% Δ CMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow growth (bottom 25%)</td>
<td>mean</td>
<td>–42.7%</td>
<td>–23.0%</td>
<td>–23.6%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Medium growth (25th–75th percentile)</td>
<td>mean</td>
<td>–49.3%</td>
<td>–40.5%</td>
<td>–43.2%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>46</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Fast growth (top quartile)</td>
<td>mean</td>
<td>–55.0%</td>
<td>–47.9%</td>
<td>–51.9%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>
Four recent studies using data from Vietnam show causal impacts of household income and/or provision of education and health services on child health, education and child labor outcomes. Glewwe and Jacoby (2004) present a dynamic model of school attainment that focuses on the role played by household wealth. They use panel data from the 1990s to show that increases in household wealth over time lead to a substantial and statistically significant increase in years of schooling, even after controlling for changes in the quality of schooling, the rate of return to education and the opportunity cost of (child) labor.

A study by Glewwe et al. (2004) using the same data examines the impact of economic growth on children’s nutritional status, as measured by their height for age. Unlike the case with education, household income by itself has little impact on children’s nutritional status. This suggests that improvements in the nutritional status of Vietnamese children in the 1990s were primarily due to improvements in health care services. Unfortunately, there are no data available on changes in the quality of health care services over time, so the paper cannot present strong evidence in favor of this conjecture. However it does present some evidence that higher-quality health care facilities lead to improvements in children’s nutritional status.

Research by Wagstaff and Nguyen (2004) examines the factors that affect child mortality rates in Vietnam. The authors find that access to safe drinking water, vaccination campaigns and access to trained medical personnel during childbirth reduce child mortality in that country. In contrast, but consistent with Glewwe’s result for child nutritional status, they find no impact of household income on child mortality.

Finally, Edmonds (2005) shows that children in better-off households in Vietnam are less likely to work, and that households whose incomes increase over time are more likely to keep their children in school and less likely to put them to work. This pattern is seen in other countries as well, as discussed in Edmonds and Pavcnik (2005).

Policies to promote child welfare
The evidence from the previous section demonstrates that both increases in households’ disposable incomes and increased government spending on social services lead to improvements in child welfare. This section reviews what developing countries can do to improve child welfare using policies that work through both types of causal pathways.

Promote economic growth
Macroeconomists and other economists have had, and continue to have, long debates on the best way to promote economic growth. For recent summaries of the evidence, see Easterly (2005), Rodrik (2005) and the World
Bank (2002a). Unfortunately, there is disagreement on what has been learned, but there are some areas of agreement. This chapter is too brief to delve into the details of the best policies to promote economic growth, but there is broad agreement among economists that economic growth is necessary for large, sustained improvements in child (and adult) living standards in developing countries.

Health policies
A recent book by Wagstaff and Claeson (2004) provides a detailed assessment of the effectiveness of health policies in developing countries. Policies that are effective in raising children’s nutritional status include improved hygiene and sanitation, dietary supplements that provide iron and vitamin A (for both mothers and children), and provision of deworming drugs to school-age children. To reduce infant and child mortality, the authors recommend improved hygiene sanitation, provision of mosquito nets that are treated with insecticide, child immunizations, dietary supplements of zinc and vitamin A, and improved facilities for childbirth.

Finally, it is important to realize that child health has large implications for children's educational outcomes. Several recent studies have found sizeable and statistically significant positive impacts of child health on education outcomes. Thus there is growing evidence of a causal impact of child health on education. Note as well that there is no clear evidence of large gender differences in the impact of child health on education outcomes. For details on the relationship between health and education, see Glewwe and Miguel (2008).

Education policies
This subsection summarizes recent research on policies that lead to increases in the years children spend in school and in the skills learned while in school. For details, see Glewwe and Kremer (2006).

Several education policies have been found to be effective in increasing the number of years that children spend in school, as well as their daily attendance during those years. More specifically, programs that reduce the costs of schooling faced by parents or provide incentives for daily attendance (either explicitly or implicitly through school meals) have sizable impacts on school enrollment and attendance. Randomized evaluations of school-based health programs, for example a deworming program in Kenya, suggest that, in some situations, these programs can be an extraordinarily cost-effective means of increasing the amount of time that children in developing countries are in school.

Evidence concerning the impact of education policies, such as provision of textbooks and additional training for teachers, on the skills that children
acquire in school is more mixed. In general, studies based on cross-sectional data suggest that most education policies have had limited impacts on the academic skills of schoolchildren in developing countries. Evidence from recent ‘natural experiments’ in middle-income countries suggests that reducing class size can raise academic achievement but that providing computers has little effect. Recent randomized trials conducted in low-income countries provide a more mixed picture.

The evidence suggests that the most effective forms of spending on education are likely to be those that respond to inefficiencies in schooling systems. Providing textbooks written with atypical students in mind will benefit only atypical students, whereas remedial education may be extremely effective in an environment in which many students fall behind and are no longer able to follow their teachers’ lessons. Providing radio mathematics education or computer-based education may be effective when teachers attend irregularly.

Schools in developing countries face significant institutional problems: distortions in education budgets often result in inefficient allocation and spending of funds; weak teacher incentives lead to problems such as high rates of teacher absenteeism; and, given the difficulties faced by these school systems, curriculums are often inappropriately matched with the level of the typical student. Yet reform initiatives can easily have unintended consequences. The details of these programs are critical for determining their effects on the incentives faced by teachers and others (principals, parents, and so on). Governance reforms and allowing school choice appear to hold more promise than simply providing monetary incentives to teachers based on test scores, but much more empirical evidence is needed on the impact of these reforms as well.

Employment policies

Of particular interest in recent years has been the issue of child labor. Clearly, children who work long hours cannot attend school, and there are many children who do work that could be directly harmful to their welfare. On the other hand, most children who work are working for their parents either on the family farm or on a rented plot of land, and their contribution towards their families’ income may have important welfare benefits for them and for other family members. This raises the issue of when, if ever, it is appropriate for governments to oppose choices that parents make for their children.

Two related policies that should reduce child labor without attempting to interfere with the choices that parents make for their children are to improve school quality and reduce the cost of attending school. Better and less-expensive schools will make schooling more attractive and thus should...
persuade parents to allocate more of their children’s time to schooling and less to child labor. In contrast, Edmunds and Pavcnik (2005) argue that attempts to ban child labor are likely to be ineffective and, if effective, may harm poor families. For further discussion, see Edmunds and Pavcnik (2005) and the references cited in that paper.

Concluding comments
Child welfare has improved in all regions of the developing world since the 1960s, and continued improvements in the future are likely. Even so, there are still opportunities to increase children’s schooling, raise their health and nutritional status, and reduce the amount of time they spend working to support their families. Policies to promote economic growth are one general avenue to accelerate progress, but there are other policies that are also effective, as briefly summarized in the preceding section. Efforts should focus on South Asia and sub-Saharan Africa, where the welfare of children is lowest.

A final important point is that improved child welfare today translates into better economic and social development in the future. Children who have higher levels of education and better health are more productive workers and thus contribute more not only to their own welfare as adults, but also to overall economic growth. Lastly, more-educated and healthier children will also be better able when they are adults to provide a better life for their own children.

References


The economic geography of ethnic and racial inequality

Virtually all countries have ethnic/racial divisions. Where those divisions are present they are universally characterized by dense gaps in economic status between the respective groups. Intergroup division and inequality occur in countries at all levels of per capita income or development more broadly construed. To the extent that intergroup inequality is significantly driven by discriminatory practices, those practices have a persistence that defy the standard prediction of conventional economics that discriminators will be driven from the marketplace because their actions are unprofitable. Furthermore, despite the strong and widely popularized claims of some scholars, there is no clear general relationship between ethnic/racial division within a country and its overall economic performance. Nor is there any sound reason to believe that causation is typically unidirectional from ethnic division to economic performance. But high rates of economic growth do not necessarily erode ethnic division.

Research that one of the authors of this chapter has undertaken with other collaborators (Darity and Deshpande, 2000; Darity and Nembhard, 2000) demonstrates the international prevalence of ethnic/racial differentiation and the strong correspondence between such differentiation and economic disparity. Darity and Nembhard (2000), in particular, demonstrate that intergroup differentiation and inequality is present in countries with large and small populations (for example the USA and Belize), countries with high and low average income levels (for example Japan and India), countries with a recent experience of rapid economic growth and slow economic growth (for example Malaysia and New Zealand), and countries with comparatively high degrees and low degrees of general inequality (for example Brazil and Australia). Indeed, even countries that are more inclusive toward women in their national political processes or that display greater gender equality do not consistently display reduced levels of ethnic/racial inequality.

Moreover, inequalities between ascriptively differentiated groups can be detected using a variety of measures of disparity. Intergroup gaps exist in income, occupational status, access to quality education, access to quality health care, wealth, and related measures of well-being. With respect to wealth, for example, in the USA, one of the world’s most affluent countries,
blacks and Latinos have approximately a mere one-tenth of the net worth of whites at the median (Kochhar, 2004).

**Ethnic and racial inequality and the Human Development Index**

In the 1994 edition of the United Nations Development Programme’s (UNDP) *Human Development Report* ethnic/racial inequality was described across a number of countries by calculating disaggregated values of the UNDP’s Human Development Index (HDI). The HDI is a measure of well-being that includes not only per capita income as a component but also indicators of educational attainment and health status in a population. Again, cross-national comparisons suggest that intergroup differentiation and inequality is evident in countries with scores at all levels of the HDI (Darity, 2002).

Within countries intergroup disparity measured by the HDI is widely evident as well. In 1992, toward the end of the apartheid era, South Africa’s overall HDI score was 0.65. However, the score estimated separately for white South Africans was 0.88, a score that would have placed white South Africa 24th out of approximately 180 countries worldwide (at the same level as Spain). In contrast, blacks in South Africa would have had an HDI score of only 0.46, placing them 123rd out of 180 countries, ranking slightly above the Congo.

Disaggregated estimates for Brazil were provided by region rather than by race or ethnicity. But the specific regions utilized in the UNDP, the south and the north-east of Brazil, provide information about racial/ethnic inequality there. The population of the south is disproportionately white, consisting largely of Euro-Brazilians, while the north-east is disproportionately black and mulatto or Afro-Brazilian. The overall HDI for Brazil as a whole was 0.76 in 1992, placing the country 63rd internationally. However, southern Brazil had a score of 0.84 which would have placed that region 43rd in the HDI rankings. North-eastern Brazil’s score was 0.55; the region would have ranked in the lower half of the world on the basis of HDI.

Similarly for Nigeria the *World Development Report* did not provide direct information on HDI scores by ethnic group. Scores were provided for each of the 19 Nigerian states. This afforded indirect information about intergroup inequality in Nigeria because each of the states are relatively ethnically homogeneous internally. It is striking that while Nigeria’s overall HDI was a very low 0.348 in 1992, ranking the country 139th in the world, the state of Bendel had a score of 0.67, higher than Sri Lanka or Cuba, while the state of Borno had an HDI of only 0.16, beneath any country in the world.

Canada’s HDI in 1992 of 0.93 was the highest in the world. Nevertheless, even there the evidence of intergroup disparity was dramatic. While it was
not possible to construct separate HDIs for Canada’s ethnic/racial groups with the available data, the data ‘do show that the “aboriginals” (the Indians, the Inuit, and the Metis, constituting 2.3% of the population) have a life expectancy 5.6 years lower than the rest of the population, and their real income is one-third less’ (UNDP, 1994, p. 100). Indigenous people were far more likely to be subjected to violence, experience depression and undergo unemployment. Their unemployment rate of 20 percent was twice the national average in Canada at the time (UNDP, 1994, pp. 25–6, 32).

The record also indicates that discrimination generally plays an important role in maintaining conditions of economic disparity, particularly discrimination in employment and housing. Plus, despite the theoretical presumption of conventional economics, the historical record gives little reason to believe that such discriminatory practices will disappear with the passage of time, even in market-based economies. In some countries, for example Brazil, there has been evidence of an increase in discriminatory differentials against blacks and mulattos over time (Darity, 1998). In others, like the USA and South Africa, where legal regimes of segregation have been overturned, there is evidence of a decline in discrimination in the period immediately following the regime change, but still high levels of discrimination persisting thereafter (Darity and Deshpande, 2000; Darity, 2002). Statistical estimates actually indicate that measured discrimination increased against Puerto Rican, Mexican and native American ancestry men in the USA between 1980 and 1990 (Darity, 2002).

Ethnic conflict, neoinstitutionalism and economic growth
A provocative and influential paper by William Easterly and Ross Levine (1997) oriented the discussion of the role of ethnicity toward an examination of its impact on economic development, rather than its impact on intergroup disparity. Easterly and Levine argued that the low rates of economic growth characteristic of African nations’ economies since the 1970s are attributable to the high levels of ethnic diversity there. Ethnic variation in a country, in and of itself, would lead to rent-seeking practices that would prove to be predatory on effective governance and contribute directly to slower growth. In the most recent versions of the argument (Easterly, 2001; Easterly et al., 2006), taking a tone very similar to Robert Putnam’s (2000) approach to conditions for community health, unity and participatory democracy, ethnic diversity becomes a critical factor undermining ‘social cohesion’ and thereby undermining the quality of ‘institutions’ that might otherwise promote economic growth.

One implication of the Easterly and Levine position is African governments would be larger – inefficiently larger – relative to the scale of their respective economies than governments in parts of the world with less
ethnic diversity. But Dani Rodrik (2000) finds that the public sectors in
African countries are not generally comparatively larger by international
standards, and he shows that there is at most a weak correlation between
the size of the public sector and the magnitude of rent-seeking activity and
corruption in a country.

Easterly and Levine focus on ethnic diversity – simply the magnitude of
ethnic variation that exists in a country. They give considerably less atten-
tion to the forms of ethnic antagonism (for a catalogue of these forms see
Bardhan, 1997) that might play out in different settings and their conse-
quences. One can easily conclude that racial/ethnic differentiation that
results in genocidal violence and/or conditions of civil war will not be good
for economic growth. But this requires more than mere diversity; it requires
high levels of group consciousness and high levels of between-group
conflict. Indeed, in the context of a model utilizing evolutionary game
theory it can be demonstrated that the continuation of group conscious-
ness is dependent upon unequal resources being associated with identi-
fication with and membership in each group (Darity et al., 2006).

During the midst of the Burundian genocide of the early 1990s – in this
case directed by the Tutsis against the Hutus – Leonce Ndikumana (1993,
p. 30) described the country as possessing ‘a rigid ethnic stratification and
unequal distribution of power along ethnic lines. While the Hutu make up
the majority of the population (about 85 percent), the minority Tutsi
(about 14 percent) control the government, the military, and the economy.
The third ethnic group, the Twa, has assumed a role of second-class citi-
zens with little integration in the economic and political system’. The
Burundian genocide at that time was associated with an ‘annual decline in
agriculture (value added) [that] reached −10.5% in 1994, exceeding the
decline experienced after the 1972 massacre (−10%), [and] total production
(GDP) . . . declined at an annual rate of over 6% in 1993 and 1994’
Ngaruko (2002) estimate that Burundian per capita income fell from $620
to $370. Without the civil wars they estimate that per capita income would
have been $667 in 1997 instead of $397.

In neighboring Rwanda, five years of civil war culminated in the 1994
genocide directed by the majority Hutus against the minority Tutsis. Gross
domestic product (GDP) fell in three of the five years; in 1994 there was an
astronomical fall in GDP of 40 percent. Only the infusion of massive
amounts of foreign aid led to a 9 percent growth rate in 1995 (US
Department of State, 2007).

Burundi and Rwanda are extreme cases. Such extreme forms of inter-
ethnic conflict are not unique to the African continent. Inter-ethnic strife
leading to genocidal violence has occurred in the former Yugoslavia, in
Indonesia in 1997 and 1998, and in Cambodia under the Khmer Rouge, all contributing to economic decline. The UNDP (1994, p. 47) identifies Turkey, the United Kingdom (especially Northern Ireland), Iraq, Iran, Israel, Lebanon, Colombia, Guatemala, Bangladesh, India, Laos, Myanmar, Pakistan, the Philippines, Sri Lanka and Tajikistan as non-African nations where ethnic conflict has turned violent with adverse consequences for economic performance. France’s ethnic violence in late 2005 also fits the model.

Generally, Robert Bates (2000) argues that the level of political violence represented by riots, demonstrations, revolts and assassinations is lower in African countries than might be expected, given the level of ethnic division there. While the short-term effects on growth of political violence, particularly genocidal violence, are negative, the long-term effects on economic growth are ambiguous:

Even a genocidal process of ethnic homogenization of population or ethnic homogenization of control over a nation’s resources is not inimical to prosperity, at least for the ‘winners’ and their descendents. Indeed, wealth seizures in the form of conquest of native peoples and appropriation of their lands, coupled with the use of captive and slave labor, laid the basis for the affluence of today’s richest nations, for example, the United States, Australia, Britain, France, Belgium and the whites of southern Africa.

Theft via conquest has long constituted an effective mechanism for achieving redistribution of wealth among groups. Industrialization, by destruction of lives of indigenous peoples, has been a commonplace event during the past half millennium. Violence is the historic adjunct to compulsory wealth redistribution across racial or ethnic lines. (Darity, 2002, p. 133)

Even in the near term, if the population being exterminated is located on an ‘undeveloped’ frontier, it may even be possible for economic activity to proceed in the ‘developed’ region of a nation undisturbed by the genocide. Indeed, do the genocidists even count the persons who are being eliminated as part of their relevant national population when they are computing the economic consequences of their actions? It is also possible that if the genocidal violence reduces total population sufficiently, in principle, it could offset a decline in national output sufficiently to produce an increase in per capita income. This perverse possibility reinforces the importance of Amartya Sen’s (1997) warning that one should not confuse increases in income with improvements in human well-being.

The significance of forced intergroup wealth redistribution for the long history of economic growth and uneven development is the core theme of Eric Williams’s (1994 [1944]) now classic study *Capitalism and Slavery*. Acemoglu et al. (2001) and Acemoglu (2003) edge onto this story but back away with a neoinstitutionalist explanation of the divide in the world.
between rich and poor nations. In his paper ‘Root Causes’, Acemoglu (2003, p. 27) proposes that there are two principal explanations of ‘the fundamental causes of prosperity between countries . . . geography and institutions’. For Acemoglu (2003, p. 27):

Good institutions [that promote economic development] have three key characteristics: enforcement of property rights for a broad cross section of the society, so that a variety of individuals have incentives to invest and take part in economic life; constraints on the actions of elites, politicians, and other powerful groups, so that these people cannot expropriate the incomes and investments of others or create a highly uneven playing field; and some degree of equal opportunity for broad segments of society, so that individuals can make investments, especially in human capital, and participate in productive activities.

Presumably, one must assume that industrialization in the United States in the midst of the consolidation of a regime of legal segregation was not a sufficient violation of the institutional conditions that Acemoglu lauds to invalidate his hypothesis.

Precisely why he settles on these two as the central explanations is not made clear. In the process Acemoglu eliminates from consideration the Williams perspective that emphasizes the hothouse effects on European economic development produced by the Atlantic slave trade and the slave plantation system in the Americas. Williams’s perspective places the stress on the role of colonialism in explaining variations in prosperity in the world economy today. There is at least one additional explanation which, thankfully, Acemoglu does not invoke as an option – variations in national cultures.

When all is said and done, Acemoglu (2003, pp. 27, 29) contends that institutional strength trumps geography in the following key passages:

if you look around the world today you'll see almost no wealthy country achieves this position without institutions protecting the property rights of investors and imposing some control over government and elites.

geography neither condemns a nation to poverty nor guarantees its economic success. If you want to understand why a country is poor today, you have to look at its institutions rather than its geography.

Note that the ‘equal opportunity’ feature of ‘good institutions’ in Acemoglu’s comment about ‘look[ing] around the world today’ is notably absent.

But what determined whether a country developed ‘good institutions’ on two out of three dimensions? For those regions of the world that underwent the process of colonialism, Acemoglu argues that it is the type of colonialism they experienced that set the path for the quality of their institutions. Colonies where Europeans established ‘extractive societies’
tend to have a poor institutional framework today, while colonies where Europeans established ‘settler societies’ tend to have a positive institutional framework. Hence the places where resource extraction took place, but Europeans did not migrate in large numbers, tend to be poor while the places where Europeans relocated tend to be more affluent. Certainly this would be fuel for the Eurocentric cultural determinist, but Acemoglu avoids that trap. Instead he argues that the form that colonialism took shaped the incentives faced by local elites for institutional development that pushed the two types of colonial systems on distinct long-run paths.

From Eric Williams’s perspective the distinction between colonies of extraction and colonies of settlement is artificial. In the latter there consistently were forms of extraction that took place – expropriation of the land by the settlers, the exploitation of native and slave labor, as well as other forms of coercion. The division of the world into rich and poor countries went hand in hand with the racialization of the colonial process. The colonizers were enriched whether they remained in their home country or moved to a site of settlement; the colonized were impoverished. These are the parallel elements of uneven development – development for some and underdevelopment for others. Simultaneously, ethnic divisions were crystallized and sustained to make it possible for the ‘winners’ of the colonial game to continue winning. The fascinating sixth chapter of Williams’s (1942) *The Negro in the Caribbean*, entitled ‘The Middle Class and the Racial Problem’, provides a rich explication of the emergence of the interaction of color and class stratification as a consequence of racialized colonialism.

In another paper Acemoglu et al. (2005, pp. 546–7) highlight the centrality of Atlantic trade in the period between 1500 and 1800 as the foundation for European economic development. They contend that countries best able to take advantage of the growth opportunities afforded by the Atlantic economy were those ‘with relatively nonabsolutist initial institutions, most notably in Britain and the Netherlands [i]n contrast [with] countries where the monarchy was highly absolutist, such as Spain and Portugal’. According to Acemoglu et al., non-absolutist states that did not experience rapid growth, like Venice and Genoa, did not have adequate physical access to the Atlantic to gain from the cross-oceanic trade.

This argument should imply that Britain, in particular, should have demonstrated rapid economic growth from the point at which significant constraints were imposed on absolutism, the signing of the Magna Carta in 1215, a full three centuries before the interval that Acemoglu et al. (2005) identify as the period of the ‘rise of Europe’. An alternative to the neoinstitutionalist account is the argument that the British and the Dutch were simply the winners of the game of mercantilist rivalry (Darby, 1990).
Indeed, many of the pecuniary benefits of the Portuguese and Spanish colonial systems were transferred to Britain via intra-European trade (Darity, 1990). For example, it has been estimated that during the eighteenth century trade surpluses with Portugal brought 50,000 pounds of bullion into London weekly (Birnie, 1935, pp. 175, 180).

Acemoglu et al. (2005) are aware that their analysis touting the importance of the Atlantic trade for European economic development could connect directly with arguments, like Williams’s, about the importance of ‘the associated profits from colonialism and slavery’. But they immediately seek to sever the connection with the following observation:

> It is undoubtedly true that colonial relations with the New World and Asia contributed to European growth. Nevertheless, quantitative analyses, for example, Engerman (1972), Engerman and O’Brien (1991) [sic: O’Brien and Engerman (1991)], O’Brien (1982), and Bairoch (1993, ch. 5), suggest that the volume of trade and the profits generated by the Atlantic trade appear to be too small to account for much of European growth directly. (Acemoglu et al., 2005, p. 562)

Instead, they characterize the institutionally promoted gains from trade for European development as indirect effects: ‘the rise in Atlantic trade enriched and strengthened commercial interests outside the royal circle and enabled them to demand and obtain the institutional changes necessary for economic growth (Acemoglu et al., 2005, p. 550). They dismiss the direct effects by invoking what one of the co-authors of this chapter has dubbed the ‘small ratios’ argument.

They do not appear to be aware that the ‘small ratios’ argument has been critiqued in a counter set of quantitative analyses to such an extent that the opposite position can be sustained. The volume of trade and profits from the Atlantic economy, particularly the slave trade itself, were enormous by comparative historical standards (Solow, 1985; Bailey, 1986; Darity, 1990; Cuenca Esteban, 1997; Inikori, 2002). Even O’Brien and Engerman (1991) concede that the available trade statistics indicate that colonial trade was of paramount importance for England, at least in the eighteenth century, if not the seventeenth century as well.

Easterly et al. (2006) also subscribe to a variant of neoinstitutionalism, but offer a characterization of ‘good institutions’ closer to Putnam’s notions of social cohesion, and a quite different set of factors as contributors to ‘good institutions’. One of the factors that they hypothesize will make for ‘bad institutions’ is, again, ethnic diversity. However, in the earlier paper on Africa, Easterly and Levine (1997) attribute the extreme ethnic diversity that they claim is present there to colonialism – perhaps providing a basis for rapprochement with Williams. However, they never explore or consider the direct enrichment effects on Europeans of the colonial
process, the task which Williams pursued in depth, from the crucible of the Atlantic slave trade to the mid-twentieth century.

**Measuring ethnic diversity or ethnic polarization?**

Easterly and Levine (1997) presume that ethnic diversity necessarily has a negative impact on economic growth. In contrast, Alesina and La Ferrara (2005) propose that there are ways in which ethnic diversity might prove beneficial for economic growth. They acknowledge that there are potential costs to diversity including ‘[c]onflict of preferences, racism, and prejudices . . . lead[ing] to policies that are at the same time odious and counterproductive for society as a whole [and] [t]he oppression of minorities may[be] lead[ing] to political unrest or even civil wars’ (Alesina and La Ferrara, 2005, p. 762). But they also say that ‘a diverse ethnic mix also brings about varieties in abilities, experiences and cultures that may be productive and may lead to innovation and creativity’ (Alesina and La Ferrara, 2005, p. 762). Robert Bates (2000) has also argued that ethnic group social capital can promote human capital formation among their own to levels that would not have taken place in the absence of group identification and membership – with potential positive effects on the society as a whole. Paul Collier (2000) has suggested that the effects of ethnic diversity on growth are negative in non-democratic societies and positive in democracies. Ultimately, Alesina and La Ferrara (2005, p. 763) conclude that the impact of diversity on economic performance is an empirical question since, at the theoretical level, it is unclear whether ‘the benefits of “diversity”’ will outweigh ‘the costs of heterogeneity of preferences’.  

To address this question, an appropriate measure of ethnic diversity is needed. This is the point where a ferocious debate has raged among social scientists examining the relationship between ethnicity and economic development. The debate is simultaneously about which measure is most appropriate from a conceptual standpoint and which measure has the best technical attributes.

Easterly and Levine (1997) initiated the empirical work in this area by deploying the ethno-linguistic fractionalization (ELF) index. Using the Herfindahl concentration formula, a country’s ELF score represents the likelihood that any two people chosen at random from the population will speak different languages. The likelihood has a minimum value of zero when there is no ethno-linguistic variation in a country and a maximum value of 100 when the odds are perfect that any two persons drawn at random will speak a different language. They find that the ELF index (expressed as a percentage) is positively and statistically significantly related to their measures of unproductive macroeconomic policies.
Still, some peculiarities are immediately evident with this measure; the extent of ethnic tension and violence definitely is not reflected in the index score. Haiti’s ELF score, for example, is 1 in Easterly and Levine’s (1997) study. This is due to the high degree of linguistic homogeneity, providing no indication of the high level of conflict conducted by the social elite and the military toward the majority of civil society (UNDP, 1994, pp. 41–2). Burundi’s ELF score of 4 is attributable to the near universal use of French. The index could not have been used effectively to forecast the long cycle of genocidal violence there. The same is true of Rwanda.

Daniel Posner (2004) criticizes both the design of the ELF index as a measure of ethnic diversity and its use in growth regressions. The ELF index was built on coding decisions made by a Soviet ethnographer in the 1960s. Insofar as ethnic identities undergo some fluctuation in importance and intensity, definitions now a half-century old may be misleading. More significant, the identification of ethnic groups solely on the basis of linguistic differences ignores other critical markers of distinction, such as phenotype, religion, attire and cultural practices. Furthermore, the ELF index does not take into account the spatial distribution of ethnic groups within a country, nor does it incorporate any information that would capture the depth of ethnic division.

Nonetheless, Posner’s primary objection to the ELF index is the absence of consideration of political organization and contestation by the relevant racial/ethnic groups in a society. The original Easterly and Levine (1997) hypothesis that greater ethnic diversity leads to slower economic growth rests on an intermediate causal step involving the formation of macroeconomic policies. Greater ethnic diversity is supposed to increase interest group polarization, motivating rent-seeking, overspending and financial repression and subverting the provision of public goods. Posner argues that the ELF index is an inappropriate measure for testing this hypothesis because there is no necessary relationship between the existence of ethnic divisions and the forms of political competition that lead to an institutional environment that inhibits growth. He contends that many ethnic groups do not operate as independent political actors because they lack the political strength either to influence policy directly or to mobilize as part of larger coalitions and parties.

Posner proposes a revised measure of ethnic diversity called the Politically Relevant Ethnic Groups (PREG) index. Using secondary sources, Posner identifies those ethnic groups that have been politically active and have engaged in competition over social policy in each of four decades. He applies the Herfindahl formula to his new measure of ethnic groups and develops index values for 42 African countries in each decade. Although correlated with other comparable indices, the PREG index
generates important differences in the rank ordering of the most- and least-fractionalized countries.

The most common technique for analyzing the economic effects of ethnic diversity, in the aftermath of the Easterly and Levine (1997) study, is to include a measure of ethnic fractionalization as an explanatory variable in a cross-country growth regression. Posner compares the performance of the ELF index with the PREG index in explaining variations in economic growth across the African countries. First, a series of policy measures are regressed separately on the ELF and PREG indices. Surprisingly, although the ELF index has a statistically significant effect on many of the policy variables in the world sample, the effect disappears when the analysis is restricted to the African sub-sample. In contrast, the PREG index has a statistically significant relationship with the black market premium and with fiscal surpluses in the anticipated direction.

Posner then regresses growth rates separately on the ELF and PREG indices. The early Easterly and Levine (1997) results can be replicated with the global sample, but it appears to have no effect on the growth performance of the countries in the all-African subsample. The PREG index does have a statistically significant effect on African growth rates in the expected direction, that is, the higher the PREG value, the lower the country’s growth rate.

Finally, Posner regresses rates of growth separately on the ELF and PREG indices with other policy variables included as controls. Peculiarly, the PREG index loses significance in the presence of these additional controls, while the ELF index now has a significant effect, even in the African subsample. Posner’s interpretation is that the PREG index successfully captures the indirect effect of ethnic diversity on economic growth via macroeconomic policies but does not have a direct effect on growth. The ELF index, in contrast, has a direct effect on growth independent of other policy practices. Posner speculates that the ELF index may be picking up some other mechanism through which ethnic diversity affects economic development. In the end, Posner concludes that ethnic fractionalization is negatively and strongly related to economic growth in Africa, and offers his results as evidence that macroeconomic policies are an important channel through which ethnicity influences economic development.

James Fearon’s (2003, p. 198) construction of an alternative measure to the ELF index attempts to locate ethnic groups based upon ‘what people in the country identify as the most socially relevant ethnic groupings’. As Alesina and La Ferrara (2005, p. 792) point out, Fearon’s basis for development of his index – Fearon’s determination of what people in a given country decide are ‘the most socially relevant ethnic groupings’ – depends significantly upon Ted Gurr’s (1996) Minorities at Risk Project at the University of Maryland.
And, indeed, the criteria that Fearon utilizes predicated upon Gurr’s database comes far closer to capturing inter-ethnic tension-cum-violence than does the ELF index. Gurr (1996) defines a ‘minority’ (not necessarily a numerical minority but any group that is less than 100 percent of the population) at risk as a communal group that: (1) faces political and/or economic discrimination; and (2) acts on its own behalf collectively in the political process. The entire population consists of minorities at risk in Burundi, Chad and South Africa on the Gurr criteria. In a global sample Gurr finds that the African continent has the largest share of population comprised of minorities at risk, a result potentially inconsistent with Bates’s observations about the levels of political violence in Africa relative to the levels of ethnic differentiation.

Alesina et al. (2003) show that the Fearon index, based largely upon the Gurr scale, is closely correlated with their version of the ELF index, an extended version that includes ethnic groups defined by other characteristics such as skin color. The ‘more comprehensive’ version of the ELF index captures salient ethnic differences in Latin America that the language-only index would not capture:

In [Latin America], the language index shows more homogeneity because the language of the former colonizers (Spanish, Portuguese, English) is often spoken by most, but the index based on skin color or ethnic origin (say black, mulattos, white, mestizos, Indian, etc.) shows more heterogeneity. (Alesina and La Ferrara, 2005, p. 792)

Alesina and La Ferrara seem to find the extended ELF and the Fearon indices both to be quite satisfactory, although quite different in design.

Jose Montalvo and Martha Reynal-Querol (2005) have proposed the polarization index as an alternative to the ELF measure that is conceptually quite distinct from the others discussed here. The ELF index presupposes that a country is more fractionalized, the greater its number of separate groups. A country with two similarly sized groups facing each other in a cauldron of hostility would not be depicted as highly fractionalized by the ELF index. The Montalvo and Reynal-Querol index reaches its peak value when a country consists of two equally sized groups, and then declines in value as the number of groups increases, departing from the half-and-half split. Alesina and La Ferrara (2005, p. 793) describe the comparative assessment of the polarization index with the ELF index as follows:

[Montalvo and Reynal-Querol] show that this index is highly correlated with ethno-linguistic fractionalization (ELF) at low levels of ELF, uncorrelated at intermediate levels, and negatively correlated at high levels. In a cross-country
regression analysis, they find that ethnic polarization has a positive impact on the likelihood that a civil war occurs and a negative effect on a country’s growth rate. They do not find an independent effect of ethnic fractionalization. Using a different data set, Alesina et al. (2003) compare the results of the polarization index RQ and the fractionalization index ELF, and find that fractionalization works slightly better as a determinant of policies and economic outcomes. While the apparent inconsistency between the two sets of results may be due partly to different parameterization and partly to different data sources, it is between the two measures at low levels of fragmentation.

Tade Okediji (2005) focuses his criticism on the dimensional limitations of the ELF index. He argues that linguistic differences are only one of several possible cleavages that are associated with ethnic division. Racial and religious identities also form the basis of ethnic differentiation, and the ELF index neglects such groupings if they do not coincide with linguistic differences. Moreover, many countries are characterized by complex interactions of racial, religious and linguistic fractionalization, leading to overlapping identities and variation in the salient factors that distinguish one group from another. To counter these limitations of the language-only ELF index, he advances an alternative index that is quite similar conceptually to the more comprehensive ELF index developed by Alesina et al. (2003). But he then performs a comparison with the language-only ELF index that is quite original and has interesting implications.

Okediji proposes the Social Diversity Index (SDI), a measure of fractionalization intended to capture the multidimensional nature of ethnicity – a sort of Human Development Index for ethnicity. Also using secondary sources, he partitions each country into primary racial, religious and linguistic groups. His measure enables him to classify each individual in a country according to all three characteristics, although the person can only belong to one category within each characteristic. When compared with the ELF index, the SDI has a higher mean and a lower standard deviation across the same sample of countries. Okediji concludes that ethnically diverse societies are far more common than the ELF index would suggest, and that the variation in the degree of ethnic fractionalization across countries is much smaller than suggested by Easterly and Levine in 1997. Thus, Okediji returns the discussion to the point of origin of this chapter – the near universal presence of salient ethnic/racial differentiation across the nations of the world and the near universal presence of ethnic/racial inequality.

The research on the relationship between ethnic fractionalization and economic performance using cross-country regressions invariably treats ethnic fractionalization as driving economic performance. But it is quite plausible that the overall economic performance in an economy might
affect ethnic antagonisms positively or negatively. Perhaps conditions of slow economic growth can inflame inter-ethnic tensions while prosperity may relax them? Developments in Malaysia and in Indonesia seem to connect economic crisis to increased communal violence (Darity, 2002, p. 135). The possibility of simultaneous causation has not informed research in this area to any significant degree, although Alesina and La Ferrara (2005, pp. 772–3) are well aware that any number of the ‘independent’ variables used in regressions of this type are subject to the endogeneity complaint.

Similarly, it can be asked whether the general level of inequality in a society drives the level of intergroup inequality, rather than the level of intergroup inequality shaping the general level of inequality (Darity and Deshpande, 2000). There may be no general answer; indeed, the particular answer is probably contingent on the particular structure and history of each country or region.

And what about the fundamental relationship between ethnic/racial conflict and economic inequality between groups? At the very start of his 1985 monograph, *Racial Conflict and Economic Development*, W. Arthur Lewis raised the following question:

Is economic equality necessary for social peace? From one standpoint every so-called racial conflict is sustained (or even initiated) by an economic conflict, covert or open. What poses as a conflict between a dominant and a subordinate group, we are told, is really only a way that an exploiting minority recruits supporters of its case from people with whom it has only racial ties. The proposition is somewhat doubtful. It is probably true that every dispute, racial or not, has, or acquires some economic edge, but this is not the same as saying that all disputes originate in economic conflict. If the economic conflict were mitigated by movement toward equality, would the racial conflict be lessened automatically?

The difficulty is that, far more often than not, effective movement toward economic equality requires disrupting the economically privileged position of the socially dominant group. Its members will resist or act to destabilize the policies that have been adopted to push the society toward greater intergroup equality, whether it is affirmative action, school desegregation or a program of reparations. The very effort to achieve greater intergroup equality will exacerbate racial conflict – a backlash effect – from the group that sees its position of privilege as being threatened. The threat arises precisely because the members of that group have a material benefit from maintaining their identity as a group. To eliminate racial conflict, the benefits associated with racial division would have to be removed. But to remove those benefits typically unleashes racial conflict. In that sense, economic equality is necessary for social peace, but the process of achieving economic equality between ethnic and racial groups is invariably far from socially peaceful.
References


PART VIII

THE STATE, INSTITUTIONS AND DEVELOPMENT
The proper roles of states and markets in fostering and sustaining economic prosperity comprise one of the oldest and most debated topics in the history of economic thought. A focus of Adam Smith’s attack on mercantilism in the late eighteenth century, the roles of states and markets were debated by the historical, institutional and early neoclassical schools of economics a century later. More recently, the issue enjoyed a place of prominence in late twentieth century debates over reasons for the economic growth of Japan and other East Asian economies, the most notable success stories of that period.

It should not surprise us, perhaps, to find disagreement over the roles of states and markets in the economy, because the two institutions have histories of both synergy and rivalry (Putterman and Rueschemeyer, 1992) stretching back to ancient times. Some 4000 years ago, populous societies marked by increasingly complex divisions of labor gave birth to the first city states and empires, and while those societies probably saw unprecedented expansions of market activity, their economies were also in some cases state-dominated, and all featured attempts by the new states to enrich themselves and to support larger armies and coteries of officials by controlling sources and flows of wealth. The interplay between traders and officials differed from one society to another, and in given societies across periods of time. While extractive, centralized states may have constrained economic prosperity in some instances, in others the absence of central authority constrained it. For example, the tenuousness of law, order and safe commercial routes in much of Europe following the collapse of Roman rule probably contributed to economic stagnation. It seems noteworthy that modern capitalism and the nation state arose in tandem in Europe after 1500. Since then, no prosperous modern economy has emerged in a society lacking a well-ordered state.

The roles of markets and states
Despite controversy, most economists agree that markets, permitting competition among independently managed enterprises free to select their products and methods of production, have been central to economic
progress and crucial to the growth of productivity, technological know-how and living standards since the Industrial Revolution. Independent enterprises can thrive, these economists would argue, only when free to interact with suppliers and customers in relatively uncontrolled labor, raw material, capital goods and product markets. At the same time, most economists also agree that market systems function poorly if at all without the protection of property rights, rule of law and availability of a stable currency, providing which have been central economic roles of states. Further, economists recognize domains in which competing enterprises cannot be expected to bring about optimal results, most prominently the cases of market power, public goods (including some key trade-facilitating infrastructures), and environmental externalities. Where debate exists is over the scope of the set of public goods (for example, whether they include health care and education), and over the degree to which government remedies (for instance, in the case of monopoly) tend to ameliorate rather than worsen unregulated outcomes. Full agreement is also lacking as to the macroeconomic responsibilities of governments, and whether inequalities in the distribution of income and wealth should be viewed as market failures, also calling for government interventions. A point of particular relevance to this chapter is whether governments can promote economic development by formulating strategies to promote growth, or whether the ideal role of government is simply to create a stable institutional environment and then allow the market to 'work its magic'.

Although properly speaking markets are the loci of interactions between buyers and sellers, sometimes although not always associated with specific institutional frameworks or locations, economists use the phrase ‘the market’ to refer not only to the settings or sets of those interactions but also to the decentralized economic mechanism as a whole, including the presence of autonomous enterprises that purchase inputs and transform them into goods and services. A ‘market economy’ (sometimes called ‘the market’) in this broader sense is said to allocate scarce resources efficiently among competing needs and wants, and to stimulate technological progress, for several reasons. First, the market mechanism causes both final consumers and intermediate users of goods, services and resources to signal the values they attach to those goods, and so on, as a by-product of the exchange of offers to buy and sell. Second, the pressure that enterprise owners feel to maximize net returns, lest they be driven from business or at least suffer serious financial losses, induces them to attempt to produce goods of maximum value to consumers using the least-cost combinations and quantities of resources, while also striving to satisfy the needs of buyers in terms of quantity, quality and variety. Third, individual workers, including prospective managers and specialists of other kinds, are motivated to
invest in demanded skills due to the higher earnings which market competition assigns to those with scarce capabilities. Fourth, owners of non-labor resources can be expected to steer those inputs toward the uses most valued by society, responding to price signals which indicate, for example, the value of a parcel of land as an orchard versus its value as a grain field or a parking lot. Finally, the rents that accrue to innovation, and the fear of falling behind competitors, are viewed as major causes of the high rates of technological change observed in market capitalist economies.

But markets do not exist in social and political vacuums. Market interactions are embedded in social systems (Granovetter, 1985) and many trades would be impossible without social norms that increase trust and facilitate contract enforcement (Putnam, 1993; Greif, 1994). Although state power and the formal legal systems to which states lend their force may offer protections of final recourse, market interactions rely more directly and extensively on widespread adherence to such norms by ordinary members of society. Recent research suggests considerable cross-country variability in trust and social capital, and this variability correlates with differences in levels of economic development and growth (Fukuyama, 1995; Knack and Keefer, 1997). It can be argued that the existence of a stable political order, honest administration and a non-corrupt judiciary may be important facilitators of social trust and norm abidance (although the relationship probably goes both ways). State actions that contain the extremes of inequality, poverty and neglect of worker health and safety may, while intervening in the full expression of the logic of competition, end up helping the market by helping to stabilize the polity and society and to increase the social acceptability of leaving most economic coordination to the market.

Lessons from planned economies
While economists’ convictions about the efficacy of markets derive in part from classical and neoclassical theory as well as from observation of market economies, many draw lessons also from experience with alternative economic systems and policies. The premier experiment in operating large modern economies with almost no state-tolerated role for markets is the one that began in Russia in 1928 and lasted into the 1980s there and in numerous other countries eventually ruled by communist parties. In those centrally planned economies, prices were set administratively and the allocation of resources and determination of production plans was to take place not under the influence of market forces, but rather under the aegis of a planning bureaucracy directly weighing leaders’ political goals and perceptions of societal needs. Planners ignored notions of comparative advantage and were guided instead by the goal of building industrial
economies, which they hoped to achieve by directing huge investments into the capital goods sectors, ordering low-price crop deliveries from farmers, and exploiting natural resources without regard for opportunity cost and environmental impact. Planners paid only limited attention to consumer goods production, and even less to the provision of services other than health care and education.

According to observers (for example, Nove, 1983; Kornai, 1992), planners in the Soviet-type economies struggled with the problem of providing effective incentives to, and the need to elicit information from, enterprise managers. Quality, variety, spare parts and maintenance were perennial problems. Enterprises integrated vertically to avoid relying on the planning bureaucracy for the inputs they needed. Considerable amounts of resources were diverted into black market activities. Innovation proved difficult to engender at levels comparable to those of industrialized market economies, except perhaps in the military sector. Considerable activity took place outside of the approved plans, with some observers going so far as to argue that the claim that such economies were primarily plan-based is inaccurate.

For a time, rapid structural change and achievements in health and education sectors made the model attractive to some outsiders, especially in the developing world. The Soviet Union’s industrial output growth rate exceeded that of the United States during the 1930s and again from World War II until the late 1970s. Most estimates suggest that China achieved a higher rate of industrial growth than India from the late 1940s to the late 1970s, despite ideological excesses and political upheavals. Life expectancy in China exceeded that in India by some 12 years in 1978, and in general planned economies achieved higher life expectancy, lower infant mortality and higher literacy rates than non-communist countries at similar income levels. However, the curtailment of individual freedoms and comparisons with incomes and consumer good availability in neighboring countries like West Germany and Taiwan bred dissatisfaction with the system, ultimately leading to the system’s demise in both Europe and Asia. In addition, the planned economies’ growth was unbalanced, and much of the capacity put in place by their economic system had little value when the countries in question adopted market-oriented reforms and became more open to international trade.

State roles elsewhere
Less extreme in their departures from free market principles are the numerous cases in which market pricing and exchange were permitted but with key government interventions ‘distorting’ the price system. Most developing countries of the late twentieth century controlled foreign exchange transactions, usually overvaluing their country’s currency. By doing so,
they inadvertently discouraged exporting and necessitated combinations of import restrictions, tariffs and borrowing to deal with trade imbalances. Interest rates on bank loans were often subject to regulatory caps, and import licensing and tariff arrangements made some capital goods less expensive, creating artificially low prices for certain producer goods while the cost of capital to small-scale borrowers, including farmers, remained high. These and other interventions, described by some economists as ‘getting prices wrong’, led to limited growth of bank deposits and other forms of financial mediation (‘shallow finance’), excessive capital intensity and limited job creation in a few modern sector activities, and capital starvation and underemployment in other parts of the economy. By discouraging exports and domestic savings while encouraging imports, they also contributed to the growth of unsustainable burdens of debt at the national level.

Although these examples convince many of the virtue of markets, pure free market economies are textbook abstractions to which no modern national economy adheres in reality. After the 1930s, the ideal of a fully self-regulating economy was abandoned by most economists and politicians in industrialized countries, with macroeconomic stabilization being considered a responsibility of states. Other government roles were also growing. The proportion of national income used by governments to pay civil servants and to support various functions grew steadily until close to the end of the twentieth century. Governments were asked to respond to market failures, for example to set and police environmental standards, and to monitor the safety of workplaces, foods and pharmaceuticals. Governments invested in roads, bridges, maintenance of waterways and harbors, and rail lines. Due to some combination of market failure and distributive concerns, governments also provided unemployment benefits, food and health subsidies, pensions and other social benefits. The notion that the market and autonomous firms could be engines of production and technological progress, but that the distribution of social benefits could be partly separated from that of market rewards, was mainstream in the polities if not in the economics professions of the world’s most prosperous and technologically advanced nations after World War II.

At the same time as the role of the state was growing in industrialized mixed economies, economists were beginning to re-examine their depiction of government as a benevolent agent that could be counted upon to follow the prescriptions of normative economic theory regarding the correcting of market failures. Mainstream political economy viewed governments as being composed of individuals who might promote the well-being of citizens if imbued with social concerns or held accountable by an engaged public, but who might also be as motivated by self-interest as other
individuals. If monitoring by the public is costly and if formally democratic political institutions can be captured by groups with concentrated interests in particular areas, state officials and politicians might be poor servants of the public as a whole. Grievous policy errors could also result from simple misunderstanding of the effects of instruments like exchange controls and interest rate ceilings. The idea of ‘government failure’ entered the lexicon of economics, alongside the term ‘market failure’, and the possibility was raised that even when there exist imaginable state interventions that can increase social welfare if effected, actual government involvement might worsen rather than ameliorate some market failures. Efforts to help the poor might also have effects at odds with that aim in the long run if they resulted in reduced incentives to invest in human and physical capital and thus lower rates of growth.

The special problems of less-developed countries

Some arguably distinct aspects of the state–market relationship in those countries lagging far behind the most industrialized economies have been the subject of separate discussions at various points in time. In the nineteenth century, development strategies were proposed and to some degree adopted by national governments in then-lagging countries including the United States, Germany, Japan and Russia, usually including tariff protection against a range of importable manufactured goods and government assistance or active participation in the accumulation of investment funds. The so-called Great Industrialization Debate in the Bolshevik-ruled Russia of the 1920s would be re-examined by the advisors to leftist Third World governments like those of Mozambique and Tanzania as late as the 1970s. The post-World War II era in which the Bretton Woods institutions were formed to help manage the economic problems of less-developed and especially newly decolonized countries saw the growth of a new literature of development economics.

Early post-World War II writers such as Ragnar Nurkse (1953), W. Arthur Lewis (1954) and W.W. Rostow (1960) argued that the central problem of a developing economy was to raise the share of capital formation in national product to a level sufficient to fuel the growth of modern-sector activities and, in the formulation of Lewis and of Fei and Ranis (1964), to ‘drain off’ the pool of surplus labor underemployed in the traditional, mainly agricultural, sector by absorbing it into modern employment. As a result of such thinking, government development plans identifying the gap between domestic savings and investment targets were formulated and used as bases for seeking investment financing from international financial institutions and foreign governments. During the 1960s and 1970s, there was still considerable tolerance among Western development specialists and
advisors for using inflation as a tax to finance government investment, using tariffs to protect domestic industries deemed promising, and adopting other measures that would even then have been considered inadvisable in a developed-country context.

Governments were also thought to have a role to play as coordinators of the overall push for development. Although enterprises were mainly privately owned and prices determined by supply and demand, it was argued by some that national economic planning could still play a crucial coordinating role. The theory espoused by advocates of planning in post-World War II France, for example, was that firms might be reluctant to invest if they could not be confident that complementary investments were being made by others up and down the relevant production streams. Complementary investments in infrastructure and manpower training might also be called for. The market might be a suitable coordinator of short-term production decisions, the theory went, but the large-scale investment decisions needed to bring about true structural change might be too lumpy to overcome the hurdles of uncertainty in an unplanned economy. The argument was stretched by some to the point of suggesting that government itself had to do the investing, even in industrial enterprises and mines, to overcome private sector hesitation. But for the most part, it was used in favor of a government coordinating and facilitating role, not state ownership. The role of governments in Japan and later Korea were often understood in this light, but so too were the national plans typical in many other developing countries.

The ‘international division of labor’ had a central place in many discussions. Free trade among nations, it was argued, was more beneficial to rich than to poor countries. Europe’s poor former colonies in Asia, Africa and Latin America had been brought into the world economy for the benefit of their colonizers as sources of cheap raw materials and foodstuffs and as markets for European and North American manufactured goods. It was in the ‘core’ or ‘metropolitan’ countries’ interests that the ‘peripheral’ countries’ comparative advantage remain one based on unskilled labor and raw materials. ‘Free trade’ would perpetuate this because manufactured goods would remain less expensive for poor countries to import than to produce, so they would have no chance to learn by doing and to move towards the international frontier of industrial capability.

In response to such concerns, economic moderates called for the leveling of the playing field by reducing discrimination against developing-country agricultural exports, devising mechanisms to stabilize and maintain the prices of tropical commodities like sugar cane and coffee, and fully opening developed-country markets to Third World manufactures – a ‘New International Economic Order’. Believing that positive steps were required to foster structural change away from the old reliance on primary product
exports, many also believed that selected industries in poor countries should be protected from foreign competition by tariffs of sufficient magnitude to let domestic manufacturers obtain a foothold – the groundwork of the import substitution industrialization approach. More radical commentators called for de-linking developing economies from the world trading system, substituting (especially for smaller economies) links with neighboring countries and/or with Communist states (for a discussion, see Diaz-Alejandro, 1978). In either case, national strategies, which only governments could put forth and implement, were seen as requirements for escaping the self-perpetuating status of underdevelopment. Thus, a key role was assigned to the state; leaving things to market forces would only perpetuate underdevelopment and dependency.

**More recent discussion**

The 1980s were a watershed decade for policy and professional opinion on the problems of economic development. The decade was marked by slowdowns of growth in the industrialized market economies, a still more pronounced slowdown in productivity growth in Communist countries, the accumulation of unsustainable debt levels by many middle- and low-income developing countries, and increased international recognition of the remarkable growth achievements of a number of East Asian economies. These developments helped to fuel a conservative backlash against the ‘welfare state’ and state-owned industries in the West; the beginnings of radical economic reform in China and a last decade of reform experimentation in the Soviet bloc; initiation of structural adjustment programs in developing countries in the wake of their debt crises; and the fall from respectability of import substitution industrialization as a policy approach. Moved partly by the necessity of accepting International Monetary Fund (IMF) conditions for urgently needed loan programs, partly by the perception of the relative merit of East Asia’s more outward-focussed orientation, most developing-country governments devalued their currencies, reduced spending, began reversing the trend of nationalization, and attempted to make their economies attractive to foreign investors. Going into the 1990s and the start of the twenty-first century, increasing flows of foreign direct investment and international bank lending, growth of trade volumes, and vigorous participation in world trade by China, India and other developing economies, became hallmarks of the intensification of international trade, investment, knowledge and cultural flows that was dubbed ‘globalization’.

Even though policies were trending in this period towards liberalization in comparison with prior decades, it would be incorrect to describe the approaches of most developing-country governments as laissez-faire. Nor would it be accurate to suggest that the prevailing view of the state’s role in
development had become the minimalist one favoring protection of property rights and of a stable currency as the only valid economic functions of government. The 1997 *World Development Report* of the World Bank stated that: ‘[a]n effective state is vital for the provision of the goods and services – and the rules and institutions – that allow markets to flourish and people to lead healthier, happier lives. Without it, sustainable development, both economic and social, is impossible.’ Government investments in the health and education sectors were viewed as important both to immediate well-being and to economic growth. The role of governance, especially the rule of law and the absence of corruption, was accorded considerable importance in World Bank publications and was the focus of a number of studies including Kaufmann et al., 2000.

China’s quarter-century growth spurt after 1979 occurred in an economy in which, until the mid-1990s, most industrial enterprises were owned by local or higher levels of governments. More importantly, even the China of the early 2000s, when the private sector had come to play a more important role, resembled more the Japan of the 1950s and 1960s with its government-guided industrial policy, including special subsidies and incentives to sectors accorded important roles, than it did laissez-faire. Until its admission to the World Trade Organization (WTO) in 2001, China maintained a dual trade regime, with substantial tariff protection of most domestic industry but also a more liberal regime for imported inputs applying only to export-oriented industries originally restricted to special economic zones and coastal ‘open cities’ (Naughton, 2007). These distinctions were only gradually dismantled during the early WTO years.

Opinion regarding the role of the state was also strongly influenced by the experience of countries in the former Soviet bloc transitioning from state socialism to market capitalism. Whereas initially the emphasis of economic advisors was on ending governments’ involvements in the economies of the countries concerned, it was soon recognized that healthy market economies could not exist without strong institutional supports, including legal protection of property rights, low tolerance for corruption and adequate monitoring of financial market institutions. Furthermore, the advantages exhibited by China since 1978, a little earlier by Taiwan, Korea and Singapore, and more recently by India, were seen to be in part the result of substantial public investments in education, transportation, communications and health. Comparative political stability has also been a major asset of these countries. Although the relatively closed nature of the Chinese and Indian economies before the 1980s has often been viewed as an error that retarded development, it is difficult to prove that those initial closed periods were not of some benefit to their economies, creating a breathing space during which indigenous capacities and skills could be incubated.
Efficient states, old states
With or without agreement on what governments need to do to facilitate development, there is evidence of a general correlation between more efficient and capable government and better economic outcomes. A number of studies have found a correlation between measures of government quality and rates of economic growth. Mauro (1995) found that countries with more corrupt governments had lower rates of investment and economic growth. Evans and Rauch (1999) found that developing countries whose government administrative structures exhibited more of the classic bureaucratic features of meritocratic recruitment and predictable long-term careers achieved higher rates of economic growth even after controlling for initial GDP and human capital. Kaufmann et al. (2000) found several governance measures to be correlated with rates of economic growth.

The fact that countries with relatively capable and stable governments have better growth records than others does not prove that good government is an independent cause of economic growth. The same propitious factors may give rise both to economic growth and to a well-ordered state. Interestingly, one of those factors may be a long history of large-scale political organization. Recent studies suggest that regions that saw earlier formation of kingdoms, states or empires, especially ones not subsequently disrupted by large population shifts due to colonization, enjoyed faster growth between 1960 and 2000 (Bockstette et al., 2002; Chanda and Putterman, 2007). The countries hosting old states also tend to have better scores on commonly used measures of institutional quality. Old states are associated with early transitions to agriculture, and one study suggests that the time of agricultural transition is a strong predictor of current level of development (Hibbs and Olsson, 2004).

Conclusion
Although most economists agree that private enterprises disciplined by market competition are more efficient providers of most goods and services than are government-owned enterprises, there is also general agreement that neither a prosperous national economy nor a sustained process of economic growth are achievable in the absence of a well-functioning state. States are needed to secure property rights, manage currencies and provide the civil order without which commerce and investment become excessively risky propositions. States help to create competitive markets by regulating and breaking up monopolies and promulgating rules for the operation of banking systems and financial markets. States help to address market failures in such areas as environmental quality and workplace safety. And states can pursue macroeconomic policies that contain fluctuations in price and employment levels.
More controversial is whether state guidance or strategic planning is either necessary or desirable for a country to be launched on a path of economic growth. Some governments, especially in Asia, appear to have enjoyed success with government activism in this respect, but there is as yet no consensus about these cases, and misguided interventions can be harmful. Building state capacity while focusing on education, health, transportation, and a legal and tax environment conducive to investment may be tall enough orders for some governments. These tasks, in any case, constitute a starting point that all governments should strive for, and that the people of every country should actively demand of their governments.

References
Any survey of the issues confronting monetary policy in developing countries must first address several basic questions. First, should the central bank target development as one of its objectives, or more broadly, can it contribute to development indirectly, through for instance maintaining financial stability, ensuring international competitiveness or delivering low inflation? Second, is there any scope for, and value to, monetary independence, either because monetary policy is under the thumb of fiscal policy, or because the country concerned is small and open, or because a foreign currency circulates domestically and currency substitution is rampant? Finally, assuming that there is some scope for an independent monetary policy, what should be the operational guide for policy-setting, and is it likely to be different depending on the degree of financial development or other structural features of the economy?

This chapter will be mainly about the third question, namely the way monetary policy should be set for the short to medium run, and the institutions and policy regimes that support and implement that policy, while recognizing that developing countries are very diverse and that ‘one size does not fit all’. In discussing the issue, the position will be taken that monetary policy cannot be separated from exchange rate policy – the two have to be considered together. This is not to say that there may not be instruments that under some conditions have differential effects on domestic monetary conditions and the exchange rate (open-market operations and sterilized intervention, for instance). However, for many countries – in particular the ‘emerging economies’ with access to world capital markets and few controls on capital flows – the scope for differential effects is small (for instance, because of constraints on the cost and effectiveness of sterilized intervention). Over time, the extent of capital controls has declined, as their effectiveness has been blunted and their distorting effects on economic decisions have become manifest. Thus, countries are ill advised to use monetary policy to target domestic objectives while targeting a fixed value for the nominal or real exchange rate.

As for the first question, the position taken here is that development should not be an explicit goal of monetary policy. Development is a longer-term, structural issue relating (primarily) to the real economy while monetary neutrality precludes long-run effects of the money supply on the level
of real output. Forty years ago, this position would not command a consensus; indeed many then advocated deliberate use of monetary policy to keep interest rates low (below world real rates), to channel credit to particular sectors, to undervalue the real exchange rate in order to stimulate exports, or to redistribute income (via inflation) to those with higher saving rates. However, a policy of deliberately keeping interest rates low to stimulate investment is likely in all but financially repressed economies to produce accelerating inflation rather than higher growth, and inflation quickly gets anticipated, blunting any favorable real effects. Following McKinnon (1973) and Shaw (1973), the pervasive inefficiency of financially repressed economies has been recognized. Long-run monetary neutrality does not preclude monetary policy from contributing to short-run stabilization, or the long-run level of inflation from having real effects, but the time horizons for the effects on development are so long and complicated that they make targeting development an inadequate guide for monetary policy setting. This leaves open which of the various possible intermediate targets for monetary policy (the exchange rate, the rate of inflation, nominal income or external competitiveness, among others) is most appropriate and will most contribute to development and welfare. This is issue three above.

Turning to the second question, for a discussion of monetary policy to be interesting, it must be assumed that there is some scope for monetary independence. At the same time, it will be noted below that in some developing economies the limits on that independence are tight indeed, and this may influence the choice of operational targets for monetary policy. In that respect, the size of countries and their structural characteristics matter in considering their appropriate monetary policy. Thus, monetary policy questions for developing countries are somewhat distinct from those for developed countries, for which there is an extensive literature. There are reasons to expect that countries with higher per capita gross domestic product (GDP), more developed financial sectors and stronger institutions face different policy trade-offs. In what follows we shall pay considerable attention to a policy regime – inflation targeting (IT) – that is increasingly finding favor among both industrial and emerging market economies, while considering whether the two sets of countries differ with regard to IT’s suitability as a monetary policy regime and its implementation.

Interdependence of monetary and exchange rate policies
The tight linkage between monetary and exchange rate policy is most clearly expressed in the doctrine of the ‘impossible trinity’: a country cannot at the same time maintain a fixed exchange rate and an independent monetary policy in a context of perfect capital mobility. Put another way,
a credible peg would not allow a country to run different interest rates from those in the anchor country, provided arbitrage was free to operate and unlimited in quantity. In practice, of course, capital is not perfectly mobile, leaving some scope for monetary policy independence even with an exchange rate target. Moreover, as Frankel (1999) has persuasively argued, countries can trade off some exchange rate fixity for some monetary independence, consistent with adopting an intermediate exchange rate regime.

The constraints on countries operating a pegged exchange rate have increased as capital has become more mobile. Increased capital mobility has occurred for essentially two sets of reasons. Increasing financial sophistication, technological advances and increased financial wealth have vastly increased the volume of capital flows and hence the resources that can be marshaled to speculate against a currency peg. And a growing consensus over the last three or four decades that liberalization increases economic efficiency (subject to the cautions expressed by Diaz-Alejandro, (1985), which were confirmed by the emerging market crises of the 1990s) has meant that few countries attempt to maintain tight government regulation of capital flows.

In contrast, during the early post-World War II Bretton Woods period of fixed but adjustable exchange rates there were pervasive controls on capital flows as well as extensive domestic financial regulation, or even ‘financial repression’. In this context, monetary conditions could be set with little concern for external consequences and interest rates were kept low to stimulate investment. Some envisaged the deliberate use of inflation to raise the relative price of capital goods, lower real wages, shift income to those with higher saving propensities (that is, the rich), and call forth increased output for development. For instance, a respected text on development, Higgins (1959), discusses this strategy, advocated by Martin Bronfenbrenner; Higgins accepts the argument in principle, but argues that the optimal rate of inflation is likely to be less than 5 percent per year. Curiously, the incompatibility with the prevailing system of fixed exchange rates is not mentioned, nor does ‘exchange rate’ appear at all in the index of this comprehensive text. At most, there is a reference to the fact that inflation would make export industries increasingly unprofitable, aggravating balance-of-payments difficulties (Higgins, 1959, p. 464). The context was one where constraints on monetary policy came not from the fledgling capital account but rather through the competitiveness of the trade account.

**Targets for monetary policy**

Corden (2002) has provided a useful taxonomy of what he calls ‘approaches’ for exchange rate policy, which for the reasons above we will
adapt for our discussion of monetary policy: the ‘real targets’ approach, the ‘nominal anchor’ approach, and the ‘exchange rate stability’ approach. He distinguishes these approaches from what he considers to be the three (not two) polar regimes: an absolutely fixed exchange rate, pure floating, and the fixed but adjustable exchange rate regime (FBAR), which was the regime under the Bretton Woods period. Thus, his classification of regimes neatly sidesteps the bipolar classification of proponents of a ‘hollowing-out’ of intermediate regimes. But Corden maintains that the FBAR is not simply a compromise between the other two regimes. Because of its credibility problems – it involves an explicit or implicit commitment to a peg, without a corresponding assignment of the instruments needed to ensure maintenance of the peg – it differs from other intermediate regimes which do not promise so much, such as managed floating or crawling pegs or bands.

The ‘real targets’ approach presupposes nominal wage rigidity, so that in the short run, if not in the long run, expansionary monetary policy (or nominal exchange rate depreciation) can affect real output, employment and the real exchange rate. The value of using monetary policy in this fashion is greatest when the economy faces negative real shocks and fiscal policy is prevented from operating in a stabilizing way. However, the assignment of monetary policy to real targets suffers the disadvantage that it does not provide an anchor for the price level or the rate of inflation. Moreover, it relies on a degree of money illusion that is endogenous, and is likely to shrink drastically if monetary policy systematically tries to exploit it.

Hence the increasing emphasis, among central banks of the world, on ‘nominal anchors’ for monetary policy – a rigid link or a target for a nominal quantity or price which is intended to prevent the economy’s overall price level or rate of inflation from wandering off. Nominal anchors can be divided into domestic variables (a monetary aggregate, nominal GDP or the rate of inflation itself), and external anchors, in particular a peg to another currency or to a world commodity (such as gold). In principle, either a domestic or foreign nominal variable can anchor the price level and produce long-run nominal stability, but different targets can yield different short-run outcomes for output and inflation. They also differ in terms of institutional requirements and expose the economy to a crisis in varying degrees.

The ‘exchange rate stability’ approach is to be distinguished from the desire to put in place a nominal anchor: it postulates that exchange rates left to themselves simply add noise to the world economy, perhaps because of self-fulfilling expectations and destabilizing speculation (Williamson, 2000). Though in some cases exchange rate flexibility could facilitate adjustment, movements in exchange rates are dominated by short-run volatility unrelated to economic fundamentals and by medium-term misalignments. In this
view, a system of credibly fixed rates would clearly improve welfare compared to exchange rate flexibility. Williamson has argued that an intermediate regime with an explicit exchange rate target (for example, a band-basket-crawl – or BBC – regime) would have some of the same advantages of anchoring expectations and taming volatility.

**The exchange rate as a nominal anchor**

Exchange rate-based stabilization (ERBS) has been used, with varying degrees of success, in reducing a high initial rate of inflation. The advantage of the exchange rate as nominal anchor is that it is visible, easily explained to the public and requires little institutional credibility – hence its attraction for countries suffering from high or hyperinflation, as Chile in 1979, Argentina in 1991 and Brazil in 1994. In these and other similar cases, the only way to achieve a modicum of monetary policy credibility is to tie the hands of the central bank, since unlike the inflation rate or a monetary aggregate, the exchange rate is a variable which the public observes directly and continuously. However, the principal drawback of this strategy is that it requires an eventual exit from the peg unless the economy is to undergo a severe deflation to remove loss of competitiveness resulting from the accumulated inflation. Indeed, because of the stickiness of inflation, bringing it down to industrial-country levels in the space of even a few years still leaves embodied in the price level the integral of the inflation gaps incurred during that time. Eliminating them would require an extended period of deflation, involving output losses that few governments would willingly incur. Therefore, pegs associated with ERBS become increasingly non-credible over time, even in countries which are successful in achieving low inflation.

The problem is compounded in the presence of sufficient capital mobility that investors can take positions against the currency large enough to exhaust the authorities’ foreign exchange reserves. Thus, a speculative attack could force the authorities to devalue or float. Paradoxically, then, the attempt to gain credibility by using an external anchor sows the seeds of its own downfall. The trick is to ride the wave long enough to benefit from initial credibility gains without getting locked into a strategy that will eventually throw you up on the beach. Unfortunately, while the strategy is successful, there is little pressure on politicians to change, while when the peg is under attack, it is too late: exits in a crisis usually have dire consequences (Eichengreen et al., 1999).

There is a category of countries, however, for which monetary policy independence is of little use and hence a hard peg is credible. These are small, open countries with a high export concentration on a single commodity or service priced in an international currency, or a dominant trade
partner. In particular, countries in the Caribbean which rely heavily on
tourism and banking services have long-standing and credible pegs to the
US dollar. Other countries in this category are the small neighbors of South
Africa – Lesotho, Namibia and Swaziland – which participate in the
Common Monetary Area. This arrangement allows the countries to have
their own currencies, exchangeable at par with the rand. Nepal is an
example in Asia; that country pegs its currency to the Indian rupee. Finally,
some countries – among which are Ecuador and Panama – have simply
adopted a foreign currency, that is, have a regime of official dollarization.

It is relevant to examine the empirical determinants of changes in regimes.
In Masson (2001) and Masson and Ruge-Murcia (2005), exchange rate
regimes are divided into fixed, intermediate and flexible. Using data from
1975–97 for as many as 168 countries, the probability of changing regimes
was related to macroeconomic determinants, foreign exchange reserves
divided by GDP, and trade openness. The probability of abandoning any of
the regimes was greater, the higher the rate of inflation and the lower real
GDP growth (Masson and Ruge-Murcia, 2005). The intuition is clear:
regimes are abandoned in bad economic times, not good ones. For instance,
high inflation makes a peg increasingly precarious, but also makes it more
likely that countries that are floating will use ERBS in order to reduce it.

Analysis of exchange rate regime transitions also permits testing for-
mally the ‘hollowing-out’ hypothesis (Eichengreen, 1994). Hollowing-out
requires transitions away from intermediate regimes, but not towards them
from the poles of hard fixes and free floats. Using a constant transition
matrix, that hypothesis can be rejected (Masson, 2001). Moreover, the exis-
tence of continuing negative shocks that produce high inflation and slow
growth suggests that there will be continued cycling among regimes when
the transition probabilities are endogenous, as described above. The idea
that hard fixes such as currency boards are immune from crisis was deci-
sively proved wrong by Argentina’s abandonment of its Convertibility Law
in January 2002, and the floating of the peso. Despite having a credible
ERBS based on institutional guarantees, the severe recession suffered by
Argentina during 1998–2001 made maintaining the exchange rate strait-
jacket difficult. Argentina also illustrates the need for support from fiscal
policy to make any monetary regime successful – but especially so for a
fixed rate regime. Earlier fiscal adjustment would have allowed Argentina
to avoid a debt crisis and would have helped maintain international com-
petitiveness and current account balance.

**Domestic nominal anchors**
The principal choice of domestic nominal anchor is between a monetary
aggregate and inflation targeting. Monetary aggregates have the advantage
of being relatively easy to measure and of being a financial variable that, at least in the era of highly regulated financial systems, was relatively easy to control.³ It was postulated that there was a stable relationship between those financial and real variables, principally taking the form of a stable (and simple) money demand equation.

However, there is accumulating evidence that money demand is unstable. Greater access to other financial assets associated with liberalization has changed its nature, and probably increased the interest elasticity of demand for non-interest-bearing deposits and cash. Technological changes have also allowed greater opportunities for conserving on transactions balances. Finally, partial dollarization – the circulation of a foreign currency – provides another source of instability in the demand for the domestic currency. As a result of all these factors, targets for monetary aggregates, abandoned by almost all industrial countries, are now also increasingly being abandoned by developing countries.

Emerging-market countries have in a number of cases adopted inflation targeting. Chile was the precursor, announcing a target for consumer price index (CPI) inflation in 1991, albeit accompanied initially by an exchange rate target band. Other developing countries having forms of inflation targeting regimes include Brazil, Colombia, the Czech Republic, Israel, Korea, Mexico, Peru, Poland, South Africa and Thailand (see Mishkin and Schmidt-Hebbel, 2002, for details).

**IT in developing countries: prerequisites and experience**

Inflation targeting as a monetary policy regime has to be distinguished from a situation in which the central bank merely expresses a desire to lower the rate of inflation to a particular level or maintain it there. Price stability in some form is always part of a central bank’s mandate, but IT aims to enhance the credibility of the central bank’s commitment to price stability by improving its accountability. Announcing targets which are not met because they have no effect on policy or because they are over-ridden by other objectives does nothing to improve the credibility or effectiveness of monetary policy.

Two basic prerequisites for putting in place an IT regime are that the central bank, which is charged with implementing monetary policy, be given a reasonable degree of ‘instrument’ independence to carry out that task; and the absence of commitment to a target for another nominal variable (Masson et al., 1997). Hybrid regimes are of course possible (see below) in which countries have targets for both the rate of inflation and the money supply or the exchange rate; in practice, several developing countries have operated such a regime for a transitional period. In this sense, the advocates of IT who deride a ‘prerequisite approach’ to IT (for example,
Sterne, 2002) have a point: one can start doing ‘baby steps’ as a way of learning how to walk. However, from the standpoint of clarity it is important to be clear what the ultimate objective should be, so as to put in place the capabilities needed to achieve it.

Subject to the above two prerequisites, the IT regime needs to involve the following elements of a framework for monetary policy: quantitative targets for the rate of inflation, over a specified horizon; a commitment to those targets as overriding objectives for policy; a clear methodology for making inflation forecasts; and a transparent way of translating the possible expected deviations from target into changes in the instruments of monetary policy.

How successful are developing countries in meeting the prerequisites? First, IT is likely to be a candidate regime primarily for middle-sized or large middle-income countries – roughly speaking, the ‘emerging market economies’. Smaller, very open economies may well choose to peg – a credible monetary regime for them; and countries with low incomes would typically not have the financial development or institutional capacity to implement inflation targeting. Second, countries differ greatly as to the degree of central bank independence. In particular, where there is fiscal dominance, an independent monetary policy is impossible. Compared to industrial countries, seigniorage – used to finance fiscal deficits – in many developing countries is high. Third, de jure independence may not guarantee that the central bank is able to carry out its mandate in the face of lack of public support. A constituency in favor of low inflation has not developed in many emerging market economies. And de jure independence has sometimes been overridden, as in Argentina, where a central bank governor was summarily replaced for disagreeing with the Minister of Finance. Fourth, administered price changes and centralized negotiations (for example the Pacto in Mexico) that determine a large fraction of the economy’s annual rate of wage increase may interfere with the central bank’s ability to control inflation unless they are coordinated with the inflation target. Finally, forecasting inflation is difficult in many emerging-market countries because of highly unstable macroeconomies and lack of solid econometric relationships (due, for example, to an insufficiently long or homogeneous data sample).

These obstacles are not insurmountable, however. The experience of emerging-market economies that practice IT is generally favorable, since countries have typically met their inflation targets. Moreover, the inflation-targeting countries have had a better experience of avoiding the balance-of-payments crises than other emerging-market countries since 1994. Thus, taking as given that emerging-market countries face a more challenging environment for monetary policy, this does not imply that the choice of
regime should necessarily be biased away from IT. Instead, Fraga et al. (2003) suggest that IT be operated somewhat differently than for industrial countries, perhaps by using wider and ‘softer’ bands around the inflation target, higher targets so as to accommodate bigger shocks without risking deflation, and adjusting targets more flexibly in response to shocks (while making special efforts to communicate the reasons to the general public).

Another feature of the early years of inflation targeting in developing countries has been its association with exchange rate targets. Thus, Chile, Israel and Poland, for instance, for a time targeted a band for the exchange rate (crawling, with the band width also adjusted depending on circumstances). Such a policy had the advantage of easing into the IT framework, allowing experience with it to be built up while retaining a backstop which might prevent instability should the inflation target give the wrong signal. The disadvantage of combining the two targets, as described in Israel’s case by Bufman and Leiderman (2000), was lack of transparency and the danger of conflicting signals which could add to the public’s uncertainty about monetary policy. In practice, given the greater immediacy of the exchange rate variable (continuously observable and widely publicized), it tended to dominate the inflation target if they conflicted, and this then required an explicit change in the exchange rate band if inflation was to be given its proper weight. Over time, the bands for the exchange rate were widened, as inflation declined and confidence with the new regime increased. Israel now has no exchange rate target. A similar progression occurred in Chile’s case. At the time of the Russian crisis in August 1998 concern about external developments led the central bank to increase the short-term interest rate, causing the economy to go into a severe recession (Morandé, 2002). Sole emphasis on the inflation target would have allowed some easing of monetary policy. The exchange rate target was abandoned in September 1999, and the Bank of Chile at the time of writing in 2006 sets its monetary policy to maintain inflation within a 2–4 percent target range.

Unresolved questions concerning inflation targeting
While the verdict so far in emerging-market countries is positive, the track record is quite short. For most countries, it dates from the end of the 1990s. The environment has been relatively benign, with low inflation and low interest rates prevailing in the industrial world; this has permitted all developing countries, whether inflation targeters or not, to reduce inflation. Moreover, this period has not seen any major contagion from emerging-market currency or balance-of-payments crises. It remains to be seen whether a major world inflation shock would be weathered well by the IT regimes in place. As noted above, monetary and exchange rate
regimes have been periodically adopted and then abandoned – will the same be true of IT? Since IT is a less rigid regime – often characterized as ‘constrained discretion’ – it is likely to be less fragile than strict exchange rate targets. The danger remains however that the credibility gains that have accrued to IT central banks may be dissipated by persistent overshoots or indications that inflation does not provide an effective guide for month-to-month policy-setting. Thus, IT will face several challenges going forward.

An important question that should condition views of the advantages of IT is whether in fact it is subject to speculative attack. Kumhoff (2002) argues that the regime can be attacked, and that moreover it behaves much more like a fixed exchange rate regime than is often claimed. In the face of an unsustainable fiscal policy, it might have to be abandoned. Favero and Giavazzi (2004) provide a formal model in which such a fiscal policy could lead, through increasing default risk, to severe constraints on monetary policy’s ability to deliver on its inflation commitment. Contrary to the analysis of Fraga et al. (2003), they conclude that for at least a short period in 2002, Brazil’s economy might have tipped into a regime of fiscal dominance that, had it continued, would have doomed the IT regime. Thus, they are not convinced that the regime was ‘stress tested’.

The continued spread and popularity of IT may be affected by a parallel trend toward monetary unions, following in the wake of the successful creation of the Eurozone. Already, European integration has eliminated two inflation targeters, Finland and Spain, through their joining the Eurozone; and several among the new EU members, in particular the Czech Republic, Hungary and Poland, are expected to join the Eurozone within a few years. Sweden and the United Kingdom, prominent inflation targeters, could also conceivably join. While reducing the number of central banks implementing inflation targeting, of course should the European Central Bank (ECB) clearly adopt that regime, then the economic area it applied to might increase.

Turning to other continents, Africa, the Middle East and Asia are also considering regional monetary integration that might lead to a common currency. In Masson and Pattillo (2004) the argument is made in the context of Africa that the European example does not translate well to other regions with less broad-based integration projects and less strong regional solidarity. Thus, the success of an African single currency seems doubtful; instead, greater monetary integration could arise around regional poles such as South Africa, which already has a successful IT monetary regime. Since the Eurozone, the United States and Japan are likely to want to retain their monetary independence and exchange rate flexibility, regional currency blocs, if they are created, are unlikely to anchor their
exchange rate to any single reserve currency. On balance, then, inflation targeting among emerging market (EM) countries is likely to remain the regime of choice, whether or not regional integration proceeds.

When considering implementation of inflation targeting, a still unresolved issue is the weight to give to other variables, in addition to the inflation forecast. While of course other variables may influence the inflation forecast (for instance, the current output gap or the actual exchange rate) it seems that most countries give some additional weight to those variables – for instance, lowering interest rates if activity is weak, provided the inflation forecast is within the target range. Thus, the influence of other variables on policy may be asymmetric, but non-zero; this is consistent with Mervyn King’s dictum that central banks are not ‘inflation nutters’. Developing countries in particular are loath to ignore the exchange rate, and as Calvo and Reinhart (2002) show, exhibit a ‘fear of floating’. While intervention may smooth some exchange rate fluctuations, central banks may also want to use interest rate policy for that purpose – subject to the caveats evoked by Chile’s experience in 1998.

Finally, should the target for the (long-run) inflation rate be higher in emerging markets, as suggested by Fraga et al. (2003), because of larger shocks facing those countries? More empirical evidence is needed to resolve this issue. While the Balassa–Samuelson effect would suggest higher average inflation in faster-growing developing countries (provided one wants to avoid falling tradable goods prices in domestic currency), the argument for avoiding deflation really requires a greater understanding of the nature of downward rigidities. And the cross-country literature on costs of inflation is not very precise on when inflation becomes costly. These remain important issues for monetary policy in developing countries, whether they have adopted IT or not.

Notes
1. I am grateful to Max Corden, Frederic Mishkin and Miguel Savastano for comments on this chapter.
2. Even long-run neutrality does not command unanimity, however, and recent models suggest that transitory effects may last longer than previously thought (for example, Mankiw and Reis, 2001).
3. Programs supported by the International Monetary Fund typically included targets for the central bank’s net domestic assets.
4. However, Fraga et al. (2003) consider that Brazil ‘stress-tested’ its IT regime during 2002, when it faced a negative capital account swing of about 6 percent of GDP, missing its inflation target but not suffering a permanent loss of credibility for the regime.

References
Frankel, Jeffrey (1999), ‘No Single Currency Regime is Right for All Countries Or At All Times’, Essays in International Finance, No. 215, Princeton University.
Introduction
Fiscal policy plays an increasingly important role in many developing countries. Decisions on fiscal policy, especially if properly synchronized with monetary policy, can help smooth business cycles, ensure adequate public investment and redistribute incomes.

The four main components of fiscal policy are: (1) expenditure, budget reform; (2) revenue (particularly tax revenue) mobilization; (3) deficit containment and financing; and (4) determining fiscal transfers from higher to lower levels of government. Fiscal policy works through both aggregate demand and aggregate supply channels. Changes in total taxes and public expenditure affect the level of aggregate demand, whereas the structures of taxation and public expenditure affect, among others, the incentives to save and invest (at home and abroad), take risks, and export and import goods and services.

This chapter is organized as follows. It first analyses some basic fiscal issues confronting developing countries. Then it considers budgetary deficits and evaluates norms for tax and expenditure reforms. The final section concludes.

Tax and expenditure profiles of developing countries
Developing-country fiscal systems have three basic characteristics: (1) low tax–GDP and expenditure–GDP ratios compared to developed countries, even though developing countries need more public expenditure; (2) the fiscal stance is often procyclical; and (3) tax resources are more volatile than those of developed countries.

There are manifold pressures for high and growing government expenditure in developing countries. Because of their low per capita incomes and high poverty, developing countries face an urgency to raise growth rates. This places a heavy burden on policy whereas, concurrently, the limited efficacy of policy instruments and governance inadequacies constrain policy. Pressures for populism through price controls and the like are considerable. The state in many developing countries is politically weak and beset with lack of consensus on what constitutes a sound fiscal policy compared to most developed countries (Heady, 2004).
Table 55.1 shows government tax revenues in developed, transition and developing countries for two time periods, 1990 to 1995, and 1996 to 2002 (domestic prices, number of countries and median values of the simple averages)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Countries</td>
<td>Median</td>
<td>Countries</td>
<td>Median</td>
</tr>
<tr>
<td>Separate samples</td>
<td>56</td>
<td>31.9</td>
<td>53</td>
<td>33.8</td>
</tr>
<tr>
<td>Combined sample</td>
<td>48</td>
<td>33.5</td>
<td>48</td>
<td>33.8</td>
</tr>
<tr>
<td>Developed countries</td>
<td>21</td>
<td>37.8</td>
<td>21</td>
<td>40.1</td>
</tr>
<tr>
<td>Transitional economies</td>
<td>14</td>
<td>34.7</td>
<td>14</td>
<td>31.4</td>
</tr>
<tr>
<td>Developing countries</td>
<td>13</td>
<td>18.7</td>
<td>13</td>
<td>19.2</td>
</tr>
</tbody>
</table>

Note: Tax revenue is computed as the sum of revenue accruing to central and local governments. For each country in each category average tax revenue as a percentage of GDP is computed. For each category of country the median value of this average tax revenue is reported in the chapter.

Source: UNPAN Statistics.

Table 55.1 shows government tax revenues in developed, transition and developing countries for two time periods, 1990 to 1995, and 1996 to 2002. In the median developing country the tax–GDP ratio was below 20 per cent whereas in the median transition economy it was 30 per cent and in developed countries 40 per cent. Unsurprisingly, on average, governments in many developing countries face a severe resource crunch.

Only three out of 21 developed countries and two out of 14 transition countries had revenues falling over the two time periods, whereas the corresponding magnitude for developing countries was five out of 13. The share of distortionary commodity and trade taxes in total central government revenue is higher in developing countries (Tables 55.2 and 55.3).

Table 55.3 shows that in the richest countries personal income taxes are the most significant and contribute more than half (54.3 per cent) of tax revenue. Next are various commodity taxes and then the corporation tax. Border taxes and seigniorage revenue are low, reflected in the low value for inflation. The informal sector is small. With falling gross domestic product (GDP) per capita tax revenue as a percentage of GDP drops and corporate taxes as a percentage of revenue rise. Income taxes remain at about 30 to 35 per cent of revenue and commodity taxes are high. High rates of inflation reflect high values of seigniorage revenue. The poorest among the developing countries raise almost a quarter of their revenue through seigniorage. The informal sector in developing countries is about twice that in developed countries.
Auriol and Warlters (2005) argue that the informal sector in developing countries is large because of the higher costs of entry into the formal economy. By keeping barriers to entry into the formal economy high, those firms and individuals who make it into the formal economy acquire large rents and hence may be easier to tax than a diffused set of small taxpayers. If this argument is correct then encouraging large formal sectors should be part of a government strategy to increase tax revenue. Data for 64 countries indicate that this is indeed the case, particularly in Africa, for example, 0.4 per cent of taxpayers account for 61 per cent of total domestic tax collection in Kenya and 57 per cent in Colombia. As general policy, Auriol and Warlters argue that developing countries should lower entry barriers and raise the size of the formal sector to raise tax revenues.

Further, rapid globalization, technological advancement and the accompanying movement of factors of production across national boundaries, and the emergence of multinational corporations as major actors have eroded many developing countries’ tax bases. Taxpayers can more easily raise income outside of conventional channels (Lao-Araya, 2003).

With inflexible public expenditures and low tax revenues government finances in developing countries are weak, with high deficits, debts and debt-servicing obligations. Consolidated figures for the finances of central and local governments together are not readily available but Table 55.4 presents these for central governments. Typically, developing countries’ revenues and expenses are lower and interest payments higher than in developed countries, although government consumption in developing countries is lower than that in developed countries (Table 55.5).

---

**Table 55.2 Central government revenue by type of tax, 1990–2002 averages (% of total tax revenue, median value of simple averages)**

<table>
<thead>
<tr>
<th>No. of countries</th>
<th>Direct taxes</th>
<th>Payroll taxes</th>
<th>Sales taxes</th>
<th>Trade taxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sample</td>
<td>139</td>
<td>27.1</td>
<td>5.9</td>
<td>34.8</td>
</tr>
<tr>
<td>Developed countries</td>
<td>24</td>
<td>34.8</td>
<td>28.5</td>
<td>28.1</td>
</tr>
<tr>
<td>Transitional economies</td>
<td>23</td>
<td>17.7</td>
<td>33.3</td>
<td>38.9</td>
</tr>
<tr>
<td>Developing countries</td>
<td>92</td>
<td>27.6</td>
<td>0.7</td>
<td>33.0</td>
</tr>
<tr>
<td>Africa</td>
<td>32</td>
<td>27.2</td>
<td>0.2</td>
<td>30.7</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>27</td>
<td>22.1</td>
<td>5.1</td>
<td>38.9</td>
</tr>
<tr>
<td>Asia &amp; Oceania</td>
<td>33</td>
<td>34.1</td>
<td>0.0</td>
<td>34.8</td>
</tr>
</tbody>
</table>

*Source: UNPAN Statistics.*
Table 55.3  Sources of Government Revenue (1996–2001)

<table>
<thead>
<tr>
<th>GDP per capita</th>
<th>Tax revenue (% of GDP)</th>
<th>Income taxes (% of revenue)</th>
<th>Corporate income tax (% of income taxes)</th>
<th>Consumption &amp; Production taxes (% of revenue)</th>
<th>Border taxes (% of revenue)</th>
<th>Inflation rate (%)</th>
<th>Seigniorage Income (% of revenue)</th>
<th>Informal economy (% of GDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;$745</td>
<td>14.1</td>
<td>35.9</td>
<td>53.7</td>
<td>43.5</td>
<td>16.4</td>
<td>10.6</td>
<td>21.8</td>
<td>26.4</td>
</tr>
<tr>
<td>$746–2975</td>
<td>16.7</td>
<td>31.5</td>
<td>49.1</td>
<td>51.8</td>
<td>9.3</td>
<td>15.7</td>
<td>24.9</td>
<td>29.5</td>
</tr>
<tr>
<td>$2976–$9205</td>
<td>20.2</td>
<td>29.4</td>
<td>30.3</td>
<td>53.1</td>
<td>5.4</td>
<td>7.4</td>
<td>6.0</td>
<td>32.5</td>
</tr>
<tr>
<td>All developing</td>
<td>17.6</td>
<td>31.2</td>
<td>42.3</td>
<td>51.2</td>
<td>8.6</td>
<td>11.8</td>
<td>16.3</td>
<td>30.1</td>
</tr>
<tr>
<td>&gt;$9206</td>
<td>25.0</td>
<td>54.3</td>
<td>17.1</td>
<td>32.9</td>
<td>0.7</td>
<td>2.2</td>
<td>1.7</td>
<td>14.0</td>
</tr>
</tbody>
</table>

Source: Gordon and Li (2005).
<table>
<thead>
<tr>
<th>Country group</th>
<th>Revenue (% of GDP)</th>
<th>Expense (% of GDP)</th>
<th>Cash surplus or deficit (% of GDP)</th>
<th>Net incurrence of liabilities (% of GDP)</th>
<th>Total payment as % of GDP</th>
<th>Interest payment as % of revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.6</td>
<td>-3.2</td>
</tr>
<tr>
<td>Middle income</td>
<td>17.3</td>
<td></td>
<td>0.8</td>
<td>1.1</td>
<td>0.9</td>
<td>1.2</td>
</tr>
<tr>
<td>Lower middle income</td>
<td>16.7</td>
<td></td>
<td></td>
<td></td>
<td>2.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Upper middle income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Asia &amp; Pacific</td>
<td>8.4</td>
<td>11.5</td>
<td>12.0</td>
<td>-2.1</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Europe &amp; Central Asia</td>
<td>31.0</td>
<td>31.1</td>
<td></td>
<td>-1.2</td>
<td>0.9</td>
<td>0.4</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>20.9</td>
<td>23.0</td>
<td>-0.4</td>
<td></td>
<td>1.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Middle East &amp; North Africa</td>
<td>28.3</td>
<td>23.5</td>
<td></td>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asia</td>
<td>13.2</td>
<td>12.4</td>
<td>15.4</td>
<td>15.1</td>
<td>-2.7</td>
<td>-3.1</td>
</tr>
<tr>
<td>High Income</td>
<td>26.0</td>
<td></td>
<td>28.9</td>
<td></td>
<td>-2.8</td>
<td></td>
</tr>
<tr>
<td>Europe EMU</td>
<td>36.3</td>
<td>35.7</td>
<td>38.8</td>
<td>38.6</td>
<td>-2.3</td>
<td>-2.3</td>
</tr>
</tbody>
</table>

Jha (2006) reports that the unweighted average of tax buoyancy (defined as \( \frac{\text{Percentage change in tax revenue}}{\text{Percentage change in tax base}} \)) for several developing countries is larger than one, indicating that an expansion of income would lead to an increase in the tax–GDP ratio. Gordon and Li (2005) argue that taxation, by its very nature, must depend on the formal economy since bank records are needed to identify taxable activity. In rich countries the intermediary services provided by the financial sector are considerable, so there is a high cost of abandoning it and conducting business in the informal sector. However, this is not the case in developing countries. Further, their tax base is likely to be narrow (biased towards capital income) and cover mostly capital-intensive firms that need the financial sector the most, and tariffs are used to protect the capital-intensive sectors and shortfalls in revenue (from public expenditures) would often be met through seigniorage.

Another important characteristic of fiscal variables in developing countries is their instability. Table 55.6 reports on key fiscal variables in 13 Latin American developing countries and 14 industrialized countries. In terms of all categories and in both nominal and real terms, computed coefficients of variation are much higher for Latin American developing countries than for industrialized countries.

Fiscal variables in many developing countries move in a procyclical fashion. Standard Keynesian models require that fiscal policy should be countercyclical, that is, during recessions taxes should be lowered and public expenditure hiked whereas during good times, taxes are raised and public expenditures lessened to reduce chances of overheating of the economy.

### Table 55.5  Government consumption as percentages of GDP, 1990, 1996, 2002 (domestic prices, median values)

<table>
<thead>
<tr>
<th></th>
<th>Number of Countries</th>
<th>1990</th>
<th>1996</th>
<th>2002¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sample</td>
<td>114</td>
<td>15.3</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>less transitional economies</td>
<td>101</td>
<td>5.2</td>
<td>14.3</td>
<td>15.6</td>
</tr>
<tr>
<td>Developed countries</td>
<td>24</td>
<td>18.9</td>
<td>19.4</td>
<td>19.1</td>
</tr>
<tr>
<td>Transitional economies</td>
<td>13</td>
<td>20.0</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td>Developing countries</td>
<td>77</td>
<td>14.2</td>
<td>12.7</td>
<td>14.0</td>
</tr>
<tr>
<td>Africa</td>
<td>26</td>
<td>15.1</td>
<td>12.8</td>
<td>14.7</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>25</td>
<td>12.9</td>
<td>13.4</td>
<td>14.6</td>
</tr>
<tr>
<td>Asia &amp; Oceania</td>
<td>26</td>
<td>12.2</td>
<td>11.7</td>
<td>13.0</td>
</tr>
</tbody>
</table>

**Note:** 1. Or latest data (2000, 2001).

**Source:** UNPAN Statistics.
In contrast, the ‘Ricardian equivalence’ hypothesis suggested by Barro (1979) suggests that since rational economic agents make decisions based on perfectly anticipated tax and expenditure policies of the government, fiscal policy should remain neutral over the business cycle and respond only to unanticipated changes that affect the government’s budget constraint.

Using a sample of 56 countries (20 developed and 36 developing) Talvi and Vegh (2005) show that in G7 countries fiscal policy follows Barro, whereas for developing countries it has been procyclical. Two plausible explanations for this phenomenon exist. The first is that tax bases are so narrow and public expenditure so inelastic in developing countries that tax revenues and expenditures rise during expansions, whereas during recessions revenues and expenditures both decline for similar reasons. Second, as Talvi and Vegh (2005) argue, since fluctuations in the tax base are much larger in developing countries than in developed countries, full tax smoothing would require large surpluses during good times which is not possible since public expenditures are inelastic and resources may be wasted in enhanced public expenditures on public sector undertakings and subsidies, instead of retiring of debt as full tax smoothing would require.

---

### Table 55.6 Coefficients of variation of key fiscal variables

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industrialized Countries</td>
<td>Latin America</td>
</tr>
<tr>
<td>Total revenue</td>
<td>0.15</td>
<td>0.55</td>
</tr>
<tr>
<td>Current revenue</td>
<td>0.15</td>
<td>0.56</td>
</tr>
<tr>
<td>Non-tax revenue</td>
<td>0.19</td>
<td>0.58</td>
</tr>
<tr>
<td>Tax revenue</td>
<td>0.15</td>
<td>0.56</td>
</tr>
<tr>
<td>Total expenditure</td>
<td>0.16</td>
<td>0.55</td>
</tr>
<tr>
<td>Current expenditure</td>
<td>0.16</td>
<td>0.55</td>
</tr>
<tr>
<td>Government consumption</td>
<td>0.15</td>
<td>0.54</td>
</tr>
<tr>
<td>Interest payment</td>
<td>0.22</td>
<td>0.63</td>
</tr>
<tr>
<td>Transfers</td>
<td>0.17</td>
<td>0.58</td>
</tr>
<tr>
<td>Capital formation</td>
<td>0.17</td>
<td>0.57</td>
</tr>
</tbody>
</table>

**Notes:**
1. There are 13 Latin American and 14 industrialized countries.
2. In the sample the Gavin and Perotti (1997) database is used for the analysis.

**Source:** Bertin-Levecq (2000).
Fiscal deficit issues
The exercise of fiscal policy in developing countries has its limits. The combination of low revenues and inelastic expenditures means that expenditures routinely, and even increasingly, outpace revenues. Jha (2004) argues that there is considerable heterogeneity in experience with respect to the fiscal deficit, between the middle- and low-income country categories and even within the low-income category countries. Indeed, the poorest among the least-developed countries are caught in an insidious resource trap and the average least-developed country economy has, since the 1970s, been exposed to adverse external trade shocks with an impact, in the worst years, approximately double the average of other developing countries (UNCTAD, 2000).

External finance is limited, especially for the poorest countries, although large, stable economies attract considerable capital inflows. Official aid has been falling and private equity flows go to the best-performing developing and transition economies. Private loans, as Harberger (1985) notes, are available at increasingly difficult terms since the domestic resource cost (often underestimated) of servicing these increases with additional borrowing. Other reasons for differences across developing countries include continuity and stability of policy regimes: Zambia, with several policy reversals, will be associated with greater risks than Mauritius, which has had a credible and stable policy regime.

Given financing constraints many developing countries have to opt for some non-bond (monetary) financing of the deficit. This establishes a direct link between fiscal policy and the monetary base of the central bank, blurs the distinction between fiscal and monetary policy, and compromises central bank independence.3 If bond financing is chosen, private investment may get crowded out.

Jha (2004), shows that in the long term public revenue and public expenditure are unrelated in many developing countries so that any excess of expenditure over revenue cannot be financed by generating budgetary surpluses over a long enough time horizon. Thus fiscal deficits are unsustainable in many developing countries. Mendoza and Ostry (2007) argue that whereas fiscal policy in most countries is responsive to budgetary deficits, high-debt countries do run a risk of having an unsustainable fiscal stance.4

Jha (2004) also shows that current account deficits are unsustainable in many developing countries. The fact that external sustainability conditions are hard to meet would imply the need for continual capital inflow in order to keep the balance of payments in equilibrium, necessitating the maintenance of a substantial rate-of-return wedge between domestic and foreign rates of return. This raises domestic interest rates substantially above global interest rates and acts as a drag on higher growth, making debt servicing harder, and exacerbates the fiscal deficit.
However, public expenditure could be productive, so whether public deficits impede or spur economic growth becomes an empirical question. In this context Adam and Bevan (2005) examine the relation between fiscal deficits and growth for a panel of 45 developing countries over 1970–99. Public expenditure is permitted to be both growth-enhancing as well as growth-inhibiting and distortionary taxes exist and fiscal deficits are permitted. They show that the impact of the deficit depends upon the mode of financing it. Deficits can be growth-enhancing if financed by limited seigniorage, growth-inhibiting if financed by domestic debt, and have opposite flow and stock effects if financed by external loans at market rates. These opposite effects define a threshold effect, before attaining which fiscal deficit has growth-enhancing effects and after which the effects of fiscal deficits are growth-inhibiting. Adam and Bevan find this threshold figure to be around 1.5 per cent of GDP after grants.

**Norms for tax and expenditure reforms in developing countries**

One of the principal aims of a meaningful tax and expenditure reforms policy would be to bolster the savings and investment rates in the economy in order to raise growth rates. A higher growth rate, it is widely accepted, is the best way to lower poverty over the medium term. Loayza et al. (2000a, 2000b) indicate that the most important determinant of savings, across both developed and developing countries, is the level of per capita income and the rate of economic growth. Thus the higher the rate of savings, the higher the economic growth rate and the higher the growth rate, the higher the rate of savings at least at low absolute levels of per capita income. Their results also point to the possibility of incomplete Ricardian equivalence, that is, a given rise in public savings is accompanied by a less than commensurate drop in private savings.

The gap between the real rate of return on savings and the discount rate is critical. Savers who are liquidity constrained may be more sensitive to such differentials compared to those who are not. As financial deepening takes place and fewer consumers remain liquidity constrained, this responsiveness may drop. However, as consumers become less liquidity constrained they might also become less risk averse and opt for investments with higher returns, boosting the savings rate. Thus the impact of the tax structure on savings is of critical importance, and distorting differences in effective tax rates across sectors and assets and tax-induced distortions that create inefficiencies and lower the potential rate of economic growth should be eliminated. This would be an important component of tax reform, the basic tenets of which are well known and briefly summarized below.

As an economy develops, reliance on indirect taxation for revenue should decline. This is because indirect taxes typically have an excess burden
associated with them (Jha, 1998). Furthermore efficient indirect taxation (one that minimizes excess burden to the representative consumer) can be quite regressive.\(^5\) Indirect taxes can be made redistributive by sacrificing some efficiency, but the extent of this redistribution is limited (Sah, 1983).

If, however, indirect taxes can be levied on final consumption alone, tax-induced changes in relative prices that characterize production taxes such as excise duties could be avoided. Then, if consumer utility functions are weakly separable between consumption and leisure, a uniform tax on final consumption goods (say a value-added tax – VAT) would approximate a lump-sum tax.\(^6\) This tax, with only few exemptions (for items consumed in disproportionately large amounts by the poor), harmonized across levels of government in federal countries and few rates, is recommended. These could be supplemented with excise duties on environmental bads or ‘luxuries’. Peak tariff duties and effective rates of protection should be reduced gradually. If the tax base admits few exemptions and there are fewer rates, costs of compliance and monitoring will fall. But the VAT requires the netting out of input costs and the exemption of exports from the tax base. This, in turn, needs sophisticated account keeping which may be absent in many developing countries. The credibility of the tax regime is also important and tax reforms should aim for a stable tax environment and be well coordinated and, at all times, be simple. Tariff cuts should be accompanied by an upward revision of VAT rates to compensate for tax revenue.

However, Emran and Stiglitz (2005) show that the standard prescription of reducing trade taxes with revenue-compensating upward revision of the rate of a broad-based VAT is welfare-improving only in an economy with no informal sector, with all production and exchange activity in the tax net – conditions typically not satisfied in developing countries.\(^7\) When only the formal sector can be taxed, the introduction of a VAT (or a hike in its rate) may end up creating a distortion between the formal and informal sectors. Even broadening the VAT base to include more of the informal sector may reduce welfare (Piggott and Whalley, 2001). Similarly Bibi and Duclos (2007) show that for indirect tax reform to be poverty-reducing it must: (1) not remove all subsidies; (2) in some cases increase taxes on already taxed commodities and, concurrently, increase subsidies on already subsidized commodities; (3) not exclusively follow efficiency considerations, since redistribution may still play an important role in poverty reduction; and (4) concentrate on reform rather than removal of subsidy. Jha (2006) presents a taxonomy of the extant literature’s view on how to fine-tune the aforementioned tax reforms in order to make the resulting tax structures distributionally sensitive.

Tax structures in developing countries are not particularly progressive. Thus Chu et al. (2004) find that: (1) only 13 of the 36 overall tax systems...
surveyed by them are progressive, seven are proportional, seven are regressive and the rest neutral or insignificant; (2) income taxes were progressive in 12 of the 14 cases studied whereas indirect taxes were broadly regressive. The progressivity of direct taxes declined over time in eight cases. This needs attention.

Another principle of tax reform is that the share of direct taxation in overall tax revenue should rise. Within direct taxation, reliance has to be shifted from corporate to income taxes. Since corporate profits are taxed at the level of personal income anyway, the rationale for separate corporate taxes is rather weak. There are only two arguments in favor of corporate taxes: (1) as a tax on foreigners’ incomes; and (2) as a tax on non-competitive profits. Within the sphere of income taxation, the rate and exemptions structures need to be rationalized. The number of tax brackets should be small, the degree of progression mild with the top marginal tax rate low. Tax reform theory advocates taxation of ‘full income’ the Haig–Simons definition of which is ‘all increases in human and physical capital during a period of time’. One cannot pick and choose the types of income one would like to tax.

Another area of importance for taxation is the conduct of commerce over the internet (e-commerce). Although e-commerce is a nascent industry it should be taxed since it would be inefficient as well as inequitable to tax goods traded through bricks and mortar stores and not tax e-commerce. A commodity that is sold in a bricks and mortar store and, therefore, subject to taxation would be deemed to be different if sold through e-commerce, and escape taxation. Further, those buying through e-commerce are likely to be rich. This exacerbates inequity. There is a rationale for zero customs duties on e-commerce in line with arguments for free trade, but not for zero taxes. A policy of not taxing e-commerce would provide another avenue for tax evasion as some US evidence shows. Further, given its projected phenomenal rate of growth, if e-commerce is not taxed there will be sharp erosion of the tax bases of governments that primarily levy sales taxes.

Another issue is the presence of tax havens. The Organisation for Economic Co-operation and Development (OECD) estimates that during 1985–94 foreign direct investment (FDI) by the G7 countries in some tax havens in the Caribbean and South Pacific increased more than fivefold to more than US$200 billion – an increase well in excess of the growth of total outbound FDI. These concerns extend to transition and developing economies and have probably worsened in recent years. ‘A race to the bottom’ may ensue with national and/or state governments using tax incentives competitively to attract FDI. Such incentives interact dynamically with the existing avenues for tax evasion (for example because some
incomes are not taxed) to reduce current tax revenues and prospects for higher future tax revenues. In the face of this tax reform, particularly direct tax reform, should have a considerable element of international cooperation.

A related issue is service taxation. Services have become the dominant sector in many developing countries but are hard to tax. Not taxing services is inefficient as well as inequitable: inequitable because it discriminates between providers of goods and services; inefficient because it has the potential of creating several distortions, thus increasing non-labour costs.

**Expenditure reform**

Tax reforms should be complemented with appropriate adjustment of government expenditures. Typically this calls for reduction of current subsidies and augmentation of subsidies for well-managed capital projects. The impact of public expenditure is usually ascertained through an *ex post* incidence analysis but we must evaluate not what does exist but what might exist — the theme of benefit incidence analysis. Such analysis is marginal (to capture differences from the status quo) and behavioural (to generate counterfactuals) and is difficult to conduct in many developing countries.

Delineating expenditure adjustments according to their effects on the poor cannot await the development of *ex ante* analysis. A good rule of thumb is to delay or reduce cuts in public expenditure on goods and services that are directly or indirectly of high importance in the poor’s budget, for example, coarser types of food, fuel and agricultural subsidies.

Within the broad category of basic services the selection of programmes needs to be sensitive to the type and severity of deprivation. If malnutrition is widespread, a programme of subsidized nutritional supplements would be more effective than an elementary education scheme. Rudra (2004) establishes that only the education component of public expenditure reduces income inequality in the face of globalization. Thus, when high inequality is a concern, expenditure on education should not be cut (van de Walle and Nead, 1995).

**Conclusions**

The role of fiscal policy in developing countries is as important as it is complex. Developing countries face the unenviable task of accelerating economic growth to reduce poverty in a short span of time even as they face greater uncertainty, in the face of globalization, about key elements of their fiscal policy such as the tax base. Furthermore, the exercise of fiscal policy is often circumscribed by increasing pressures from regulatory and exchange rate regimes in place, and subject to considerable pressure from external parameters such as competing countries’ tax rates; for example, it
would be difficult for a given developing country to have corporate tax rates very different from its competitors or to burden monetary policy with high fiscal deficits which could lead to sharp depreciation of the exchange rate.

This chapter has outlined some of the major challenges that developing countries face in some key areas of fiscal policy, particularly tax and expenditure. Even here the treatment has been selective (for example, there has been little discussion of corporate taxation and indirect tax harmonization) to provide an overview of the issues involved and an introduction to the literature on these topics.

Notes
1. I am grateful to Amitava Krishna Dutt and Jaime Ros for helpful comments on an earlier draft of this chapter. The usual caveat applies.
2. The rationale for the existence of multi-tiered governments owes much to the classic statement by Oates (1972), and has been extensively reviewed (for example, Jha, 1998). Intergovernmental fiscal relations are surveyed, among others, by Bjørgo and Smart (2001) and Bird and Smart (2002). Fiscal federalism is not considered in this chapter.
3. This may also lead to an exacerbation of inflation as de Haan and Zelhorst (1990), Easterly and Schmidt-Hebbel (1993) and Buffie (1999) show.
4. They find these countries to be Malaysia, Hungary, Ecuador, Morocco, Panama, Philippines, Indonesia, Bulgaria, Côte d’Ivoire, Egypt, Israel, Jordan, Lebanon, Nigeria and Pakistan. Clearly both transition and developing economies belong to this group.
5. Efficient indirect taxation calls for tax rates to vary inversely with compensated elasticity of demand making them regressive.
6. Separability of the utility function between goods and leisure would indicate that taxation of goods would have no implications for the labour–leisure choice.
7. CSO (2000) notes that in 1999–2000 as much as 60 per cent of India’s GDP came from the unorganized sector and this sector employed 92 per cent of the labour force.

References


Introduction
Economic development strategy in open industrializing economies is dominated by stabilization policy and structural adjustment. Developing countries are faced by an exogenous and changing set of world prices and export demand schedules on the one hand, and rationed global credit markets whose dynamic is determined by financial cycles in the core economies on the other. In consequence, domestic demand stabilization in response to unexpected temporary shocks and supply adjustment to permanent shifts in global markets determine the growth path, rather than a process of intertemporal optimization in the stable and foreseeable world of textbook economic theory.

Global economic shocks are exacerbated by armed conflicts and natural disasters, which often affect entire regions, while accumulated debt positions affect not only fiscal and current account balances but also the future expectations (and thus current behaviour) of the private sector. The political economy issues arising from accompanying changes in employment levels, wage rates and sectoral output are complicated by the key role played by international institutions – particularly the International Monetary Fund (IMF) and the World Bank – both as providers of financial resources to governments and as arbiters of ‘sound’ economic management.

These orthodox macroeconomic management criteria are based on specific models of economic behaviour for small open economies as well as a bias against active intervention in markets. In consequence the disappointing record of orthodox stabilization and adjustment policies in Eastern Europe, Latin America and Asia in the 1990s has led to a renewed interest in ‘heterodox’ methods of macroeconomic management that correspond more closely to the structure and behaviour of the emerging market economies with domestic capital markets integrated to the international financial system. However, an alternative set of underlying macroeconomic models with the generality and power of the orthodox models has yet to be constructed.

This chapter opens with a brief survey of the analytical literature that reveals an excessive emphasis on inflation targeting in stabilization policy
and a lack of attention to issues of investment and distribution in structural adjustment. It then outlines an alternative new-Keynesian approach to stabilization policy that explicitly contrasts with the standard IMF monetary programming framework. When capacity utilization and open capital account are included, inflation targeting is shown to have destabilizing consequences, requiring a return to active fiscal and monetary policy. The chapter then presents a new-Keynesian view of structural adjustment in explicit contrast to the standard ‘1-2-3’ open economy model that underpins World Bank analysis. It is shown that resource reallocation in the medium term can only take place through new investment, with significant distributional effects via employment and real wages. The chapter concludes by suggesting that emerging market authorities should and can engage in active macroeconomic intervention based on a more realistic analysis of the structure and behaviour of their economies.

**Stabilization policy and structural adjustment: the analytical debates**

I start by examining the macroeconomic programming models used by the Fund and the Bank in designing stabilization policy and structural adjustment, respectively. These two models are not entirely consistent as they are built on different assumptions as well as referring to different time horizons, despite earlier attempts to reconcile them (Khan et al., 1990), and thus must be treated separately.

The IMF Basic Financial Programming Framework (BFPF) is the standard model used by the Fund in designing stabilization programmes, the object of which is to reduce inflation to as near zero as possible and ensure debt service payments (IMF, 1987; Mussa and Savastano, 1999). The intellectual origins of the BFPF are Hicks’s interpretation of Keynes, expressed as the ‘absorption approach’ where excess domestic demand creates current account deficits and/or domestic inflation when imports are constrained by lack of external finance (Polak, 1957). The model subsequently metamorphosed into a version of the Chicago ‘monetary approach to the balance of payments’. It is concerned with the short run, where both real output and exports are taken as given, so the focus is on aggregate demand management. The budgetary balance plays a key role in this (particularly when financed by money supply) as does the nominal exchange rate because it sets import purchasing power.

Apart from the usual set of national accounting identities linking the macroeconomic and monetary variables, there are only two behavioural relationships in the BFPF: the demand for money and the demand for imports. A budget deficit beyond that warranted by output growth and the inflation target (that is, ‘seignorage’) is then reflected fully in the current account deficit if the exchange rate is fixed and foreign finance is available;
or fully in domestic inflation if the exchange rate is flexible and foreign finance constrained. Stabilization policy design then consists in the required fiscal adjustment\(^3\) in order to bring inflation down to target and restore external reserve levels to a prudent proportion of imports. This is supported by last-resort short-term lending from the Fund, which alleviates the fiscal adjustment required to meet these targets; and also provides powerful external leverage (‘conditionality’) to ensure compliance with this form of stabilization policy (Collier and Gunning, 1999). However, the behavioural relationships are clearly far too simplistic in the BFPF: for instance, interest rates and wealth (especially debt) stocks do not enter the model, while capital flows are exogenous, despite the central place of both in modern monetary theory.

The World Bank Revised Minimum Standard Model (RMSM) is used in designing structural adjustment programmes, the object of which is to restore current account stability and raise output growth (Addison, 1989 [1999]). The RMSM has Keynesian roots too: in the Harrod–Domar theory of growth constrained by savings, modified to include an external constraint reflecting the dependence of developing countries on imports of producer goods, leading to the ‘two-gap’ model (Chenery and Strout, 1966). However, it has since involved into the more neoclassical framework of a ‘computable general equilibrium’ model discussed below. The RMSM is concerned with the medium term, so aggregate supply is endogenous. Exports respond to the real exchange rate, which acts so as to allocate production factors between the traded and non-traded sectors: in other words, a relative price effect on supply instead of the income effect of the nominal exchange rate on demand in the Fund model. Investment is simply driven by the availability of savings: private saving (a constant proportion of disposable income) less the budget deficit plus external finance (‘foreign saving’).

In addition to the usual national accounting identities, the RMSM contains five behavioural relationships for the investment–growth linkage, import demand and export supply, fiscal income and private saving. Structural adjustment design seeks to relax the current account constraint on growth by raising exports through real exchange rate devaluation; and to raise the growth rate itself by reducing government expenditure and thus reversing the ‘crowding-out’ of private investment. Regulatory reforms follow the same logic, emphasizing trade and financial liberalization combined with extensive privatization to reduce the size and scope of the public sector. External finance in the RMSM plays three roles therefore: directly increasing public investment (for example in infrastructure) and output growth; reducing domestic borrowing to fund the budget deficit, and thus allowing private investment to rise; and funding more imports and thus
output. As in the case of the Fund, the Bank’s role as a leading provider of long-term official loans to poor countries, and its influence on other aid donors, ensures the adoption of this ‘sound’ approach to structural adjustment (Mosley et al., 1995).

As the RMSM is a one-sector model it is useful for macroeconomic programming, but not entirely appropriate for the analysis of structural adjustment, so World Bank policy design also has analytical foundations derived from the ‘dependent’ economy model set out by Dornbusch (1986) and Buiter (1988). This is disaggregated to generate what has now become the ‘industry standard’ with three products – exportables, importables, and non-tradable or ‘home’ goods and services – which we use as a framework in the fourth section below. This ‘1-2-3’ model has generated a wide range of applied computable general equilibrium models (Devarajan and Robinson, 1993) used by the World Bank to inform structural adjustment programmes and to link macroeconomic policy to poverty reduction strategies (Bourgignon and Morrison, 1992).

A number of significant lessons are drawn from this simple yet powerful model. One of these is the well-known ‘Dutch disease’ interpretation of the effect of an unexpected increase in world primary commodity prices or a rise in external aid flows: the real exchange rate appreciates, the non-traded sector expands, other traded sectors contract, so imports rise and exports fall, which is unsustainable in the long run. Another lesson is the effect of fiscal expansion: as government expenditure is intensive in non-traded goods and services, the real exchange rate appreciates and traded export production falls as non-traded output rises, leading to unsustainable debt problems.

However, neither the RMSM nor the 1-2-3 model are dynamic and thus do not allow for intertemporal optimization by economic agents: that is, the fact that households, firms and governments take investment, saving and borrowing decisions looking forward over many years. This is the basis of modern neoclassical macroeconomics and allows resource allocation behaviour to be endogenized. Further, they fail to reflect the elements of modern growth theory in general and the role of public expenditure in physical and human capital formation in particular. Last but not least, the simplistic view of the negative effects of budget deficits (on inflation for the Fund and on private investment for the Bank) ignores the modern macroeconomic theory of intertemporal budgetary and financial policy. Indeed from a strictly neoclassical viewpoint this persistence of the ‘financing gap’ tradition can be seen as invalidating the proposals from the Bank and the Fund on additional lending and debt forgiveness (Easterly, 1999).

These orthodox models have also been subjected to a much broader critique from non-neoclassical standpoints. Four such lines of argument are:
(1) the Keynesian critique of the failure to understand the exogenous nature of cycles in open economies; (2) the structuralist critique of the neglect of supply constraints in developing countries; (3) the Kaleckian critique of implausible assumptions on investment and savings; and (4) the Fabian critique of the exclusion of poverty reduction from macroeconomic strategy. These critical theoretical views have been fuelled by the evident failure in most cases of stabilization policy to get beyond inflation reduction, and of structural adjustment to achieve sustained growth (Williamson, 1997).

The underlying assumption in the IMF model that output is unaffected by demand and indeed that the economy operates at full factor employment is clearly implausible. Excess capacity in the Keynesian sense is often present in practice, as well as chronic underemployment in the Lewis sense. Moreover, the central issue in monetary policy for most developing countries today is not inflation as such but rather countering the effects of externally generated cycles exacerbated by inherited debt positions (Ocampo, 2000). In the upswing of a cycle the interest rate declines and the exchange rate appreciates, but any attempt to counter the boom attracts still more funds and the exchange rate appreciates still further. In the downswing, markets push for devaluation but this forces up interest rates and exacerbates production declines, promoting further capital flight and debt default. Moreover, the budgetary dependence on foreign borrowing makes the fiscal stance automatically procyclical. The application of the standard IMF policy model during these financial crises worsens economic recessions and further destabilizes capital flows (Stiglitz and Greenwald, 2003).

A central feature of the standard theory of structural adjustment is that any imbalance between traded and non-traded sectors in the dependent open economy is a result of distorted domestic relative prices (that is, differing from ‘world’ prices): thus the emphasis on real exchange rate correction and trade liberalization. However, this analysis rests on the twin assumptions of full employment of labour and capital on the one hand and the perfect substitution of existing factors between sectors in response to relative prices on the other; which is clearly unrealistic. Indeed, excess capacity and immobile factors explain much of the lack of supply response to structural adjustment (Taylor, 1993). Moreover, the assumption that domestic prices are not affected by exchange rates (and thus that devaluation is not ‘passed through’ into inflation) is similarly implausible for small open economies (Taylor, 1988). These supply response failures are exacerbated by the lack of business liquidity caused by restrictive monetary policies, because even under normal circumstances credit rationing prevails and output as well as prices are affected by interest rates (Blinder, 1987).

More generally, there is a clear parallel between orthodox structural adjustment theory and neoclassical trade theory because the internalization
of world prices is intended to bring about an intersectoral resource reallocation in line with comparative advantage. Trade liberalization raises the return to the abundant factor of production (assumed to be unskilled labour in developing countries); and because primary exports are taken to be more labour-intensive than tariff-protected industry there should also be a net employment creation (Obstfeld and Rogoff, 1997). However this does not often occur in practice: either because exports are based on natural resources in which case rents rise and little unskilled employment is generated; or because skilled labour is the scarce resource and export expansion opens up wage differentials (Wood, 1994). Moreover, there is in consequence no theoretical reason to believe that income distribution will necessarily improve with structural adjustment.

Central to the theoretical approach of both Bank and Fund is that private saving is a fixed proportion of disposable private income, and that private investment (and thus growth) is determined by private savings less the budget deficit plus external finance. However the large fluctuations observed in the savings rate for developing countries and the empirical evidence of the influence on private investment of other factors such as profit rates, credit conditions, public infrastructure, debt overhang, regulatory change and political stability all suggest that in developing countries at least investment is not constrained by private saving (FitzGerald, 2003). Indeed, the policy uncertainty caused by violent and unpredictable stabilization and adjustment episodes is among the most depressive influences on investment in developing countries (Rodrick, 1991). Nonetheless, successful structural adjustment and sustained growth require high rates of investment so that production capacity can change and thus the desired structural adjustment takes place.

This process cannot simply be considered as an overall proportionate expansion based on a fixed savings rate (augmented as necessary by external funds) once domestic resources have been reallocated, as the RMSM does; nor as a smooth process of reallocation of labour and capital between sectors in response to changing relative prices as the 1-2-3 model does. Both modern intertemporal macroeconomics and traditional Keynesian theory tell us that the investment process has its own dynamic based on future profitability, and this has profound implications for adjustment policy. Further, the financing gap theory used in both Bretton Woods models assumes that extra external finance always contributes to growth, by simply and directly adding to investment funds: but it is well established that capital inflows often lead to increased consumption (Jansen and Vos, 1997).

Last, but far from least, the neglect of distributional considerations in both the Bank and Fund models is not only inconsistent with their institutional commitment to poverty reduction but also leads them to
underestimate the political economy constraints on macroeconomic policy. There exists a long-standing critique of adjustment policy in this respect in terms of the negative effect on social service provision of fiscal expenditure cuts as the central macroeconomic policy tool (Cornia et al., 1987). Targeted poverty reduction programmes, while desirable in themselves, do not redress the effects of macroeconomic policy design on employment and wages, which are more significant in determining the welfare of the majority of the population. These effects in turn determine social support for economic policy, and thus its political sustainability.

Stabilization policy, inflation targeting and monetary autonomy

We have seen that the open developing macro-economy works in a different way from that which the Fund model supposes – the role of domestic credit rationing and external capital flows being crucial in the short-run context. This section sets out, therefore, a model with a formal framework similar to that in IMF (1987) except that: (1) output can be below capacity and is determined by the level and costs of credit; and (2) the interest rate and exchange rate are related though arbitrage across the capital account.

The standard inflation-targeting model can be set out as follows (IMF, 1987). As this is a short-run model exports (X) and real output (Q) are exogenous, as are the capital flows, net of debt service, that determine the net change in external liabilities (\( \hat{F} \)). The level of domestic debt (D) and foreign exchange reserves (R) are set according to fixed prudential rules. The endogenous variables are thus domestic aggregate income (Y), the level of imports (M) and the demand for money (B) and for credit (H) from the private sector. The nominal exchange rate (E) floats under the current Fund doctrine, and thus is also endogenous. The target variable is the price level (P) and the policy instrument is the interest rate (i).

I start with three accounting identities. Nominal income (Y) and inflation (p) are:

\[
Y = Q \cdot P
\]

\[
p = \frac{\dot{P}}{P}
\]  

(56.1)

The balance of payments (denominated in foreign currency) is the familiar:

\[
X - M = \hat{R} - \hat{F}
\]  

(56.2)

and the domestic monetary balance (Khan et al., 1990, p. 158) is:

\[
B = D + H + E \cdot R
\]  

(56.3)
There are three behavioural equations in this standard model, each of which reflect a key aspect of aggregate private sector macroeconomic behaviour. Import demand in nominal terms is a proportion \((m)\) of aggregate income \((Y)\):\(^{14}\)

\[
M\cdot E = mY \tag{56.4}
\]

Deposits in the banking system (that is, ‘demand for money’) depend on income \((Y)\) and the interest rate \((i)\) for a given velocity of circulation \((v)\) and positive interest ‘elasticity’ coefficient \((\alpha)\):

\[
B = Yv + \alpha i \tag{56.5}
\]

The credit (and cash) requirements of the private sector – that is, the supply of money – have a similar form because the Fund model assumes that the authorities always accommodate the monetary needs of the market (that is, passive rather than active monetary stance) and that the impact \((\beta)\) of the interest rate on this demand is of course negative:

\[
H = Yu - \beta i \tag{56.6}
\]

The ‘prudential rules’ for domestic debt \((D)\) and reserves \((R)\) are:

\[
\dot{D} = \lambda Y \\
R = \theta M \tag{56.7}
\]

The reserves rule \((\theta)\) is based on a specific degree of ‘liquidity’ in the form of import coverage;\(^{15}\) while the domestic debt rule \((\lambda)\) effectively constrains the fiscal deficit as a proportion of GDP.\(^{16}\)

This model is simple to solve because it can be distilled down to two reduced-form equations based on (56.2) and (56.3). The domestic price level \((P)\) is determined from the domestic monetary balance by substituting (56.5), (56.6) and (56.7) into (56.3) using (56.1) and (56.4) to yield:

\[
P = \frac{D_{-1} - (\alpha + \beta)i}{Q\{v - u - \lambda - m\theta\}} \tag{56.8}
\]

from which it is clear not only that higher interest rates \((i)\) reduce the price level and thus inflation, but also that a key determinant of inflationary pressure is the domestic debt overhang \((D_{-1})\), as indeed is the prudent fiscal deficit \((\lambda)\), thus the emphasis on fiscal retrenchment in Fund stabilization programmes.
An alternative formulation of the reserves rule that has recently found favour in the Fund with the spread of full currency convertibility is that there should be maintained a constant proportion (\(\kappa\)) of the money supply \((H)\). This gives a similar result:

\[
ER = \kappa H
\]  
\[
P = \frac{D_{-1} - \{\alpha + \beta(1 + \kappa)\}i}{Q(v-u(1+\kappa)-\lambda)}
\]

The external foreign currency balance found by substituting (56.1), (56.4) and (56.7) into (56.2) and using (56.8) serves to determine the nominal exchange rate \((E)\) for a given domestic price level under the floating exchange rate regime espoused by the Fund:

\[
E = \frac{mQP(1+\theta)}{X+(\theta M_{-1}+\bar{F})} = \frac{\{D_{-1} - \{\alpha + \beta\}i\}}{(v-u-\lambda-m\theta)} \frac{m(1+\theta)}{\theta M_{-1} + \bar{F}}
\]

This in turn implies that the real exchange rate \((e)\) using world prices as numeraire\(^{17}\) – and thus export competitiveness in the medium term – is endogenous and appreciates (that is, \(e\) falls) with positive external shocks such as capital inflows \((\bar{F})\) or increased commodity export income \((X)\) because:

\[
e = \frac{E}{\bar{F}} = \frac{mQ(1+\theta)}{X+(\theta M_{-1}+\bar{F})}
\]

However, note also that although from (56.9) inflation targeting will affect the nominal exchange rate (higher interest rates leading to appreciation) the real exchange rate in (56.10) remains unaltered. As a whole, therefore, the policy stance applied by the Fund model is procyclical because not only is the impact of external shocks on the economy unmitigated by active domestic fiscal or monetary policy, but also any unexpected decrease in output \((Q)\) must be met by higher interest rates to keep \(P\) (in 56.8) stable and vice versa in the inflation targeting approach.

I now adapt the model to allow for the two characteristics noted at the beginning of this section, which are essential in order adequately to describe middle-income ‘emerging market’ economies and the larger low-income countries with a domestic capital market open to foreign investment. I retain the same basic modelling framework in order to facilitate comparison between our new-Keynesian approach and the orthodox model.

The first modification is to the private credit channel. The relationship in (56.11) appears superficially similar in form to (56.6) but in fact causality
has been reversed: in a credit-rationed economy monetary policy determines the level of real output \((Q)\) as this responds to credit supply \((H)\) within the limit of capacity \((AK)\).\(^{18}\) This response comes about from both consumer credit expansion affecting demand and working capital availability affecting supply, and is a more realistic representation of emerging market economies than the Fund model. The authorities can alter the supply of money by straightforward monetary emission as an alternative to debt issue for fiscal deficit finance, by varying reserve requirements on banks or by changing the mode of financing foreign exchange reserve holdings. Note that this relationship means that raising interest rates \((i)\) will reduce output. Only with full capacity utilization is the demand effect felt on prices \((P)\) rather than output, and only then will reductions in money supply or higher interest rates reduce inflation. In contrast, my bank deposit function \((B)\) function is similar to that of the Fund model:

\[
\begin{align*}
Y &= H(\phi - \beta i) \\
Y &= QP \\
Q &= H(P(\phi - \beta i)) \text{ when } Q < AK \\
P &= H(\alpha - \beta i) \text{ when } Q = AK \\
B &= Y(v + \alpha i)
\end{align*}
\]  

(56.11)

The second modification is to open up the capital account of the balance of payments in (56.2) by expressing capital flows – changes in external liabilities \((F)\) – as a function of domestic interest rates \((i)\) and changes in the nominal exchange rate \((E)\).\(^{19}\) This contrasts with the Fund model where capital flows (limited to aid and foreign direct investment, FDI) are entirely exogenous:

\[
\dot{F} = \gamma(i - \frac{E}{E})
\]  

(56.12)

Note also that foreign investor risk appetite or world interest rates (both reflected in the parameter \(\gamma\)) can shift suddenly in practice, and that a sufficient imbalance between interest rates and exchange rate changes can lead to a capital outflow \((\dot{F} < 0)\).

Reflecting observed practice in emerging market economies, the ‘prudential rules’ for the management of external reserves and domestic debt in our model need to be adapted to these structural characteristics. Thus instead of the import coverage rule in (56.7), the central bank maintains a reserve level adapted to the external debt position \((F)\) as a form of insurance against external capital account shocks:\(^{20}\)
And the debt solvency rule in (56.7) is applied in an intertemporal context, changing with the primary fiscal deficit ($Z$) and limited to a given ratio ($\lambda$) of output at full capacity:\textsuperscript{21}

$$D = Z + (1 + i)D_{-1}$$

$$\frac{D}{PK} \leq \lambda$$ (56.14)

Finally, although the emerging market economy is exposed to external shocks even if the nominal exchange rate is allowed to float, the real exchange rate in our model is no longer determinate, unlike the Fund model (56.10), because it is affected by the domestic interest rate through the capital account. This endows the monetary authorities with a degree of freedom both to ensure that exports remain competitive and to respond adequately to external shocks. The primary deficit ($Z$) thus becomes the policy instrument affecting inflation, while interest rates set the real exchange rate and credit levels determine capacity utilization.

We can see how this policy framework works in practice by condensing the model into three reduced-form equations.\textsuperscript{22} The first is that real credit supply ($H/P$) be set so as to ensure full capacity utilization. From (56.11) we have this condition as:

$$\frac{Q}{K} = \frac{H}{P} = \frac{K}{\phi - \beta i}$$ (56.15)

The second is the balance of payments identity (56.2), re-expressed in terms of the exchange rate, interest rates and output by substituting in equations (56.1), (56.4), (56.12) and (56.13). Assuming that the target of stabilizing the real exchange rate is in fact achieved (that is, $e = 0$ and thus $p = \hat{E}/E$), inserting expressions for the real exchange rate (56.10) and inflation (56.1) yields the real exchange rate ($e$):

$$X - \frac{mPK}{E} = (\pi - 1)\gamma(i - \hat{E})$$

$$e = \frac{E}{P} = \frac{mK}{X + \gamma(1 - \pi)(i - p)}$$ (56.16)

In other words, the policy instrument that determines the real exchange rate in an emerging market economy with an open capital account is the real interest rate ($i - p$). The higher the real interest rate, the more the real
exchange rate will appreciate (that is, $e$ falls). The desirable policy target is thus clearly to maintain the real exchange rate at a stable and competitive level, keeping real interest rates low and adjusting them actively according to world market conditions.

The third reduced-form equation is derived from the monetary balance (56.3): substituting in (56.5), (56.13), (56.14) and (56.15) yields an equation for the price level ($P$) in terms of the other targets and instruments:

$$P = \frac{Z + (1 + i)D_{-1}}{K\left((v + \alpha i) - \frac{1}{\phi - Bi}\right) - e\pi\{F_{-1} + \gamma(i - p)\}}$$  \hspace{1cm} (56.17)

Because the interest rate ($i$) instrument is already employed to stabilize the real exchange rate ($e$), and the credit level ($H$) is already set so as to stabilize output ($Q = K$), the budget deficit ($Z$) plays the key role of price stabilization. However from (56.17) it is clear that a low level of inflation does not mean that the budget should always be in balance ($Z = 0$) or even a fixed proportion of aggregate income ($Z = \lambda Y$), but rather should compensate for exogenous shocks to international asset demand ($\gamma$) or production capacity ($K$) even when international price fluctuations are smoothed by the real exchange rate.

In sum, in this type of open economy integrated – albeit asymmetrically – to international capital markets, a decision by the central bank to raise the interest rate in order to curb inflation from (56.17) will actually have three undesirable effects: inflation will rise due to the effect of debt service on the budget deficit; the exchange rate will appreciate through the capital account effect; and real output will fall from the credit channel effect. My more realistic new-Keynesian model for the emerging market economy thus shows not only that this is a misguided stabilization policy, but also that a wider range of policy instruments should be used to achieve multiple stabilization targets, with particular emphasis on trade competitiveness and full employment. These instruments include low real rates of interest, a balanced budget over the cycle and above all strong prudential control of bank credit.

**Structural adjustment, sectoral investment and income distribution**

We have also seen that emerging market economies work in different ways from that which the World Bank model supposes – the existence of surplus labour and sector-specific installed capital being crucial. This section sets out, therefore, a medium-term model with a framework similar to the 1-2-3 model used as a formal analytical basis for the RMSM, except that: (1) although production capacity may be fully utilized, this does not involve
the full employment of the labour force; and (2) shifts in production patterns are not possible except through new investment.

In the canonical dependent economy model\textsuperscript{23} there are two sectors producing traded ($T$) and non-traded ($N$) goods from homogeneous production functions with sector-specific labour ($L$) that is intersectorally mobile, so that the nominal wage ($W$) is equalized across sectors. Total labour supply is fixed and fully employed at the equilibrium wage. For the two sectors ($j$):

$$Q_j = Q_j(L_j) \quad (56.19)$$

and the real exchange rate ($e$) is now defined as the ratio of traded ($P_T$) to non-traded prices ($P_N$):

$$e = \frac{P_T}{P_N} \quad (56.20)$$

Each sector employs labour up to the point where the marginal product of labour is equal to the single intersectoral wage,\textsuperscript{24} which is thus equalized between the two sectors. Defining the real wage ($w$) in terms of non-traded prices ($P_N$) we thus get:

$$w = \frac{W}{P_N}$$
$$Q'_T(L_T) = \frac{w}{e}$$
$$Q'_N = w \quad (56.21)$$

It follows that the ratio of the marginal products of labour in the two sectors is equal to the real exchange rate ($e$) at equilibrium:\textsuperscript{25}

$$\frac{Q'_N(L_N)}{Q'_T(L_T)} = e \quad (56.22)$$

Demand for labour is the inverse function of the product wage and there is full employment of the total labour force ($\bar{L}$):

$$L_T(ew) + L_N(w) = \bar{L} \quad (56.23)$$

The real exchange rate thus drives labour allocation across the two sectors, and hence output. The sectoral supply functions become:

$$Q_j = Q_j(e/w)$$
$$Q'_T > 0, \; Q'_N < 0 \quad (56.24)$$
And generally an inverse relationship between the real wage ($w$) and the real exchange rate ($e$) is also implied by this result, because with full employment in (56.23) and exploiting the implicit function rule:\footnote{26}

$$w' = -\frac{wL_N'}{L_T}eL_N' < 0$$  \hspace{1cm} (56.25)

Disaggregating traded goods into exportables ($x$), importables ($m$) and non-tradable or ‘home’ goods and services ($h$) generates the 1-2-3 model (Devarajan and Robinson, 1993), which we use as a framework in order to facilitate comparison. The prices for exportables ($P_x$) and importables ($P_m$) are determined by world prices multiplied by the nominal exchange rate ($E$), while home goods prices ($P_h$) depend upon supply conditions, as the domestic market must clear to ensure full capacity utilization.

However, in addition to the unwarranted assumption of full employment of labour, the notion in the 1-2-3 model that capital can simply be moved between sectors with a constant elasticity of transformation (CET) production function is obviously implausible. In fact the key issue in structural adjustment is investment behaviour because installed capital cannot be shifted between traded and non-traded sectors, while labour cannot easily be substituted for capital when technology is largely imported. The sectoral production functions are thus better seen as separate, limited by the installed capital stock in each.\footnote{27}

World prices for exports and imports ($P_f$, $P_m$) and thus terms of trade ($\tau$) are exogenous, while the unit import ($m$) and labour ($l$) input coefficients are technologically fixed. As before there is a single nominal wage rate ($W$), but in our model it is set either institutionally or by the reserve price of labour from the household sector, rather than by labour market clearing, and unemployment persists. Sectoral employment ($L_j$) is thus determined by output and there is excess labour supply, due to the Leontief fixed technical coefficients:

$$Q_j \leq A_jK_j$$

$$G_j = Q_j(P_j - Wl_j - m_jEP_m)$$

$$L_j = l_jQ_j, \sum_j L_j < \bar{L}$$

$$\tau = P_f/P_m$$ \hspace{1cm} (56.26)

Traded exportable prices ($P_x$) are as before, but home goods prices ($P_h$) are formed by a mark-up ($g$) on production costs, where the nominal wage as well as the exchange rate plays a central role. I use home goods prices ($P_h$) as the numeraire in order to define the real exchange rate ($e$) and the real wage rate ($w$):
I assume as before that the export sector produces at capacity because world demand is infinitely elastic to supply by our country. The demand for home goods is a proportion \( (a) \) of the aggregate factor income, but the domestic market clearing condition will not now determine price as in the standard model, but will rather determine output within the capacity constraint. Substituting the income (56.26) and price (56.27) equations into the demand function for home goods gives the output level, therefore:

\[
P_x = EP_f \\
P_h = (1 + g)(Wl_h + Em_h P_m) \\
e = \frac{EP_f}{P_h} \\
w = \frac{W}{P_h} \tag{56.27}
\]

Note the so-called ‘Dutch disease’ effect of an improvement in the terms of trade \( (\tau) \) leading to an expansion of the home goods sector as incomes rise.

Taking the home good price \( (P_h) \) as the numeraire and rearranging (56.27), we now have real sectoral profit levels \( (\bar{G}_j) \) in terms of the real exchange rate \( (e) \), the real wage rate \( (w) \) and sectoral output \( (Q_j) \):

\[
\bar{G}_x = Q_x \left[ e \left(1 - \frac{m_x}{\tau} \right) - wl_x \right] \\
\bar{G}_h = Q_h \left[ 1 - \frac{em_h}{\tau} - wl_h \right] \\
\bar{G}_x(e) > 0, \quad \bar{G}_h(e), < 0 \quad \bar{G}_j(w) < 0 \tag{56.29}
\]

The channel through which the real exchange rate and the real wage rate affect profits is now evident and the implications for investment decisions and income distribution can be explored properly – something that is not done in the standard model.
I now examine in detail the investment decision for the case when \( e \) rises (that is, real depreciation) but the obverse is simple to work out; as are the consequences of other shocks such as changes in labour productivity \((t)\). I assume for convenience that in previous periods firms have been able to adjust their capital stocks to the desired level (that is, \( Q_j = A_j K_j \)). From (56.29) real depreciation will raise real profits in the traded export sector and reduce them in the non-traded home goods sector; but any shift in output depends on investment, which is irreversible. The two-period \((0, \ 1)\) problem for firms in each sector is whether to invest or not. If they do not invest, then capacity (and thus production) falls by the amortization rate \((\delta)\). If they do invest then they must do so at the level that maximizes its present value \((V)\) discounted at the interest rate \((i)\) where the installation cost \((J)\) expressed in home goods prices is an increasing function\(^{29}\) of real investment \((I)\) using imported equipment at the world price \((P_m)\) and the corresponding import coefficient \((m_k)\):

\[
\begin{align*}
V_{j,0} &= \bar{G}_{j,0} + \frac{\bar{G}_{j,1}}{1+i} - J_j \\
J_j &= I_j (e m_k P_m + \frac{1}{2} \sigma I_j) \\
I_j &\geq 0 \\
K_{j,1} &= K_{j,0} (1 - \delta) + I_j
\end{align*}
\]

The optimal investment level \((\bar{I})\) is simply found by differentiating \(V\) with respect to \(I\) in order to maximize the present value of the firm:

\[
\frac{dV}{dI} = \frac{1}{1+i} \frac{dG}{dI} - e^{-\frac{m_k}{r}} - \sigma I = 0
\]

\[
\bar{I} = \frac{1}{\delta} \left\{ \frac{1}{1+i} \frac{dG}{dK} - e^{-\frac{m_k}{r}} \right\}
\]

Because the capital stock in the home goods sector \((K_h)\) was adjusted to the previous real exchange rate so as to maximize profits, from (56.29) no investment takes place in the sector (that is, \(I_h = 0\)) and capacity declines by the amortization rate:

\[
K_{h,1} = K_{h,0} (1 - \delta)
\]

In contrast, the export firms do invest as real profits have risen and:
Note that improved terms of trade or real depreciation raise export sector investment, while higher wages or interest rates reduce it. Traded output capacity thus rises while non-traded output capacity falls, due to the changes in their respective capital stocks. Aggregate real output \(Y\) only rises if the net output capacity shift is large enough:

\[
Y = eQ_x + Q_h
\]

\[
Y_1 > Y_0 \text{ if } eA_x (K_{x,1} - K_{x,0}) > \delta A_h K_{h,0}
\]  

(56.34)

Moreover, even if the adjustment is sufficient to cause aggregate real output to rise, there is no necessary reason why net employment should do so as well. The general condition for this to happen is found by substituting the employment functions from (56.27) into (56.33) to yield:

\[
L_1 > L_0 \text{ when } Y_1 > Y_0 \text{ if } l_x > eI_h
\]

(56.35)

This result has three interesting characteristics. First, there is no guarantee that total employment will increase with real devaluation because this depends on the labour intensities \((l)\) of each sector as well as the investment outcome: clearly only if the export sector is the more labour-intensive \((l_x > l_h)\) is this likely to occur. Second, the ratio of the real wage rate to the real exchange rate \((w/e)\) is clearly critical to the outcome. If real wages fall then employment will rise, although this trade-off is not the result of factor substitution along the constant elasticity of transformation (CET) curve, but rather of investment incentives. Third, an increase in the interest rate \((i)\) will reduce the employment gain due to the investment disincentive (56.33): this is the reverse of the factor substitution effect textbook theory would predict. Moreover, for investment to take place, private investors must have confidence in future profits and be provided with sufficient credit and infrastructure.

As we have seen in (56.25), the dependent economy model implies that depreciation of the real exchange rate will reduce the real wage rate, which with full employment implies a deterioration of the overall distribution of income. The new-Keynesian model reveals a more complex relationship. Rearranging (56.27) we have:

\[
\tilde{I}_x = \frac{1}{\sigma} \left\{ \frac{A_x}{1 + \tau} \left[ e \left( 1 - \frac{m_x}{\tau} \right) - w l_x \right] - e^\frac{m_x}{\tau} \right\}
\]

\[
K_{x,1} = K_{x,0} (1 - \delta) + \tilde{I}_x
\]

\[
\tilde{I}_x'(e) > 0, \tilde{I}_x' (\tau) > 0, \tilde{I}_x' (w) < 0, \tilde{I}_x' (i) < 0
\]

(56.33)
which has interesting characteristics: there is again a negative relationship between real wages and the real exchange rate; but the home goods profit mark-up also plays an important part in income distribution, and productivity in the home goods sector (the inverse of the labour input coefficient \( l_h \)) is also a key determinant of real wages. This last point should remind us that while traded investment is the central focus of adjustment policy, improved living standards require a greater supply of wage-goods.

Further, if I define workers’ living standards \((\omega)\) as the nominal wage deflated by the cost of living \((P_c)\) defined by the mean domestic product prices weighted by home goods consumption propensity \((a)\) in (56.28):

\[
\omega = \frac{W}{P_c}
\]

\[
P_c = aP_h + (1-a)P_x
\]

Substituting the relevant definitions from (56.29) into (56.37) then gives:

\[
\omega = \frac{W}{a + (1-a)e}
\]

Here the welfare consequence of the fall in the \(w/e\) ratio in (56.33) required to raise investment and permit structural adjustment is clearly revealed as a deterioration in workers’ living standards.

In other words, the targeting of the real exchange rate necessary in order to maintain export competitiveness is in fact an ‘incomes policy’. If employment expands then overall income distribution can improve, but this requires active intervention in order to raise investment rates as we have shown. Monetary policy should be geared to low real interest rates and producer credit provision on the one hand, and an active fiscal stance geared to damp exogenous macroeconomic shocks on the other. This desirable outcome can be reinforced by a system of dividend taxation (designed to stimulate investment) and social spending specifically aimed at raising the living standards of employees’ families.30

**Conclusions**

In this chapter we have seen how the macroeconomic models conventionally used to analyse stabilization policy and structural adjustment are open to a wide range of criticism from both neoclassical and Keynesian standpoints. Specifically, the existence of excess productive capacity, mark-up pricing by firms, credit rationing by banks and open capital accounts on the
one hand; and the central role of investment in determining sectoral output and the impact of the real exchange rate on both export levels and real wages on the other; mean that the standard IMF and World Bank models can become seriously misleading as the basis for macroeconomic policy.

By including more realistic formulations of credit supply and external capital flows for emerging markets in the standard stabilization model, I have shown not only that inflation targeting using the interest rate can have serious procyclical consequences, but also that a more active monetary policy based on fiscal and credit instruments can make for higher and more stable output solutions – that is, sustainable stabilization. Similarly, by including the pricing and investment behaviour of firms in the standard adjustment model, I have shown that effective structural adjustment will not take place unless active exchange rate and monetary policies are implemented in support of traded production.

Finally, this new-Keynesian approach also differs from that of the Bretton Woods institutions in its attitude to macroeconomic intervention. The Bank and the Fund insist that inflation-reduced public expenditure and balanced budgets are essential in order to promote growth. This passive ‘rules-bound’ approach is central to their lending conditionality, and is built into significant institutional reforms such as central bank independence. In marked contrast, this chapter has shown how emerging market authorities can combine active fiscal and credit management with real exchange rate targeting in order to cope with exogenous shocks and promote longer-term export-led growth in a more purposive stabilization policy. Finally, it has demonstrated that active intervention to maintain both a competitive real exchange rate and a low real interest rate is necessary in order to promote sufficient investment in the traded sector and thus ensure not only export growth but also employment expansion so that real wage constraints do not lead to a worsening income distribution as a consequence of structural adjustment.

Notes
1. This chapter does not address issues such as privatization of public enterprise or government expenditure reform that, while central to structural adjustment and stabilization policy in practice, are covered in other chapters of this Handbook.
2. These two well-known models can be found in the official sources referenced below and are summarized clearly in Khan et al. (1990) and Agénor (2000).
3. Usually government expenditure cuts rather than increases direct tax pressure – which would depress saving, assumed to be a constant proportion of private disposable income in this model – with obvious distributional consequences.
4. In consequence, it might better be described as a ‘three gap model’ (Bacha, 1990).
5. In the sense a small open economy that is a price-taker in world markets, popularized by Dornbusch (1986), rather than the wider Latin American notion of dependencia.
6. This model is lucidly set out in Montiel (2003, Part V).
7. Because there are one country, two sectors and three products in the model.
8. See Corden (1984). In fact ‘British disease’ would be more appropriate. The Netherlands did indeed experience real exchange rate appreciation in the 1980s due to North Sea gas finds, which rendered manufactured exports uncompetitive; but the fiscal resources were reinvested in infrastructure and skills, with subsequent growth based on advanced services exports. The Thatcher administration used the North Sea royalties to reduce UK profit taxes, stimulating consumption and depreciating the real exchange rate; but the consequences are now visible in deteriorating public transport and education.


12. As opposed to the justifiable concern with hyperinflation in the 1980s.

13. Strictly speaking, bank deposits in the Fund’s BFPF model.

14. This implies unitary price and income elasticities of import demand, which are much higher than those found empirically and means that import contraction through reductions in real demand \( \frac{Y}{P} \) become the main channel for stabilizing the current account of the balance of payments.

15. The conventional rule of thumb is three months’ cover (that is, \( \theta = 0.25 \)).

16. As in the ‘Maastricht Criteria’, where \( \lambda = 0.03 \).

17. This corresponds to the IMF definition of the ‘effective real exchange rate’ as the nominal rate divided by the ratio of the domestic price level \( P \) to the weighted mean of the price levels in trading partners (unity in our case). The alternative definition of the real exchange rate in terms of the ratio of traded to non-traded prices is discussed below. On both definitions, see Dornbusch and Helmers (1988) and Montiel (2003).

18. Which in turn is given in the short run but depends on investment in the medium term, a point taken up below.

19. This is a simplified form of the full international demand function for emerging-market assets, which itself can be derived from standard portfolio theory (FitzGerald, 2006).

20. The most cautious position would be to maintain reserves equal to short-term external liabilities, commonly known as the ‘Greenspan rule’. The value of \( \pi \) will then depend on the maturity structure of external liabilities \( F \).

21. See Missale (1999) for a discussion of optimal debt models in a full intertemporal context, from which this familiar rule is derived.

22. Formally, Walras’ Condition is satisfied because the model has 11 equations and 11 variables (eight endogenous and three targets); while Tinbergen’s criterion is met by having three instruments \( (H, i \text{ and } Z) \) with which to hit the three targets \( (Q, e \text{ and } P) \).

23. The model is very well explained in Chapter 2 of Agénor and Montiel (1999).

24. That is, \( P^e_0L_0 = W \).

25. Note that this is the formulation used to derive the Harrod–Belassa–Samuelson model of real exchange rate trends in the long run.

26. If \( F(x, y) = \text{constant} \), then \( dx/du = -F_x/F_y \).

27. In the standard 1-2-3 model set out above, there is a domestic production function with a constant elasticity of transformation \( \sigma \)

\[
Q = A\left[\gamma Q^\alpha_x + (1 - \gamma) Q^\alpha_y\right]^{1/\sigma} \quad 0 < \sigma < \infty
\]

where the convention is to adopt a Cobb–Douglas unitary elasticity \( \sigma = 1 \). Below I adopt the more realistic Leontief form \( \sigma = 0 \) to reflect the fact that once installed, capital is entirely immobile.

28. This form is in fact the constant elasticity of substitution (CES) consumption function used in the canonical 1-2-3 model discussed above, but with unitary own-price elasticity and zero (Leontief) substitution between home and export goods.

29. This is a standard representation of the intertemporal optimisation process for the firm, which as part of the ‘AK’ model underpins endogenous growth theory – see Heijdra and van der Ploeg (2000, Chapter 2). In developing countries this can also be seen as reflecting limited local project implementation capacity.
30. This topic lies beyond the scope of this chapter, but see FitzGerald (1993) for a further discussion of such a policy based on dividend taxation and social expenditure; and FitzGerald (2002) for the derivation of an optimal profits tax to fund infrastructure provision.

References


Bourguignon, F. and C. Morrison (1992), Adjustment with Equity in Developing Countries, Paris: OECD.


Economic planning in developing economies

Bill Gibson

Introduction

‘Planning’ is a term that generally has fallen into disuse. Todaro defines development planning as ‘the conscious effort of a central organization to influence, direct and in some cases even control changes in the principal economic variables (such as GDP, consumption, investment, savings, etc) of a certain country or region, over the course of time in accordance with a predetermined set of objectives’ (Todaro, 1971, p. 1). Planning connotes, but does not logically imply, command-and-control mechanisms by which authorities issue directives for which compliance becomes a matter of administrative law.

Development planning was attempted in the Soviet Union and Eastern Europe and to a degree in India, Cuba, Egypt and Tanzania and other countries in the immediate post-World War II period. Indeed, it was largely the success of the Soviet Union in raising per capita incomes in the first half of the twentieth century that demonstrated the existence of a practical alternative to market allocation. Soviet performance impressed policy-makers in developing economies who had come to see the market as inadequate to the task of industrialization. In non-communist countries, planning without enforceable command-and-control mechanisms was widespread in the immediate post-World War II period. The United Nations and other sources even withheld development aid unless a plan was in place and as a result, planning ministries became commonplace throughout the developing world. Planning models that demonstrated how foreign aid could be coordinated to achieve maximum impact on growth and development were especially popular. Despite its increasing technical sophistication and theoretical appeal, planning in the post-World War II period led to widespread disillusion and rejection by even formerly ardent supporters. By the end of the 1970s, Chowdhury and Kirkpatrick noted that many economists were talking openly about the failure of planning, and as early as 1965, Waterson had concluded on the basis of a study of 55 country experiences that ‘the majority of countries have failed to realize even modest income and output targets’ (Chowdhury and Kirkpatrick, 1994, p. 2).

Since the 1970s explicit plans in developing countries have largely been abandoned. Many of the problems planning was designed to confront are
still present, of course, and the need for some kinds of planning persists. As a result, planning has re-emerged in a more market-friendly variant, development policy management, with emphasis on the price mechanism, incentives and schemes such as ‘cap-and-trade’, that rely heavily on decentralized implementation.

This chapter critically reviews planning as applied to developing countries. The next section addresses the general question of the relationship of planning to the market. Economy-wide planning models and techniques are discussed in the following section, while the subsequent section turns to microeconomic planning and cost–benefit analysis. A concluding section discusses the uses of surviving planning models in the current context.

Plan versus market
Plans may either be economy-wide or partial. Heal reviews the theory underlying the economy-wide planning procedures (Heal, 1973). He notes that much of the early writing on planning, by distinguished economists such as Lange, Lerner, Arrow and Hurwicz, sought to establish that an efficient centrally planned system would employ the same marginal equalities as in the Walrasian system, with the central planning board playing the role of the auctioneer. Plans in which individual preferences are constitutive of the social objective function, therefore, yield the same pattern of resource allocation as would a competitive market. In other words, there is nothing inherently inefficient about planning. This conclusion is today widely accepted inasmuch as planners’ preferences often proxy a social welfare function under the assumption that a freely functioning competitive market mechanism would produce an identical allocation of scarce resources.

In partially planned economies, planning is generally conceived as a response to market failure, including externalities, informational asymmetries and public goods. If market failure is widespread then it follows that central planning can serve as a substitute for the market; if not, then planning can, in principle, resolve allocational issues related to market failure.

In addition to concerns about market inefficiencies, equity was also considered a legitimate objective. The Coase theorem holds that efficiency and equity are separable, but the distinction in the early days of planning was not clearly recognized. Force-draft industrialization had achieved rapid modernization in the Soviet Union, but at the great expense of a debilitated agricultural sector. The First Five Year Plan under Nehru in India in the early 1950s explicitly prioritized reducing unemployment and poverty over maximizing the rate of economic growth. The principle that income could be redistributed without disturbing the price-guided marginal equality of social costs and benefits was ignored.
There is significant disagreement as to the extent to which government can improve outcomes by realigning social and private costs. In standard theory, a properly tuned set of taxes and subsidies could repair markets that failed and public sector institutions could fill in when markets were missing altogether. In developing-country practice, however, public policy often did not improve outcomes and the term ‘government failure’ gained currency to describe counterproductive intervention by states. The necessity of a one-to-one relationship between policy objectives and policy instruments, originally due to Tinbergen, shows how precarious is the entire mission. The collapse of earlier planning initiatives was in part due to a mismatch in this relationship, with goals grossly exceeding the number of instruments, other than command-and-control, available for implementation.

Killick (1976) provides a comprehensive discussion of government failure in development planning. He argues that the plans failed because their creators assumed that politicians see the planning problem ‘essentially as economists do’. The assumption that governments are composed of ‘public spirited, knowledgeable and goal oriented politicians . . . clear and united in their objectives, choosing policies which will achieve the optimal results for the national interests’, is unwarranted. Anticipating much of the subsequent public choice literature, Killick argues that politicians should be seen as rational, self-interested, acting to maximize the short-term probability that they will be re-elected. The implicit assumption of the existence of a ‘benevolent despot’ was at variance with both reality and the ‘liberal-individualist’ tradition of Western civilization.

Modern public choice theory does indeed suggest that planning will be undertaken for the benefit of the planners themselves or their clients, and that command-and-control directives will give rise to rent-seeking behavior and other principal–agent problems that deprive a country of needed resources and talents. In democratically organized societies, a major problem arises when costs of a directive are widely distributed, while benefits accrue to a smaller set of individuals. Significant pressure to change course can develop as a result, with powerful groups lobbying effectively to push the economy away from its socially optimal path.

Planning, *ipso facto*, could never have resolved these deeper issues of policy-making. Problems of coordination, incentives and the trade-off between efficiency and equity are at the root of the problem of underdevelopment and were beyond the reach of technocratic planners and their tools. Planning was abandoned within a broader current of change that involved rethinking the role of the state generally. Much of the planning literature reviewed above is now seen as archaic, and to say that planning is ‘out of fashion’ is an understatement. As demand management, automatic stabilizers, incomes policy and the Phillips curve gave way to rational expectations,
new public choice theory and inflation targeting, planning became caught up in a generalized retreat from *dirigisme* and ascendancy of the market mechanism. Killick, who had so thoroughly excoriated the planning process earlier on, came later to wonder if there had not been ‘a reaction too far’ in moving so decisively away from planning toward the market. We shall return to this issue in the concluding section.

### Economy-wide planning models

Planning models can be classified in several different categories: aggregate, main sector, multisectoral, regional and project-specific models. Economy-wide models include the first three categories, but not the last two, and may be static or dynamic. They typically reflect the accounting regularities and conventions of national income and product accounts, balance of payments, and income and expenditure balances of the public sector (Taylor, 1979). These can be simulation models or more traditional econometric constructs. The former employ informal calibration procedures, while the latter are estimated formally, using statistical theory under the usual assumptions. The simulation approach does not rely on statistical theory, but rather on whether the model captures salient features of the economy (Gibson, 2003).

### Input–output models

Economy-wide planning models have their roots in the model first described by the young Harvard economist W. Leontief in the 1930s and 1940s. The inter-industry or input–output approach pioneered by Leontief, and first implemented with his help in the Soviet Union, served a means by which consistent intersectoral plans could be drawn up. Input–output models have their roots in Quesnay’s *Tableau Economique*, a Physiocratic device that was the first effectively to separate real from nominal resources flows.

Input–output models are used to analyze the impact of a change in final demand on the levels of production. Let $A = \{a_{ij}\}$ be the coefficient matrix such that each $a_{ij}$ describes the use of input $i$ for the production of one unit of output $j$, and $X = \{X_j\}$ be a column vector of gross outputs, including intermediate goods. Factors of production, labor, $L$, and capital, $K$, are treated separately, usually with fixed coefficients under the assumption that factor prices remain unchanged. Together with the labor and capital coefficients, the $A$ matrix represents the technique by which goods are produced.

Final demand is denoted by $F = \{F_j\}$, a column vector of outputs, and may be disaggregated into consumption, government spending, exports and imports as needed. The essential equation of input–output analysis, known as the material balance, is then:
In a one-commodity world, say corn, let the output of six ears require the input in the form of seed corn of one ear. In order to consume 500 ears of corn, we must then produce \( X = (1/6)X + 500 \) or \( X = 600 \) ears to make sure that there is enough for both final, 500 and intermediate, or seed, of 100. Since we could just as easily produce 1200 ears with 200 ears of seed, the model evidently assumes constant returns to scale.

So-called dual variables can also be defined and interpreted as prices, denoted here by row vector \( P = \{p_j\} \). The equation dual to the material balance is then:

\[
P = PA + V_A
\]  

(57.2)

where \( V_A \) is a row vector of value added, and may be disaggregated into wages, profits, imports, taxes and rents as needed.

As a result of a linear production technology, input–output models are relatively inexpensive and easy to formulate and run. Since prices were often administered and incentives less relevant, the absence of a functioning price mechanism in the model was unimportant. They were, consequently, enormously popular in early development planning.

The model can be made dynamic if investment, \( I \), is first disaggregated from final demand, \( F \), and then used to determine the time path of capital stock, \( K \). This is done by way of the ‘stock-flow equation’:

\[
K_t = K_{t-1}(1 - \delta) + I
\]

where \( \delta \) is the depreciation rate. Consistent forecasts of intermediate demand, labor and foreign exchange requirements, for example, could then be made, contingent on a forecast for investment.

The framework just presented is the ‘open Leontief model’, but a closed version is available in which all elements of final demand and value added are made dependent on \( X \). It was left to von Neumann to show that a maximum rate of sustainable growth is well defined by the model. Despite the elegance of the von Neumann model itself, there were limited direct practical implications of the closed Leontief model for planning. One reason was that prices played virtually no role whatsoever; the technical coefficients, whether for capital or feeding labor, determined the entire balanced growth path. ‘Turnpike optimality’, as exhibited by the von Neumann model, was intellectually appealing, but offered little insight into the nature of the far bumpier road on which developing countries were traveling.
Linear programming models

The chief limitation of all linear models is that they do not allow for substitution in response to changing prices of goods and factors. A partial solution to this problem is provided by the linear programming approach. Introduced by Dantzig for the US Air Force in 1947 and popularized in a classic text by Dorfman, Samuelson and Solow, linear programming models allowed for prices to have a limited impact on the allocation of resources (Dorfman et al., 1958). Unlike their more rigid input–output counterpart, linear programming models could be set, for example, to maximize employment by choosing a sectoral pattern of output consistent with a foreign exchange constraint or some other supply-side limitation (Blitzer et al., 1975).

In a typical linear programming model, there is usually more than one feasible solution. The feasible solutions are then ranked according to an explicit objective function that depends on prices of goods and factors, or some other methods of valuation. An optimal primal solution satisfies all constraints and provides a maximum of the objective function. Like input–output models, there is a dual solution which minimizes the value of the dual objective function. A powerful and fundamental duality theorem of linear programming establishes complementary slackness, which holds that if a constraint in the primal solution of a linear program does not bind, that is, it is satisfied only as an inequality, then the corresponding dual variable is zero. In less formal language, an additional unit of a resource that was already in excess supply could have no effect on social welfare.

The impact of complementary slackness on development planning cannot be overestimated, since at once the notion clarified the relationship of a ‘social optimum’, however planners wished to define it, to factor abundance and the related production technology. From linear programming and complementary slackness, practitioners derived the idea of a ‘shadow price’, or the change in the value of the social objective function with respect to a change in the quantity of a specific binding resource. Now the size of the wedge between the social and private cost of resources could be computed. The application was immediate: in economies with surplus labor, the shadow value of unskilled labor was effectively zero and thus planners would be justified in substituting a lower than market wage when computing the social cost of any particular project or policy intervention.

While linear programming models allow for choice of technique, they do not allow for smooth substitution and infinite divisibility between discrete techniques. This may in practice be more realistic but does give rise to jumps in the values of the solution variables. Data permitting, smooth substitution can always be approximated to any degree of accuracy by
increasing the number of available techniques of production. Moreover, since linear programming models are special cases of non-linear programming models, computer software available for the solution of the latter, for example, General Algebraic Modeling System (GAMS), Matlab, Mathematica, and so on, also compute solutions for the former. Specialized packages exist for linear programming problems, such as Lindo, that are fast, efficient and give highly detailed computational results.

That linear programming would show only how one analyzes resource constraints given the objective function, rather than the deeper problem of how social objectives are themselves to be defined, would ultimately lead to its undoing. But for a while, the technique enjoyed immense popularity, and still does in many specialized applications. Moreover, that it could neatly separate the role of the policy-maker, who determined the coefficients in the objective function, from that of the economist or planner, who designed, built and ran the model, only enhanced its scientific patina.

**SAMs and CGE models**
Social accounting matrices (SAMs) extend the usual conceptual categories of input–output frameworks to account for more detailed expenditure and distributional categories. The constructs are not properly referred to as ‘models’, but rather serve as a database to which behavior equations can be calibrated.

Just as linear programming generalized input–output analysis, so computable general equilibrium (CGE) models take the next step in integrating price signals in more fundamental ways (Gunning and Keyzer, 1995). They are usually multisectoral, economy-wide models, calibrated to SAMs. They may be static or dynamic with short-run coverage of one to three years, three to seven for the medium run, and long-term models that extends beyond a decade. Static models compare two points in time without explicit attention to the path connecting these points, while dynamic models trace out a locus of points with explicit stock-flow adjustment processes. The models may exhibit a wide range of adjustment mechanisms, from closed, purely competitive, Walrasian models to macro structuralist models in which foreign exchange availability determines the level of output in some key sectors.

The structure of a typical CGE model can be briefly sketched as follows. Beginning with the material balance, in equation (57.1), the model links the various elements of final demand to goods prices and incomes. Factor demand equations determine factor prices when supply is binding, but this need not be the case and some other mechanism might be introduced to determine nominal factor prices. CGE models (CGEs) can be constructed
in real or nominal terms, but it is a characteristic feature of structuralist CGEs that equations are given for nominal quantities, which are then converted into real terms by the price vector that results from general equilibrium. This implies that money or some other nominal quantity be fixed exogenously and thus, inflation can be modeled in dynamic systems.

There is, of course, no need to specify supply and demand equations in CGEs since the underlying determinants are modeled directly. Production functions combine labor, \( L \), and capital, \( K \), so that equation (57.1) can be expressed as:

\[
X(K, L) = AX + F(P, Y)
\]

with final demand written as function of income, \( Y \), given by:

\[
Y = V_A X.
\]

Since value added depends on factor supplies, equation (57.2) should be re-expressed as:

\[
P = PA + V_A(K, L).
\]

Unlike the input–output or linear-programming models, both \( X \) and \( P \) must be solved for simultaneously. Prices appear throughout the model in a more integral way, causing substitution of both goods and factors and determining incomes.

It follows that the elasticities of substitution must be carefully calibrated for each application. Overestimating these elasticities implies a failure to recognize structural rigidities that may be present in the actual economy. Models in which response elasticities are too high underestimate the effect of policies, since the model allows adjustment in both production and consumption to be smoothly and easily accomplished. In the real economy, there may be significant transactions costs associated with substitution, and thus policies may be more effective in the real economy than in the model.

Dynamic CGEs are more cumbersome, and to the extent that they are designed to reflect Walrasian dynamic adjustment mechanisms, with perfect foresight, are less realistic than models which depend on an explicit investment function. The latter can employ parameters that are econometrically estimated to enhance realism. Dynamic CGEs can be calibrated to historical time series in the same way as large econometric models can be, and provide much more detailed and consistent information than typical time-series models.
Environmental planning models

Computable general equilibrium models in theory can be extended to address a range of related policy problems, such as environmental components. So long as stable contaminant coefficients can be found and linked to production and consumption levels, the models can generate an endogenously determined estimate of environmental quality along with its forecasts for production, consumption, investment and international trade. There are several important problems of implementation, however, the first of which is that contaminant levels can vary significantly between two industries that have been aggregated into a larger category, and even within an industry pollution levels can vary between two firms. Moreover, the coefficients presently in use are derived from studies of US manufacturing firms and one can only guess how these coefficients would need to be adjusted to conform to conditions in developing countries.

Environmental policy analysis thus requires considerable sophistication. Without some detailed microeconomic analysis built into the model, it might become difficult or impossible to judge how firms would react to the introduction of tradable emissions permits, that is, pollution rights that can be bought or sold in a specified market. Earlier planning models could adequately capture a command-and-control system that targeted output levels, but would fail to capture more nuanced response to cap-and-trade policies, such as the time-phasing of investment in compliant technologies. Moreover, models that do not include a feedback loop from the toxic contaminants to price or output levels would also fail to capture reality. While environmentally augmented CGEs have been employed in a small subset of developing countries, they are in their infancy.

Growth and long-term planning models

Even if resources are efficiently allocated statically, a sequence of Pareto-optimal states need not be Pareto optimal when viewed as a sequence (Dorfman et al., 1958). Hence markets may function well to allocate resources over space, yet do a poor job over time. This is especially difficult when the allocation problem stretches over generations, some of which are not yet born. Heal has argued recently that markets systematically err in valuing the future. Thus, inadequate capital accumulation due to uninsurable risk, credit rationing, asymmetric information and other imperfections is related to, though not the same thing as, imperfections that block trades between agents who happen to be alive at the same time. In this limited but important regard, the coefficients in the planner’s objective function may be more accurate than market-determined weights.

Growth models have a distinguished history in planning, stretching back to the 1920s when issues of capital accumulation were first addressed in the
Soviet Union. Following Fel’dman’s work in the USSR, Indian statistician P.C. Mahalanobis in 1953 developed a two-sector model that examined the allocation of investment between capital-good-producing and consumption-good-producing sectors and implied that investment needs to be allocated to the former sector to increase the rate of growth of the economy. By implicitly ignoring the agricultural sector (under the assumption that cheap food could be extracted from the agricultural sector through favorable terms of trade), it was a real failure of planning inasmuch as agriculture stagnated and ‘cheap food’ became expensive, sometimes prohibitively so. The models eventually fell out of fashion.

Following the emergence of the one-sector Solow model in the 1950s, gap models, essentially aggregate growth models with both savings and trade constraints, became popular planning tools. Gap models continue to be used to resolve issues of whether faster growth will be self-canceling by stimulating imports to the point that a balance-of-payments crisis develops.

Under very restrictive conditions, dynamic planning models can be used to determine optimal accumulation paths far into the future. One of the most well-known early models in economics, due to Ramsey, employs the calculus of variations to find the optimal savings rate, the one that maximizes the discounted value of future consumption. Despite their technical sophistication, these optimal growth models, like the von Neumann model, never guided real planning exercises in any important or practical way. Similarly, endogenous growth models have been current since the 1990s, but neither have they gained much traction for development planning.

**Regional models**

Regional models comprise a final subcategory of planning models. Since data requirements are hefty and data availability is sometimes scanty, regional models have lagged in application. The exception was in Eastern Europe, where data were more abundantly available, even if fabricated out of legal necessity. It is clear, however, that in the case of India and China, which together comprise almost half the developing world, regional models are not merely desirable, but unquestionably necessary. Combining regions in China could be as misleading as aggregating North and South America, and therefore aggregate models could grossly distort the true state of economic activity.

**Micro-level planning**

In the 1970s there was an explicit attempt to integrate micro planning into comprehensive models that were used to check consistency and direct and indirect effects of policies. The two best known were the Organisation for Economic Co-operation and Development (OECD) manual written
by Little and Mirrlees (1974) and the *UNIDO Guidelines for Project Evaluation* by Dasgupta et al. (1972).

**Cost–benefit analysis**

Public sector projects for electrification, hydrological development or transportation and communications infrastructure are key components of any development plan. Costs and benefits of projects are optimally evaluated using an hierarchical methodology in which the project is sequentially evaluated at ever higher levels of aggregation. Eventually, of course, the model may not ‘see’ the project, simply because the project is too small to matter at the aggregate level.

The private sector criterion for project acceptance is either that the present discounted value of costs and benefits as they are distributed over time should be positive, or that the internal rate of return of these same costs and benefits exceeds the cost of capital to the firm. Because externalities are so prominent in developing countries, however, the private project selection procedure has long been considered inadequate for use by development planners. While the present value template itself is appropriate, it is social – rather than private – costs and benefits that must be reconciled. Shadow, rather than market, prices are then used to evaluate project costs and benefits.

As discussed above, shadow prices are intended to reflect the marginal social benefit of available resources. Computation of these shadow prices, however, is fraught with controversy due to the large number of assumptions required for their determination. Projects that would utterly fail a private screening can, perhaps, be accepted using one method of computing shadow prices, but not another. Since shadow prices purportedly measure the marginal impact of aggregate welfare, the whole procedure had now to be vetted by the political process. This dulled the technocratic gloss that project evaluation had acquired under the direction of the authors cited above.

As noted, the linear programming approach imputes a shadow value of zero to factors of production in excess supply. Much of the early literature was devoted to calculating shadow prices in specific markets: labor, both skilled and unskilled, foreign exchange, and capital markets. The United Nations Industrial Development Organization (UNIDO) guidelines developed an extensive analysis consistent with optimal accumulation paths in surplus labor economies, all done in an analytically rigorous fashion (Dasgupta et al., 1972). Recently, more elaborate economy-wide simulation models have been used to calculate shadow values, but have not escaped intense methodological criticism since so much depends on the objectivity of the price scheme.
In retrospect, it is hardly surprising that the less analytically demanding scheme of Little–Mirrlees became dominant. It is the approach to shadow pricing most widely accepted today (Little and Mirrlees, 1974). The economy is divided into traded and non-traded goods markets and there is a competitive primary factor market as well. The shadow price of traded goods is simply the border price, since the import border price is the clearest measure of what the country is willing to give up in order to secure an additional unit of a good. Similarly, if foreigners are willing to pay the border price for our exports, that stands as the next-best alternative to any domestic use. It is a straightforward application of the basic principle of opportunity cost and requires no political justification, defense or intervention.

Non-traded goods are still difficult to shadow price. If there happens to be a separate factor of production for every traded good, and input–output relationships are known, it would be possible to solve for the shadow prices of non-traded goods and factors as a function of the known traded goods prices. If the number of factors is greater than the number of tradables, then the indeterminacy must be removed by additional information. If, for example, it is possible to deduce the foregone output of a traded good upon removing a unit of unskilled labor, then we would have a measure of the shadow value of unskilled labor that could be used to reduce the number of unknowns. If the number of factors were less than the number of tradables, the system would be overdetermined and there would exist two shadow prices for the same good.

Eventually shadow prices would be calculated directly from computable general equilibrium models, but this did not fully resolve the problem either. Model structure clearly matters and moreover, shadow prices are sensitive in general equilibrium models to how projects are financed. If a project is offset by an increase in lump-sum taxes, then the effect on aggregate welfare is the simplest to calculate. But since these tax vehicles are not usually available in developing countries, one immediately has to contend with distortionary mechanisms like income or sales taxes, which add another assumption-laden level of complexity to the analysis. Other complications include economies with segmented goods (traded and non-traded) and labor markets (which may also be regulated), large informal sectors, credit rationing, an inadequately developed or captured regulatory apparatus, and the like (Squire, 1996).

Projects that do not represent Pareto improvements, since they may easily imply a loss of welfare to some members of society while others gain, can be accepted. Income distribution need not, however, be taken into account in project appraisal if an appropriate scheme of taxes and subsidies is available to compensate losers. This is a big ‘if’ however, and some
authors have tried to incorporate distributional concerns directly into the procedures for project evaluation. Government policy-makers may choose to redistribute income from current to future generations or within the current generation from one class of households to another. As Chowdhury and Kirkpatrick note, distributional weights applied to utility representations of individual households is an explicitly subjective exercise, which varies across both time and space (Chowdhury and Kirkpatrick, 1994, p. 2). Efficiency calculations are rarely of such magnitude that they cannot be reversed by small changes in weights in the aggregate welfare function. For this reason, planners have been reluctant to mix concerns of equity and efficiency.

Public investment in infrastructure projects including electrification, telecommunications, transportation and marketing facilities would seem to address problems of static and dynamic market failure. Oddly, it has been argued that there in fact has been too much investment in infrastructure. Project evaluation techniques – even when undertaken by competent economists, such as the staff of the World Bank – fail to account properly for the welfare loss in cost recovery. On the other hand, welfare losses per dollar of public revenue raised are typically calculated using static computable general equilibrium models and therefore cannot account for the dynamic market failure of the underproduction of public goods. Getting prices wrong ultimately means they will not be used for any politically sensitive decision. Planning succumbed in large measure because, in democratically organized societies, only the market has been able to claim objectivity in determining shadow values.

**Current uses of planning models**

Planning and planning models may be out of fashion, but they can still serve a useful purpose. The most obvious use is that they allow policy-makers to form quantitative estimates of the various trade-offs in preparing development policies. They can be used to comb out inconsistencies in the ways in which policy-makers believe the economy is working. The models also enhance internal communication, adding clarity to discussions within the policy establishment as well as between these individuals and politicians, the public and other interested parties, such as non-governmental organizations (NGOs). Planning models also serve as a means of external communication. The models communicate the thinking about how resources are employed and the explicit assumptions (behavioral parameters, elasticities and the like) underlying the model can be reviewed and evaluated by outsiders. Models can signal to donors that contributed resources will be used wisely and in ways consistent with broad development objectives. Finally, planning models with sufficient structural detail
also can be used to counterbalance any undue influence of generic, one-
size-fits-all models.

Proper incentives were often ignored in early planning and this was
reflected in the models themselves. More recently, CGE models explicitly
incorporate the incentive structure. They derive their strength from the
comprehensive picture they paint of the economy and can account for the
combined effects of numerous simultaneous policies, from labor markets to
exchange rates, taxes and transfers. Planners can conduct realistic ‘what if’
experiments, refining their understanding of the various channels by which
adjustment processes unfold. Some, although not all, unintended conse-
quences are likely to be anticipated, allowing for corrective policies to be
put in place.

Planning, as an institution throughout the developing world, has not
to entirely disappeared but rather has changed forms in significant ways.
Policies often have unintended consequences, most often when they are
blind to the implicit incentive structures they erect. Consequently, planning
ministries have given way to development policy management offices. The
latter explicitly strive to enhance market outcomes. Rather than having to
anticipate the various ways in which the private sector may try to evade the
planners’ directives, modern theory suggests that a market-driven approach
can yield more satisfactory results. Planners set broad overall planning
objectives and then encourage the private sector to maximize their own
interests subject to these imposed constraints. Decision-making is decen-
tralized and the social cost of compliance is minimized.

This enlightened approach takes much of the conflict out of planning
and the negative connotation associated with command-and-control is
thereby lessened. As states abandoned coercive methods, fewer trades were
blocked, and economic efficiency automatically increased. This is not plan-
ning ‘lite’, but rather a different approach that tries to exploit fully the infor-
mational content of prices rather than issue legally binding directives.

Note
1. Thanks to Diane Flaherty and the editors of the Handbook for many constructive com-
ments and criticisms.

References
Blitzer, C., P. Clark and L. Taylor (1975), Economy-Wide Models and Development Planning,
Dasgupta, P., A. Sen and S. Marglin (1972), UNIDO Guidelines for Project Evaluation, New
Dorfman, R., P. Samuelson and R. Solow (1958), Linear Programming and Economic


Introduction

The 1960s and 1970s were characterized by the rapid expansion of the public sector in both developed and developing economies. But from the 1970s there was a breakdown of the social-democratic Keynesian consensus of the preceding 30 years about the strong social and economic role of the state. Ideas of market failure came to be replaced by ideas of government failure. The 1980s therefore witnessed widespread attempts by policymakers to curb the state’s economic role. Privatization was a significant element of these attempts. Moreover, the fiscal crises faced by many developing countries in the 1980s considerably curtailed the capacity of the state to invest in state-owned enterprises (SOEs). The consequent attempts at reforms were sometimes part and parcel of structural adjustment programmes that emphasized speedy privatization. In particular, countries facing foreign exchange constraints which had to approach the international financial institutions – the IMF and the World Bank – for assistance were given such assistance subject to certain conditionalities, including privatization of SOEs.

It is interesting to note that the term ‘privatization’ was originally coined by Peter Drucker and replaced the term ‘denationalization’. The first large-scale ‘denationalization’ programme of the post-World War II era was launched by the Adenauer government in the Federal Republic of Germany. In 1961, the German government sold a majority stake in Volkswagen in a public share offering. The offering was heavily weighed in favour of small investors. Four years later another larger offering took place. Both offerings were initially received favourably. However, a later cyclical downturn in share prices forced the government to bail out many small shareholders.

The next major European initiative came 20 years later with the successful British Telecom initial public offering in November 1984. This was followed by increasingly massive share issue privatizations in the late 1980s to early 1990s. As a result, the share of SOEs in British gross domestic product (GDP) reduced from around 10 per cent to almost 0 per cent in the space of 18 years. France, Italy, Germany and Spain followed with their own programmes. Typically these were public share offerings, often launched by socialist governments.
Among Asian countries, Japan has sold only a relatively few SOEs, mostly via share issue privatization. Some of these have been truly enormous. The $40 billion Nippon Telegraph and Telephone offer in November 1987 remains the single largest security offering in history. In China, numerous small privatizations have taken place, but relatively few outright sales of SOEs. The reason may be that most Chinese SOEs are burdened with social welfare responsibilities. It will be difficult to implement a very large privatization programme since that will seriously undermine the state’s economic role.

India has undertaken a ‘disinvestment programme’ since 1991, but the targets have consistently exceeded the actual proceeds realized, and major controversies have erupted over some of the SOEs privatized. In Latin America, many countries have undertaken large-scale privatization – Chile, Mexico and Brazil being the most prominent. In sub-Saharan Africa, privatization can be characterized as a ‘stealth economic policy’ (Megginson and Netter, 2001) since few governments have openly adopted an explicit divestment strategy. But there has been more privatization than is commonly believed. For example, in South Africa the African National Congress came to power on the planks of nationalization and redistribution of wealth, but the Mandela and Mbeki governments have almost totally refrained from nationalization and have even sold off several SOEs.

In Central and Eastern Europe, privatization efforts are part of a broader effort to transform from command to market economies. Two imperatives facing these countries are the lack of financial savings and the infeasibility of foreign purchases of divested assets because of political considerations. These two imperatives compelled the launch of ‘mass privatization’ programmes that involved the distribution of vouchers to the population. The vouchers could be used to bid for shares in companies being privatized. Such mass privatizations resulted in massive reduction in state ownership. Later this technique became very unpopular, especially in Russia where a perception grew that privatization had led to robbery by the old elites and new oligarchs.

It has been estimated that cumulative value of proceeds raised from privatization exceeded $1 trillion in mid-1990. Annual proceeds peaked at $160 billion in 1997.

**Definition and techniques**

Privatization involves a transfer of ownership and control from the public to the private sector. Privatization can be accomplished in two ways. The government can sell its assets to private buyers. Or the government can stop providing a service directly and rely on the private sector to deliver the
service. Given that a government often does not unload 100 per cent of its holdings in a company, the distinction between public and private entities can become nebulous. The problem is compounded by our inability to identify precisely the point at which control passes to private agents.

The public enterprises relevant here are revenue-generating entities originally owned or controlled by the state. An example of a revenue-generating entity is a municipal corporation that gets property taxes.

There are a number of techniques of privatization (Lopez-Calva, 1998). The most well known among these is the sale of equity to the general public. This is called ‘divestiture’ (divestment or disinvestment), and may be complete or partial. Divestiture can be done through both direct sales and equity offerings. While the developed countries have often utilized equity offerings as a technique for privatization, developing countries have found the process to be difficult for a number of reasons. The inadequacy of national stock markets and the lack of domestic capital in these countries have sometimes led to a shortage of local buyers, while foreign investors, unable to obtain sound information on the enterprises offered, often lacked sufficient interest. Even in developed countries, the direct sales approach may be costly and slow, owing to the complexity of preparing each state asset for sale individually, and then ensuring that buyers observed all contract provisions.

One special type of direct sales is a management–employee buyout. Shares of an enterprise are sold or given to some combination of managers and other employees. Well-structured management–employee buyouts can sometimes lead to efficient results, since the people who know best about an enterprise, that is, the employees and managers, become the owners. It is also rapid and easy to implement. Nevertheless, experience shows that these buyouts suffer serious disadvantages. Yielding to insider interests often entails large costs in inefficiency and poor management. Insiders may also lack many of the skills necessary to function in a market-oriented economy. Further, the process is seen to be inequitable, handing employees, rather than the population at large, most of the benefits.

In a number of former socialist countries, restitution has been employed to privatize SOEs. Restitution refers to the return of state assets to their former private owners in situations where the government’s original acquisition is seen as unjust, such as uncompensated seizure. Restitution, in such cases, it is argued, is essential on moral grounds. Opponents of restitution counter that the process is necessarily selective, and therefore an unsatisfactory way of achieving justice retroactively. Moreover, private claims can often be complicated and drawn out, bogging down privatization unnecessarily. In practice, the transition countries have seldom used restitution, except for Estonia and, to a lesser extent, the Czech Republic.
In such economies, another important technique has been mass privatization. In mass, or equal-access, voucher privatization, the government generally gives away, or sells for a nominal fee, vouchers that can be used to purchase shares in enterprises. This technique has proved to be popular particularly in the Czech Republic. Voucher privatizations can not only help to overcome the shortage of domestic capital, but they are also politically popular because they address the perceived unfairness of other approaches and avoid the charges of a sell-out of national assets to foreigners. The main risk is that a dispersed ownership structure will lack the focus and power to direct effective corporate management. These problems have been partly addressed by pooling ownership interests in investment or mutual funds. The funds, however, do not always have adequate management, control and supervisory powers. In such cases, voucher privatization becomes merely ineffective absentee ownership.

Contracting out or leasing out of government services can be another technique of privatization. For example, a municipal corporation can contract out the task of garbage collection to a private party.

Associated with privatization usually are processes of liberalization and deregulation. Liberalization refers to the introduction or promotion of competition in a traditionally monopolized industry. Deregulation refers to the abolition of statutory barriers to the operation of market forces. For example, the government of India controls the prices of many commodities through the administered pricing mechanism. If some commodity is taken out of the purview of this mechanism, then this is a deregulatory measure, because the price will now be determined by market forces.

**Reasons for the establishment of SOEs**

State owned enterprises (SOEs) were created for a number of reasons. It was believed that nationalization of private sector enterprises and establishment of SOEs would provide governments with access to much-needed revenues. These profits or surpluses could then be channelized to develop the priority sectors of the economy. Implicit in this line of thinking was the assumption that the private sector would not help in the rapid and sustained development of the economy if left to itself. There was a need to control the ‘commanding heights’ of the economy, that is, the strategic industries. If the government controlled these industries, it would be able to steer the economy in the right direction and overcome critical bottlenecks. National security reasons were sometimes added as justifications to the above, particularly in the context of heavy industries.

In many developing countries, lack of private entrepreneurs was also a major concern and forced the state to take an active role in the process of industrialization. Local private entrepreneurs often were in short supply.
Even if they did exist, they might not have access to adequate capital, partly because stock markets were not well developed. In some countries, private entrepreneurs came from unpopular minorities or were linked to foreign powers.

Distributional considerations also played a role. In India, SOEs were set up in backward regions to lessen regional inequalities. SOEs were also used to increase employment generation.

In political terms, SOEs constituted important resources for state elites – politicians and bureaucrats. They could be used to provide jobs to potential voters and service constituencies (for example, a railway minister could order railway officials to provide better links to his or her constituency).

**Why privatize?**
Over time, in many countries, the performance of SOEs turned out to be, by and large, unsatisfactory. They incurred losses, or did not make as much profit as they should have, given that they had privileged access to capital, various subsidies and protection from domestic and foreign competition. The main reason for this failure seems to have been the problems of gathering appropriate information and devising appropriate incentives for the pursuit of public interest. The incentives for serving customer interests and controlling costs were usually weak. Management was given ill-defined objectives and these frequently conflicted with political objectives. ‘The accounting systems were not appropriate to the information needs required for efficient pricing, i.e. setting prices equal to long-run marginal costs and investment projects were often appraised against technical criteria rather than economic hurdle rates of return’ (Jackson and Price, 1994, p. 2) As a result, many governments started considering privatization of SOEs as a solution to these problems because it was felt that markets provide better incentives to participants and use information more efficiently. Privatization would provide greater incentives for cost minimization, encourage more effective managerial supervision and stimulate greater employee effort.

We next examine the possible impacts of privatization in economic terms. In addition to efficiency, distribution and stability factors also need to be taken into consideration.

**Fiscal impact**
When a SOE is sold off to the private sector, the government gets the sales proceeds. Further, if the SOE had been making losses and was being subsidized, then these subsidies come to an end, which further helps the government. Thus the immediate generation of revenues is supplemented by reduction in recurrent expenditures.
But does the government really gain? In the simplest case, the buyer will be willing to pay only so much as the SOE is expected to bring in the future. The discounted sum of the future stream of returns from the SOEs is what a buyer will pay (Van de Walle, 1989). The government would have got the same revenue had it not sold the SOE. Therefore, it would seem that privatization does not have any real impact on the government’s finances.

There are two reasons why privatization might still make a difference. First, a privatized firm might be expected to be more efficient than a SOE. Hence, the sum of discounted returns will be higher than that under government ownership. Secondly, the government, when it privatizes, is getting funds immediately. This added liquidity might be desirable for a number of reasons: for example, because the government might want to spend on education or infrastructure.

It is interesting to note that in theory, for a loss-making SOE the price might be negative. This is not very far-fetched. Governments have sometimes given so many concessions to the buyer to induce them to buy loss-making concerns that in effect the price has turned out to be negative.

One pertinent question here is whether selling bonds is a better means of raising revenues than equity sales. Suppose that privatization via equity sales does not change the earnings prospects of the firm in question. In countries with liquid bond markets, selling bonds might involve lower transaction costs than privatization. Moreover to garner popular support for privatization, equity is often underpriced. However, countries constrained in their ability to sell bonds, for example those facing debt crises, or those that are trying to limit borrowings to commit to an anti-inflationary policy, might be compelled to sell equity. In addition, the perceived risk of default on bonds might be higher than equity risk.

Efficiency gains
The argument for privatization often rests on the supposed superiority of the private sector in attaining the goal of economic efficiency – both allocative efficiency and productive efficiency. Proponents of privatization have argued that a change in ownership can have an important effect on economic efficiency. In SOEs, prices sometimes did not reflect scarcities properly. For example, if the government gives a subsidy for an input used by an SOE, the SOE would tend to overuse that resource. Or, if a SOE is a monopoly, then it can set its own price. SOEs therefore would not attain allocative efficiency.

It has also been argued that SOEs are likely to exhibit greater internal inefficiencies than private firms for various reasons. Public managers are given numerous and inconsistent objectives. Instead of control by
shareholders who are interested in profit-maximization, there is bureaucratic control which puts more emphasis on ‘playing it safe’. Suppose that productive efficiency requires use of an input that is not available in a competitive market. The manager in an SOE is required to obtain competitive quotations for almost everything, and hence may have no way of using this particular input because there is only one seller.

Whether such efficiency will indeed be achieved depends in turn on the goal(s) that managers pursue in private enterprises vis-à-vis public enterprises. In theoretical literature, it is often assumed that a manager of a private enterprise maximizes economic profit, while the manager of a public sector has the liberty or is forced to pursue a more diffused agenda. If this assumption is correct, then of course ownership per se becomes a determinant of efficiency, and there is no dearth of empirical research looking into the ownership issue. However, there are reasons to believe that the monitoring system and incentive system in private enterprises may not always work perfectly, and to a large extent, the efficacy of these systems depends on the market structure and the regulatory policy. Then not only is it necessary to turn our attention to a comparative evaluation of the incentive structure in these two types of enterprises, but we are forced to consider other alternatives for achieving efficiency.

In many economies today, the incidence of owner-managers is declining and most large firms are coming to be characterized by the separation of ownership from control. Even if the shareholders can be assumed to have profit-maximization as the overriding objective, managers can pursue their own goals because of the existence of asymmetric information: in general the managers have much better knowledge about market conditions, technology and their own effort levels than shareholders. Since shareholders cannot control managers, various instruments for curbing managerial discretion have been suggested. These include: product market competition which imposes a Darwinian survival requirement of profit maximization; labour market signalling which leads to diminution of market value of non-performing managers; the threat of hostile takeover and use of incentives like employee stock option plans to align the interests of managers with those of shareholders.

All these instruments have their own requirements to be successful. In many markets, competition is muted. The Old Boy network bypasses the signalling effect of non-performance. In response to the threat of hostile takeovers, managers have developed a host of takeover defences to discourage or foil such bids. Moreover, as Grossman and Hart (1980) have pointed out, a free-riding consideration which makes shareholders refuse to part with their shares in the anticipation of an imminent rise in share prices can abort attempted takeovers. The experiences of offering stock
option plans have been mixed, mainly because the design of these plans often insures managers against the downside risk.

Of course, in addition, market failures from externalities and public goods nature of certain commodities may mean that profit-maximization does not lead to efficiency.

For public sector enterprises, on the other hand, there is no market for shares and hence no market for corporate control. How does government monitoring compare with monitoring in private enterprises? In theory, the government has the ability to correct for all types of deviations between social and private returns in goods and factor markets (Yarrow, 1986). The dissatisfaction with government control stems from the fact that the ‘market’ for political control is highly imperfect. The voting public controls the government, but its knowledge about specific enterprises may be very weak, and normally it cannot vote separately on the issue of running public sector enterprises (though when the public sector constitutes a very large part of the economy, poor performance by such enterprises can become a significant political issue in elections). The resources at the government’s command can also lead to the soft budget constraint – continuing support of even non-viable and inefficient enterprises to keep them going. If it is argued that it is necessary to keep these enterprises alive to attain distribu-tional goals (for example, provision of cheap products to poorer sections), then one is faced with the question whether there are alternative and better ways of attaining these goals. Again, sometimes, a fraction of the shares of an SOE are sold to the public, and then outsiders can monitor the enterprise’s performance to some extent. However, it still remains unclear how much divestment must take place before the outsiders can exert a significant oversight on the running of such enterprises.

As already mentioned, product market competition can be one instrument for checking managerial discretion. Product market competition forces firms to minimize costs and maximize profit for long-term survival. Moreover the observance of the performance of competing firms provides shareholders and governments with additional information about managerial inputs and the firm’s true opportunities. Thus in any year, profitability will depend on the levels and the quality of managerial inputs as well as a host of other factors outside of the manager’s control. It then becomes difficult to employ profitability as a correct indicator of managerial inputs and base rewards on such a measure. However, when the profitability of other firms can be observed, this to some extent aids shareholders in disentangling managerial contribution from random factors in a firm’s performance.

It is clear that for SOEs operating in competitive markets, prices would better reflect scarcities and therefore allocative inefficiency would be less. Then the gains from privatization would also be less. On the other hand,
transforming a public sector monopoly into a private sector monopoly would also not lead to increases in allocative efficiency. We can expect large increases in allocative efficiency to be achieved when a public sector monopoly is privatized and the market opened up to other players.

_Distributional impact_

Privatization can have rather significant effects on income distribution. One can discuss distributional issues either by looking at the sources of distributional changes or by identifying the major groups who might be affected by privatization. Perhaps the most important sources of redistribution effects are the changes in the prices of privatized assets and the pricing of commodities after privatization. If, for example, water and electricity supplies are privatized and water and electricity charges go up after privatization, this can affect large segments of the population adversely. If privatized assets are sold at discounted prices, there is a transfer of wealth to the new owners from the wider public and taxpayers in particular (Vickers and Yarrow, 1988). Such sales at discounts are sometimes politically attractive because the risk of shares being unsold is minimized, and ‘because the beneficiaries tend to be more aware of their gains than the losers feel the losses’ (Yarrow, 1986). In all these instances, it is clear that different groups are affected differently. If privatization is followed by layoffs, employees are affected. If privatization leads to higher profits, shareholders gain.

Sometimes the goods and services made available by the SOE to the poor may become less accessible after privatization. For example, a privatized airline may choose not to fly on unprofitable routes. On the other hand, if privatization is accompanied by a more competitive environment, then greater varieties of commodities may be available at lower prices. Especially in the telecommunications sector, privatization has given access to new and cheaper services to the population at large.

While privatization can have some adverse distributional implications, one can also argue that the SOEs have not had a very successful record of reaching the poor and the disadvantaged sections of the population. For example, in India, it has been repeatedly shown that the Public Distribution System does not do a good job of enabling essential commodities to reach the poorest sections; it is more useful to the richer sections.

_Some key issues_

_Implementation issues_

Worldwide experience shows that implementation of privatization programmes has lagged well behind stated intentions. Barring a few countries,
privatization has been limited to small SOEs of the manufacturing and the services sector. There are quite a few problems faced by countries trying to privatize SOEs. Firstly, in some developing countries, there is a lack of well-established, competent management consulting groups, accounting firms and investment bankers. These are needed to provide technical advice and valuation of SOEs. As a result, in some instances, foreign experts have been brought in.

Secondly, a valuation of the SOE has to be carried out before it can be offered for sale and the valuation exercise has faced severe problems. Valuation is a sensitive subject politically, because governments want to get high sales prices and at the same time the valuation process might raise questions about past public management and investment decisions. There have been inordinate delays in valuation. The problem is aggravated when poor records are maintained by SOEs.

Thirdly, once the valuation has taken place, administrative capacity is needed to assess buyers’ bids, arrange finance and insurance, and deal with a host of complex legal issues. Sometimes, a comprehensive rehabilitation plan for the SOE has to be designed, evaluated and financed before privatization is possible. Moreover, appropriate regulatory structures may not exist and may have to be set up, particularly when privatization leads to the creation of a monopoly.

Fourthly, capital markets in many developing countries are typically weak and poorly regulated. Large investments in equity are quite unusual. SOEs are some of the largest firms in the country and the private sector may not be in a position to fund the purchase of large assets. The private sector may also be suspicious about the government’s intentions, given the record of nationalization in the past. On the other hand, the government may not be willing to sell assets to foreign investors.

**Political constraints**

Generally, the costs of privatization are borne by a small group of people, for example the workers of the enterprise who may lose their jobs, or the suppliers who may lose favoured contracts. The benefits, however, are spread out over a large number of people, sometimes a very large section of the population. Public choice theory suggests that in such situations, it will be easier to organize opposition to the privatization programme than support. Experience tells us that in many countries, privatization programmes fail to mobilize popular support and in fact give rise to strong opposition.

Trade unions, in particular, tend to react strongly against privatization. Trade union power is often concentrated in the public sector and the public sector provides a base for such power. Unions oppose privatization, not
only because of the direct effect on employment, but also because of a fear that trade union power will be reduced in the private sector. The restructuring process generally involves laying off part of the workforce. Usually, forced dismissals are politically infeasible and only generate more opposition to privatization. Governments therefore try to adopt some kind of voluntary approach. Components of voluntary approaches that have been tried out include monetary compensation (for example through voluntary retirement schemes), retraining and redeployment. Cash-strapped governments may find it difficult to cover the cost of laying off workers. Sometimes the government agrees to accept a lower price for the enterprise in return for an assurance from the new owner that employees will be retained even after privatization. In the East German privatization programme, there is an instance where an enterprise was sold for 1 Deutschmark, because the bidder promised to retain all the workers.

Should restructuring occur before or after sale of the unit?  
Most SOEs will not fetch a good price if they are sold in their current condition. For historical reasons, many have excess workers, are burdened with obsolete machinery and technology, and often are run bureaucratically. One option before the government is to restructure these enterprises before placing them on the market, for example, by laying off excess workers, by inducting new workers with appropriate skills, by selling off non-strategic parts of the business, computerization of operations, and so on. These enterprises will then become attractive to private investors who would be willing to pay high prices for them. On the other hand, proponents of speedy privatization (the ‘big bang approach’) argue that the attempt to restructure these enterprises before sales will inevitably lead to delays and the entire momentum for privatization will be lost. Further, it is doubtful whether the governments are at all adept at restructuring.

The evidence  
Historically, it appears that SOEs have contributed quite significantly to the gross fixed capital formation of many economies. They have played an important role even in the highly successful East Asian newly industrializing countries. Moreover, there is no clear evidence of a negative correlation between the size of the SOE sector in an economy and its economic performance in terms of the rate of growth. There are also acknowledged outstanding cases of efficient SOEs, for example the giant Korean state-owned steel enterprise POSCO.

There are two separate empirical issues that must be kept in mind. One is the question of whether SOEs are necessarily less efficient than comparable private sector enterprises. The other is the question of the success of
privatization programmes. The first question can be summarized as ‘ownership matters’, that is, regardless of all other factors, the mere act of transferring a SOE to the private sector will increase its efficiency. The discussion in the section on the reasons for privatization has demonstrated that there is no a priori theoretical reason to believe that ownership alone matters.

Empirical investigations of these two questions run into several problems. For one, using profitability as a performance indicator for comparing SOEs with private enterprises misses out the point that SOEs are often established for reasons other than making profits. Therefore, if this measure is used it will tend to flatter privatization if under state ownership non-profit goals had been pursued. Even studies using profitability do not establish that SOEs are invariably inefficient. Moreover it is difficult to control for the effects of factors other than ownership which might affect a firm’s performance. Chang (2003) notes that there might be country-specific, industry-specific and firm-specific factors that are the determinants of performance differentials rather than ownership per se. Thus, countries with successful records of privatization sometimes underwent substantial macroeconomic changes that created a climate conducive for realizing microeconomic efficiency gains. So far as industry-specific results are concerned, the evidence of successful privatization in the telecommunications sector, to take one example, cannot be generalized to other sectors. Rapid technological changes in this sector have increased competition and reduced regulatory problems.

Parker and Kirkpatrick (2005) note that to assess the impact of privatization, in addition to using the correct performance measure, there is the problem of taking into account relative price changes with spillovers into other sectors of the economy, and redistribution effects on different socioeconomic groups. Their review of the empirical literature leads them to conclude that: ‘The studies vary in terms of the financial and economic performance measures and show that privatization measures can lead to widely differing results.’

**Conclusion**

Privatization in essence represents a reduction in the role of the government in the economic activities of the nation. The process of privatization has been impelled by ideology as much as by necessity. Even if we restrict ourselves to efficiency concerns, there seem to be no convincing theoretical reasons for arguing that privatization per se will improve efficiency. This is especially important in developing countries where a number of preconditions needed to carry out successful privatization programmes are often missing. It is therefore no wonder that the empirical evidence is also mixed, particularly in the context of developing countries.
References
The attention given to corruption and its causes and consequences for economic development has expanded enormously in recent years. While in the 1970s and 1980s corruption was often treated in the economics literature as a benign redistribution of economic rents, since 1990 there has been increasing recognition in both academic and policy circles that high levels of corruption can have a large detrimental impact on economic growth and development. The change in thinking reflects the increasing focus of the economics profession on the role of institutions in economic development, as a rapidly growing body of economic research in the 1990s examined the implications of different institutional arrangements on the functioning of markets and the supply of public goods. This change also partly reflects political developments – most notably the end of the Cold War, the transition from communism in Central and Eastern Europe, and the break-up of the Soviet Union – as the softening of superpower competition created conditions in which international organizations and donor agencies could be more forthright in challenging corrupt regimes in developing countries. The opening up of political regimes in other parts of the world, including increasing democratization in Africa, Asia and Latin America, also provided a window for citizens to organize and put increasing pressure on their governments to fight corruption.

Corruption can be defined as ‘the use of public office for private gain’, or it can be defined more broadly also to encompass the abuse of positions of trust in the private sector for personal gain. For the purposes of this chapter the term is limited to the narrower definition pertaining to the public sector, around which most research and policy work has focused. Fraud and abuse in the private sector is most often addressed by literature on corporate governance.

Measuring corruption
Before the mid-1990s it was generally assumed that corruption could not be measured – both because it was difficult to define and because perpetrators wanted to keep it secret – and thus it was a phenomenon that could only be discussed in generalities. Since the mid-1990s, however, survey tools have been developed to measure levels and patterns of corruption. While significant inaccuracies no doubt remain, these survey tools have added...
immensely to economists’ ability to analyze the phenomenon and its impact on other economic variables.

There are two general types of survey tools: polls of ‘experts’ and surveys of economic actors. In each case the questions can be focused either on opinions and perceptions or on actual experiences. The first type of tool asks questions of people who are considered to have expert knowledge on a particular country, and tends to focus primarily on their opinions and perceptions on various issues, including the incidence and severity of corruption.\(^2\) The second type of tool surveys economic actors – whether enterprise managers, citizens or public officials. While such surveys may also contain questions about opinions and perceptions, surveys of economic actors often try to gather information about actual experiences in dealing with government – for example, the amount of bribes paid in obtaining a business license, getting access to medical care or interacting with tax officials.\(^3\) A third type of tool, pioneered in Transparency International's Corruption Perceptions Index, is an attempt to combine all surveys on an individual country into one composite indicator that facilitates rankings and comparisons among countries.\(^4\)

Economics literature and applied policy analyses have utilized these various survey tools extensively since the mid-1990s to understand the extent and patterns of corruption and their implications for economic development. Aid donors have also relied increasingly on these tools to help focus aid in countries with lower levels of corruption.\(^5\) As the number of types of measurement tools have increased and begun to have greater practical impact in recent years, it has also become increasingly important to understand the pros and cons of different approaches and how they compare to each other.\(^6\)

**Levels and types of corruption**

The surveys undertaken since the mid-1990s have proven what casual observers already know – that levels and patterns of corruption vary widely among countries. It is indeed not really possible to talk about ‘corruption’ as one phenomenon, as public office can be abused for private gain in a myriad of different ways. Some literature distinguishes between ‘petty’ and ‘grand’ corruption, with the difference between the two being largely a function of the size of the bribe and the status of the briber. Large bribes paid to senior officials for major public contracts are seen as ‘grand’ corruption, while small bribes paid to the traffic police, for example, are classified as ‘petty’. More recent literature distinguishes between ‘state capture’ and ‘administrative corruption’, with the former referring to corruption in the formation of laws and regulations (for example bribes to legislators or regulators to tilt lawmaking in certain directions) and the latter
referring to corruption in their implementation (for example bribes to get goods through Customs or enroll children in favored schools)\textsuperscript{7}. State capture is often seen as the more pernicious, because it affects the rules by which markets function. Both state capture and administrative corruption come in various forms, depending on who the parties to the corruption transaction are and what is being bought, and both can occur on a large or small scale. ‘Unbundling’ corruption into its various types can give analysts and policy makers a more nuanced sense of its economic and political causes and consequences.

The costs of corruption

High levels of corruption can have devastating impacts on an economy and a society. Among the most pernicious are bribes that allow people to evade laws that protect public safety. Importers, for example, may bribe customs officials to allow dangerous or ineffective drugs into the market, or builders may bribe regulatory agencies to erect buildings that do not meet safety codes. Also costly to public welfare is corruption that affects access to public services, as when parents must bribe doctors or teachers for medical care or education for their children. In all cases, widespread corruption undermines citizen trust and rule of law, and thereby impedes the arm’s-length transactions among strangers that are so fundamental to a market economy.

A large body of academic analysis has been undertaken since the mid-1990s to try to measure the economic impact of corruption more precisely, with various dimensions of economic impact being studied.\textsuperscript{8} One of the first studies was the 1995 analysis by Mauro of investment in a cross-section of 67 countries, which found that corruption has a significant negative impact on the level of investment in relation to GDP.\textsuperscript{9} Further studies in the late 1990s using a variety of corruption indicators reinforced this overall finding. One such study differentiated among types of corruption and found that in settings in which bribes had less predictable outcomes – that is, where bribers were less confident about getting what was paid for – corruption had a stronger negative impact on investment.\textsuperscript{10} Another study differentiated between ‘centralized’ (or coordinated) and ‘decentralized’ (or uncoordinated) bribe-taking, arguing that the economic impact of the latter is likely to be more severe, and using the model to illustrate the increased cost of corruption in post-communist Russia as compared with the centralized monopolistic corruption imposed by the Communist Party.\textsuperscript{11} Analytic work has also focused on the negative impact of corruption on foreign direct investment (FDI). Wei found that corruption at the level found in Mexico was equivalent to a 20 percent tax as compared to that found in Singapore.\textsuperscript{12}
While corruption and gross domestic product (GDP) per capita are highly correlated, the direction of causation is more difficult to untangle. On the one hand, corruption can hamper growth by reducing the efficiency of public spending and the effectiveness of public service delivery. On the other hand, poorer countries have a more difficult time tackling corruption, both because bribes may be more tempting when public sector salaries are low, and because it takes resources to fund ‘watchdog’ groups needed to prevent corruption, such as the press, accounting and auditing services, and police and other investigative and law enforcement agencies. A large body of recent literature attempts to unravel the effects of corruption on either the level or the rate of economic growth. While many methodological difficulties make this line of research difficult and skepticism remains about the findings of individual studies, the broad consensus supports the view that corruption – rather than ‘greasing the wheels’ of commerce – has a substantial negative impact on economic growth.

Other economic costs and consequences have also been identified through research in recent years. Corruption has been found to increase inequality, and the reverse – that higher inequality leads to more corruption – has also been shown. Corruption reduces public revenues, leads to lower quality in public investments and public services, and tends to skew public spending away from education to other types of spending – such as large infrastructure projects or military procurement – where bribes are more lucrative. Corruption has detrimental effects on the environment by reducing the effectiveness of environmental regulation, and it is positively correlated with a country’s rates of inflation and crime.

Most fundamentally, corruption reduces citizens’ trust in government and the political system, which undermines the ability of governments to cooperate with the citizenry in the formulation of policies and enforcement of laws. A dysfunctional and mutually reinforcing equilibrium of citizen distrust, poor government performance and corruption can result, undermining economic growth and development in the many ways outlined above.

**Tackling corruption**

In its simplest form, corruption is facilitated by the ability of a public official to exercise discretion in the carrying out of his or her duties without having to answer to formal organs of accountability. Hence the formulation that \( C = M + D - A \) (corruption equals monopoly plus discretion minus accountability). This simple formulation points to several variables that can affect the level of corruption in an economy. A large role for government and greater discretion for individual public officials opens avenues for potential corruption, and such corruption is more likely to materialize...
if formal institutions of government accountability are weak. Corruption is also more likely when valuable and sought-after assets are under government control, as, for example, in countries with abundant state-owned natural resources (for example, oil- and gas-producers) or in transition countries with large-scale privatization programs. The simple lesson would be to minimize government’s role and public officials’ discretion while strengthening oversight institutions in an economy.

Reality is much more complex, however. First of all, although people may disagree at the margin about the appropriate roles for government, such roles will never be eliminated entirely, as there are certain public goods (for example, defense, law and order, basic infrastructure and education, environmental protection) that governments have an important role in supplying. Furthermore, giving discretion to public officials is often critical to getting good results, particularly in more complex areas of public policy. Building in mechanisms to ensure transparency and accountability in public decision-making – for example through public hearings on draft laws, ‘freedom of information’ and publication of government decisions, or internal or external audit procedures – is always desirable, but these can be costly or difficult to implement, particularly in poor countries with few public resources and a severe shortage of skills. The challenge in any particular case is to understand these trade-offs and try to tailor the role of government, the discretion granted to officials, and efforts to strengthen accountability to the particular needs and characteristics of each country situation.

Controlling corruption also requires the active commitment and involvement of the citizenry. Formal public law enforcement is expensive and necessarily operates only at the margin in any country; most laws are enforceable in practice because citizens willingly obey them and are willing to take action when someone else breaks them. If, in contrast, people lack confidence that laws can be enforced, this can become a self-fulfilling prophecy, as they will neither respect and follow the laws themselves, nor bother to report others who break them. One can envision two situations of equilibrium: one where most people follow the law and expect others to follow it, and thus where transgressions are few and can be managed by formal law enforcement; and the other where most people do not follow the law nor expect others to, and transgressions are too numerous for formal law enforcement to handle. Both equilibrium conditions are common in today’s world: corruption is the rare exception in some countries, but it is systemic and widespread in others. The challenge that many systemically corrupt countries face is how to move from one equilibrium state to the other.

In practice, governments that are strongly motivated to reduce corruption have many policy and institutional levers they can use. A multi-pronged
approach to tackling corruption would address many institutional dimensions, as shown in Figure 59.1.23

State capture tends to thrive when the private sector is monopolized, and economic conditions and policies that enhance private sector competition and transparency will tend to reduce incentives and opportunities for corruption. These include, for example, policies that promote international trade, low entry barriers for new firms, a stable macroeconomic framework with low inflation, and well-designed regulatory, corporate governance and anti-monopoly laws.24 Competition and transparency in the political realm can also help, including clear and binding rules for political party financing or asset declaration and conflict of interest rules for senior government officials.25 More fundamentally, government structures that limit power and create horizontal or vertical checks and balances – for example, through legislative or judicial oversight of government decisions or through multi-tiered layers of government27 – can greatly enhance political competition and accountability. On a day-to-day basis governments can take steps to improve public sector management by adopting rules and enhancing incentives and skills for meritocratic civil service staffing,29 honest tax and customs administration, and transparent public procurement and budgeting procedures. As there are almost always losers as well as winners in these reforms, governments must make intensive efforts to build public support and publicize early ‘wins’ to overcome countervailing pressures. Governments can also restrict opportunities for corruption by taking steps to reduce their role in an economy – for example, by privatizing commercial firms and reducing regulations on business activity. While it is true that cross-country regressions do not show a statistically significant correlation between the overall size of government and the level of corruption,30 it is also true that opportunities for corruption can be reduced in any particular country by shedding unnecessary activities in the public sector. Finally, governments can help create the legal basis and political openness for a vibrant and independent media and a strong civil society to function. Both are critical counterweights to government power and are essential to the control of corruption.

Reforming the political system to increase transparency and accountability is part of the challenge, as noted above. However, the presence or absence of democracy per se is not necessarily the determining factor. Young democratic systems with limited transparency and accountability can be highly corrupt, and pay-offs often increase during election years as politicians trade bribes for votes or private parties buy jobs or favors from new governments. However, political reformers and the international community need to persist in promoting democratic values, because there is clear evidence that a longer exposure to democracy lowers corruption.32
Figure 59.1 Addressing corruption requires action across many fronts
And when there is a window of opportunity to change the design of the democratic system, reformers should note that parliamentary systems appear to do a better job than presidential systems in controlling corruption, at least as long as political parties are generally oriented toward the public interest. The Westminster system of democracy found in the UK – with a long and unbroken history, direct election of candidates in lieu of proportional representation, strong and accountable political parties, high electoral participation, and voting districts that are not unduly small – is considered by many to be the model of democracy most likely to control corruption.

The question of whether decentralization of government power reduces corruption is a subject of active debate, and many aid donors are pressing for further decentralization and community participation in countries with weak governance. Decentralization advocates argue that moving government closer to the citizenry will force greater transparency and accountability, while more cautious observers question the capacity of local governments (particularly in poor countries) and the likelihood of ‘capture’ of decentralized power by local elites. On the one hand, it is clear that some of the best-governed countries in the world (Denmark, for example) are also among the most decentralized. On the other hand, it is also clear that state and local governments in many settings (the United States, for example) have been more prone to corruption scandals than national governments. The extensive research on this issue is inconclusive and suggests that the details of each country’s historical, cultural, political and economic situation are likely to be more important than the extent of decentralization per se in determining how decentralization will affect governance and corruption.

Progress is possible
Experience around the world shows that strong and committed leaders can indeed reduce corruption in government. Among the more advanced countries, for example, both the United States and the United Kingdom undertook far-reaching civil service reforms in the nineteenth century that helped to reduce corruption. More recently, corruption in state governments in the USA is widely thought to have been far worse 50 years ago than it is today. And, more recently still, many of the former communist countries of Central and Eastern Europe appear to have made significant strides in reducing corruption since the mid-1990s. In-depth surveys of businesses undertaken in 1999, 2002 and 2005 indicate that the frequency of bribery in many types of public services and the constraint that corruption imposes on doing business declined significantly in many countries over this period. A combination of better economic policies, stronger
economic and political institutions, faster economic growth and a more vibrant civil society combined to put pressure on governments for greater honesty and accountability.

Eliminating all corruption is not feasible. It exists and will continue to exist in every government in every country in the world. However, a much more feasible – and in the end more important – goal in every country should be to reduce the level of corruption and its negative social and economic impacts. For developing countries with widespread and systemic corruption, the goal is to move to an equilibrium where corruption is the exception rather than the norm. There is little doubt that declining corruption, enhanced government effectiveness, greater citizen trust and increased private investment are mutually reinforcing phenomena that together spur economic growth and development.

Notes
1. See, for example, North (1990), Bardhan (1997a) and World Bank (2002).
4. Examples include the Transparency International (TI) indicators (www.transparency.org) and indicators put together by the World Bank Institute (www.worldbank.org/wbi/governance).
5. For example, both the level of resources made available to poor countries by the World Bank’s International Development Association (IDA) and the eligibility of countries to access grants from the US Millennium Challenge Account (MCA) are dependent in part on countries’ scores on various governance and corruption indicators.
7. For further explanation of the concept of state capture, see World Bank (2000) and Hellman et al. (2006).
8. For a synopsis of some of the vast literature on corruption and development, see Bardhan (1997b) and Lambsdorff (2006).
13. Some recent research that attempts to use instrumental variables to address these simultaneity problems finds that corruption leads to lower GDP per capita. See, for example, Hall and Jones (1999) and Kaufmann et al. (1999).
17. Tanzi and Davoodi (1997) and Gupta et al. (2002).
21. La Porta et al. (1997).
24. Research points to the value both of low international trade barriers and of a uniform (rather than highly diversified) tariff structure in reducing opportunities for corruption: Gatti (1999) and Ades and Di Tella (1999). Wei (2000a) argues that a country’s ‘natural openness’ (that is, small size and non-remote location) is more important than trade policy per se in increasing competition from trade and thereby reducing corruption.
26. With regard to electoral rules, there is some support for the view that corruption is less likely in countries where voting districts are larger (allowing greater political competition) and where citizens vote for individual candidates rather than party lists (which increases individual accountability of politicians to the electorate). Persson et al. (2003).
27. For an analysis of the importance of an independent judiciary in controlling corruption, see World Bank (1997) and Ades and Di Tella (1997).
28. See fuller discussion on decentralization below.
29. Evans and Rausch (2000). Raising civil service salaries can also help to reduce corruption (van Rijckeghem and Weder, 2001), but, above a certain reasonable minimum, raising salaries is not as important as promoting merit-based hiring and promotion.
30. It is not surprising that cross-country regressions show no significant correlation between government size and the level of corruption, given the two-way nature of causation. While an overextended public sector may create more opportunities for corruption, corruption (and the low level of trust and public sector effectiveness it causes) may also make it more difficult for governments to collect public revenue. Thus only the more effective governments may have the capacity and citizen trust to grow and remain large. Indeed, many of the world’s least corrupt countries – most notably in Scandinavia – are also among the countries with the largest public sectors. When the Scandinavian countries were omitted in one study, size of government was found to be positively correlated with the level of corruption. LaPalombara (1994).
34. With regard to culture, Husted (1999), drawing on Hofstede (1997), finds that more hierarchical, more materialist, and more risk-averse cultures are likely to be more corrupt.
35. A large number of studies have been undertaken in recent years to try to assess the links between decentralization and corruption, with some supporting the view that greater decentralization is correlated with lower corruption and some refuting that view (generally finding that the correlations disappear when the sample changes or when decentralization is measured in a different way). Another strand of this literature looks at federal and unitary states, with some studies finding that federalism is correlated with higher levels of corruption and others showing no effect. See Lambsdorff (2006) for further discussion.

References


Melbourne: Cambridge University Press.
599–617.
Monetary Fund Working Paper 97/139*.
van Rijckeghem, C. and B. Weder (2001), ‘Bureaucratic Corruption and the Rate of
Rule of law

In his manuscript *Lectures on Jurisprudence*, based on his lectures at Glasgow University in the early 1760s, Adam Smith stated that a factor that ‘greatly retarded commerce was the imperfection of the law and the uncertainty in its application’. This is still one of the main messages of the Law and Economics literature as it pertains to development. Law and Economics is a thriving subject in the USA, and it is now being widely adopted in other countries, including in law schools of developing countries. But its Chicago origins and the general American mold may have given a particular slant to the development of the subject, which is not always quite appropriate for these countries. In this chapter I shall focus very generally on some of the special issues that arise in the context of developing countries that the literature on Law and Economics needs to address if it is to be applicable there. These special issues arise primarily because the institutional, political and behavioral context in these countries is different from the usual context of the literature.

In this literature as well as that of recent Institutional Economics the major emphasis is on contract law and security of property rights. In the pervasive context of incomplete contracts the emphasis is rightly on the residual rights of control, and the security and predictability of these property rights are crucial for economic performance and long-term investment. Throughout history in any time-separated activity – for example, if the seed planter cannot be secure in reaping the harvest, if a trap-setter cannot claim the trapped game, or a lender is uncertain of being repaid – economic life is hampered by insecurity of property rights. North and Weingast (1989) trace the success story of development in English history to the King giving up royal prerogatives and increasing the powers of the Parliament in 1688, thus securing private property rights against state predation and allowing private enterprise and capital markets to flourish. The more recent empirical literature has tried to quantify the effect of these property rights institutions – or what they call in this literature the ‘rule of law’ variable (one standard measure combines indices of effectiveness and predictability of judiciary, enforceability of contracts and incidence of crime) – on economic performance from cross-country aggregative data. Since these institutions may be endogenous (that is, economically better-off countries may have
more of those institutions, rather than the other way round), the literature tries to resolve the identification problem by finding exogenous sources of variations in those institutions. See Acemoglu et al. (2001, 2002). Rodrik et al. (2002) use similar data to show that once the property rights institutions are accounted for, the role of other factors like geography or openness to trade in explaining cross-country variations in per capita income is minimal.

What is often ignored in this literature is that the ‘rule of law’ actually involves a whole bundle of rights, and we need to ‘unbundle’ it. Even for security of property rights, different social groups may be interested in different aspects of these rights. For example, the poor may be interested primarily in very simple rights like land titles, and also, to a very important extent, in protection against venal government inspectors or local mafia; to them that is the most salient aspect of security of property rights. For the richer investors, however, a whole range of other issues like protection of the minority shareholders in corporations, oversight of capital markets against insider abuse, bankruptcy laws, and so on, loom large; these are what investors emphasize when they talk about security of property rights. As different groups are thus interested in different aspects of security of property rights, these rights may have differential political sustainability, depending on how politically influential the corresponding groups are in a given polity.

‘Rule of law’ should also include other rights, some quite different from mere security of property rights. For example, one part may involve various democratic rights of political participation, association, mobilization and expression of ‘voice’. An analysis of cross-country variations in human development indicators (which includes education or health variables like mass literacy or life expectation) shows that an institutional variable measuring ‘voice’ or participation rights is just as important as that measuring security of property rights as an explanatory variable; – see Bardhan (2005), Chapter 1. In other words, the part of ‘rule of law’ that refers to democratic participation rights explains a significant amount of variation in human development indices across countries. Those who emphasize property rights often ignore the effects of participatory rights, and there is some obvious tension between these two types of rights included in the standard package of ‘rule of law’.

The idea of security of property rights has been extended to the case of intellectual property rights for the preservation of incentives for innovation. Since innovations are the main source of economic growth, laxity in the enforcement of international patents and copyrights in developing countries for products that are knowledge-intensive or require expensive investments in research and development is often regarded as harmful for
long-term economic growth. This has been the rationale for the incorporation of TRIPS (Trade-Related Intellectual Property Rights) in World Trade Organization (WTO) rules, when developing countries accepted these rules under some pressure from rich countries. While keeping incentives alive for new research and innovations is extremely important, the question from the point of view of a developing country is usually whether the enormous costs (including the often exorbitant monopoly prices charged by the patent holder for a prolonged period) are always worth the benefits, and whether there are better alternative ways of encouraging research. It is recognized now by many scientific researchers that existing patents often act as an obstacle to further research that tries to build on earlier findings (in developing countries this includes research for adapting new technology to the special conditions there). This is linked with the question of the optimal patent breadth, which is about how broadly the protection of existing innovations ought to extend to related innovations in the future. The alternative method of subsidizing research inputs (rather than rewarding research output with temporary monopoly) has the advantage of encouraging information-sharing and collaborative research. Of course, upfront funding carries with it the moral hazard problem that researchers, once having secured funding, may be tempted to pursue activities or lines of research other than those most desired by the public sponsor. This problem may be mitigated if researchers expect to apply for public funding in future.

The problem of international patents in life-saving drugs in poor countries recently caught public attention in connection with the controversies about the prices of anti-retroviral drugs for AIDS patients in Africa. The major problem in corporate drug research is that only a tiny fraction of what the companies spend on finding new diet pills or anti-wrinkle creams is spent on drugs or vaccines against major killer diseases of the world’s poor, like malaria or TB, and the situation has not changed with the onset of TRIPS and is not expected to change even with a more stringent enforcement of TRIPS in poor countries. So alternative avenues of encouraging such research have to be sought. There are now the beginnings of some international attempts to make credible arrangements on the part of international organizations like the World Health Organization (WHO) in collaboration with non-governmental organizations (NGOs) like Médecins sans Frontières, private foundations (like the Gates Foundation) and donor agencies and governments to a commitment to purchase vaccines to be developed by pharmaceutical companies against some of these diseases. For a discussion of the incentive issues in vaccine purchase commitments, see Kremer (2001). For other diseases (like diabetes or cancer) which kill large numbers of people in both rich and poor countries, the incentive argument for enforcing patents in poor countries is weak, since that research will
be carried out by the transnational drug companies in any case as the market in rich countries is large enough (provided resale can be limited).

We have earlier commented upon the different kinds of property rights being relevant for different social groups. In the case of intellectual property rights as well, the transaction costs may limit the symmetry of access of different groups to those rights. Khan and Sokoloff (1998), in a historical comparison of the patent systems in the USA and Britain in the first half of the nineteenth century, show that while the British system used to effectively limit access to intellectual property rights to the relatively wealthy and well-connected, access in the American system was much more broad-based, and this contributed to a much more vigorous and wider spread of patenting activity in the USA in that period.

**Beyond formal legal institutions**

While nobody will deny the importance of innovations in the process of economic growth, in the case of manufacturing technology in most developing countries the problem is really in adaptation of technology theoretically available elsewhere. Much of the effective use of that technology particularly in these alien circumstances is not codified, but implicit or tacit, and cannot be just transplanted from abroad. Learning by doing and domestic efforts to adapt and assimilate are critical, costly and time-consuming, and in this, government investment in market-supporting infrastructure and in research and training and extension are quite important. Just putting in place a legal system facilitating private efforts may not be enough. As Pack (2003) points out, in recent years many developing countries have liberalized domestic and international trade regulations but have not realized high total factor productivity, in the absence of a set of institutions constituting a national innovation system and extension services that facilitate appropriate training and technology absorption.

There are also corresponding implications for the inadequacy of just a legal framework in developing credit and equity markets or the requisite financial infrastructure in general. Investment in learning by doing is not easily collateralizable and is therefore particularly subject to the high costs of ‘imperfect information’. At an early stage (which can be prolonged in poor countries) when firms are not yet ready for the securities market (with its demands for codifiable and court-verifiable information), there is often a need for some support and underwriting of risks by some centralized authority (with, of course, its attendant dangers of political abuse). There is also the problem of interdependence of investment decisions with externalities of information and the need for a network of proximate suppliers of components, services and infrastructural facilities with large economies of scale. Private financiers willing and able to internalize the externalities
of complementary projects and raise large enough capital from the market for a critical mass of firms are often absent in the early stage of industrialization. Historically, the state has played an important role in resolving this kind of coordination failure by facilitating and complementing private sector coordination – as the examples of state-supported development banks in nineteenth-century France, Belgium and Germany, and more recently in Japan, Korea, Taiwan and China, suggest. There are, of course, many examples of state failures in this respect and politicization of financial markets in other developing countries. In much of the literature on Law and Economics, as in Institutional Economics, the importance of the state is recognized only in the narrow context of how to use its power in the enforcement of contracts and property rights, and at the same time how to establish its credibility in not making confiscatory demands on the private owner of those rights. The history of the successful as well as failed cases of the state as a coordinator of technology assimilation and financial market development has lessons which should be analyzed in a framework that goes beyond this narrow context.

Why does a society not always adapt its legal and institutional set-up to facilitate productivity-enhancing innovations? Such innovations have gainers and losers, but in most cases the gainers could potentially compensate the losers. The problem is that it is politically difficult for the gainers from a change to commit credibly to compensate the losers ex post. As Acemoglu (2003a) puts it, there may not be any political Coase Theorem, whereby politicians and powerful social groups could make a deal with the rest of society, give up some of their control on existing rules and institutions that are inefficient, allow others to choose policies and institutions that bring about improvements in productivity, and then redistribute part of the gains to those politicians and groups. Such deals have severe commitment problems; those in power cannot credibly commit to not using this power in the process, and others cannot credibly commit to redistribute once the formerly powerful really give up their power for the sake of bringing about new rules and institutions.

A central issue of development economics is thus the persistence of dysfunctional regulations and institutions over long periods of time, as we discuss in Bardhan (2005), Chapter 2. In particular, the history of underdevelopment is littered with cases of formidable institutional impediments appearing as strategic outcomes of distributive conflicts. Acemoglu and Robinson (2002) develop a theory where incumbent elites may want to block the introduction of new and efficient technologies because this will reduce their future political power; they give the example from nineteenth-century history when in Russia and Austria-Hungary the monarchy and aristocracy controlled the political system but feared replacement and so
blocked the establishment of rules and institutions that would have facilitated industrialization. These replacement threats are, of course, often driven by extreme inequality in society.

In explaining the divergent development paths in North and South America since the early colonial times, Engerman and Sokoloff (2002) have provided a great deal of evidence of how in societies with high inequality at the outset of colonization rules and institutions evolved in ways that restricted to a narrow elite access to political power and opportunities for economic advancement. Initial unequal conditions had long-lingering effects, and through their influence on public policies (in distribution of public land and other natural resources, the right to vote and to vote in secret, primary education, patent law, corporate and banking law, and so on) tended to perpetuate those institutions and policies that atrophied development. Even in countries where initially some oligarchic entrepreneurs are successful in creating conditions (including securing their own property rights) for their own economic performance, as long as that oligarchy remains powerful, they usually get away with regulations that raise entry barriers for new or future entrepreneurs, and this blocks challenges to their incumbency and thus sometimes new technological breakthroughs. See Acemoglu (2003b) for a theoretical analysis of this kind of dynamic distortion in oligarchic societies even when property rights are protected for the initial producers. The classic example of inefficient rules and institutions persisting as the lopsided outcome of distributive struggles relates to the historical evolution of land rights in developing countries. In most of these countries the empirical evidence suggests that economies of scale in farm production are insignificant (except in some plantation crops), and the small family farm is often the most efficient unit of production. Yet the violent and tortuous history of land reform in many countries suggests that there are numerous roadblocks on the way to a more efficient reallocation of land rights, put up by vested interests for generations.

Inequality in power distribution in society also influences the social legitimacy of laws enacted or decreed by the powerful, and the degree of commitment of the general population to the rule of law. When the state is captured by a narrow clique, or when the state is weak so that there is an ‘oligopoly’ of coercion and authority (as opposed to the ‘monopoly of violence’ that Max Weber attributed to the state) shared by various protection rackets and corrupt officials (police, judges, bureaucrats), there is usually a big gulf between laws that are in the statute books and their enforcement, and, most importantly, a deficiency in every citizen’s expectations about others’ compliance, which form the foundation of the rule of law. Along with the underlying power distribution and enforcement mechanisms in society, some overarching social norms and political commitments provide
the main structure within the confines of which the formal legal system operates, and compared to the former the latter – which is the focus of much of the Law and Economics literature – is often in a secondary role.

These important elements of the institutional, political and social framework are ignored in a recent burgeoning of empirical literature on the effects of legal origins of a system. La Porta et al. (1997, 1999) have called attention to the superior effects, across countries, of the Anglo-Saxon common law system based on judicial precedents over the civil law system based on formal codes, on the corporate business environment both in terms of more flexibility with changing needs of business and in terms of better protection for external suppliers of finance to a company (whether shareholders or creditors). Apart from some doubts about the establishment of causality in these cross-national studies, one can also question the historical evidence even in the rich countries themselves. Lamoreaux and Rosenthal (2005) have done a comparative study of the constraints imposed by their respective legal systems on organizational choices of business in the USA (with its common law system) and France (with its civil law codes) during the middle of the nineteenth century around the time when both countries were beginning to industrialize. They conclude that there was nothing inherent in the French legal regime that created either a lack of flexibility or a lack of attention to the rights of creditors or small stakeholders. Many of the rules in the USA for minority shareholder rights actually came after the insider scandals of the Great Depression period. Franks et al. (2003) point out that in the UK it was not until as late as 1948 that the Parliament began to enact limited legislation to protect minority shareholder rights. Rosenthal and Berglof (2003) also question the primacy of legal origin in explaining institutions of investor protection; drawing upon the legislative history of US bankruptcy law they show how the USA, with an English common law legal origin, ended up with a bankruptcy regime quite different from that in the UK, and how political and ideological forces shaped financial development. Several legal scholars – see, for example, Roe (2003) – have pointed out how the nature of corporate governance even in American large firms depends more on socio-political factors than on the form of corporate laws.

In any case, as we have indicated earlier, the importance of the legacy of the formal legal system is rather moot where much too frequently in developing countries the enforcement of whatever the laws are in the statute books is quite weak, and the courts are hopelessly clogged and corrupt. Take the two largest developing countries, China and India. India has inherited the English common law system, and being a democracy, legal rights there are more well defined and the legal system is less subject to political discretion than in China under the monopoly control by a Communist Party. And yet, according to the World Bank Report on Doing
Business in 2005 (World Bank, 2005), it is China which seems less disad- 
vantaged in most indicators of regulatory and judicial effectiveness in 
business matters. For example, registering property requires 67 days and 
costs about 14 per cent of property value in India, whereas in China it is 32 
days and 3 per cent of property value. In enforcing debt contracts it requires 
425 days and costs about 43 per cent of debt value in India, whereas in 
China it is 241 days and 26 per cent of debt value. On closing an insolvent 
business it takes about ten years in India, in China 2.4 years.

In many developing countries the efficiency of courts as mechanisms of 
resolving disputes or enforcing contracts is shaped by a rather warped 
system of incentives: judges, even when they are not corrupt, do not care 
about delays, lawyers earn more when court proceedings are prolonged, 
appeals are too easy and some defendants deliberately seek continual delay 
in judgment. Courts are congested because of too-lengthy procedures and 
built-in incentives for over-litigation, apart from administrative delays in 
appointments of judges. Such low judicial effectiveness in commercial law, 
apart from raising transaction costs all around, has important effects on the 
size and structure of firms. This is because the more effective the judicial 
process, the more you can have relatively complex contracts, larger firms 
can thrive and more complex goods be produced.

Social and behavioral presumptions
Finally, I am going to comment on some of the broad presumptions of the 
Law and Economics literature which may need to be changed or made more 
flexible if it is to be applied to developing countries. One relates to the scale 
of economic activity. In small peasant communities where the scale of eco-
nomic activity is not large, informal relational contracts may be more 
efficient than rule-based contracts supported by elaborate legal–juridical 
procedures. Breaches of relational contracts are often observable by other 
community members even when not verifiable by courts, and punishment 
is usually through social sanctions and reputation mechanisms. Another 
advantage is flexibility and ease of renegotiation. But as the scale of eco-
nomic activity expands, as the need for external finance becomes impera-
tive, and as large sunk investments increase the temptation of one party to 
renege (and as increased mobility and integration with the outside world 
raise exit options), relational contracts and reputational incentives 
become weaker.6 As Li (2003) points out, relation-based systems of gover-
nance may have low fixed costs (given the pre-existing social relationships 
among the parties and the avoidance of legal–juridical and public infor-
mation and verification costs of rule-based systems), but high and rising 
marginal costs (particularly of private monitoring) as business expansion 
involves successively weaker relational links.7
Of course the transaction costs of legal–juridical systems are asymmetric in their incidence on the rich and the poor as they try to get legal remedies, and it is not surprising that the legally handicapped poor often feel that the law is just another ‘stick’ with which the resourceful rich can beat them. In small face-to-face communities what anthropologists call the ‘politics of reputation’ may provide some modest measure of protection for the weak against the strong; as long as all parties belong to what is perceived to be the same ‘moral community’ in terms of which reputation is defined, there are some accepted limits and symbolic sanctions against the kind of ruthless exercises of power that sometimes accompany the cut-throat impersonality of the legal system enforced by the gendarmerie of the state.

It also needs to be recognized that in a world of highly imperfect information and the interlinked and multiplex nature of traditional informal contracts, the establishment of market relations enforced by the legal system in one market can crowd out implicit contracts in other related markets. Kranton and Swamy (1999) show in a study of the impact of the introduction of civil courts in British India on the agricultural credit markets of the Bombay Deccan that while it led to increased competition in the credit market, it reduced lenders’ incentives to subsidize farmers’ investments in times of crisis, leaving them more vulnerable in bad times, with insurance markets largely absent. In the context of environmental management of the village commons, Seabright (1993) has pointed out that as contracts are necessarily incomplete, attempts to enforce private property rights may weaken the mechanisms of cooperation that previously existed among the resource users, who may have shared implicit non-contractual rights in the common property resource.

The Law and Economics literature has inherited from mainstream economics the latter’s behavioral postulate of rational self-interested individuals. This postulate is being increasingly questioned in the branch of economics that is now called ‘behavioral economics’, but there may be special reasons for questioning it in the context of poor countries. In traditional communities where your conformity to community norms is at a special premium, we may have to pay particular attention to social preferences (‘other-regarding’ as opposed to self-centered, or ‘process-regarding’ as opposed to simply outcome-oriented) which may go beyond the narrow interpretation of self-interested behavior. For example, social reciprocity (individuals going out of their way to reward helpful actions by other members of the community, or taking revenge for perceived unfair or nasty behavior on the part of others at some considerable cost to the revenge-taker – ‘honor killings’ in many traditional societies being the extreme but not uncommon case) is often a foundation stone of community norms,
which define the informal institutional framework within which particular legal rules can be implemented.

It is also a questionable presumption of the Law and Economics literature that individuals always behave in their best interests. Common observations of myopic, weak-willed, procrastinating and time-inconsistent behavior fly in the face of the inexorably rational economic man of our textbooks. This may be a special problem in poor countries where public information media are weak, many people are uneducated and superstitious, and there is a surfeit of touts, middlemen and operators trying to manipulate people to make hasty uninformed decisions. The innate psychological characteristics of people may not be different in poor countries, but their circumstances and information sources are often quite different, and capacity for complex calculations is an acquired trait, honed only as transactions become more complex. Also, people often internalize their constraints and by all accounts the constraints are much more severe in the case of poor people. All this may sometimes call for more paternalistic regulations than are admitted in the rational-choice framework of Law and Economics. For example, consumer protection regulations in food labeling and health warnings, publicizing of information about often the exorbitant implicit interest rates charged in instalment purchases of durables from retailers and pawnbrokers, publicizing the odds of winning lotteries (which are very popular, as most people systematically overestimate their chance), are all instances of paternalistic regulations that are particularly important in poor countries. One, of course, has to be wary of the slippery slope here that may easily end up in heavy-handed regulations or regulatory capture, but one cannot deny that the sovereignty of the rational consumer is a particularly egregious myth in such contexts.

Furthermore, the Law and Economics literature, particularly through its Chicago origins, has inherited a presumption about voluntary contracts that one may have to be careful about. Milton Friedman and others have repeatedly asserted that if parties enter into a transaction voluntarily (without adverse effects on third parties), legal rules should not interfere; they should play only an enabling or facilitating role in that transaction. There are, however, many cases, particularly in poor countries, where it is possible to show that one party in this transaction would have been actually better off if the law intervened to take out certain options from the choice set. Take the case of ‘bonded labor’. Genicot (2002), in describing what she calls ‘the paradox of voluntary choice’, constructs a case where the strategic interaction between the landlord and the local credit institutions can constrain the poor peasant to ‘choose’ a bonded labor contract, whereas if bonded labour were banned it would have resulted in welfare-enhancing credit opportunities for the peasant. Basu (2000) models a
somewhat similar case of a woman choosing a ‘sexual harassment contract’ where she would have otherwise been better off if such contracts were disallowed. Similar cases can be argued for legally taking out the option for a poor worker to work in unsafe or hazardous conditions. These are all cases for interventionist regulations in the context of extremely unequal but ‘voluntary’ contracts.

Let us end with a comment on a fashionable attitude to the rule of law in the context of development that is sometimes expressed at the opposite end of the political spectrum. We have indicated earlier in this section as well at the beginning of this chapter that the rule of law is often an instrument in the hands of the propertied, ruling over and restricting the activities of the propertyless. This undoubted fact sometimes leads commentators to dismiss the rule of law merely as an instrument of class oppression or as part of a modernizing elitist project that rides roughshod over the ‘subaltern’. In the face of such tendentious simplifications we can do no better than to quote here from the far more nuanced historical analysis of E.P. Thompson. At the conclusion of his 1975 book, *Whigs and Hunters* (which shows how a political oligarchy in eighteenth-century England invented callous and oppressive laws to serve its own interests) Thompson writes:

We reach, then, not a simple conclusion (law = class power) but a complex and contradictory one. On the one hand, it is true that the law did mediate existent class relations to the advantage of the rulers . . . On the other hand, the law mediated these class relations through legal forms, which imposed, again and again, inhibitions upon the actions of the rulers . . . In a context of gross class inequalities, the equity of the law must always be in some part sham . . . We ought to expose the shams and inequities which may be concealed beneath this law. But the rule of law itself, the imposing of effective inhibitions upon power and the defence of the citizen from power’s all-intrusive claims, seems to me to be an unqualified human good. To deny or belittle this good is . . . a desperate error of intellectual abstraction. (pp. 264–6)

**Notes**

1. For a discussion of the limitations of such exercises see Bardhan (2005, Chapter 1).
2. Even when the original patent is about to run out, the transnational company holding the patent often has various ways of effectively extending it: by slightly changing the composition of ingredients in the product and then taking out a new patent, bribing or intimidating the potential producers of the generic substitute, and through high-pressure advertisement keeping many of the customers hooked on to the original brand.
3. For a discussion of some of these issues see the papers by Gallini and Scotchmer (2001) and Kremer (2001).
4. For a review of the theoretical political economy literature on credibility of commitment see Bardhan (2005, Chapter 4).
5. For example, among developing countries many French legal origin countries are in Africa or Latin America and it may be standing as a proxy for other (unmeasured) deficiencies in state capacity in several of these countries.
6. Some of the pros and cons of relational contracting are empirically studied in the case of Vietnam’s emerging private sector by McMillan and Woodruff (1999).

7. For a formal treatment of the subject see Dixit (2003).

8. The standard argument that ‘irrational’ behavior is weeded out in the evolutionary process is much too limited. Other-regarding cooperative behavior may be more successful in many cases. Evolutionary success in replication and the economist’s narrow conception of efficiency may not go together if pay-offs to adherence to particular behavioral rules depend on adherence by others, or if there are positive and negative interactions of different behavioral rules.

References


Introduction
The New Institutional Economics (NIE), which has expanded rapidly since the mid-nineteenth century, has placed the issue of property rights at the forefront of its intellectual priorities. This is not surprising since, in the absence of well-established and well-enforced property rights, trade transactions may involve considerable transaction costs which have the effect of slowing market development. Perhaps to the dismay of those who believe in the absolute superiority of private ownership, a transaction-cost analysis aimed at assessing the relative efficiency of different property regimes does not necessarily point to the desirability of private property. In this chapter, we elucidate the kinds of circumstances under which private property is likely to dominate alternative property regimes, and this is done by limiting our attention to land and other natural resources (forests, lakes, pasturages, and so on). The demonstration proceeds in two steps. First, we compare a regime in which there are no property rights (the so-called open access regime) to alternative regimes in which such rights exist. Thereafter, we discuss the advantages and disadvantages of a regime of collective or community ownership vis-à-vis a regime of private ownership. The final section offers concluding remarks.

Open access versus property rights
A simple but basic principle taught by economics since Adam Smith is that a commodity, whether a good, service or asset, carries a low value if its supply is abundant relative to the demand expressed for it. The value can even be nil if the commodity is so plentiful that there is actually no competition between potential users. The next step is to recognize that, if a resource has no value, users have no incentive to seek to establish and acquire property rights which would assure them exclusive and free disposal of a portion of that abundant resource. The above proposition applies very well to land and other natural resources. Thus, in an environment characterized by low population density, access to land and natural resources is so easy that a portion of them can be occupied and exploited by individuals or collective entities without there being any need for them to defend against possible encroachments from outside.
Problems arise when an increasing number of individuals want to use a resource that is protected by no exclusive property rights acting as barriers to entry. More exactly, the regime under which no property rights exist (known as *res nullius*) begins to be problematic as soon as the entry of an additional user in the resource domain causes a damage to those users who were exploiting it before, and there is no penalty that this additional entrant is required to pay for such negative externalities. There are in fact two different sorts of externalities involved here.

The first type is known as congestion externalities. They manifest themselves in all the cases where the use of a natural resource necessitates that it is divided into distinct portions individually exploited by the users. When pressure on this resource increases, the activity of an individual user then unavoidably entails ecological costs for the other users, especially those located in the immediate neighbourhood. For example, a farmer who has cut off trees in the upper part of a hill in order to open a new field can cause erosion that will result in the silting of the lower parts.

The second type of externalities are rent-dissipating. They are observed in conditions where a natural resource is jointly, rather than individually exploited. And the problem arises only insofar as the resource is subject to decreasing returns, which implies that there is some degree of population pressure. In these conditions, through his additional harvesting effort, a new entrant causes a fall in the average productivity that is felt by all the previous users. A rational individual will enter into a resource domain, or continue to increase its activity therein, as long as the benefit he obtains exceeds the cost. What needs to be emphasized is that he will so act even though his income, or part of his income, is earned at the expense of previous users of the resource whose incomes are diminished as a result of his additional efforts. It is even possible that the additional effort of the new entrant does not give rise to any increase of output (the marginal productivity of effort is nil), yet the new entrant finds it individually profitable to apply this effort.

From the social viewpoint, such a decision is evidently undesirable since the costs thereby incurred are borne in vain, that is, they are entirely unproductive. A social waste of scarce production factors (labour, capital, fuel, and so on) occurs to the extent that the same aggregate output could have been obtained without applying additional amounts of these factors. The natural resource can then be said to be inefficiently managed. If the unproductive factors are used, it is because they enable the individual who owns them to participate in the exploitation of the resource and to obtain the average product which decreases as the number of harvesters rises. In the open access equilibrium, the resource rent is entirely dissipated (the costs eat the product), an absurd outcome since the natural resource is scarce.
(subject to decreasing returns), and should therefore have a positive value reflected in the rent.

In fact, the root cause of this inefficient use of a natural resource lies in the absence of property rights. Indeed, had property rights prevailed, the addition of an effort unit to be applied to the resource would have been decided by the owner only if the expected additional product exceeded the cost. In other words, it is the marginal product instead of the average product that would have been compared to the cost in order to decide whether an additional unit of effort must be applied. The marginalist rule guarantees the efficient use of a resource and, which amounts to the same thing, the maximization of the rent associated with it. It has this property because it allows a perfect internalization of externalities (see Demsetz, 1967; Alchian and Demsetz, 1973, for a first formulation of this well-known principle).

Dynamic losses must also be taken into consideration when assessing the pros and cons of a given property regime. More precisely, in the same way that he has no reason to be concerned with the damage caused to other users by causing a fall in their incomes, an individual harvester is not induced to forsake present benefits in order to ensure better the long-term conservation of the resource. For example, a fisherman has no incentive to throw back juveniles into the sea so that they can grow to mature size and be caught in their adult state at a later time. His reasoning is, indeed, that if he does not keep those juveniles when they are caught in his net, other fishermen will not hesitate to kill them with the result that they will anyway be unable to reach mature size. Had private property rights existed, the owner would have seen to it that conservation measures are adopted, since the future benefits of present sacrifices would have accrued to him rather than to other users.

**Private property versus communal property**

**General considerations**

In the absence of transaction costs, communal property allows rightsholders to internalize externalities as effectively as private property. Under this condition of zero transaction costs, the two property regimes are thus strictly equivalent (Platteau, 2000, Chapter 3). As a matter of fact, a group of people who own a resource will make exactly the same decisions as an individual owner regarding the amount of effort to be applied. Such an outcome results from the fact that, like an individual, a group seeks to maximize the rent or the surplus, that is, the difference between the value of the flow of produce extracted from the resource and the capital and labour costs. In both cases, the marginalist decision rule is applied.
But the assumption of zero transaction costs is totally unrealistic: it is useful to set a theoretical reference point, but not to describe reality. Once transaction costs are taken into account, the members of the property rights school contend, private property appears superior to any system of collective or communal property. The underlying argument can be summarized as follows: while an individual owner forms an organically integrated decision unit (he need not discuss with anyone else than himself in order to reach decisions regarding the use of the resource), so as to create a one-to-one relationship between individual actions and their effects, a collective owner must achieve an agreement between its members to decide how to use it. An agreement will not necessarily emerge from intra-group discussions. And, even if an agreement can be eventually reached, the negotiation process will necessarily entail non-trivial costs, such as the opportunity cost of the time spent in meetings, transport expenses, the costs of communicating the time and place of the meeting, and so on (Baland and Platteau, 1998b; Platteau, 2000, Chapter 3). Lastly, assuming that an agreement has been found and that the details of its modus operandi have been worked out, there remains the delicate question as to how it will be effectively enforced. At the very least, one may fear that its implementation will not remove all the possible sources of inefficiency, something which single private ownership is apparently capable of achieving. In the following, all the costs plaguing group ownership will be referred to as governance costs.

The determinants of governance costs

The feasibility of an agreement regulating the use of a natural resource at village level is itself dependent on a certain number of factors, among which the size of the user group and its degree of heterogeneity stand foremost. Regarding the first factor, it is evident enough that the smaller the number of rightsholders the lower the negotiation costs involved in the process of devising the regulatory agreement. If their number is too high, on the contrary, one may fear that no regulatory mechanism will be put into place for lack of an internal agreement.

The impact of heterogeneity is also evident. Thus, it is easy to understand that divergences between group members regarding the intended uses of a natural resource will make an agreement more difficult to reach. Illustrations are numerous, as attested by the prolonged conflicts between farmers and herdsmen around land areas claimed by each category for their own specific purpose. Herdsmen want to maintain their customary rights to large grazing areas at least during a part of the year (so that animals can feed themselves on crop residues after harvest time), while farmers are increasingly eager to win exclusive rights over well-delineated zones so as
to be able to practice more intensive forms of agriculture requiring continuous cultivation and long-term land improvements.

Income or wealth inequality constitutes another form of heterogeneity that tends to make regulation of the use of natural resources difficult to achieve. It has indeed been shown that the more unequal the distribution of income between members of a village community or any user group, the harder it is to find a regulatory scheme that satisfies all the people concerned. Moreover, if such a scheme exists, the efficiency gains that it will yield compared to a situation with no regulation diminish as income distribution becomes more unequal (Baland and Platteau, 1998a; 2003). The underlying intuition is simple: when there is a need to regulate the use of a resource, the group involved must not only determine the extent to which the intensity of use must be reduced to approach efficiency, but also the manner in which the effort reduction will be shared among the various users. If users are relatively identical, the latter problem is unlikely to be serious: a uniform reduction of individual effort levels appears as the natural solution. If, on the contrary, the users are different in terms of wealth or income and these differences are reflected in different rates of resource use, the problem of sharing the burden of effort reduction obviously becomes more complicated. In particular, the efficient solution might well imply that the larger part of this burden be borne by the less productive or the more impatient (those more preoccupied with subsistence constraints) users who often are the poorer members of the community.

Let us nonetheless assume that an agreement can be found that entails efficiency gains and allows each resource user to improve their situation compared to the status quo state of no regulation (only access rights exist). There remains the tricky issue of the enforcement of the regulatory scheme: once the rules are decided and agreed upon, each user has an incentive to violate them. The problem is especially serious because, by rendering the resource more valuable than before, effort restriction has the effect of increasing the benefits which can be obtained by exceeding one’s allowed quota while other users follow the rule. To put it in another way, the outcome of the agreement is to enhance the temptation for individual users to free-ride on the sacrifices incurred by fellow users.

In addition, again assuming that an agreement is feasible, there is the question as to who will bear the costs of formation of collective action, understood as the costs involved by the very process of creating collective mechanisms for both decision-making and enforcement. Again, the characteristics of the user group – its size and heterogeneity, in particular – influence the extent to which this problem can be surmounted (Baland and Platteau 1997; 2003). To begin with, the impact of group size is identical to
the one observed when the problem consists of devising a regulatory scheme: the smaller the size of the community, the more likely the costs of formation of collective action will be actually incurred and, therefore, the more likely the agreed rules will be designed and applied. The main argument here is known as the incentive dilution argument (Olson, 1965). Each individual makes a personal calculation when he decides whether or not to contribute to the production of a local public good (or, in our context, to the creation of a collective mechanism, on the one hand, and to the implementation of a regulatory scheme, on the other hand). The individual compares the gains from abstaining from contributing to the collective effort with the cost. When a group is smaller, the cost of withdrawing participation obviously rises relative to the benefit.

In addition, when a group is smaller, members tend to know each other better and, therefore, reputation effects are more important. More attention will be paid to the future consequences of opportunistic behaviour in order to avoid punishment in the form of exclusion from the group or denial of the right of access to the resource. Furthermore, not only is communication facilitated within a smaller group, but the formation of collective identity feelings is also easier and, as a consequence, individuals are more induced to take into account the effects of their decisions on the other members (Baland and Platteau, 1996, pp. 75–8).

The impact of heterogeneity is more ambiguous than the impact of group size. In fact, it cannot easily be predicted a priori. Let us examine, in particular, the impact of inequality of income or wealth on efficiency in the production of a public good, such as the formation of a regulatory body or framework. Two effects are at work which run into opposite directions. On the one hand, a great inequality creates a situation in which the rich guy internalizes a large part of the externalities created by his particular contribution to the public good, thereby inducing him to apply the required effort. Yet, on the other hand, those who are at the lower tail of the income distribution find themselves in exactly the opposite situation: they will draw only minor benefits from the collective good and hence they will have weak incentives to contribute.

It is thus impossible to predict in a general manner whether a higher degree of income inequality will actually result in an increase or a decrease of the aggregate amount of contributions to the local public good. The aggregate amount will rise only if the increased contributions of the village elite (who better internalize the externalities) exceed the reduced contributions of the common people. What is certain, however, is that an extreme inequality corresponding to a total concentration of all the wealth in the hands of a single villager will lead to an efficient provision of the public good (see Baland and Platteau, 2007).
The evolution of governance costs

As is evident from the above discussion, communal property gives rise to serious incentive problems, especially when it involves the regulation of the use of natural resources besides the setting of access rules. A reasonable position therefore consists of admitting that inefficiencies are bound to persist under this ownership regime, whether in a static form (the dissipation of part of the resource rent) or in a dynamic form (the lack of investment to conserve the resource, including actions to fight against predators who threaten its long-term stock).

A straightforward consequence of such a situation is the following: the people’s ability to cooperate in the management of common access resources determines the profitability of jointly held resources compared to their profitability when they have been individualized, and it also influences the allocation of resources between various uses. Inasmuch as this cooperation ability varies from one area to another, one must expect to observe geographic variations in the uses and rates of profitability of local-level natural resources. For example, it has been shown that in Mexico, when cooperation fails in the management of collectively grazed pastures, more land is allocated to crops than under successful cooperation and less to pastures, while the stocking rate on pastures is increased. This results in too much land in extensive crops and too many animals per hectare of pasture (McCarthy et al., 1998).

The point that we want to make now is that efficiency losses caused by externalities are likely to grow with the value of the resource, hence the frequent emphasis in the literature on the unit value of natural resources as one of the main determinants of its privatization (division). To illustrate, in his classical study of the Swiss Alps, Netting contrasts the lowlands of the valley which are fertile and therefore tend to be privately appropriated with the more arid highlands which are used as communal (summer) pastures under the authority of the village council (Netting, 1976; 1981).

Population pressure bears upon the efficiency gains of division in two different ways. For one thing, by increasing the number of users per unit area, it creates more room for external effects and, thereby, the governance costs and the inefficiencies involved in the joint exploitation of the resource rise. For another thing, by making the resource increasingly scarce, population pressure enhances its value and therefore makes for increased aggregate losses from collective exploitation. In other words, the amount of the rents foregone by not dividing the resource tends to increase with population. This is especially true when population pressure involves a transformation of the pattern of resource use, such as a shift from extensive to intensive agricultural or grazing practices, since intensive practices have the
The role of direct transaction costs

When the problem of choice of ownership regime is considered from the exclusive standpoint of governance costs, the balance sheet is unmistakably favourable to the division and private appropriation of local-level natural resources. But this is only one side of the balance sheet. So far, indeed, we have implicitly assumed that the establishment and protection of property rights are costless operations. Such an assumption is manifestly unrealistic and we need to remove it now. What appears then is that, compared to communal property, private property is costlier from the viewpoint of direct transaction costs, which primarily include set-up and protection expenses. There thus exists a trade-off between two kinds of costs that have a different impact according to the ownership regime considered: the governance costs.
that affect communal but not private property, on the one hand, and the direct transaction costs that are smaller under the former than under the latter regime, on the other hand.

The latter conclusion follows from the fact that it is less costly to fence, demarcate and protect a territory of a given size than to do those things for divided portions of that territory. As a corollary, when the surface area of a resource domain is larger, the per capita direct transaction cost of privatizing it increases. To put it in another way, direct transaction costs increase with the physical base of the resource: the more spread the resource base (or the less concentrated the resource) the higher the costs of delimiting and defending the resource territory. Other things being equal, therefore, the more spread a resource is, the less profitable it is to privatize it, and the more compact the resource, the more attractive is its privatization.

A consequence of the above is that we expect private property rights to be established over high-density resources and communal property rights over resources with the opposite characteristic. It is good news for economic theory that such a prediction is systematically verified in reality. Yet, at the same time, one must reckon that practically it is often difficult to disentangle the impact of the density of a resource from that of its value on the probability of privatization. As a matter of fact, high-value resources – for example, fertile lands that are susceptible to being irrigated owing to their favourable location – tend to be divided more often than low-value resources – for example, semi-arid lands that are hardly suitable for any other purpose than extensive grazing. In other words, there exists a strong correlation between the density or compactness of a resource and its value. In fact, in many instances the second characteristic causally determines the first one. What empirical evidence reveals is that compact resources with a high value (for example, intensively cultivated lands or fertile lands located near an important market town) are generally held under private property while resources that stretch over large areas and carry a low value (for example, the immense low-quality grazing areas in Mongolia or the Maasai Mara in Kenya) are jointly held by a local user group or community. Our analytical argument simply shows that these two correlated characteristics of a natural resource – its high value per unit area and its high density – tend to make its privatization relatively profitable.

Two remarks are in order. First, there exist natural resources, the division of which would entail prohibitively high direct transaction costs under the present state of technology. For example, the open sea – or, more exactly, the fish stock contained in it – presents insuperable difficulties for private appropriation. The enforcement of exclusive property rights to individual patches carved up in the ocean would, indeed, be infinitely costly. This is especially evident when fish species are mobile and move over large water
spaces, since exclusive rights are too costly to establish and enforce whether over the resource or the territory in which it moves. The example of wildlife reserves also comes to mind.

Interestingly, even in the case of maritime fisheries, privatization may sometimes be a viable solution. This tends to happen when species are rather sedentary (for example, lobsters, shellfish, molluscs, seaweeds) and live in relatively compact and well-delimited spaces, such as when the fish are found around islands (the Pacific islands, the Shetlands in Scotland, and so on) or in relatively well-sheltered aquatic zones (for example, in deltas or in backwaters) where fishing locations can be easily demarcated and protected against external encroachments. In these conditions, fishing spots are frequently assigned to individuals or families for their exclusive use and these private rights can generally be inherited by future generations as long as they are used effectively (see Platteau, 2000, p. 85, for references to the anthropological literature).

Second, the direct costs of resource division are not exogenously fixed. In the above, we have considered that these costs are determined by the inherent characteristics of each resource, it being understood that a resource may take on various forms and characteristics depending on the precise location and environment in which it is found. Here, we want to point out that, in fact, direct transaction costs may fall with technological progress. One well-known example is the discovery of the barbed wire which proved to be a decisive step in the reduction of the cost of protecting property rights through cheaper fencing of agricultural fields (North, 1981). Another example is the introduction of modern borehole drilling facilities in arid and semi-arid areas where this has the effect of facilitating the privatization of common grazing areas. Before this invention, in a country like Botswana, water extraction was subject to important scale economies as a result of which grazing lands were always the collective property of herders' communities (Peters, 1994). Private appropriation is thus not only facilitated by factors which contribute to enhance the value of a resource, but also by factors which have the effect of reducing the direct cost of partitioning.

The role of opportunity costs

Two types of opportunity costs appear to play an important role in this respect: scale economies and insurance benefits associated with collective ownership.

Let us first consider the impact of scale economies. Resources offering multiple products tend to be subject to scale economies to the extent that they form part of an overall ecosystem. This multiple product character of the resource is a reason often mentioned to argue against the parcelling out
of forest areas into individual holdings. In the case of hunting, on the other hand, wild animals require large territories to survive and reproduce, so that division of a hunting domain into smaller parcels would imply the destruction of the resource.

When discussing the role of scale economies, it is important to bear in mind that they may be present not in the resource itself but in complementary factors. The obvious advantage of coordinating the herding of animals so as to economize on shepherd labour in extensive grazing activities is probably the best illustration of the way scale economies in a complementary factor may prevent the division of a resource domain. Another illustration can be taken from fisheries (maritime or inland). In many cases, indeed, the guarding of privately apportioned fishing spaces is subject to strong scale economies. As a consequence, it may make sense for several individuals or families to get together to enforce a de facto right of collective property over a given fishing ground. This is actually what many traditional fishing communities have done in the past when competition around scarce fish resources began to develop, particularly in inland fisheries.

Opportunity costs of privatization also come into the picture when returns to a resource are highly variable across time and space. The need to insure against such variability is then a consideration that militates against resource division (McCloskey, 1976; Dahlman, 1980). When a resource has a low predictability (that is, when the variance in its value per unit of time per unit area is high), indeed, users are generally reluctant to divide it into smaller portions because they would thereby lose the insurance benefits provided by the resource kept whole. In the words of Nugent and Sanchez: ‘the lower the quality of land or the more variable the weather, the more important it is that the land be held in communal, that is, tribal form’ (Nugent and Sanchez, 1993, p. 107).

The example of extensive grazing and also that of maritime fishing again provide us with good illustrations of the above. Herders (fishermen) typically need to have access to a wide portfolio of pasture lands (fishing spots) insofar as, at any given time, wide spatial variations in yields result from climatic or other environmental factors. Assuming that the probability distributions are not correlated too much across spatial groupings of land or water and that they are not overly correlated over time, a system offering access to a large area within which rightsholding users can move freely appears as highly desirable from a risk-reducing perspective.

It must nevertheless be pointed out that private property rights over portions of the resource could apparently solve the predictability problem. Consider the case of extensive grazing again. Due to the unpredictability of rain-induced growth of grasses within any small region, what is important for herders is to be able to move over large ranges of land and rapidly
change location when the need arises. By holding exclusive ownership rights over widely dispersed patches of pasture lands, they would therefore achieve their objective of risk reduction. Clearly, to account for the maintenance of communal property, one needs to appeal to transaction-cost considerations and the high exclusion costs of a spread resource base (Platteau, 2000, p. 88). Because it would be prohibitively costly to enforce exclusive rights over widely dispersed and infrequently visited ranch patches, the division of the resource domain turns out to be infeasible. The same situation actually obtains in many fisheries.

The additional (direct) transaction costs implied by the necessity to insure against income fluctuations in the event of division or privatization of a resource increase with the variability of incomes and the surface area of the domain of this resource. As a matter of fact, the higher the variability of incomes the larger the number of resource portions that a particular user needs to insure himself and, hence, the higher the costs of establishing and protecting private property rights. On the other hand, the more stretched is the resource base the higher the (direct) transaction costs caused by the demarcation and the guarding of a given portion of the resource.

Conclusion and final considerations about the evolutionary approach to institutions
Two central conclusions emerge from our analysis. First, the transaction-cost economic theory of institutions leads us to expect that an increasing number of village-level natural resources will be divided and individually held as they acquire more value under the combined impact of population growth and market penetration. The important role of governance costs associated with collective ownership goes a long way toward explaining this gradual shift from corporate to private forms of ownership. Nonetheless, and this is our second conclusion, certain resources possess characteristics that make their division and their private appropriation especially costly. The costs involved are those required to establish and protect private property rights (direct transaction costs), or opportunity costs resulting from the loss of benefits provided by communal property.

The second conclusion raises a thorny issue. Indeed, in order that communal property be viable in conditions where private property is infeasible owing to prohibitively high direct transaction costs, or undesirable owing to high opportunity costs, it is essential that governance costs remain within tolerable limits. If this condition is violated, the natural resource concerned will not be regulated in a satisfactory manner and efficiency losses will be significant, perhaps considerable. It will either become a non-regulated common property (meaning that it is characterized by access
rules while rules of use are absent), or it will fall under the open access regime. If the number of users is large, these last two regimes will produce more or less equivalent results in the form of grave inefficiencies both in the static (rent dissipation) and in the dynamic senses (destruction of the stock following a lack of conservation investments or a lack of control of the extraction efforts applied by users).

**Note**

1. ‘A primary function of property rights is that of guiding incentives to achieve a greater internalization of externalities.’ (Demsetz, 1967, p. 348).

**References**


Introduction
Are some cultures more prone to development than others? Does economic development require cultural change? Which is more important, economic development or cultural integrity? These are among the core questions that swirl around the debate on the relationship between culture and development. The debate was initiated in its modern form by Max Weber’s classic work, *The Protestant Ethic and the Spirit of Capitalism* (1958), and it has continued ever since. Social scientists past and present have staked out positions on all sides of the issues. Traditional modernization theorists take it for granted that economic development is primary, but they differ on whether cultural change leads development or vice versa. Some see culture functioning as either an instrument of or an obstacle to development, while still others see culture as irrelevant. Post-development localization theorists, on the other hand, disparage modern economic growth and argue that culture is paramount. They are highly resistant to Western-style corporate-led globalization, which they believe destroys local culture. Inbetween are many intermediate positions, like those who hope to maintain cultural integrity while also allowing for the development of market economies, the expansion of trade and economic growth.

Historical background
To begin the exploration of the role of culture in development, it is worthwhile to review briefly the rise of the modern economy in the Western world. Two facts stand out from an examination of the history of modern capitalist development. First, capitalism has succeeded in producing quantities of goods and services unprecedented in history; second, it has done so in a temporally and spatially uneven manner. The economies of some nations take off into a self-sustaining growth, with other nations eventually catching up. Still others, the loosely called ‘developing countries’, seem to be left hopelessly behind. This pattern occurs across nations as well as across regions within nations.

One of the great economists of the twentieth century, Joseph Schumpeter, captures this dynamic process in his concept of ‘creative destruction’: ‘The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers’ goods, the new methods
of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates . . . [These developments] incessantly revolutionize the economic structure from within, incessantly destroying the old one, incessantly creating a new one’ (Schumpeter, 1950, p. 83).

In both industrial and poor countries this creative-destructive process of development has created socio-political tensions both because of its uneven nature and because of its challenge to traditional values and ways of life. Today this is true particularly in countries with strong Islamic roots and in countries where readily identifiable minorities control the wealth of the society.

A simple policy of free markets and free elections may not advance the welfare of the poor in the world, and it must be remembered that the Western world did not develop that way itself. Development was slow, its major transformation spread over more than a century. For much of that period democracy was limited and countervailing policies were gradually enacted to curb the divisive effects of growing inequality.

Most historians, whether their orientation is political, cultural or economic, recognize that the eighteenth century was a turning point in the nature of the Western world that saw momentous movements and events – intellectual, political, military, social, cultural and economic. The Enlightenment, with its emphasis on reason, natural law and progress, and its avant-garde – the philosophes and physiocrats – opened new vistas, even though most of the population of Europe and the remainder of the world scarcely glimpsed those vistas.

In economic affairs, the eighteenth century began with Francois Quesnay’s campaign against mercantilism and ended with the completion of the campaign by Adam Smith. In the process, the classical school of economics, a new social science, came into being. Finally, in the course of the century, the agricultural and commercial revolutions of the previous two centuries initiated the Industrial Revolution in England that formed the basis of our modern economies.

Now that self-regulated market capitalism had arrived, what were its characteristics? In its textbook purity, a capitalist market economy is controlled, regulated and directed by markets alone. Socially beneficial outcomes in the production and distribution of goods are entrusted to this self-regulating mechanism, based on the expectation that human beings behave so as to achieve money gains.

In contrast, during preceding historical periods (and even today in many of the poor countries), markets were never more than accessories of economic life. Instead, the economic system was embedded in the social-cultural system. In places like Babylonia and Greece the local markets
(trading centers) were compatible with the established social way of life; markets did not expand at the expense of the society. Even under the mercantile system of the previous two centuries, where markets had expanded to involve a large part of the nation, they were not free markets, for they were subjected to centralized administration. Karl Mannheim argued that the move to self-regulating markets entailed a transformation from a regulated and socially controlled mechanism into the very organizing principle of society itself (Mannheim, 1950, p. 191).

Moreover, such a market-first institutional pattern cannot function unless other aspects of a society’s life are subordinated to its requirements, which is what happened over time in today’s developed economies. A market economy can only exist in a market society, and the requisite process of social, cultural and institutional change evolved in conjunction with the transition to a market economy. Nations wanting to catch up economically are thus naturally led to the question of how such a market society can be created in countries far removed from Western culture and in a time span shorter than the century or two that was required in Europe. On the heels of this question, another follows close behind, which is whether the creation of such a market society should even be a goal of development. Needless to say, there is substantial disagreement over these questions.

Nailing down a working definition of culture is itself problematic, in part because culture is easier to identify in others than it is in oneself and one’s own society, just as most people think others speak with accents, not themselves. That in itself makes talking about, and especially evaluating, culture potentially invidious, because in discussing cultures people are inevitably discussing the ways of life of others. Going way beyond people’s culinary, musical preferences, dress and traditions, Rao and Walton understand culture to be:

about relationality – the relationships among individuals within groups, among groups, and between ideas and perspectives. Culture is concerned with identity, aspiration, symbolic exchange, coordination, and structures and practices that serve relational ends, such as ethnicity, ritual heritage, norms, meanings, and beliefs. It is not a set of primordial phenomena permanently embedded within national or religious or other groups, but rather a set of contested attributes, constantly in flux, both shaping and being shaped by social and economic aspects of human interaction. (Rao and Walton, 2004, p. 4)

Culture is thus about our deepest beliefs, values, sense of identity, ways of life and longings, which makes it unsurprising that the discussion of culture and its significance for development generates controversy.

For economists the interest in culture has centered on its support of traits that contribute to economic growth, that is, thrift, hard work and
reinvestment by the middle class; hard work, obedience and contentment for the working class. The key is to discover the historic role of culture in generating the capitalist spirit of entrepreneurship, which for Schumpeter is so basic. For example, what were the psychological conditions – the capitalist spirit – that accompanied and aided the development of a thriving capitalist economy? Greed and the pursuit of riches are nothing new. Money-lending, commercial trading, piracy, plunder and other forms of unrestrained avarice are as old as history. But a way of life based on the rational, calculated pursuit of pecuniary profit through Smith’s ‘truck, barter and exchange’ and its organization into an economic system using free wage labor is a modern phenomenon.

It was only after centuries of struggle that capitalism established its claim to legitimacy, for it involved a code of economic behavior and a system of human relations sharply at variance with traditional religious customs and values. Originality, self-confidence and tenacity of purpose were required to initiate and carry on this struggle. This was the role of entrepreneurs. They emerged partly because changing economic conditions helped the Reformation succeed and helped shape the development of new theologies and creeds. In turn, the emerging religious beliefs helped direct and shape the subsequent economic development. Economic reasons alone are insufficient to account for the extraordinary power of entrepreneurship and rational profit-seeking in the modern world.

Alternative views on the role of culture in development
How does this historical experience of the rise of capitalist economies in the now developed world impact upon our theory and practice of development today? Mainstream views of development today continue to follow in the modernization tradition that arose in the aftermath of World War II. On this view the primary measure of development continues to be a self-reinforcing tendency to economic growth and material progress. It is from this basic developmental reality that other positive features of a good society spring, like better health care, improved education and democratic governance. As Benjamin Friedman says at the close of his recent moral defense of economic growth: ‘Only with sustained economic growth, and the sense of confident progress that follows from the advance of living standards for most of its citizens’, can a nation hope to achieve ‘an open, tolerant, and democratic society’ (Friedman, 2005, p. 436). Over time it has become clear that an equitably growing economy requires financial capital, the adoption of efficient technologies, investment, human capital, entrepreneurship and market-promoting policies and institutions, which begs the question of how to get nations to build and use these necessary attributes. For many theorists who see a connection between culture and development,
inquiries into this question lead directly to issues of how cultural adaptation similar to that which occurred historically in the West can be facilitated so as to accommodate the needs of modern market economies throughout the world.

On the other hand, it needs to be said that many economists believe that culture is essentially irrelevant to economic development. Instead, the road to development is paved with market-promoting economic policies and institutions. Hernando De Soto, for example, argues that efforts ‘to explain why capitalism fails outside the West remain mired in a mass of unexamined and largely untestable assumptions labeled “culture”, whose main effect is to allow too many of those who live in the privileged enclaves of this world to enjoy feeling superior’ (de Soto, 2000, p. 225). Along with economist Mancur Olson, De Soto argues that people of all cultures respond similarly when property rights and contract-enforcing institutions are established (Olson, 2000). In his high-profile treatment of twenty-first-century development challenges, Jeffrey Sachs calls the cultural thesis a myth, saying that cultures often follow rather than lead economic change and that culture-based arguments ‘are usually made on the basis of prejudice rather than measurable evidence’ (Sachs, 2005, p. 317). In his critique of Lawrence Harrison’s emphasis on child-rearing, for example, Sachs points out that children are taught the value of hard work more consistently in Nigeria, South Africa and Tanzania than they are in the United States (Sachs, 2005, p. 318). For William Easterly too, the issue is not culture, but the proper structuring of incentives (Easterly, 2006). Summarizing these lines of thought, David Throsby points out that: ‘mainstream texts in economic development have no time for culture; taking three such texts more or less at random, an inquisitive reader can find no reference to culture in the subject indexes of any of them’ (Throsby, 2001, p. 67). For such theorists, it is no surprise that nations of greatly diverse cultures, like Ireland, South Korea, France, the United States, and now China and India can all enjoy the fruits of economic progress without fundamental changes to their cultures.

And yet there are perhaps a greater number of development theorists, like Weber, who find culture to be of central importance. For many of the early thinkers in this tradition, like economist Bert Hoselitz and sociologist Talcott Parsons, transitions from traditional to modern patterns of life required nothing short of major cultural overhaul. Major efforts to discover potential sources for such cultural change led David McClelland to identify the significance of a people’s psychological ‘need for achievement’, what he calls ‘n-achievement’. McClelland argues that modern societies have been built by innovative entrepreneurial types with high n-achievement. N-achievement, however, is so deeply embedded in people’s psyches, which develop gradually
during child-rearing and other social and cultural nurturing practices, that he despairs of offering practical advice for how actually to promote development. He notices, for example, that religious and ideological changes, like the rise of Protestantism in some Mexican communities, were associated with a rise in n-achievement. But he is doubtful this awareness of the relationship will itself lead to acceptable development policies. And though he believes education might be of some help, he doubts it can make a large contribution, because people’s personalities are too deeply formed by prior and ongoing child-rearing practices. Ultimately, the best he can do is to encourage policies that facilitate the interaction among entrepreneurs from developed countries with scarce but vital entrepreneurial types engaged in business in ‘underdeveloped countries’ (McClelland, 1961).³

In a similar vein, Everett Hagen, recognizing both the importance of entrepreneurship and the role of psychological formation and traditional cultural patterns in creating people resistant to progressive change, identifies one major source of change to be a socially deviant group that finds a psychological outlet in violating traditional patterns of life and thus creates its own identity through entrepreneurial change and success in business. Subdominant or threatened minorities, or marginalized but progressive immigrant communities, might be able to provide an impetus to change that breaks through traditional cultural patterns and points the way to modernity (Hagen, 1962).

Still today, many contemporary theorists within the modernization tradition continue to focus on the fundamental importance of cultural change. Perhaps most well known among these is Lawrence Harrison, who after long experience in development work in Latin America has developed his version of the cultural thesis in a series of books written over a twenty year period starting in the mid-1980s.⁴ According to Harrison, the basic thesis is that ‘values, beliefs, and attitudes are a key but neglected factor in understanding the evolution of societies and that the neglect of cultural factors may go a long way toward explaining the agonizingly slow progress toward democratic governance, social justice, and prosperity in so many countries’ (Harrison, 2006, xiii). In Harrison’s view, progress in such basic areas as life, health, liberty, prosperity, education and justice depends on the adoption of a democratic capitalist way of life, which in turn depends on cultural orientations.

Following many of his forbears who have developed lists of the contrasting cultural characteristics of traditional and modern societies, Harrison too identifies cultural traits that either inhibit or advance progress. Over the years, his list has expanded to 25 core cultural traits that make societies either progress-prone or progress-resistant (Harrison, 2006, pp. 36–7). These include religious orientations (for example favorable or
non-favorable attitudes toward material pursuits), values (for example how flexible a society’s ethical code is), economic behavior (for example whether people have entrepreneurial inclinations) and social behavior (for example the radius of trust). Societies that value competition instead of fearing it as a threat to equality, for example, are more likely to progress, as are those that focus on success in this world over their place in the next. If progress is to come about in poor countries, then it is culture that must ultimately be changed.

Harrison is the name most often associated with the cultural thesis, but he is not alone. David Landes, in a sweeping study of development patterns across the world, both historical and contemporary, is drawn toward cultural explanations, ultimately assigning a major causal role to culture. Landes argues that: ‘just because markets give signals does not mean that people will respond timely or well. Some people do this better than others, and culture can make all the difference’ (Landes, 1999, p. 522).

Recent studies lend the cultural thesis some degree of empirical support. In a study of various factors that influence growth, David Weil finds that openness to new ideas, an inclination to work hard and to save, and the level of trust play a significant role in explaining economic growth (Weil, 2005, p. 427). Other studies analyze World Values Survey data with a special focus on the role of religion and find that religious faith and traditions definitely matter. Luigi Guiso et al. conclude that: ‘on average . . . religion is good for the development of attitudes that are conducive to economic growth’, and that ‘on average, Christian religions are more positively associated with attitudes that are conducive to economic growth, while Islam is negatively associated’, and that Protestants and Catholics have different mixes of positive and negative factors’ (Guiso et al., 2003, p. 280). In neither the Weil nor the Guiso study is the direction of causality firmly established. In an attempt to address this shortcoming, Robert Barro and Rachel McCleary find that religious beliefs, especially as regards the existence of heaven and hell, seem to play a causal role in the achievement of higher economic growth. Like Weber’s theories about the role that Calvinism played in Europe’s drive toward capitalism and industrialization, Barro and McCleary conjecture that such ‘religious beliefs stimulate growth because they help to sustain aspects of individual behavior that enhance productivity’, like thrift and a greater work ethic.5

Social capital
Another line of thought that has arisen since the 1990s is focused on the causal role played by social capital and civil society. In the economics literature, social capital has come to mean that ‘social relations’ are important factors in the economy. Francis Fukuyama and Robert Putnam have each
studied the extent to which social relations promote group cooperation, civil society, good governance, trust and productive economic activity. Fukuyama emphasizes interpersonal trust as a key cultural aspect, arguing that: ‘one of the most important lessons we can learn from an examination of economic life is that a nation’s well-being, as well as its ability to compete, is conditioned by a single, pervasive cultural characteristic: the level of trust inherent in society’ (Fukuyama, 1995, p. 7). In a similar vein, Putnam shows that the cultural predilection to work together cooperatively in civic groups is a major well-spring of democratic governance and economic well-being (Putnam, 1993). Fukuyama and Putnam both argue that cultures that foster trusting working relationships outside of narrow family interests, referred to by Fukuyama as ‘weak ties’ and by Putnam as ‘bridging capital’, are much more prone to establish successful democratic capitalist societies.

The recent attention to social capital, combined with the renewed focus on religious beliefs, has given rise to the concept of spiritual capital, which refers to the spiritual or religious resources that contribute to a well-functioning community. Putnam argues that religion is by far the largest generator of social capital in the United States, contributing to more than half of the social capital in the country. For developing countries, where religious commitments are generally stronger than they are in the economically developed countries, religious sources of social capital may be even more important. Theodore Malloch claims that: ‘In the ultimate sense spiritual capital is the missing leg in the stool of economic development, which includes its better known relatives, social and human capital’ (Malloch, 2003, p. 2).

Social capital studies have become common in development economics due, at least in part, to the World Bank which has been working on the concept since the 1990s. On one of its websites it says that: ‘Social capital refers to the norms and networks that enable collective action. Increasing evidence shows that social cohesion – social capital – is critical for poverty alleviation and sustainable human and economic development.’ Still, the extent to which social capital and spiritual capital are intellectually valid conceptual categories for economic development is carefully explored in a World Bank-supported volume edited by Anthony Bebbington et al. The presumption throughout the book is that the character of a society’s social capital is a culturally defined reality that is central to the empowerment, participation and inclusion of people in the development process. The purpose of the book is to study efforts at the World Bank, populated as it is with technically trained economists, to figure out ways to include social capital in their analysis and in their programming (Bebbington et al., 2006).
In the hands of economists, social capital is typically integrated into neoclassical microeconomic theory, making it a behavioral property of individual actors. The impact of social capital is thus seen to flow from the rational choices of those individuals. The most common ways to model social capital are: (1) as a preference in a utility function; (2) as an individual resource owned by persons or firms; and (3) as an instrument to reduce risk.

Against the instrumentalist view of culture
The neoclassical tendency to consider social capital as a resource for development has generated some strong criticism. Using the term ‘capital’ is seen as misleading and even ideological, and the empirical support is also seen as weak and circular. As Van Staveren and Knorringa argue: ‘one of the most central flaws . . . is a circular explanation of social capital: a group’s success is attributed to its social capital, but social capital is measured by group success’ (Van Staveren and Knorringa, 2007, p. 110). Moreover, if social capital is a resource for development, then social relations are stripped of their own value and become nothing more than instruments in the service of economic growth. One of the troublesome implications of such a view is that cultures can thus be judged worthy or unworthy, good or bad, based on their ability to generate economic growth. Not surprisingly, such invidious views are not readily accepted by people who have learned to value their own cultures for more than their pecuniary potential.

Amartya Sen, for example, argues that a focus on whether cultures are ‘good or bad’ fosters prejudicial attitudes, sometimes leading to a ‘blame the victim’ mentality that can cause great harm. English responses to Irish famines in the nineteenth century, for example, were thought by the British to be the result of Irish cultural deficiencies, which, if true, meant that direct assistance would only exacerbate the problem. What the Irish really needed were civilizing influences, a cultural makeover. Thousands died. By contrast, economic downturns in England were seen by the British as resulting from events beyond the control of the citizenry. Fast forward to the present day and one finds Catholic Ireland’s economy growing much faster than Protestant England’s. The same is true for India, which with its caste system and other-worldly religious views, was always assumed to have a culture resistant to development. Like China and Ireland, India is now among the fastest-growing countries in the world. Did their cultures change? Or was something else at work? (Sen, 2004, pp. 37–58).

Even some neoclassical economists have trouble with this tendency to instrumentalize culture. Oliver Williamson, responding to the definition of trust as rational expectations of the behavior of others, says ‘calculation-ness will devalue the [social] relations’ because it ‘may well be destructive of
atmosphere and lead to a net loss of satisfaction between the parties’ (Williamson, 1993, p. 481).

Such disrespect for local culture is even more vehemently rejected by the post-development school, which, with leaders like anthropologist Arturo Escobar, is much to the political and philosophical left of Sen. Escobar, in a postmodern vein, argues that the discourse dominating development thought and practice today is of Western origin and thus embeds the superiority of its cultural orientations in everything it says and does.7 When the international development discourse emanates from rich-country governments, universities, non-governmental organizations (NGOs) and multilateral organizations like the World Bank, all dominated by the institutions of Western and/or Northern culture, is it any surprise that its democratic capitalist ways of life and thought are privileged over others? When international economic, political and social institutions are structured on the basis of powerful democratic capitalist countries, is it any wonder that other cultures fail to compete successfully? In almost diametrical opposition to cultural modernizers, who believe Western-style cultural reforms are the way to developmental salvation for the poor world, Escobar believes they will bring destruction, for in disrespecting and ultimately undercutting their own cultures, the requisite cultural changes will cause local cultures to lose their identities, meaning systems and control over their own lives, resources and communities. There is in post-development a strong environmental orientation as well, for often people are dispossessed of their property and their livelihoods to make way for dams, agro-export products and mass production techniques that destroy traditional ways of life. Modernizers like Harrison might think these to be progressive developments, but Escobar believes they lead to domination, dispossession, violence, cultural chaos and poverty.

Unlike modernization theorists like Harrison, whose prescription is to criticize local culture and to enact policies to bring people into the modern world, Escobar exalts local culture, insisting on local solutions, discovered and implemented through locally developed institutions, and understood in terms of local languages and ways of life. Escobar and the post-development movement are thus among the main theoretical supports to the contemporary localization movement, of which David Korten is one of the most prominent supporters, but which also includes popular authors like Wendell Berry and novelist Barbara Kingsolver (Korten, 2001; Kingsolver, 2003).

In addition to the modernizers who discount culture entirely or treat it like one of any number of instruments, and post-development scholars who have a tendency to sacralize local cultures, there are a good number of scholars who take culture seriously while not being so judgmental about it. In these perspectives, culture is seen as a vital and, sometimes, revered
aspect of a people’s identity, and as representing the integral and holistic nature of a society that changes in response to internal and external pressures. Culture is not apart from or outside of economic life, but is instead integral to and interactive with it. Lourdes Arizpe, in reference to the UN Commission on Culture and Development, says: ‘it is not culture that is embedded in development; it is development that is embedded in culture’ (Arizpe, 2004). Gunnar Myrdal was among the early economists to consider culture in this way, emphasizing that the development of a society needed to be fundamentally based on the people’s own choices, which would in turn be based on their own values (Myrdal, 1968).

Another leader of this third way was Denis Goulet, who advocated an understanding of development that respected local cultures while at the same time recognizing the need for cultures to change. Goulet thus tried to find a way out of what he termed ‘The Cruel Choice’, which forced cultures outside the Western mainstream to choose between keeping their local cultural traditions and staying poor, or opting to join modernizing trends and losing their identity and sense of meaning (Goulet, 1971, 1980). His solution was to work within cultures and to find the ‘latent dynamisms’ that allowed cultural groups to respond constructively to the challenges of modernization. His approach coincides with the views of anthropologist Mary Douglas, who thinks the question of which cultures are ‘better’ is misguided and dangerous (Douglas, 2004). The common practice of evaluating which religious traditions are more prone to progress, for example, is the wrong way to approach the cultural question. Instead, she promotes a theory that sees every culture as a mixture of four groups and tendencies: the hierarchical, the entrepreneurial, the dissenting and the apathetic. Each of these first three plays crucial roles in the maintenance, protection and growth of cultures in their ever-changing environments. The hierarchical types, often government and religious leaders, want to keep traditions as they are and thus ensure social stability. Entrepreneurs, often from the world of business, are change agents who test the limits of their cultures by trying out and promoting new ways of doing things. Dissenters are typically idealists and visionaries who provide checks and balances on both groups. The apathetic tendency arises out of practices and patterns of life that marginalize whole groups of people or isolate them from the circles of power and decision-making. As the size of the apathetic group grows, so too does cultural distress. In many societies where poverty is prevalent, a study of the interaction of these four groups will reveal much about who holds power, how it is used and how whole groups are consigned to poverty. The analysis is not so much about whether a culture is good or bad as a whole, but about the way power is held and used by the different groups within the culture.
Interaction of development and culture

The notion that economics and culture make up integral parts of a large whole has led David Throsby to hint at the possibility of a new development paradigm that manages to bring these ‘two disparate fields closer together’, for ‘culture is in fact central to and inextricable from the development process, providing both the context within which economic progress occurs and the very object of development from the perspective of individual needs’ (Throsby, 2001, pp. 164, 165). This is the project of the two World Bank-spawned volumes cited earlier, and it seems also to be the project of Amartya Sen, who has been a highly influential voice in the development of the Human Development Index that is reported in the United Nations’ annual *Human Development Reports*.

As an economist, Amartya Sen is oriented toward matters of efficiency, growth and distribution, but, like Goulet, believes that any decisions to change culture must arise from the people themselves, in processes that ensure and facilitate their active and informed participation. Sen has been in the forefront of a new perspective on development known as the ‘capabilities approach’. Based on the idea of individual freedom within cultural contexts, this perspective respects culture by insisting on the legitimacy and inclusion of every voice by allowing the people themselves to evaluate their own cultural institutions and ways of life.8 Understanding culture as varied and complex, Sen affirms that culture and economic development are tightly interwoven, arguing also that both cultures and economies evolve in mutually reinforcing patterns. Whether and how economies and cultures should change, however, can only legitimately be determined by the people themselves, which is why Sen is so insistent that the capabilities to make such decisions be widespread throughout society. The capabilities approach thus assesses more whether people have both the individual capabilities (for example, education) and the necessary participatory and inclusive social structures for making informed choices about their own development path, than it does their achievement of predetermined ends, like high gross domestic profit (GDP) growth rates. While respectful of culture, Sen also points out that culturally based arguments are often improperly employed by powerful leaders who invoke the culture argument to defend oppressive systems. He is not persuaded, for example, by the argument that Asian values are more oriented to authoritarian styles of governance than European or other sets of values. Instead, he finds it is the leaders and holders of authoritarian power rather than the masses who support so-called Asian values. He argues that the real purpose of the Asian values argument is not to support local culture, but to legitimize the leaders’ hold on power.

For Sen, generalized poverty is often explained by the fact that the poor are disempowered and marginalized and lack key individual and social
capabilities. He is famous for noting the fact that famines do not occur in democratic countries, which is a consequence of the breadth and depth of participation in social, political and economic institutions. Where wealth is widespread, on the other hand, there will also tend to be a widespread diffusion of education, opportunity and voice, which combine to define capability.

Sen’s influence at the UN is especially transparent in the 2004 Human Development Report entitled Cultural Liberty in Today’s Diverse World (UNDP, 2004). It respects culture, recognizes that cultures can and should change in the development process and identifies widespread participation as the legitimizing principle for choosing. The report begins with the assertion that: ‘cultural liberty is a vital part of human development because being able to choose one’s identity – who one is – without losing the respect of others or being excluded from other choices is important to leading a full life’. In contrast to some of the studies cited above, the report argues that ‘there is no evidence from statistical analysis or historical studies of a causal relationship between culture and economic progress or democracy’, thus rejecting the Weber hypothesis. The report goes on to argue in favor of multicultural democracies that ensure the broadest possible participation in the social choices that affect people, identifying such choices as fundamental human rights. Furthermore, the report does not accept the reality of Goulet’s ‘cruel choice’, for participation guarantees that there need be no trade-offs between cultural integrity and human rights, democracy and economic improvements.9

Yet many would argue that Sen and the Bank are too optimistic. Modernizing development strategies have always meant conflict with traditional cultural institutions. Goulet pointed out in one of his last works: ‘Under the banner of development, powerful standardizing forces dilute cultures and relegate them to purely ornamental, vestigial or marginal positions in society’ (Goulet, 2005, p. 23). He highlights three of these standardizing forces: (1) technology, especially media technology, which spreads the values of individualism, instant gratification and consumerism; (2) the modern state which centralizes everything, including ideas and values; and (3) the managerial ethos which spreads cost–benefit analysis to every sphere of activity as the best way to make decisions.

Moreover, the pervasiveness and the creative destructiveness of these standardizing forces have fomented cultural resistance in many parts of the world, probably best seen in Muslim countries. The present style of development and globalization threatens to generate a whirlwind of cultural and political backlash.

Culture can both hinder and/or aid development, and development can in turn both harm and/or help culture. What drives the change, and whether
development institutions like the World Bank, national aid agencies and NGOs ought to promote such change, are fundamental and controversial questions. Regardless, cultures must and will change. How they will change becomes the issue. Will the changes be destructive, resulting in people losing what gives them meaning in life? Or will people become part of a change process that allows them to adapt gradually and intentionally with minimal social and cultural cost, and without violent and regressive backlashes? Over 30 years ago Jim Lamb pointed out the road that needed to be traveled if that cultural and political backlash is to be avoided: ‘Development should be a struggle to create criteria, goals, and means for self-liberation from misery, inequity, and dependency in all forms. Crucially, it should be the process a people choose, which heals them from historical trauma, and enables them to achieve a newness on their own terms’ (Lamb, 1973, p. 20).

Notes
1. The basic sources for this section are: Tawney (1926) and Weber (1958).
3. See especially the final chapter.
5. See Barro and McCleary (2003, p. 37). The paper can also be found at http://www.economics.harvard.edu/faculty/barro/papers/Religion_and_Economic_Growth.pdf. Interestingly, Barro and McCleary also find that though certain religious beliefs have a causal impact on economic development, economic development itself also generates a decline in overall religiosity, thus supporting to some degree the secularization thesis.
7. See Sen (1999) for his most complete explanation of the theory. Another main voice in developing the capabilities approach is philosopher Martha Nussbaum. See Nussbaum (2000) for a discussion of how she integrates the legitimacy of local cultures with an attempt to discover universal values.

References


Easterly, William (2006), *The White Man's Burden: Why the West’s Efforts to Aid the Rest have Done So Much Ill and So Little Good*, New York: Penguin Press.


Introduction
The incidence of violent conflict among poor countries is high: seven out of ten of the poorest countries have recently experienced some sort of civil war. Since conflict has a serious adverse impact on development, as well as causing massive human suffering, efforts to promote development and reduce poverty must include policies to prevent conflict and to protect populations during conflict. Moreover, as poorer economies are more likely to experience conflict and conflict inhibits development, a vicious cycle can ensue – of underdevelopment–war–underdevelopment – which it is essential to break if either peace or development is to be sustained, yet to do so is very difficult. This review covers both sides of this cycle: the following section analyses economic causes of contemporary conflicts; the subsequent section explores economic and social consequences.

Economic explanations of violent conflict
While some attribute contemporary conflicts to fundamental differences arising from ethnicity or religion (for example Huntington, 1993), such differences are evidently insufficient as an explanation since many multi-ethnic or multi-religious societies live peacefully while others are at peace for decades before experiencing conflict. In fact, the vast majority of multi-ethnic societies are at peace (Fearon and Laitin, 1996). Therefore, we need to look beyond ethnicity to issues of power and economics to understand conflict (Cohen, 1974). Below we consider four explanations that have dominated recent economic analysis of conflict: group motivation and inequalities; private motivations; a failed ‘social contract’; and environmental pressures (‘greenwar’).

Group motivation
Political conflicts consist in fighting between groups – groups that wish to gain independence or take over the state, and others that resist this (Horowitz, 1985). Such groups bring individuals together with a common purpose. While individual motivation is also important, this perspective argues that group motivation and mobilization underlie most political conflicts.

Groups engaged in internal conflict are often united by a common ethnic
Since 1945, the proportion of conflicts attributable to ethnic violence has been steadily increasing (Figure 63.1). While such conflicts are generally presented in religious or ethnic terms, and such identities provide a powerful source of mobilization and unity, underlying differences in access to economic or political resources are generally also present, providing both leaders and followers with a strong motive to fight. Gurr (1970, 1993) terms such group differences ‘relative deprivation’ and Stewart (2000, 2008) defines differences in groups’ access to economic, social and political resources as ‘horizontal inequalities’. Horizontal inequalities consist of inequalities in access to resources between groups differentiated by racial, ethnic, linguistic or religious characteristics, in contrast to vertical inequality which measures inequality among individuals or households. The horizontal inequalities explanation of conflict is based on the view that when such cultural differences coincide with economic and political differences between groups, this can cause deep group resentments that may lead to violent struggles.

Note: The data from which these figures were calculated – the list of ‘major episodes of political violence’ compiled by Monty G. Marshall of the Centre for Systemic Peace – provide a ‘magnitude’ score for each episode, ranging from 1 (mildest) to 10 (severest). Total magnitude here is calculated simply as incidence weighted by magnitude.


Figure 63.1 Ethnic violence as a proportion of ‘major political violence’, 1946–2004
Empirical evidence is accumulating that horizontal inequalities constitute a significant cause of violent conflict. Cross-sectional quantitative analyses have shown a significant relationship between various dimensions of socio-economic inequality and conflict. Mancini (2008) shows that differences in infant mortality rates – a broad proxy for levels of socio-economic deprivation – between ethnic or religious groups among districts in Indonesia help explain the location of the communal conflicts that occurred after 1998. A similar relationship – between ‘spatial’ horizontal inequalities and the intensity of insurgency – has been found in Nepal (Gates and Murshed, 2005). Although multi-country studies have been hampered by poor data, supporting evidence has been found by both Østby (2004) and Barrows (1976). There is also substantial case study evidence: Stewart’s (2002) review of the experiences of nine countries shows not only that severe socio-economic horizontal inequalities preceded the emergence of violent conflict, but that reductions in socio-economic horizontal inequalities – such as occurred in Northern Ireland during the 1980s – may contribute to the conditions for a peaceful resolution of such conflicts. However, some societies show severe horizontal inequalities without experiencing conflict. Political inclusiveness is one reason that some societies avoid conflict despite severe economic horizontal inequalities; other reasons are lack of unity among the deprived groups, and state repression.

Political horizontal inequalities – the exclusion or under-representation of groups within the political structure of a state – can provoke violent conflict, especially when they change abruptly. In Côte d’Ivoire, three decades of post-independence rule by Félix Houphouët-Boigny avoided significant conflict, largely due to the policy of balancing representatives of the major groups in positions of importance in the government and bureaucracy. Following Houphouët-Boigny’s death and the introduction of multi-party elections in the early 1990s, political leaders sought to mobilize ethnic sentiments to enforce their grip on power and thus undermined Houphouët-Boigny’s careful balancing act, leading to a spiral of ethnicization, xenophobia and, ultimately, civil war (Langer, 2005).

It is important to note that relatively rich groups may instigate conflict, as well as the relatively poor. The relatively rich do so to preserve their riches (and/or power), while the relatively poor do so out of a sense of injustice with the intention of achieving some redistribution.

**Private motivation**

People who fight are, of course, individuals with their own private motivation as well as being members of a group. War confers benefits as well as costs on some individuals. Political sociologists (Keen, 1998; Duffield, 1994), and economists (for example Collier and Hoeffler, 2001), have
emphasized private or individual motivation as the fundamental cause of conflict, arguing that the net economic advantages to individuals motivate them to fight. In this approach, which has its basis in ‘rational choice’ assumptions, group identities are not an independent factor but are instruments, created to help fulfil the private motives of those who fight, especially leaders (Hirshleifer, 1994).

Keen lists many ways in which war confers individual benefit on particular categories of people: it permits people, especially uneducated young men, to gain employment as soldiers; it offers opportunities to loot, to profit from shortages and from aid, to trade arms and to carry out illicit production and trade. Where alternative opportunities are few, and the possibilities of enrichment by war are considerable, wars are likely to be more numerous and longer. Conflicts may persist because some powerful actors benefit through the manipulation of scarcity, smuggling, and so forth and have no interest in resolving the conflict. An oft-cited case used to support this view is the role of ‘conflict diamonds’ in the prolongation of the civil war in Sierra Leone (Collier, 2000, p. 5).

However, case studies suggest that even where natural resources are abundant, private maximizing motives are rarely the full explanation. A study of seven countries in conflict concluded:

very few contemporary conflicts can be adequately captured as pure instances of ‘resource wars’ . . . Economic incentives have not been the only or even the primary causes of these conflicts. (Ballentine and Sherman, 2003, pp. 259–60)

In most cases of conflict, the risk of death or debilitating injury are high, so that ‘rational’ actors might be likely to choose another option before engaging in rebellion. This argument may not apply to leaders, who are less often killed or injured, while followers may be coerced into fighting, or persuaded to fight by leaders playing up religious or ethnic differences and grievances:

Grievance is to a rebel organization what image is to a business . . . [A] sense of grievance is deliberately generated by rebel organizations . . . [rebels supporters] are gulled into believing the discourse which self-interested rebel leaders promote. (Collier, 2000, p. 5)

At this point the group explanation and individual explanation of conflict come together. Grievances are hard to sell to the extent of people risking their lives if they are not genuine (that is, unless there is some sort of exclusion or economic horizontal inequalities), while leaders are motivated by political exclusion (that is, political horizontal inequalities) which denies them access to resources and power.
While generally not a sufficient explanation of conflict, it is clear that expected rewards often play a role in the decision to rebel. Econometric evidence confirms that conflict incidence is higher in resource-rich areas (Humphreys and Varshney, 2004). The gains (and motivation) in resource-rich areas may be individual or group, or both. As Collier notes, citing the cases of Aceh (Indonesia), Biafra (Nigeria) and Katanga (Zaire), separatist rebellion often emerges in resource-rich areas of a country (Collier, 2000, p. 10). Yet all these conflicts were framed in ethnic terms. Moreover, in many cases the leaders of the rebellions left lucrative and safe positions to instigate rebellion. For example, Hassan di Tiro left a secure position at the United Nations to instigate the Acehnese uprising. In the case of Colombia, often depicted as a ‘greed’-motivated conflict, interviews with both leaders and those mobilized to fight show that generally their economic position worsened as a result of participating in the conflict – most put forward ideological reasons for their actions, especially the issue of land reform (Gutierrez Sanin, 2004).

There are also examples of separatist movements in regions with poor resource endowment such as Eritrea, Bangladesh (then East Pakistan) and the Tamil rebellion in Sri Lanka. Moreover, it is not possible to create an identity out of nothing (Smith, 1991). A common history, language, culture or religion is generally required to generate felt identities powerful enough to mobilize people for conflict.

**Failure of the social contract**

A third theory of violent conflict derives from the view that social stability is implicitly premised on a social contract between the people and the government. According to this hypothetical contract, people accept state authority so long as the state delivers services and provides reasonable economic conditions in terms of employment and incomes. With economic stagnation, or decline, and worsening state services, the social contract breaks down and violence results. Hence high (and rising) levels of poverty and a decline in state services would be expected to cause conflict (Nafziger and Auvinen, 2000). High vertical inequality might also be associated with such a failure, unless accompanied by populist measures to compensate the deprived. Conversely, political institutions that are able to channel and respond to socio-economic discontents strengthen the social contract, thus reducing the risk of conflict.

Considerable evidence from econometric studies shows that conflict incidence is higher among countries with lower per capita incomes, life expectancy and economic growth (Elbadawi and Sambanis, 2000; Nafziger and Auvinen, 2000; Collier and Hoefler, 2001). Many analyses have found an inverted U-curve relationship between the extent of democratization in
a country and the risk of conflict (for example Ellingsen, 2000), with the usual interpretation being that ‘stable’ democracies are able to avert violent conflict through a strong social contract, while strongly authoritarian regimes are able to suppress conflict. However, Reynal-Querol (2002) has argued that it is the particular type of democracy – whether majoritarian, presidential or proportional representation – that affects propensity to conflict, rather than the level of ‘democracy’ per se.

‘Greenwar’ and environmental scarcity
The fourth explanation of violent conflict, associated with the work of Homer-Dixon and the ‘Toronto Group’ (for example Homer-Dixon, 1994; Percival and Homer-Dixon, 1998), is the ‘greenwar’ or ‘environmental scarcity’ argument. The essence of this perspective is that contest for control over declining natural resources, often intensified by population pressures, is a major cause of violent conflict around the world. Poorer societies are more at risk because they will be ‘less able to buffer themselves’ from environmental pressures (Homer-Dixon, 1994, p. 6). Three dimensions of environmental scarcity are identified which may lead to conflict: ‘supply-induced scarcity’, linked to the ‘depletion and degradation of an environmental resource’; ‘demand-induced scarcity’, linked to population growth and the consequent extra pressures on existing resources; and ‘structural scarcity’, which ‘arises from an unequal distribution of a resource that concentrates it in the hands of a relatively few people’ (Percival and Homer-Dixon, 1998, p. 280). Homer-Dixon thus predicts ‘an upsurge of violence in the coming decades that will be induced or aggravated by scarcity’ (Homer-Dixon, 1994, p. 6).

However, while it is clear that pressures arising from environmental scarcity may play an important role in many conflicts, the environmental scarcity hypothesis is – and really does not claim to be more than – a partial theory that contributes towards our understanding of a set of conflicts, but not the general conditions under which conflict is more likely to arise.

The environmental scarcity hypothesis overlaps substantially with the other hypotheses discussed here. It overlaps with the social contract hypothesis in viewing poverty as the root cause of conflict, although it points to specific environmental causes of such poverty. It also often overlaps with the group motivation approach, as environmental pressures usually lead to conflict where there are ‘groups with strong collective identities that can coherently challenge state authority’ (Percival and Homer-Dixon, 1998, p. 280). Indeed, the ‘structural scarcity’ dimension of the greenwar approach is very similar to the group motivation hypothesis, albeit restricted to a particular dimension of inequality.
The environment scarcity view has been criticized by Fairhead (2000) who argues that it is environmental riches, not scarcity, that is associated with conflict because people fight to control lucrative natural resources, as in the Congo. This view fits well into the private motivation or greed hypothesis. In fact both environmental poverty and environmental riches may cause conflict, for different reasons and in different circumstances.

The theories outlined above appear, in their extreme formulations, to be diametrically opposed – as manifest in the ‘greed versus grievance’ debate (Ballantyne and Sherman, 2003). But, as we have seen, proponents of one perspective usually accept in part the insights of other perspectives. Some conflicts fit neatly into one of the explanations, some into others, and some clearly have multiple causes. One rather simple conclusion, therefore – that qualitative analysts of conflict are mostly aware of, but that quantitative analysts tend to overlook – is that each of the broad causal theories discussed above involves a degree of oversimplification and generalization. The causes and dynamics of any single conflict are typically complex, sometimes contradictory, and involve aspects of many, if not all, of the perspectives discussed above.

Yet it is important to understand which explanation dominates in a particular case, since this has important implications for appropriate policy prescriptions for the prevention and resolution of the conflict. There is not space here to discuss policies in detail. In brief, where group motivation is a fundamental cause, policies need to be inclusive and correct large horizontal inequalities (Stewart, 2008). To the extent that private motivation is key, policies need to reduce the profitability of illicit war-sanctioned activities (like drug production and smuggling), and to offer combatants income-earning opportunities as an alternative to fighting. If a failed social contract is the fundamental cause, then the aim should be to improve the functioning of government in relation to security, the economy and the provision of social services. Greewar conflicts require an attack on the fundamental causes of environmental pressure, both from demand and supply perspectives. Such policies should be applied to all societies at risk of war, which includes all low-income countries, any country which has experienced conflict in the past few decades, and any economy suffering severe horizontal inequalities.

Two points are worth making about this policy set. Firstly, none of these policies are a central part of the current development agenda of the international financial institutions; secondly, they are all desirable in themselves, quite apart from their impact on conflict prevention.

**The economic consequences of war**

War in general, and civil war in particular, is one of the main causes of human suffering and economic underdevelopment. Most of the human and
economic costs of war do not result directly from battle deaths and injuries, but indirectly from the loss of livelihoods caused by the dislocation of economy and society. An important implication of the substantial indirect costs is that policy might be able to reduce costs substantially if appropriately designed, even while conflict is ongoing.

The economic effects of war are the result of a complex interaction between the particular processes of war and the economy in which it takes place. Figure 63.2 provides an overview of the main relationships likely to be affected, with arrows indicating the direction of causality and expected direction of impact. It is helpful to distinguish between the direct economic consequences of the conflict, and the compensating behaviour of economic agents in their attempt to moderate or offset the negative impacts of war.

Direct effects include: output loss as people move from their place of work because they join the fighting, are killed or flee; the destruction of capital and consequent loss of output; disruption of transport links due to physical destruction; a loss of trust among economic agents, reducing market transactions; disruption of international markets due to frontier closure or embargoes; reduced foreign investment and the diversion of foreign exchange from economic and social needs to military uses.

These effects will tend to reduce aggregate levels of output. Labour markets will be disrupted as many of the unskilled of prime working age become military recruits and some suffer violent deaths, while much skilled labour is likely to leave the country. Reduced agricultural output, and disrupted internal and international markets, are likely to affect exports particularly heavily. Reduced foreign exchange availability for productive inputs results, leading to a shortage of imported inputs and to a further fall in output and exports.

Compensating behaviours which can moderate the negative impact of the effects of war include: increased capacity utilization and import substitution; the provision of international credit or aid, offsetting loss of foreign exchange; the emergence of new forms of social capital compensating for loss of trust in formal institutions – such as enhanced cooperation and trust among members of a group on the same side of the war; and rapid government or community action to reconstruct facilities destroyed by belligerents.

To understand the total impact of these mechanisms on individuals one needs to go beyond money incomes to explore how they affect different types of individual entitlements, including market entitlements (acquired largely through employment and self-employment), direct entitlements (subsistence production), social entitlements (provided by the state), civic entitlements (provided by the community and NGOs) and extra-legal entitlements (see Stewart et al., 2001a, Chapter 1).
Figure 63.2 Predicting changes in the economy and entitlements during conflict
There are serious methodological problems in estimating the costs of war (Stewart, FitzGerald and Associates, 2001a, Vol. 1, Chapter 1). Below we summarize results adopting a variety of methods.

**Macroeconomic consequences**

All studies find negative effects on gross domestic product (GDP) growth. One study found a negative impact on GDP per capita growth in 13 out of 14 countries suffering the worst conflicts between 1975 and 1995, with considerable variability in magnitude. The worst losses occurred in long and pervasive wars (Stewart et al., 2001a). Regression analysis of 92 countries, 1960–89, showed an annual loss of 2.2 per cent during the war and in the immediately following years, compared with a no-war situation (Collier, 1999). Other cross-country regression analysis for 1960–99 came to similar conclusions, with an average loss of growth of 2.4 per cent per annum (Hoeffler and Reynal-Querol, 2003), although Imai and Weinstein (2000) suggest somewhat lower costs. The wide range of estimates indicates how dependent they are on the methodology adopted. Nonetheless, all give negative results. Evidence shows that wars with more widespread geographic coverage have a more negative impact (Stewart, Fitzgerald and Associates, 2001; Imai and Weinstein, 2000).

Reduced economic growth is the result of capital destruction, lower investment and disrupted markets. All types of capital stock are eroded or destroyed in war. Physical facilities suffer direct attack – roads, ports and energy plants are often targeted. In Mozambique, Brück estimates that there was a two-thirds reduction in operational dams and plant nurseries, with 40 per cent of rural facilities destroyed or eroded. Social infrastructure is also commonly destroyed – again in Mozambique, almost 60 per cent of primary schools were closed or destroyed (Brück, 2001, pp. 64–7). Human capital is killed, or flees. About half the doctors and 80 per cent of the pharmacists left Uganda in the late 1970s (Dodge and Wiebe, 1985). The spread of AIDS that results from the sexual activities of combatants further reduces human capital. Institutions are destroyed – in Uganda, the agricultural extension system virtually disappeared (Matovu and Stewart, 2001). Social capital is weakened, with a severe loss in trust, particularly across groups. Yet new forms of institution and social capital emerge, for example informal banking systems.

Rates of investment and savings fall due to increased uncertainty. Evidence shows falling domestic saving rates (Stewart et al., 2001a; Brück, 2001), while capital flight accelerates. The proportion of private wealth held abroad rose from 9 per cent to 20 per cent in the course of civil wars according to Collier et al. (2004). Voluntary private lending from abroad tends to fall with increased uncertainty, but changes in official foreign lending
depend on political factors. ‘Forced’ foreign savings may occur, as countries renege on debt servicing obligations. In fact, aggregate foreign savings seem to hold up more than might be expected, evidenced by the huge accumulation of foreign debt during conflict (Stewart et al., 2001c).

Private (domestic and foreign) investment is adversely affected by uncertainty, rising costs of transport and difficulties in securing finance. Foreign investors are likely to be concerned about the safety of their personnel and equipment, and the increased foreign exchange risk. Country studies show a fall in foreign direct investment as expected – indeed this was the most important macroeconomic cost of Sri Lanka’s conflict (O’Sullivan, 2001). Imai and Weinstein (2000) show a strong negative impact of civil war on private investment. Government investment is likely to be negatively affected by reduced revenue, and diversion of expenditure to military uses – Ra and Singh (2005) estimate that development expenditure in Nepal fell by one-third, 2001–04. Country studies show that aggregate investment on balance does not fall as much as domestic savings, possibly due to buoyant small-scale investment as the informal sector expands.

Exports are also negatively affected, as a result of the general fall in production, a shift towards domestic markets and disruptions in international markets. In most cases, imports hold up much better than exports, financed by foreign debt. The share of imports going to military items and food imports rises, however, leaving a much smaller share for inputs into the productive sector.

Government revenue is likely to fall absolutely and as a proportion of gross national product (GNP) as the government finds it more difficult to collect taxes and major sources of revenue (for example, from export taxes) fall away. There are sharp divergences across countries. In Uganda (1979–80), revenue as a share of GDP fell dramatically, but in both Mozambique and Nicaragua in the 1980s, the revenue ratio rose. Budget deficits increase as government expenditure rises faster than revenue (Stewart et al., 2001c).

Inflation is expected to accelerate, as governments resort to deficit financing to finance the conflict and other essential services and public confidence in the currency declines. In recent wars, there seems to have been only minor acceleration in price inflation (Stewart et al., 2001c).

**Meso-economic consequences**

In general, there is a shift from tradeable to non-tradeable sectors, as a consequence of market disruptions, including undermining of formal organizations such as banks, reduced trust and failures of the transport system. One aspect of this is a switch to subsistence and informal activities, including simple production (even arms) and trading (particularly smuggling).
For example, Mozambique experienced rapid growth in the urban informal sector; while the ability to shift from producing marketed crops to subsistence agriculture also helps protect food production and nutritional standards – it did so, for example, during the Amin era in Uganda.\(^1\)

The share of government expenditure going to military items invariably increases, making it difficult to sustain social and economic expenditure. On average, it is estimated that military expenditure rises from 2.8 per cent of GDP (average for developing countries in peacetime, 1995) to 5.0 per cent of GDP (Collier et al., 2003). One consequence is likely to be a fall in the share of social expenditure. In Uganda, social expenditure fell by roughly 9 per cent annually during conflict. In exceptional cases such as Nicaragua in the 1980s, countries give increased priority to social expenditure and succeed in increasing public entitlements. Where government revenue collapses there may be a dramatic decline in public entitlements: this was true of Uganda in the late 1970s and mid-1980s and Afghanistan in the early 1990s.

Civic entitlements can substitute for reduced government provision. There was extensive foreign non-governmental organization (NGO) provision of social and economic services, for example, in Afghanistan in the early 1990s (Marsden and Samman, 2001). In Sri Lanka, local NGOs and communities and the Tamil rebel forces provided significant support (O’Sullivan, 2001). But in the worst-affected areas, communities disintegrate as people flee, and NGOs are able to do little – examples are Uganda, Southern Sudan and Sierra Leone.

**Human costs**

In addition to deaths and injuries, flight and ensuing psychological trauma, human costs result from the changing economy, with worsening entitlements of most types:

- Market entitlements decline as household incomes fall with worsening employment conditions, while the main earners may leave the household to fight or flee. However, extra-legal entitlements rise, with big gains for some households profiting from types of illegal production, but losses for others, subject to theft and looting.
- Worsening nutrition arises from reduced incomes and agricultural output. Calorie consumption fell in over 70 per cent of the countries worst affected by conflict, in 1970–95, falling to below 1700 calories per person per day in Ethiopia, Liberia, Mozambique, Sierra Leone, Somalia and Afghanistan (Stewart et al., 2001c, p. 90). The agricultural sector is typically badly hit in civil wars, as people are forced to move. Rising food prices can have a devastating impact on access to food. Millions of deaths in the Bengal famine of the 1940s have been
attributed to war expenditures leading to food price increases (Sen, 1981). Speculative traders can also contribute to rising food prices (Ravallion, 1987). The impact on nutrition may be offset by food subsidies, food distribution and rations (including food aid). In Nicaragua, nutrition actually improved in war-affected zones as a result of such measures (Utting, 1987), while in Mozambique, food aid seems to have prevented mass starvation (Stewart and Samman, 2001).

- Social entitlements, including health and education services, worsen as a result of destruction of facilities, reduced government resources and flight of personnel. Government expenditure on health and education is estimated to have fallen in ten of 14 war-affected countries (1970–95), by over 40 per cent in Angola, Liberia, Uganda, El Salvador and Iran (Stewart et al., 2001c, p. 87). Primary school enrolment fell substantially in some countries – notably Angola and Mozambique. Doctors per person also fell significantly in about half the war-affected countries. Civil society in some situations is able to substitute for government services but it too disintegrates in the worst cases.

- Health is affected by increased infection rates associated with the mass migration that often accompanies war – as many as one-third of the people in Mozambique were forced to move, while one-third of the Afghan population left the country in the 1990s. The spread of AIDS has also become a particular feature of war, as soldiers are notably highly infected and infect others, including sometimes through mass rape. HIV prevalence in the military was estimated to be 40–60 per cent in Angola, the Democratic Republic of Congo and Sierra Leone in the late 1990s (Collier et al., 2003). The extent of health costs is indicated by rising infant mortality rates: in Uganda additional infant deaths, compared with non-war regional trends, amounted to over 2 per cent of the 1995 population. Econometric estimates across countries show an increase in infant mortality rate (IMR) of 13 per cent during conflict (Hoffler and Reynal-Querol, 2003). Moreover, some case study evidence suggests increases in adult mortality rates sometimes exceed increases in IMR (Guha-Sapir and Van Panhuis, 2002). World Health Organization (WHO) estimates of disability-adjusted life years show a major loss due to increased infectious diseases (Ghobarah et al., 2003), which persists in the post-conflict era.

In summary, while the direction of impact on most variables is demonstrably negative, and most types of entitlement worsen, there are large variations in the magnitude of costs across countries and on the burden of
entitlement loss across groups within a country. Variations in the social and economic impacts of war arise from differences in the nature of the war, the structure of the economy, the character of the government, the response of the international community and people's own actions. Moreover, while the evidence on worsening GDP and average consumption levels suggests that the net situation is one of loss, there are also gains from conflict for particular groups which can form one of the private motives for war or its prolongation as discussed above.

The economic consequences are obviously highly dependent on the nature of the war itself. Firstly, and above all, its duration is important. In a long war, reserves become exhausted, so vulnerability is increased. The geographic spread of the war is also relevant. When confined to one part of the country the war may have only small direct effects on the economy as a whole – conflict in Northern Uganda, which persisted over decades, had much smaller economic effects than the conflict in the mid-1980s which was centrally located. The extent of foreign involvement in the war is another factor affecting vulnerability, since external support may compensate for lost export earnings; it may also, however, contribute to prolonging the conflict.

Secondly, the structure of the economy helps determine the costs. An economy heavily dependent on the agricultural sector will be especially badly affected by widespread disruption of the sector, but may be less affected by reduced imports, since it is possible to retreat into subsistence if markets break down; an inflexible economy with a sizeable industrial sector may be particularly vulnerable to foreign exchange loss, although this can be offset by external grants or credit. An economy with a flexible industrial sector operating at less than full capacity may suffer little, as it is able to make up for loss of imports by domestic production – the UK in World War Two is a classic example.

A third critical factor is the nature of the government: a government that is or becomes very weak loses the ability to provide essential services and relief. Strong governments can sustain services, but they will only do so if they are relatively benevolent, wishing to provide for all the people, despite the war. Some strong governments may deliberately reduce food and basic services to ‘enemy’ territory, as was the case in Sudan.

Fourthly, international actions contribute to variations in costs. The large supplies of food aid in Mozambique undoubtedly reduced death rates, but in Sudan in the 1980s food aid was too little, too late and its delivery and use were distorted by government policy (Keen, 1994). In Afghanistan in the 1990s, massive support for the refugees in Iran and Pakistan greatly lessened the human costs. In contrast, the international community did little to offset costs in Sierra Leone in the 1990s, while the
USA increased the burden of civil war for vulnerable groups in Nicaragua by trade and aid embargoes.

Finally, people’s own actions are important in moderating both human and economic costs. In almost all cases, people found new economic possibilities – many created by the war – which enabled them to survive. The burgeoning of the informal sector in Mozambique is one example (Chingono, 2001). People also protect themselves by fleeing, relocating within the country, or emigrating. People are rarely completely passive victims, but in the worst situations, there is little they can do to protect themselves.

Policies towards economies in conflict
The analysis above suggests that economic and social policies of both governments and donors can be designed to reduce the economic and human costs of conflict, even during the conflict.

An overriding aim of economic and social policies towards countries in conflict should be to maintain entitlements of the vulnerable, especially to food and health services, if possible in a self-sustaining way. In addition, the policies should also aim to tackle the causes of war, following the analysis above.

There can be no generalization about policy irrespective of the actual situation, since it is essential to understand the major cause of entitlement collapse, whether it is loss of employment opportunities, escalating inflation or destruction of key assets; and to understand the nature of the authorities, whether they are so weak that for short-run action they must be bypassed, or are strong but ill-intentioned so resources channelled through them will not reach those in need, and whether there are alternative structures (for example local authorities or rebel authorities) that can handle projects.

Monitoring is essential to identify appropriate actions before the situation becomes critical. Official monitoring is often weak (and politically biased). Greater use could be made of NGOs, local and foreign, who can be well situated to monitor developments and provide early warning of impending adverse changes in human conditions, for example, distress selling of assets, small movements of people or adverse changes in nutrition. In the Sudan, a major reason for the huge rise in death rates in the early 1980s was the failure to take early action, and this partly stemmed from the development community being slow to note early warning signs (Keen, 1994).

Successful macro-policies are generally much more effective in maintaining essential entitlements than direct relief. The aim should be to sustain the economy – which may require external aid and support for export markets – and to prevent escalating inflation. It is vital to sustain revenue to support public entitlements, compensating for declines in the
normal revenue base by devising taxes on war-related activities. The sale of food aid, for example, can be an important source of revenue.

Meso-policies need to be directed towards sustaining the share of aggregate output going to supporting entitlements of the vulnerable. On the social side, the objective is to ensure that everyone has access to adequate food and to sustain public entitlements to basic health and education. Preventive health measures, notably immunization, are particularly vital in war because unusual movement of people causes infections to spread rapidly. Expenditure on basic health and education accounts for only a fraction of social expenditure, so that strong prioritization of these services can ensure their maintenance even if the total is being cut. Yet the problem, of course, is not only one of money. Teachers and doctors may flee, and facilities can be destroyed. A flexible approach is needed. For example, Mozambique introduced mobile clinics and classrooms when Renamo was targeting health and education buildings.

Ensuring food security requires that food prices are monitored and escalating prices prevented, through some combination of increasing supplies (via food aid), controlling prices and rationing – policies which advanced countries have adopted when themselves at war. For the rural population, a combination of ensuring adequate agricultural support (seeds, fertilizer and so on), employment schemes, and the provision of food in schools and clinics, can achieve wide food access.

As well as domestic policies, there are important spheres of international action which can help (or worsen) the situation – including long-term policies to support development (such as improved terms of trade and aid flows), and short-term policies to reduce human costs (such as welcoming refugees, and providing food aid), as well as policies to reduce the financing of conflict (including policies towards trade in conflict commodities: for example, diamonds). International policies need to take into account the impact on the livelihoods and survival of poor populations. Economic sanctions, for example, are frequently harmful to the poor and often ineffective in achieving political objectives (Clark, 1996; O’Sullivan, 2003).

Conclusion
This overview of the economics of war and development suggests three major conclusions. Firstly, economic analysis of causes and consequences of conflict is essential to design appropriate policies. Secondly, both causes and consequences are diverse, varying across countries, so in-depth knowledge of the particular case is vital. Thirdly, there is a rich menu of appropriate policies which can help reduce the incidence and costs of conflict – yet in most cases these policies currently form a part of neither the normal development agenda nor the normal relief agenda.
Note
1. The burgeoning of subsistence and informal activities means that official statistics can greatly understate production, so that the aggregate costs of conflict may not be as great as they appear from official data.

References


PART IX

ECONOMIC DEVELOPMENT EXPERIENCE
Introduction

According to conventional economic logic the surest way for poorer countries and communities to achieve lasting gains in economic welfare is by getting rid of state-created ‘distortions’ that impede market forces in allocating existing, and mobilizing future, resources. In the recommended reform package, opening up to international markets and firms offers the surest way of bringing about the right set of incentives to help realize those gains. The liberalization of international trade will enable countries to exploit their comparative advantages better, with big gains expected in the South given their legacy of protectionist policies. Where finance is a constraint on growth, the liberalization of international capital markets will ensure that investment funds flow from the capital-abundant North to the capital-scarce developing countries. Attracting foreign direct investment (FDI), including through the sale of state-owned assets, will help gain quick access to new technologies and management practices as well as providing ready-made export opportunities. Entrusting development to these international market forces is seen as having the additional advantage of their being less vulnerable to capture by local interest groups and rent-seeking coalitions. A plethora of econometric studies have backed up this case for hitching development to open markets, and a string of popular economic pundits have been ready at hand to explain just how their ‘win–win’ logic has already begun to flatten out the global economy, raising expectations of a swift eradication of extreme poverty, a narrowing of income gaps, and the emergence of a truly global middle class.

There is little doubting that, in the wake of the debt and development crisis of the 1980s, such thinking, oftentimes with the strong backing of the international financial institutions, affected a radical shift in policy-making in many developing countries. During the 1990s, new technologies and international business practices (closely identified, if not synonymous, with globalization) were added to the reform mix, ending, on many accounts, any further debate on development strategy. Efforts to direct globalization were deemed futile and resistance would result only in marginalization, or perhaps worse.¹ Developing countries were, accordingly, advised to adapt
to this new reality by fully relinquishing economic sovereignty to mobile capital and the forces of international competition.

This chapter questions such advice. It argues that economic trends during the past quarter-century do not support the utopian pronouncements of many globalization enthusiasts; that the destructive impulses released by the radical turn in policy after the debt crisis have, in many countries, outweighed the creative impulses; and that expunging the historical detail and structural diversity from the catch-up process is unlikely to provide the direction on policy advice and institutional reform needed to narrow income gaps worldwide.

Openness, accumulation and structural change in an interdependent world
Most development economists agree that a strong productivity performance is essential to any successful catch-up growth path, not only because it translates (though not always directly) into rising living standards, but also because it enables poorer countries to manage better the various adjustments, trade-offs and distributional conflicts that taking such a path will generate. Most would also agree that this performance is more about galvanizing dynamic economic forces than it is about maximizing the static gains from an improved allocation of existing resources.

Certainly, recent efforts to revive the idea of market-driven convergence for a globalizing world have tried, by introducing a broader conception of capital (including human capital and other less tangible wealth-creating assets, and an expanded role for foreign capital), more variegated technological trajectories and additional behavioural parameters, to accommodate more dynamic forces in their blend of (old) trade and (new) growth models. The empirics of this new convergence literature have already provoked much debate and controversy and methodological questions surround the idea of ‘conditional’ convergence. However, what is more revealing about this literature is its enduring commitment to a set of standardized policies derived from the identification of distortionary market impediments within mathematically tractable equilibrium models. This approach continues to trump historical detail and structural differences in the design of development strategy (Kenny and Williams, 2001).

An alternative place to begin thinking about development strategy is with the empirical regularities identified by economic historians and classical development economists linking industrialization, and more particularly its manufacturing component, to strong productivity and income growth. This leading role is due less to any uniquely intrinsic qualities of industrial activity and more to the confluence of growth impulses that accompany its evolution, including the presence of increasing returns whether at the plant, firm or industry levels, high elasticities of supply and demand for
manufactured goods, strong externalities linked to a high rate of technological innovation, and complementarities between production and consumption. Successful industrialization paths have, moreover, been closely identified with a ‘well-filled input–output matrix’ with an increasingly dense set of links between sectors (a high level of sectoral articulation between, for example, rural and urban, and consumer goods and intermediate goods), and a structure of demand such that a high proportion of domestic production is sold to domestic wage earners (Wade, 2003, p. xlviii).

In the interplay of elements making up this strong growth regime, capital accumulation plays a pivotal role. A given pace of accumulation can of course generate different growth rates, depending on its nature and composition, as well as the efficiency with which production capacity is utilized. This is one of the main reasons why econometric studies have failed to establish a one-to-one relation between the rate of investment and economic growth. Unfortunately, systematic discussion of the forces that govern the process of capital accumulation has long been a stumbling block in the development literature (Hirschman, 1958, p. 35), and the recent growth literature is no exception. In particular, the neglected role of profits for financing investment in developed and developing countries alike, cuts off conventional economic analysis from a careful examination of how dynamic industrial activity can provide abundant opportunities to create rents whose reinvestment is key to perpetuating a dynamic growth regime. This profit–investment nexus has been identified in the recent fast-growth experiences in East Asia (Akyüz and Gore, 1996; Singh, 1999; Ros, 2000) and provides an initial guide to why and how policy interventions might make a difference in stimulating catch-up growth (UNCTAD, 1997).

While industrialization can generate the kinds of cumulative impulses that describe a successful growth process, these are not automatically self-sustaining. Various constraints, traps and coordination failures can upset the process, making it generally impossible to rely on market forces to establish the linkages associated with internal integration and to move economies through the various stages of industrialization. This provides further rationale for policy intervention and institutional learning. The variety of linkages and their local specificity are now much better appreciated. Among these, insufficient domestic demand to absorb the growing industrial output can be just as important as failures on the supply side. The close links between industrialization and ‘external’ integration have, of course, been a familiar feature of the development process since Adam Smith saw the size of the market as a constraint on the division of labour. By broadening the size of the market, exports allow scale economies to be exploited; they also provide the foreign exchange needed for capital accumulation, in view of the dependence of most developing countries on
imported capital and intermediate goods. At the same time, investment improves export potential by adding to production capacity and improving competitiveness through productivity growth. As such, a successful industrialization path is usually characterized by rising investment, exports and manufacturing value added, both in absolute terms and as shares of gross domestic product (GDP).

The historical evidence provides little support for an independent role for rapid liberalization of foreign trade and finance and the deregulation of domestic markets in stimulating and sustaining catch-up growth. In most cases, the state has provided a necessary complement to, and sometimes a corrective influence on, the market, particularly by promoting a rapid pace of capital accumulation and technological progress linked to expanding industrial output, employment and exports. Accordingly, establishing a robust nexus between investment, profits and exports remains key to the design of development strategy, helping identify the cumulative links that describe a successful industrialization path, as well as providing a framework in which to explore the institutional and policy challenges thrown up by the catch-up process.

The international economic environment: open for business

Following the debt crisis of the early 1980s, deregulation of domestic economic activity and its opening up to international firms and market forces became the leitmotif of economic policy design in many developing countries, more often than not with the overt support of the international financial institutions. Success has tended to be measured in terms of monetary and fiscal discipline, an increasing volume of international trade and capital flows, and rising ratios of trade and FDI to GDP. On these measures, many poorer countries, and the world economy more generally, have since the late 1980s registered a good deal of success. Trade has consistently outpaced global output, with the pace of expansion much faster in the 1990s, and with developing countries in the vanguard (UNCTAD, 2003, pp. 41–4). As a result there has been a rapid and ubiquitous rise in the share of exports and imports in GDP in developing countries, as well as a rapid increase in the share of these countries in global trade – from about 24 per cent of total exports in 1980 to 28 per cent in 1995 and 34 per cent in 2004.

The rise in capital flows has been even more dramatic. The global stock of financial assets rose more than elevenfold between 1980 and the end of 2004, from $12 trillion to $136 trillion; in 1980 they were roughly equal to global GDP but by the end of 2004 were more than three times higher. Much of the increase in flows has been among developed countries, but the 1990s saw a strong surge of financial flows to developing countries, following a sharp dip
in the 1980s (UNCTAD, 2003, pp. 23–31). The greater part of these flows consisted of short-term investments: bank loans, equities and short-dated government securities, inter-bank and other deposits. However, beginning in the early 1990s, FDI in developing countries also rose sharply, more than doubling between 1995 and 2000, when it reached over $250 billion, rising from a quarter to close to one-third of global flows.

These numbers do not fully capture the changes in the workings of the international economy. In a world where a good deal of trade takes place between affiliates, where technology transfer is tightly controlled from corporate headquarters and where credit is extended by firms to their customers, corporate governance has become a much more prominent factor coordinating international economic relations. Innovation has also characterized financial institutions as a growing share of their activity is distributed across multiple locations and these institutions tend to offer a variety of services and undertake multiple activities to an extent that breaks with past practice.

Multilateral institutions, rules and arrangements have, over this same period, also gone through a series of mutations in line with a more open and liberal system of trade and capital movements. The International Monetary Fund (IMF), since the mid-1970s, has abandoned its objective of seeking to ensure stable exchange rates in an orderly international financial system and, instead, openly encouraged the opening up of capital accounts as a way of attracting international financial flows. The World Bank has also given up its emphasis on longer-term infrastructure project lending, concentrating instead on adjustment lending and poverty reduction. With the Uruguay Round negotiations, the governance of international trade has moved towards a single-tier system of rights and obligations, in which trade liberalization has been given priority over economic growth and full employment, and linked to a series of ‘related issues’ which take the multilateral rule-making process much deeper into the workings of national economies.

Trade liberalization and the investment climate

Conventional trade theory does not distinguish between different activities in terms of their differential impact on economic growth. In practice, however, most economic policy-makers worry about the composition of trade for this reason. There have been some positive trends in recent years. In particular, the sharp rise in the share of manufactures in exports from developing countries, from one-fifth in the early 1980s to around 70 per cent by 2004, points to an important change in global trade dynamics and one that most proponents of market-led globalization have been quick to highlight as a further measure of its success. However, this is hardly an
exhaustive picture of the changes in the trading system. In the first place, the rise of manufacturing exports is down to just a handful of countries, largely from East Asia. Indeed, simply taking out the first-tier newly industrialized economies (NIEs) from the developing-country group, their share of world trade drops from around 22 per cent in the early 1980s to below 15 per cent by the end of that decade, but has now (in 2005) climbed back to the earlier figure; however, taking out China leaves the developing countries with less than one-sixth of total world trade. By implication, other developing regions have exhibited a stagnant or declining share since 1980, with exports very heavily skewed towards natural resources; these account for more than 30 per cent of export earnings in over 80 countries, often from just one or two products. Together these trends imply that many countries remain stuck with undynamic trading regimes.

Another, and less reported, trend in the trading system is that while many countries have been trading more, they have been earning less from doing so. In a recent study of 127 developed and developing countries, Dowrick and Golley (2004) found that between 1960 and 1980, increased trade helped productivity to grow in poorer countries at double the rate in richer countries, but that this gain was reversed in the period of more open trade between 1980 and 2000, when the marginal impact of trade on productivity growth favoured the richer countries, and indeed turned negative for poorer countries. Wacziarg and Welch (2003), using the Sachs–Warner methodology for determining openness, found that its links to faster growth were period-sensitive, with much weaker links in the 1990s than in earlier decades, and with more open economies if anything benefiting less than relatively closed economies. Finally, increases in the developing countries’ share of world manufactured exports since 1980 have not been matched by a corresponding rise in their share of global value added, and in a number of cases shares of global manufacturing income have actually fallen over the past decade or so, even as their share of world exports of manufactures was rising, while for others it has risen by much less than that share.

One possible explanation for these trends is biases in the liberalization process which have prejudiced growth prospects in developing countries by discriminating against sectors where they had or could build strong export sectors, even as domestic producers were forced to confront strong competitive rivals on home markets (UNCTAD, 1999; ILO, 2004). However, the fact that many countries were trading more but earning less during the 1990s suggests more deep-seated biases in the operations of the trading system (UNCTAD, 2002). These seem likely to be linked to the lopsided reliance on external demand as the basis of sustained growth (Heintz, 2003).
The risk of falling export prices, resulting from too many producers chasing too few markets, has of course been familiar to commodity exporters, where returns have long been asymmetrically skewed through organized markets in rich countries and more disorganized and fragmented markets in poorer countries. However, the structure of some key markets for developing-country manufactured exports seems to point to similar pressures emerging even where trade expansion has been rapid (UNCTAD, 2002, pp. 121–4). In this respect, the suggestion that in the new era of globalization developing-country exporters of some high-technology products have been able to bypass technological constraints is often misleading. Even when the final product has been classified as high-tech, many developing countries have in reality only been involved in low-skill assembly activities, using imported capital and intermediate goods, and whose contribution to value added is determined by the cost of the least-scarce and weakest factor, namely unskilled labour.

The confusion over what constitutes a dynamic export is linked, in part, to the increase in FDI flows through international production networks (IPNs), in which corporations slice up their value chain, relocating or outsourcing the different parts, from product design to final delivery, in a way that enhances overall profits. In some cases, production is organized by large transnational corporations (TNCs) producing a standardized set of goods in several locations (as in electronics and transport industries). In others, production involves groups of small and medium-sized enterprises located in different countries and linked through international subcontracting (as in clothing). Under these conditions, FDI could bring macroeconomic benefits to the host through a positive impact on its investment climate and on its balance-of-payments position. Certainly the expectation after the debt crisis, and particularly following the Brady Plan, was that liberalization, particularly when export-oriented FDI was attracted, would bring such benefits (Camdessus, 1997).

Although the bulk of FDI continues to flow between advanced countries, developing countries have, indeed, experienced a fifteenfold rise in average annual inflows of FDI since the mid-1980s, which has translated into it taking a significantly higher share of capital accumulation in most developing countries and regions. Moreover, between 1990 and 2003, the share of manufacturing in the FDI stock of the group of developing countries rose from 25 per cent to 37 per cent while the share of developing countries in the global stock of manufacturing FDI increased from one-fifth to close to one-third during the same period. However, FDI surges have often failed to stimulate a domestic investment recovery and there is evidence that it has even crowded out local investment (Gosh, 2004). The extent to which these flows have eased the balance of payments constraint has varied.
unpredictably with the share of TNC profits in value added, the degree of import dependence, and the proportion of the final good sold in domestic markets (Akyüz, 2004). Certainly where the propensity to import of foreign firms has been much higher than that of domestic firms, and their export propensities similar, increased FDI has added to balance-of-payments pressures (Chudnovsky and López, 2002). In fact, the broad body of evidence points to this constraint remaining a tight bind on economic growth in many developing countries (UNCTAD, 1999).

In the absence of favourable macroeconomic effects from attracting FDI, much depends on technological and other spillovers that might strengthen internal integration. The broad body of evidence suggests that such effects are plant-, sector- and country-specific, though generally significant effects seem to depend on local absorptive capacity already being in place. As such, FDI tends to lag rather than lead the growth process (UNDESA, 2006). From this perspective, production networks may well increase the risk of TNCs being attracted into enclaves with limited linkages to the domestic economy; taking account of the different methodologies employed, there is little evidence to date of positive spillovers from the recent surge in FDI to developing countries, and virtually none on a significant scale.

Financial liberalization and capital flows
Financial markets have undergone a dramatic transformation since the early 1980s, thanks to a combination of deregulation, internationalization and innovation. While the impulse for this transformation came from the advanced countries, economic logic promised much for the world’s poorest countries. Deregulated and open financial markets would not only increase the availability of investment finance, both domestic and foreign, but they would also help create a more stable and disciplined investment climate, and free deficit countries from the unpredictable politics of ODA flows (Camdessus, 1997; Mishkin, 2006).

An upsurge in flows began in the 1990s, albeit in part a return to trend after the blighted years of the 1980s. Still, a plethora of new financial instruments promised to mitigate risk, particularly in those emerging markets that greatly excited investors after the Berlin Wall collapsed, providing arbitraging opportunities but also encouraging herding behaviour to become a more significant influence on the direction of flows. Consequently, these flows were very unevenly distributed, increasingly concentrated in a small group of 20 or so emerging markets which received over 90 per cent of total inflows of capital in the 1990s, compared to some 50 per cent before the outbreak of the debt crisis. Thus, the share of low-income countries in net private capital flows to developing countries has been steadily declining.
since the second half of the 1980s, from 20 per cent (around $6 billion) to just 6 per cent in the second half of the 1990s (around $14 billion), with an attendant rise in the share of middle-income emerging markets. In recent years more than three-quarters of bond issues have been accounted for by less than ten borrowers in Latin America and Asia and much of the syndicated bank lending has gone to half a dozen countries in Asia. These same countries were also the main recipients of international equity investment.

Despite the ever louder beating of the financial liberalization drum during the 1990s, including efforts to put capital account openness in the IMF’s Articles of Agreement, these flows proved increasingly difficult to manage in a way consistent with faster and more inclusive economic growth. For countries seeking re-entry into international financial markets after the debt crisis, higher real interest rates and a stable exchange rate were prerequisites for attracting new inflows. While the resulting financial stringency could prove attractive to foreign investors, a tight monetary and fiscal stance, oftentimes with an appreciating currency, did little to stimulate domestic investment or to improve export prospects. Indeed, increased debt servicing obligations resulting from higher interest rates, along with import surges, ran the risk of reproducing an unsustainable debt burden. In many cases, a combination of capital outflows, profit remittances and the accumulation of exchange reserves greatly reduced net inflows, and of these a growing proportion was absorbed by activities which added little to productive capacity (UNDESA, 2005). Particularly in the form of short-term loans and portfolio equity, these inflows, often highly leveraged through derivative contracts and hedge funds, could be very unstable and an unreliable source of development finance.

As the workings of financial markets became increasingly disconnected from the longer-term demands of productive investment and industrialization, unregulated financial flows triggered boom–bust cycles, which became a recurrent feature of the developing world during the 1990s. The precise circumstances in which the vulnerability to the reversal of capital inflows arose, and the subsequent impact on growth, varied from region to region. An early warning was given by the Mexican peso crisis of 1994. However, the full force of unregulated financial flows was revealed by the financial crises in East Asia, a region with a long-standing record of strong growth and fiscal discipline. As in other episodes of financial crisis and currency turmoil, the crisis in East Asia was preceded by financial liberalization and deregulation which, in some cases, constituted a major break with past practice. Moreover, the extremes of collapse were amplified by unnecessarily tight monetary policies which deepened the debt deflation process, served to depress output and employment, and caused serious dislocations in the corporate and financial sectors (Stiglitz, 2002).
Global slowdown and instability

The debt crisis and its aftermath generated a lost decade for many developing countries in the 1980s, with incomes contracting in many cases. On some assessments this has given way to a much more benign macroeconomic climate with lower real interest rates, less-volatile growth in the leading markets and diminished inflationary threats. Despite this, the slowing global trend persisted for much of the 1990s. Indeed, even as recovery in the USA turned into more sustained growth from the mid-1990s and growth in China and India accelerated, the average global growth rate for the decade was still not above that of the 1980s, and remained below that of the 1960s and 1970s.

This slowing trend has been associated with a good deal less stability in the growth performance of many developing countries. Some of this (as in the case of transition economies) can be traced to political shocks. But, if history is any guide, it should not come as a surprise that a relaxation of regulations put in place in response to earlier excesses generates ‘manias, panics and crashes’ (Kindelberger, 1984), or that these would hit weaker economies the hardest. On one estimate, during the 1990s the financial system was in crisis for 40 out of 120 months (Plender, 2003, p. 57). According to Barry Eichengreen (2002), the period since the collapse of Bretton Woods has seen a sharp increase in the incidence of financial crises, principally in the form of currency crises but also in conjunction with banking crises. Most of these have been in the developing world.

The global slowdown and increased incidence of financial crises has coincided with a declining frequency of strong growth episodes and a rising frequency of negative growth episodes. According to Ocampo and Parra (2005), in the 1960s and 1970s about 40 per cent of developing countries had successful growth episodes (with annual average rates of per capita GDP growth greater than 3 per cent over at least a five-year period) but this proportion fell to less than 20 per cent through most of the past quarter-century, while in 40 per cent of countries there were negative growth episodes compared with just 15 per cent in the earlier period. Similarly, Hausmann et al. (2004) searched for episodes of rapid acceleration in economic growth that were sustained for at least eight years and found that while there were 23 and 30 such episodes in the 1960s and 1970s respectively, there were only 14 in each decade of the 1980s and 1990s.

Fragmented development

While the influence of radical market-based reforms on policy-making in advanced countries was heavily qualified by checks and balances in their political systems (Krugman, 2007), these, thanks in no small part to the support of the international financial institutions, quickly became a fully
fledged development blueprint showing how poor countries could extricate themselves from the problems of the debt crisis and establish a new growth path. It promised to remove structural and institutional impediments to growth, improve productive capacity and trade performance, and put an end to stop-go development associated with excessive indebtedness and periodic payments crises. Above all it promised strong convergence in the global economy as growth in poorer countries outpaced that in richer countries and income gaps across the world economy narrowed sharply. The previous section has raised some initial doubts about blindly trusting in international market forces and firms to achieve this outcome. This section looks in more detail at how the interplay of external and internal integration forces has played out across the developing world.

**Growth, gaps and international inequality**

Contrary to much contemporary rhetoric, the growth in cross-border trade and financial flows since the mid-1980s has failed to stop the growing gap in real per capita incomes between rich and poor countries. Economic divergence is, here, the simple product of the higher starting income of richer countries and their faster average annual per capita growth rates (Milanovic, 2005). This has happened despite the sharp slowdown in the advanced countries since the mid-1970s, a slowdown which, as noted earlier, has resulted in a general slowing of the world economy over the same period.

Such a high level of aggregation can, however, be misleading and there are important differences across regions within the developing world. Asia has persistently maintained a much stronger growth performance than other developing regions, enjoying catch-up growth with the advanced countries since the early 1990s, and for a good deal longer in East Asia. By contrast Africa, particularly south of the Sahara, and Latin America have continued to fall further behind. Given that, among developing countries, Latin America was the richest developing region by some margin at the end of the 1960s, subsequent growth trends have generated convergence across the developing world. There are also some noteworthy differences in growth performance over time, with developing countries outpacing advanced countries in the 1970s and again over the decade since the mid-1990s, with a lost decade sandwiched between in the 1980s when developing-country growth performance was below its own historical average and that of the more advanced economies.

In both popular and scholarly accounts, the actual and projected tilting of the world’s economic axis towards Asia is often taken as synonymous with globalization. In fact, that shift has been ongoing since the 1950s, when Japan entered a period of rapid and sustained growth, joined soon
after by the small Asian tigers, notably Korea and Taiwan, China. A second tier (of more variegated economies) from South-East Asia joined in during the 1980s. However, it has been the emergence of China and India, particularly the former, that has added a new dimension to the Asian catch-up story. While both are still a long way from being middle-income countries, they have a large number of citizens who would qualify as middle class. Their emergence has already had a significant impact on global growth dynamics as well as on the wider regional performance, with growth in South Asia, since 1980, improving sharply over the preceding two decades, and high growth rates maintained in East Asia in the 1990s despite the onset of a severe financial crisis in 1997, though the pace did not match that of the 1970s. There is also little doubt that the impact is being felt beyond the region, though just how the balance between their growing demand for natural resources and their addition to the world’s unskilled labour force will play out elsewhere is a subject of much dispute.

After the lost decade of the 1980s, Latin American economies enjoyed a brief renaissance in the early 1990s when the intensification of structural reforms enabled them to return to the international capital markets; but after 1997 developments again turned sour and produced another ‘lost half-decade’ (Ocampo, 2002). Growth performance in the 1990s was similar to that in the 1960s but well below that in the 1970s. Sub-Saharan Africa, like Latin America, also suffered a ‘lost decade’ of development in the 1980s, but with a weaker (if less erratic) recovery in the 1990s; there was generally little or no inflow of private capital into the region, and no abrupt reversal at the end of the decade.

Against the backdrop of a global slowdown, tight macroeconomic policies and the increasing frequency of financial crises, many developing countries have seen a weakening of growth relative to their own past performance. Of a total of 124 developing countries, growth in 95 of them (that is, over three-quarters) was faster in the period 1960–78 than between 1978 and 1998 (Milanovic, 2002), and only a handful of countries have been able to hit the kind of growth targets needed to address their economic and social deficits: between 1980 and 2000, of 140 developing countries only 20 grew at annual rates above 5 per cent, a number that rises to 30 for the period 1990–2000, but if a rate of 7 per cent is taken as the benchmark, the number of success stories falls to just five and six respectively.

Countries at the very bottom of the income scale appear to have lost most ground. This has lead some to talk of a ‘twin peaks’ global income distribution (Quah, 1996), with a ‘hollowing-out’ of the middle-income range of countries (UNDESA, 2006b, p. 8). Milanovic and Yitzhaki (2001), for example, estimate that just 8 per cent of the world’s population
fall into that category. Polarization can, in large part, be explained by the fact that the general slowing of global growth since the 1980s has hit poorer countries particularly hard. According to Milanovic (2005, p. 5), the average annual per capita growth rate of the group of LDCs was just 0.1 per cent between 1980 and 2002, compared to 1.9 per cent in the ‘old’ Organisation for Economic Co-operation and Development (OECD) economies, and while the former figure hides a very wide dispersion of performance, there were no stellar growth performance of the kind found in East Asia. Growth collapses were, not surprisingly, much more frequent among this group of countries than other developing-country groupings.

Given the close (though not direct) connection between income growth and poverty reduction, there have been some clear differences in poverty trends across regions over the past two decades. Overall the number of people living on less than $1 a day has dropped since the early 1980s by around 400 million. However, this headline figure hides very large falls in some countries, notably China, which pulls down the regional figure for Asia, along with little change or increases in almost all other regions. Indeed, the improvement in the world poverty figure can be attributed almost exclusively to China.

Just how income and poverty trends translate into a picture of global inequality is an ongoing source of controversy among economists. The Gini coefficient is often used as a more integral measure that aims to take all members of a chosen set into account, although measurement problems abound. On this measure, a number of studies have reported a reduction in international inequality since 1980, linking this in turn to a more open world economy. However, these findings are open to serious questioning on both methodological and empirical grounds. Indeed, as Sutcliffe (2006) has noted in his balanced assessment of the debate, there is little agreement on how significant the change has actually been or just when the reversal might have begun. More importantly still, the result hinges on the performance of a single outlier; removing China from the country set reverses the trend to one of rising international inequality, and even more sharply than before 1980. The Theil decomposition of international inequality shows a similar result (UNDESA, 2006, p. 14). Given its size, China’s performance is obviously central to the analysis of global trends. However, from a comparative perspective what happens in a single country (however large) should not be used to obscure what is in fact a highly variegated picture across the global economy (Berry and Serieux, 2004).

The variety of development experiences in the 1990s
A recent assessment by the World Bank (2005, p. 30) of its own policy advice during the 1990s has acknowledged that it persistently overestimated
growth prospects in regions implementing adjustment programmes, and underestimated growth performance in those that were not. This is a belated acknowledgment of a good deal of academic research which finds little evidence that adjustment programmes have had a positive impact on growth, employment or poverty. However, there is a reluctance to accept that the adjustment path that these policies helped fashion has actually set back development prospects, in many cases by distorting the process of internal integration, focusing instead on what might have been left out of the package in terms of social policies, good governance, institutional detail, and so on.

In fact, regions that pursued extensive adjustment programmes in the 1980s found it particularly difficult to reverse the sharp drop in the share of investment triggered by the debt crisis. In some countries, this dipped to below the levels needed to replace depreciated capital, and where adjustment programmes persisted, the investment cycle remained volatile, even after the immediate disturbances of the crisis subsided. The downsizing of public investment has been a prominent part of this story, and in many cases this has crowded out domestic private investment. But, as important, has been the approach to macroeconomic fundamentals guiding policy design, which has ignored the mutually reinforcing links between aggregate demand, capacity utilization and investment. Tight monetary conditions and accompanying currency gyrations have further discouraged growth based on capital accumulation, adding to the difficulties encountered by domestic industry in introducing the kind of restructuring that would help raise productivity, even as it faced stiff competition from lower trade barriers. In part as a consequence of these difficulties in the productive sectors of the economy, and in part as a result of ongoing liberalization of the financial sector, ‘rentier’ investments (including in government bonds) and ‘capital flight’ have become much more attractive options. In many cases this financialization of the domestic investment regime has coincided with and reinforced commodity dependence, often linked to increased FDI inflows (Magalhães Prates and Paulani, 2007).

Under these conditions, the cumulative nature of industrial development can very easily go into reverse gear. Certainly, where industrial stagnation was the norm in the 1980s, as in Latin America and Africa, most countries found it difficult to reverse the trend in the 1990s, with premature deindustrialization a visible trend in some cases (UNCTAD, 2003, pp. 92–9). Even where this has not been a dominant trend, there is little sign of the diversification and upgrading which describes a healthy internal integration dynamic. This is often associated with enclaves of specialized development. In the case of many Latin American economies these enclaves have been in the primary sector, though in some cases manufacturing activities
have been able to carve out niche markets. Where productive capacities are even more rudimentary, as in much of sub-Saharan Africa (SSA), the likelihood of enclave development in the primary sector generating a natural resource curse seems to be even greater.

Moreover, in economies with declining shares of investment and manufacturing value added, a stagnant or falling share of manufactures in total exports was often the norm, even as the overall composition of developing-country exports was shifting rapidly towards manufactures, including more skill- and technology-intensive goods. Labour market performance, key to tackling poverty, has also been caught up in these divergent developments. The impact of liberalization on labour market performance is a complex matter, with the effects differing between countries depending on the domestic and international conditions prevailing when liberalization occurred, as well as on the broader development strategy being pursued by policy-makers. However, a good deal of evidence from Latin America and sub-Saharan Africa, where neoliberal policies have been pursued most intensively, suggests that increases in unemployment and/or increasing wage inequality have often accompanied rapid trade liberalization.24 In many cases, adjustment policies and the downsizing of the public sector have led to a hollowing-out of the middle class. The counterpart of these trends has been an expanding informal economy, which by the end of the 1990s accounted for anywhere between one-third and three-fifths of the labour force in Africa and Latin America (Schneider, 2002).

By contrast, the countries in East and South Asia that bucked the declining investment trend in the 1980s were able to maintain a reasonably stable or even rising pace of capital accumulation for most of the 1990s, with the share of public investment holding up or rising further (UNCTAD, 2003, pp. 65–73). But just as importantly, strong investment has supported structural change. In countries where investment and industrial output expanded in the 1980s, this continued in the 1990s. In these cases rising and fully utilized capacity levels have had a knock-on effect on productivity growth through both the demand and supply sides. Indeed, Asian development since the 1970s confirms the key role of industrialization in establishing a rapid growth path, with positive cumulative effects through strong export drives linked to rising levels of productivity (UNDESA, 2006). During the 1990s, just eight East Asian countries accounted for 70 per cent of developing countries’ trade in manufactures. Outside this group, export strategies relied on low (and in some cases falling) wages or currency depreciation rather than strong productivity growth, and while this stimulated recoveries in some countries, few were able to reach a threshold level of exports consistent with a vibrant industrialization path (UNCTAD, 2003, pp. 99–102).
Taking stock of trends since the debt crisis of the early 1980s, it is clear that by itself the degree of exposure to global market forces is not what distinguishes ‘winners’ and ‘losers’. Rather, the main difference, particularly between the East Asian NIEs and most other developing countries, was that liberalization followed the successful implementation of industrial and trade policies; protection and support were removed in large part because they were no longer needed. In the latter, on the contrary, liberalization has largely been triggered by the failure to establish efficient, competitive industries in labour- and/or skill-intensive sectors. From this perspective, the different ways in which internal and external integration have taken shape in developing countries have failed to trigger convergence since the early 1980s, but have instead created a very uneven economic landscape of diverse experiences.

*Mature industrializers* This group includes the first-tier NIEs, notably the Republic of Korea and Taiwan Province of China, which achieved industrial maturity through rapid and sustained accumulation of capital, and growth in industrial employment, productivity and output, as well as manufactured exports. In the 1990s these economies enjoyed a share of industrial output in GDP above the levels of advanced countries, exports had shifted to more capital- and technology-intensive goods, and industrial growth was starting to slow down as resources shifted towards the service sector.

*Rapid industrializers* A number of countries saw a rising share of manufactures in total output, employment and exports, based on strong investment in resource-based and labour-intensive activities, and were beginning to upgrade to middle-range technology products. This group included the second-tier Asian NIEs, but also isolated success stories from other regions, as well as the waking giants of China and (albeit more cautiously) India.

*Enclave industrializers* Some countries moved away from dependence on commodity exports by linking to international production chains, often by attracting large amounts of FDI and with heavy reliance on imported inputs and machinery. Export growth was often very fast, as in the Philippines, Mexico and, to some extent, Morocco. However, overall performance in terms of investment, value added and productivity growth was often quite weak.

*Premature deindustrializers* This group included most countries in Latin America, which had achieved a certain degree of industrialization but were unable to sustain a dynamic process of structural change through rapid
accumulation and growth. In a context of rapid liberalization, declining shares of manufacturing employment and output and a downgrading to less technology-intensive activities were common trends.

*Commodity-dependent exporters*  Many poorer economies, particularly in sub-Saharan Africa, remained heavily dependent on one or two commodity exports. In the face of relatively stagnant markets, volatile prices and declining terms of trade, investment dropped further, diversification stalled and productivity remained stagnant. In some cases enclaves of faster export growth emerged in the extractive sectors, usually tied to FDI, but with weak linkages to the rest of the economy. However, some wealthier developing countries, notably Chile, did achieve a faster pace of investment and growth based on their natural resource endowments.

**Conclusion**

The historical experience of advanced countries, including recent graduates in East Asia, establishes that a broad and robust domestic industrial base remains key to successful development, because of its potential for strong productivity and income growth. Success hinges on building such a base from local conditions and overcoming local constraints on its expansion and evolution. This is still the big policy challenge for most developing countries. The experience of the 1980s and 1990s suggests that the policy direction launched in many parts of the developing world after the debt crisis failed to establish flourishing alternatives. Despite the widely shared belief that a more open economic environment would demonstrate the benefits of unrestricted capital mobility and the superiority of markets over government intervention, the period since the collapse of the Bretton Woods system has instead been marked by very heavily concentrated flows to comparatively wealthy countries and by an increasing incidence of financial crises, and their growing virulence in terms of lost output and jobs. Recent efforts to salvage the strategy have emphasized omissions: not enough good governance, not enough market access, not enough FDI or aid. Doing so continues to overlook the damage to macroeconomic growth fundamentals and to the direction of structural change that have accompanied the turn to neoliberal policy reform.

By contrast, success stories in the 1990s built steadily on improving performance established in the 1980s or earlier. In most cases they followed what Birdsall et al. (2005) have called ‘heterodox gradualism’, using an array of policy options to manage integration into the global economy and ensure that more of the value added linked to trade stayed at home. Doing so involved experimenting with a range of more strategic measures to encourage strong capital formation, expand domestic markets and support
technological upgrading. These measures have been rooted in specific institutional settings reflecting national political and social cultures and consistent with the bureaucratic and entrepreneurial capacities of local elites. In their different ways, all have eschewed a softening of the state and instead premised their structural transformation on a harder development state, that exhibits both ‘adaptive efficiency’ and the effective creation and utilization of ‘policy space’ (Kozul-Wright and Rayment, 2007). These same features will need to be recovered in countries that have fallen back since the debt crisis, if catch-up growth is to be reignited. This will require active policies, particularly on such matters as industrial support, technological progress and public infrastructure, all of which will have to be tailored to the particular circumstances of the countries concerned.

Notes

1. In an article deploring the slow progress in the Doha round of trade negotiations, the Financial Times journalist Martin Wolf described French President Chirac’s scepticism as to the virtues of unlimited free trade as ‘foolish, even depraved’ since if the negotiations ‘should fail, disorder alone should triumph’, Martin Wolf, ‘The World has Everything to Lose if Trade Liberalisation Fails’, Financial Times, 2 November 2005.


3. The Nobel Prize-winning economist Douglas North (1994) has candidly admitted that the aversion of neoclassical economics to historical processes and structural discontinuities precludes a proper understanding of the development process.

4. These regularities are associated with the empirical work of Verdoorn, Lewis and Kaldor. For a review see Toner (1999). See also Rodrik (2006) for a recent assessment of why manufacturing still matters to the development process.


6. This conclusion applies, we believe, to much of the recent discussion of the importance of the ‘investment climate’ in developing countries.

7. Dowrick and Golley call for further research to understand these findings. Their own tentative suggestions are that the nature of technology transfer through MNCs has changed in the latter period and that the range of complementary policies that supported successful liberalization in the earlier period have gone missing in the ‘one policy fits all’ approach of the latter period. Both suggestions are in line with the arguments presented in this and subsequent chapters.

8. Certainly, the period of rapid export expansion in the advanced countries in the three decades after World War II coincided with strong wage growth and the development of deeper domestic consumer markets (Armstrong et al., 1984). However, intra-industry trade, strong productivity growth and high rates of capital formation were not the automatic outcome of market forces in these countries but were closely interrelated components of a politically fashioned socio-economic compromise.


10. Such networks are not a new development, dating back to the 1960s in parts of East Asia, and becoming a more prominent feature of the international division of labour in the 1970s, see Helleiner (1973) and Henderson (1991).

12. The annual capital inflow in the 1990s was around 5 per cent of GNP, which was roughly the level prevailing in 1975–82. If China is excluded, the ratio is actually lower than in the earlier period by one percentage point.

13. Although the call for financial liberalization was heeded across the developing world in the 1990s, the majority of countries, and particularly those in Africa, attracted little private flows, and certainly not enough to offset declining aid flows during the 1990s. FDI was also very unevenly distributed: three-quarters of the total in the 1990s went to just ten emerging market economies, and China, Brazil and Mexico together took nearly one-half. Indeed, China accounted for about one-third of all FDI in the developing countries in the 1990s and about two-thirds of the total went to East Asia. Other developing regions – sub-Saharan Africa, South Asia, North Africa and the Middle East – have only marginal shares of the total and in the case of sub-Saharan Africa a falling one, mainly in extractive sectors or in public utilities.


15. The city states of Hong Kong and Singapore are fascinating stories in their own right. However, it should be noted that their per capita incomes in the early 1950s were comparable to Japan, and higher than much of the European periphery. Recognizing this goes a long way in dissociating their transformation from the current phase of globalization.

16. China’s break with its earlier, highly erratic growth path occurred in the mid-1970s and has been building strength since the early 1980s. There are some significant differences between these two awakening giants, notably the much slower pace of urbanization in India, and with this a much slower pace of industrialization and capital accumulation. There is a good deal of dispute among economists as to which of the two development paths is likely to be the most sustainable.


18. The recent pick-up in African growth rates is largely a result of an increase in commodity prices linked in no small part to growing demand in China. According to a recent IMF Working Paper, a dozen middle-income and oil-producing African countries have been the principal beneficiaries of a stronger growth performance since 1997, see Tahari et al. (2004).

19. Obviously in light of what has already been said it would be wrong to identify any hard and fast benchmark for all developing countries. Still, many regard a 6–8 per cent target as what is needed to tackle those deficits and visibly to close income gaps with the more developed countries.

20. Identifying potential growth rates, and the related challenge of suggesting growth targets, must be seen as a somewhat hit-and-miss business. But assuming that a country’s labour force is growing on average at 2–3 per cent per year and that productivity growth needs to match that rate to maintain internal and external balance, then 5–6 per cent would be a minimum growth rate for GDP. Many, including ourselves, would regard a 7–8 per cent target as more appropriate if the policy objective is to make real progress in tackling the massive social deficits that have built up in most developing countries, and to start closing the income gaps with the more developed countries.

21. The Gini coefficient derives from the Lorenz curve which shows the cumulative share of the income received by the cumulative shares of the population, starting from the poorest income-receiving units. The coefficient measures the area between this curve and the diagonal of perfect equality and varies from 0 (maximum equality) to 1 (maximum inequality) or from 0 to 100 when expressed in percentages. Milanovic’s review of the literature notes that there is little disagreement on the degree of international inequality, with most studies finding a Gini coefficient in the range of 63 to 68, a figure that is higher than the coefficient in highly unequal countries such as Brazil or South Africa (Milanovic, 2006, pp. 140–42).

23. See further UNCTAD (2003, pp. 74–6), also Ramirez and Namzi (2003).

Bibliography
Amsden, Alice (2001), The Rise of the ‘Rest’: Challenges to the West from Late Industrializing Economies, New York: Oxford University Press.
Birdsall, Nancy, Dani Rodrik and Arvind Subramanian (2005), ‘How to Help Poor Countries’, Foreign Affairs, July/August.


ILO (2004), A Fair Globalization: Creating Opportunities for All, Geneva: ILO.


Latin America and the Caribbean

Albert Fishlow

Introduction

Economic development in Latin America, finally, seems on the verge of beginning again. After close to 25 years of poor performance, despite dramatic policy changes in fiscal policy, and hence inflation, and expansion in international trade, the region seems to be expanding. Results both in 2004 and 2005 have been better than initially foreseen. The key question is whether this process will continue over the next decades, or whether the current advance will soon collapse – as it has done in the past – with declining commodity prices, limited foreign demand for exports, less foreign investment and more inflationary domestic policy.

Much has altered within the region’s economies since the 1990s. Domestic politics have also changed profoundly, as democracy has continuously spread and deepened. Yet a high degree of political dissatisfaction has evolved, and successive elections have accurately reflected the population’s unhappiness. Although inflation rates are now virtually at a record low, benefiting substantially the bottom third of the income distribution, people do not vote for those that promise fiscal discipline. The Left has run well, and won, virtually everywhere since the beginning of the 21st century. But, with the exception of Hugo Chavez, President of Venezuela, and possibly Nestor Kirchner, President of Argentina, policies have remained responsible and even responsive to the international markets. Evo Morales’s election in Bolivia introduces a new element, control over the trade in drugs, while recapitulating once more the question of state control over energy resources that has grown in significance in recent years with the rising price of petroleum in world markets.

These contemporary events, both economic as well as political, are better understood within a historical context. Latin America was a latecomer to continuous economic expansion. Only since the last decades of the nineteenth century did economic growth become a regular process. That was the era when the region began its increasing participation in external markets, taking advantage of its natural resources, and simultaneously sought to implement, late, its own industrial revolution. That is where we start the story in the next section. It extends through the Great Depression, with its dramatic decline in external demand, up to the recovery of the world economy at the end of World War II.
Next, we analyze the rise and decline of import substitution in the three following decades until the debt crisis that began in 1982. That history simultaneously comprehends a surge of military dictatorships within virtually the entire region. At the same time, that experience encompasses a period of above-average growth in Latin America. Finally, this interval also represents growing dependence on external finance to finance the increased domestic investment that drove expansion. This combination proved too fragile to survive the successive oil shocks of 1973 and 1979.

In the next section we take up the economic performance of the region during a subsequent period of profound retrenchment, altering the previous large role of the state almost everywhere. Fiscal deficits were sharply curtailed, much privatization of state-owned activities occurred, and protection against international imports was sharply curtailed. Latin America in the 1990s gave signs of joining Asia and Eastern Europe in rapid expansion. But economic development soon proceeded much less rapidly than had been hoped, giving rise to increasing criticism of the Washington Consensus and sparking a clear move to the political Left that continues today.

A final part takes up the central problems and policy issues that confront virtually all the countries of the region today, in the midst of a record number of presidential elections. Politics and economics are intrinsically linked. That combination sometimes leads to an inability to focus effectively upon the longer run. But Latin America needs to continue to change if it is to keep up with global advance. It requires more and better educational and health outlays, both to improve the quality of the labor force as well as to improve the lamentable degree of inequality that plagues almost all countries. Macroeconomic reform has to continue, both in government tax collection as well as more efficient expenditure. And that subject necessarily includes the indispensable increase in domestic savings and investment. A third matter is the continuing increase, as well as diversification, of export earnings and thus reduced sensitivity to variation in external demand. Finally, the region will have to recognize the continuity of present economic policy as a virtue rather than a vice, even as the electoral process becomes more inclusive and civil society flourishes. Institutional change, making permanent these structural advances, requires continuing attention and additional major effort.

A brief conclusion extends the lessons of the Latin American and Caribbean experience to the broader agenda of contemporary development economics.

**Boom, bust and war recovery: 1900 to 1945**

Latin America, despite the achievement of political independence early in the nineteenth century, missed out on the initial Industrial Revolution. As
a consequence, overall, its standard of living probably improved little between 1820 and 1870. Those were years of political instability and modest engagement in international trade. Only during the period 1870 to 1913, when an initial era of globalization fueled economic activity, did the region expand its relative export position, helped by an increasing number of immigrants and much greater foreign investment. That investment underwrote great expansion of the rail infrastructure and construction of port facilities to facilitate export of wheat and meat, coffee, sugar, rubber, nitrates and other primary products demanded by the United States and Europe.

Such were the mounting receipts from exports after 1900 that not only Argentina returned to the gold standard, but Brazil and Chile also joined, then still a signal of successful developing-country adaptation within the expanding world economy. The region impressively focused on exporting primary products – something of the order of 97 percent of exports were classified as such in 1913 – even while high levels of tariff protection assured a growing market for domestic industrial production. Import substitution generally began in the larger countries before the Great Depression of the 1930s, focusing on the elemental, but substantial, areas of foodstuffs and textile and clothing production.

What is important to note is the great difference in incomes among countries of the region just before World War I. Data are presented in Table 65.1. Per capita income in 1913 in Argentina was comparable with that of Western Europe as a whole, and some three-quarters of that in the United States. Chile followed, about a third below Argentina’s pace-setting level. The rest of the region trailed much farther behind. At the very bottom was Brazil, burdened by the poor north-east that offset the rising, and more diversified, economy to the south.

These regional income differences reflected a variable commitment to exports, as can easily be seen in the same table. Argentina and Chile were the leaders in their share of income generated by exports. They were, as well, producers of non-tropical exports more in competition with developed countries than the products from other parts of the region, and hence, as W.A. Lewis has shown, able to gain higher incomes. Both of these countries also benefited from high levels of capital inflow from abroad, contributing to higher levels of domestic investment. Finally, the labor force in the case of both countries was more literate than elsewhere, and the movement toward local non-agricultural activities was equally more substantial. Bulmer-Thomas runs a regression for 13 countries, relating economic growth to exports over this interval, and finds a high degree of explanation; all seven of the countries separated correspond well.
The World War in 1914 had variable effects within the region. No longer is it viewed as an unmixed blessing. The sudden reduction of imports that the conflict imposed did offer opportunities for domestic producers in several countries in the industrial sector. Yet capacity for production of capital goods was still rudimentary, and that constraint limited possibilities for significant output gains. Foreign investment from Europe ceased, and that of the United States was still directed primarily to the north of the region. Argentina, in particular, disappointed. With lower imports almost everywhere, governmental receipts flagged, and inflation began to expand in many countries. What helped was the relative brevity of the conflict, allowing many of the countries to go back to what they had been doing before: exporting their primary products.

Thus the 1920s featured a return to previous history. Imports of manufactured products went up, as a consequence of rising primary exports. Even Chile again managed to expand its sales of nitrates by 1929 to three-quarters its 1913 real level. National performance, as before, was related closely to success in international trade. There was a new entrant with much larger income upon the stage, Venezuela, whose petroleum exports found an expanding market within the United States. Colombia and Peru also improved their positions.

---

Table 65.1 Latin America before World War I

<table>
<thead>
<tr>
<th>Country</th>
<th>Per capita income 1913</th>
<th>Growth per capita 1900–1913</th>
<th>Exports as % of income</th>
<th>Literacy %</th>
<th>% Labor force agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>3797 $ Geary–Khamis</td>
<td>38</td>
<td>0.36</td>
<td>63</td>
<td>34.2</td>
</tr>
<tr>
<td>Brazil</td>
<td>839</td>
<td>19</td>
<td>0.3</td>
<td>35</td>
<td>66.7</td>
</tr>
<tr>
<td>Chile</td>
<td>2653</td>
<td>36</td>
<td>0.54</td>
<td>56</td>
<td>37.7</td>
</tr>
<tr>
<td>Colombia</td>
<td>1236</td>
<td>27</td>
<td>0.13</td>
<td>41</td>
<td>70.5</td>
</tr>
<tr>
<td>Mexico</td>
<td>1467</td>
<td>27</td>
<td>0.17</td>
<td></td>
<td>63.7</td>
</tr>
<tr>
<td>Peru</td>
<td>1037</td>
<td>27</td>
<td>0.24</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>1104</td>
<td>34</td>
<td></td>
<td>30</td>
<td>72</td>
</tr>
<tr>
<td>Seven countries</td>
<td>1134</td>
<td>39</td>
<td></td>
<td></td>
<td>0.28</td>
</tr>
</tbody>
</table>

Exports as % of income, literacy and % labor force: *An Economic History of Twentieth Century Latin America*, ed. Rosemary Thorp, Enrique Cardenas and Jose Antonio Ocampo 2000, Vol 1, p 26
But there were also some important differences. Two stand out. First was the reality of major external change. Globalization, and British leadership, was a phenomenon of the past. There was increasing financial dominance of the United States in the world, and also at the regional level. Much of the renewed surge in investment flows went to local and state sources instead of, as earlier, to national governments. As it would soon turn out, the new American investment banks competing for business were more effective in finding a supply of finance than identifying productive applications of the money. The weakened position of Europe had consequences not only for investment, but also for international trade. The gold standard had vanished, leaving individual countries confronting hyperinflation and internal political disruption. The League of Nations, an attempt at supranational economic guidance, provided little assistance in the midst of these new circumstances.

Second, the renewed rise in export markets in the 1920s was weaker than it had earlier been. Countries varied in their ability to cope. Brazil, for example, despite losing out in exports of rubber and facing a slowing market for coffee, managed to continue a growth predicated upon diversification of its industrial structure. Others, like Argentina and Chile, saw lesser rises in expansion in the period from 1913 to 1929 than they had previously. This modest deviation from the historic export-oriented model of growth was, of course, soon to become the beginning of a new commitment to import substitution after the start of the Great Depression.

Much of Latin America definitively changed its development style in the 1930s. There was little option. Prices of the region’s raw material exports plunged with declining demand. Quantities also fell. This time, unlike the earlier decline after World War I, recovery did not occur within the industrial center. As a result, it was necessary to cease full payment of interest on the public debt, which had much risen as a proportion of export earnings. Exchange rate controls were applied to limit imports of consumption goods capable of being supplied internally. Real exchange rate devaluations occurred almost everywhere. They most frequently took the form of multiple exchange rates enhanced by direct controls.

The deterioration of the balance of payments meant a simultaneous decline in government revenues: some countries were still dependent upon import duties and export taxes for as much as half of their total receipts. As a natural consequence government deficits showed a tendency to rise, financed through increases in the money supply, and thereby stimulating domestic demand. Inflation did not increase much as domestic supply instead responded well to the lack of import competition. Special circumstances, such as the potential Brazilian civil war with a seceding Sao Paulo...
in 1932, or the Leticia War between Colombia and Peru in 1933–34, occasion ally helped in establishing such proto-Keynesian policies and contributed to earlier and stronger recovery.

With a rise in export earnings after 1932 for most countries, rising imports of intermediate and capital goods became available to satisfy increasing demand for such inputs as could not be supplied locally. These sometimes consisted of used machinery rather than that newly produced. Labor migrated to the cities to meet the demands. For this period, unlike the later decade of import substitution in the 1950s, production increases were dependent upon increased labor, rather than capital, input. In some instances, extra shifts were employed to satisfy an expanding market. Noteworthy is the greater rise in industrial output than in gross domestic product that occurred virtually everywhere in the region.6

Such expansion was not without a negative side. Military leaders emerged in many countries in response to the new economic challenges being confronted. Constitutions were rewritten, or reinterpreted. Authoritarianism had a preview before its re-enactment in the 1960s and 1970s. There was an increasing degree of state intervention and regulation, not always efficiently. Markets were not again to operate in an unfettered fashion as they had somewhat done during the years before 1930.

With the approach of a new war in the late 1930s, economic conditions began to worsen slowly. Current dollar trade growth reversed after 1937, giving further stimulus to domestic production where it was possible. When the European war emerged in 1939, Latin America, like the United States, stayed out. With Pearl Harbor in December 1941, and United States engagement, several of the countries declared war. But only Brazil was actively engaged in the conflict, with a detail of some 25,000 troops. Indeed, the southern cone countries, with a long and strong German influence upon their military, were late in their formal espousal of the Allied side. That indecision did not affect their trade flows much. Between 1940 and 1945 there was a major reversal: all countries suddenly found their exports dominantly flowing northward to the United States, with portions going to other neighboring countries within the region.

The war saw slower expansion in Latin America as a result of much lesser imports. Reserves accumulated, substantially in some countries. Inflation too accelerated domestically, leading to real appreciation of unchanged nominal exchange rates. When peace returned, repressed demand and favorable prices led to substantial import flows. There was now a domestic sector in many countries resistant to such foreign competition and a return to the experience of the 1920s. Not surprisingly, within the region a new group of young economists emerged calling for a conscious strategy of import substitution. They soon took form around the Economic
Commission for Latin America (CEPAL) in Santiago under the leadership of Raul Prebisch. Their views were given powerful reinforcement by the failure of the Latin American countries to obtain their own Marshall Plan, as had the European countries, thereby assuring integration into the wider global economy.

Table 65.2 sets out the growth experience that the principal regional countries had achieved since 1913. What these data reveal are three characteristics. First is the extraordinary rise in income per capita of Venezuela: it moves from a third of the Argentine level in 1913 to one that exceeds it by half as much again. Note as well that the second-largest rate of advance is attained by Peru. That provides the basis for those countries’ choice not to move to import substitution. Second is the significant slowing in the performance of Argentina and Chile, whose leadership had been so predominant earlier on. Third is the relative rise of Brazil, Colombia and Mexico as a result of the move toward domestic industrial production.

But, ultimately, and alas, there is the unfortunate regional relative aggregate. It moves from 29 percent of the per capita level of the United States in 1913 to 27 percent in 1950. Even with the better performance of Latin America during the 1930s, the rapid recovery of the United States during the war years and immediately thereafter gave proof to the continuing circumstance of regional backwardness.

Post-war growth
Latin America, or at least a goodly number of countries in the region, launched itself toward a continuation of industrialization, impelled by a
more active state policy in the 1950s. Helped by the rise in primary product prices in the wake of the Korean War, and thus sufficient export receipts, countries – both large and small – imposed high tariffs, quotas and even multiple exchange rates, to permit giving substantial incentives, and subsidies, to the intermediate and capital goods sectors where imports retained a large percentage of the domestic market. As the world gradually moved away from the limitations to world commerce imposed during the Depression, Latin American countries largely ignored the opportunities afforded for greater diversification as well as market expansion of their traditional export products.

In this attempt, most countries badly failed, including the two regional leaders, Argentina and Chile. Three countries, Mexico, Brazil and Colombia, managed the process better during the 1950s. They had started farther behind, with large internal markets, and managed to focus on less-inefficient undertakings. They also benefited from lesser balance-of-payments shocks during the expansion of the 1950s, assisted by more foreign investment. Two countries, Venezuela and Peru, continued their past emphasis on petroleum and mineral exports, respectively, and emerged well.

What characterized the CEPAL model were three marked deviations from standard economic theory. First, the state role in the process of economic performance was much enhanced. Priorities were a matter of public policy: planning groups were established, production was directly encouraged, special tariff protection was granted, subsidies appeared, and so on. Matters during the 1950s in Latin America were less formal than the Indian Mahalanobis scheme of targeted expansion, or the planned efforts of the Soviet Union and its followers in Eastern Europe. Nonetheless, they represented new and important interventions, then and subsequently, designed to achieve desired results within the industrial sector.

Second, at the macroeconomic level, there was clear understanding of the recurrent need for external support of the balance of payments. International markets would not by themselves, in the short run, absorb a sufficient quantity of primary exports to support the large quantity of imports of intermediate and capital goods required to allow import substitution to proceed. Neither, in the midst of the Marshall Plan and World Bank efforts directed toward Western Europe, were there public funds available. Foreign direct investment, although frequently criticized as part of the greater degree of nationalism characteristic of the period, helped in two ways: it provided the needed finance, but also a direct access to the newer technology needed for more advanced stages of industrialization to proceed. This tension was to continue through subsequent years.

Third, at the microeconomic level, all efforts to enhance industrial expansion were provided. Public investment in urban centers occurred to
meet the needs of a rapidly expanding population, many now coming from the rural areas. New highways were built, as was occurring in the United States and Europe, to extend the transportation network. Older railways, established to facilitate primary exports, languished. The national political balance was frequently transformed in a variety of countries as this new effort was launched: labor unions became more relevant, and wage policy frequently surged to the national level for decision. At this time, minimum wages were a new instrument of government policy.

These characteristics, and their strong element of directing the market, rather than reacting to it, stood in sharp contrast to the capitalist style of earlier export orientation. But the import-substitution model also gave rise to important internal contradictions that very soon – almost within a decade – represented the beginning of the end.

One of the problems with the strategy was exactly its focus upon encouraging domestic production to substitute for continuing imports. Alas, that necessarily required still other imports of capital goods and other inputs in order to accomplish it. So net savings of foreign exchange were not as great as initially were conceived. Frequently, imports were initially permitted as an incentive to establish the industry domestically. Moreover, to ensure that those imported inputs were inexpensive, an overvalued exchange rate was the rule. In turn, this ensured reliance only on export of the basic primary commodities that had been present from the start. There was no motivation, nor was there initial efficiency, enabling exports to emerge from the new manufacturing firms that had been established. Instead, there was great emphasis on import protection, equivalent to a tax on the few agricultural exports able to compete. What saved the day, as noted above, was foreign direct investment and thus greater dependence on the outside world instead of the independence so much vaunted.

Another difficulty encountered was a rising government deficit, soon resulting in higher rates of inflation. Government expenditures and employment rose, while taxes lagged behind. An increased money supply was the solution. A debate ensued in most countries, between the structuralists – seen to be progressive – and the monetarists – conservative and crotchety. It is true that one can transfer some resources to the government sector through an inflation tax of limited amount. At that time, advocates of a bolder governmental strategy were speaking only of something like price increases of 10 to 15 percent, a rate that was soon exceeded in most cases. The problem was that the process of inflation acceleration was auto-generated. Continuously larger price increases – to make possible the needed transfer – were necessary as the public reduced their holdings of money in order to limit the loss of income. Another, and important, consequence was deterioration of the income distribution. Those at the bottom
of the income distribution, half and even more of the population, were affected negatively; they had no alternative but to use cash whose value began to deteriorate more rapidly.

Additionally, the very success in establishing industry had its negative counterpart in the lack of expansion of traditional sectors. Agriculture and mining lagged behind, unaccompanied by application of new technology or capital applied to enhance productivity. Because industry was capital-intensive, the growing labor force could not find increased employment there, but rather in urban services and governmental employment. Additionally there were pressures placed on agricultural producers to keep prices of foodstuffs low for the expanding urban areas. A sectoral imbalance emerged, one that prejudiced policy in many countries. Argentina is perhaps the best example of this, with political appeals to the military to prevent relative deterioration of agriculture, whose large exports of foodstuffs were regularly impeded and reallocated to domestic consumption. Everywhere, internal relative prices favored industry and penalized agriculture.

By the end of this first decade, even CEPAL could see that its initial hopes were unrealized, and that some change in policy was necessary. Raul Prebisch opted for a slightly altered strategy, focusing this time upon import substitution at the regional level. This, by allowing trade with neighbors, extended the basis for a viable industrialization through economies of scale. Despite the creation of a Latin American Free Trade Area in 1961, that undertaking made scant progress at the time. There were ambitious, but fortunately unrealized, plans for new geographic specialization in a variety of industries. Later, in 1967, the United States even accepted waiver of the ‘most favored nation’ principle: greater protection could be legitimately applied against its exports than those of Latin American neighbors.

Another option, emerging from the Left, was for much more fundamental internal reform, and less reliance on trade with the outside world. This view, in a few years to emerge more fully in the guise of dependency theory, saw the problem as a continuing commitment to the market system. There was too little focus on wide-ranging national reforms. Deficiencies of the import-substitution period began to be recognized. But the solution was now to come through greater centralization and state engagement, not exactly of the Soviet type, but more sympathetic to planning as well as more committed to elimination of rank income inequality.

A third alternative, and the one actually pursued, but only for a short time, was the Alliance for Progress, joined by the new Inter-American Development Bank. Both were to be new sources of public foreign capital, capable, together with private investment, of carrying Latin America to a stage of higher rates of economic growth and internal reform: land
redistribution, a new more equitable tax code, commitment to wider education at the basic level, and so on. Planning agencies were made universal, and approval of national plans was necessary to receive resources. Large groups of technical personnel made their way to Latin America to offer assistance. The region, atypically and suddenly, was momentarily at the center of attention.

This bold notion of democratic reform throughout the hemisphere was a United States response to the Cuban Revolution. It had barely begun its operation before another option soon gained force and spread through much of the region. That was explicit military intervention, first apparent in the case of Brazil in 1964, but rapidly spreading to other countries thereafter. The Alliance continued for additional years, but largely in name alone.

The military role was now somewhat a repeat of the Depression decade, although the economic circumstances were far different. The major objective was to preserve capitalism in the midst of the Cold War, and to eliminate the more radical options that threatened. Not all countries in the region were affected: Colombia and Venezuela in South America, Costa Rica in Central America, and Mexico retained civilian leadership. But the decade of the 1970s was later to be recalled as a decade of infamy for the violations of human rights that occurred in the Southern Cone.

Complicating matters still more, the petroleum crisis in 1973 exposed a major weakness in the model of Latin American development. Countries were substantially dependent upon petroleum imports, even those with national oil companies. Venezuela alone was able to benefit greatly from the fourfold rise in price from $3 to $12 a barrel in that year. Others coped by relying on a great surge of borrowing. At first, because interest rates remained low while commodity prices rose, such debt appeared a sensible way to finance the increased trade deficits. Unfortunately, debt-led development soon gave way to debt-led debt, a cumulating problem whose magnitude further multiplied with the outbreak of the war between Iraq and Iran in 1979, and a doubling once more of the price of petroleum.

Some in the region initially benefited. On this positive side, Venezuela was joined by Mexico, which had, during the 1970s, expanded its production considerably. Others, like Brazil, Chile and Argentina, had immediate adverse effects, the more so as international interest rates rose to new highs, for Paul Volcker’s Federal Reserve now sought much more actively to restrain inflation in the United States. As events would unfold, it was one of the oil beneficiaries, Mexico, that first signaled an inability to pay. That happened in July 1982, when an appreciated exchange rate and lack of international reserves forced it to seek help from the International Monetary Fund (IMF) and the United States. Thereafter, soon came many
others. By the mid-1980s, more than a dozen countries in Latin America were receiving financial assistance, and advice, from the IMF. Even with that help, imports had to be drastically curtailed. The international banks that had been so eager to lend in the 1970s had completely withdrawn. Debt-led debt had come to a disastrous end.

At the virtual peak in 1981, as Table 65.3 reveals, many of the countries in the region had shown a rapid expansion from 1950: per capita income had doubled, and for Brazil, almost tripled. Venezuela retained its position as the leader in per capita income, but failed to alter its dependence on oil exports alone as the source of wealth. Argentina and Chile only grew modestly, indicative of the failures of import substitution, compared to the relative successes of Brazil, Colombia and Mexico. More impressively, average Latin American income per capita had finally gained on the United States. It is no wonder that many look back to those three decades with continuing longing for the levels of high tariff protection, for substantial government assistance and for a clear emphasis upon industrial expansion.10

Two decades of reconstruction
In the 1980s, in the midst of coping with the debt crisis, came a new beginning for the region, both politically and economically. The ubiquity of military government that had begun in the 1960s and reached its height with the Pinochet intervention in Chile in 1973, faded extraordinarily rapidly. Interestingly, the problem of managing the debt – whose accumulation had occurred under military regimes – was a decisive element in speeding their demise. New constitutions and new civilian governments appeared

| Table 65.3 Latin American per capita income (1990 Geary–Khamis dollars) |
|-----------------|-------|-------|
|                 | 1950  | 1981  |
| Argentina       | 4987  | 7655  |
| Brazil          | 1673  | 4984  |
| Chile           | 3827  | 5933  |
| Colombia        | 2089  | 4272  |
| Mexico          | 2085  | 5582  |
| Peru            | 2263  | 4292  |
| Venezuela       | 7424  | 9637  |
| Regional average| 2614  | 5528  |
| Relative to USA | 0.27  | 0.30  |

Source: Angus Maddison, Monitoring the World Economy, OECD Development Centre, 1995
throughout the region. Leaders sought support in fairer elections in which larger proportions of the population figured than had previously. Democracy, and inevitably a negative response to continuing inflation, whose inevitable consequence was redistribution of income away from the poor, emerged and strengthened over this period.

The ongoing debt problems of the region in the 1980s cried out for solution. All that initially emerged were low growth, negative capital flows, real devaluations and postponement of external payments. Finally, by the decade’s end, the principle of debtors paying less was accepted. Private banks in the USA and elsewhere had finally emerged from the danger of their own failure, and now, in conjunction with the IMF, were amenable to the substitution of new Brady bonds – so named after the US Secretary of the Treasury – that saw reductions of some 40 percent of initial value. Mexico was the first country to proceed along these lines; Brazil was the last. Commercial banks got out of the business of extensive lending for capital investment, and back into providing shorter-term finance for international trade and other needs.

As this solution of the external problem evolved, there emerged a parallel recognition that fiscal deficits were the prime causal force in explaining Latin American inflation. Only if that governmental excess were credibly, and visibly, curtailed would prices stop their continuing acceleration. Some stabilization plan was necessary, and one that was more immediate in impact than past exercises. National outcomes varied in their particular style, their timing and the number of attempts required. Sometimes, as in Argentina and Brazil, multiple efforts were necessary. Sometimes, as in Nicaragua and Peru, populist measures were vainly instituted as alternative strategies. And sometimes, as in Bolivia in 1985, after achieving rates as high as an annualized 50,000 percent, results were virtually immediate after curtailing government deficits.

This was the essential key. In almost every instance, a firm price anchor was initially provided by a fixed exchange rate, and abetted by freer access to imports that helped to restrain future price increases. Ultimately, however, the key variable to brake inflation expectations remained the fiscal surplus, and when that was not realized, very high real interest rates and reliance on external capital flows were forced temporarily to substitute. These could work, but only for a short period. That is what we learned from the ‘Tequila’ crisis in Mexico at the very end of 1994, the Brazil crisis in 1999, and the Argentine collapse at the end of 2001.

What impresses is the extent to which past inflationary experience had to be overcome. Cutting back dramatically on inflation in the short term was not too difficult. The problem was ensuring continuity. Modest governmental deficits – by comparison with those of Europe, Japan and the
United States – gave rise to very high rates of price increase in Latin America. Internal credibility was slow to be achieved. Today, despite the variety of political leadership found in the region, the battle against inflation seems to have been achieved. That success had a notable and positive effect on the distribution of income, even when other factors intervened to worsen the aggregate.

Concern about fiscal deficits had a direct spillover upon two related subjects. The first of these was privatization. The impulse to sell off substantial state assets in energy, telephones, railways, steel and other intermediate sectors, petroleum exploration and production, airlines, banks, and so on came less from a philosophical commitment than from financial necessity. Initial levels of tax revenue, post-reform efforts to stem inflation, were generally insufficient to cover continuing current expenses. Virtually the only feasible solution was massive disposition of state property; the revenues thereby gained immediately eased the problem. When the need for revenue was very large, the terms of privatization were necessarily more favorable to the private buyers. Employment was almost always directly affected. There was excess labor, at higher wages than paid in the private sector, and almost every instance of privatization saw reductions in jobs, reflected in increased productivity, but also increases in formal sector unemployment.

Additionally, there was renewed attention to collecting larger revenues on a regular basis. Reform of the public sector to reduce current expenses was always announced, but rarely implemented. As a result, revenue collection generally increased much more than outlays were reduced. New taxes were frequently imposed: Brazil is at the regional extreme, with a total inflow of something like 37 percent of product, compared to an initial imposition of 25 percent. In other countries, beginning with Chile in the 1980s, attention turned to privatization of the social security system as a means of reducing future expenses. But these conversions required a surplus of current revenues as an enabling mechanism, thereby providing further incentives to ensuring adequacy of tax receipts.

Still a third area of great change during this period was a movement away from domestic protection and toward greater engagement with enhanced trade flows. Import flows, actual and potential, were an important competitive force capable of checking domestic price increases. Liberalization was therefore frequently utilized as part of the anti-inflationary effort. As countries privatized, external inflows to purchase, partially or entirely, the former nationalized assets became substantial. Currencies appreciated. The familiar problem of balance-of-payments limitations to growth soon reasserted itself. Numerous countries experienced serious crises as a consequence. There are the cases of Mexico in 1994, Brazil in 1999 and Argentina in 2002. But the commitment to openness, although occasionally frayed,
has remained. Tariff levels declined from an average in excess of 60 percent at the end of the 1980s to something between 15 and 20 percent in the mid-1990s.

These dramatic changes in the region inspired much criticism from the Left. As they progressively took place over the decade from 1985 to 1995, there was vocal opposition that took form in the electoral process. But newly elected presidents, frequently seemingly critical of such neoliberal policies, soon adopted the same measures. There was an initial positive consequence upon economic growth almost everywhere. But the new circumstances of the Tequila crisis, declines in Asia, the strong US dollar, failure in Russia and finally, Brazilian devaluation in 1999 took their toll: the promise that such reforms – termed neoliberal by their critics – might permit a resumption of sustained expansion was frustrated.

This is no way better seen than by viewing the data in Table 65.4. For virtually every country, with the prominent exception of Chile, and to some degree Mexico, the lost decade of the 1980s has transformed into a virtual repetition in the 1990s and even beyond. Note especially the dramatic decline in the ratio of income relative to that of the United States. Over the course of the entire twentieth century, there has been a dramatic decline; instead of convergence, there has been a disappointing divergence.

The present
Political opposition to these ‘Washington Consensus’ reforms has continued to increase in recent years. Much of that opposition – within, as well as outside, Latin America – emanates from intellectuals. This opposition

Table 65.4 Latin American per capita income (1990 Geary–Khamis dollars)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>8206</td>
<td>6834</td>
<td>6436</td>
<td>8005</td>
<td>8544</td>
</tr>
<tr>
<td>Brazil</td>
<td>5198</td>
<td>4917</td>
<td>4923</td>
<td>5296</td>
<td>5556</td>
</tr>
<tr>
<td>Chile</td>
<td>5738</td>
<td>5168</td>
<td>6402</td>
<td>8612</td>
<td>9841</td>
</tr>
<tr>
<td>Colombia</td>
<td>4265</td>
<td>4282</td>
<td>4840</td>
<td>5418</td>
<td>5096</td>
</tr>
<tr>
<td>Mexico</td>
<td>6289</td>
<td>6218</td>
<td>6119</td>
<td>6027</td>
<td>7218</td>
</tr>
<tr>
<td>Peru</td>
<td>4205</td>
<td>3631</td>
<td>2955</td>
<td>3505</td>
<td>3686</td>
</tr>
<tr>
<td>Venezuela</td>
<td>10139</td>
<td>8521</td>
<td>8313</td>
<td>8950</td>
<td>8415</td>
</tr>
<tr>
<td>LA average</td>
<td>5412</td>
<td>5052</td>
<td>5053</td>
<td>5460</td>
<td>5838</td>
</tr>
<tr>
<td>Relative to USA</td>
<td>0.29</td>
<td>0.24</td>
<td>0.22</td>
<td>0.22</td>
<td>0.21</td>
</tr>
</tbody>
</table>

starts from a legitimate concern about highly unequal income distributions and excessively high rates of unemployment that have been the counterpart of low growth. They then place blame upon a macroeconomic policy that has seemingly been too much in search of price stability, and too little in pursuit of economic expansion. The IMF has become an easy target for its insistence upon stabilization: reduced government outlays, higher rates of taxes, but above all, high real rates of interest. Privatization has been lamented, and its reversal sought, largely ineffectively. Now the argument has shifted to demands for greater regulatory control over private operation and decision. Greater protection against imports has been sought by industrial producers, eager to re-establish their former degree of dominance in internal supply. Finally, the initial aim of the 34 countries of the hemisphere – excluding Cuba – for a comprehensive free trade area, opening trade in goods and services, has not met its goal. Instead, there have been only an increasing number of bilateral agreements with the United States, with overt opposition from Venezuela, Mercosur and Bolivia.

But outright reversal of the great changes since the 1990s hardly guarantees the future development of Latin America and the Caribbean. Much that has occurred has been necessary, and much overdue. Market forces and wider trade are now a prominent feature almost everywhere in the world. Ignoring them has a real cost. Better to build upon what has been accomplished, and to recognize that more reforms, not fewer, are needed. Recent high rates of growth in the region since 2003 provide a singular opportunity to seize the moment.11

As essential first step is a substantial and ongoing commitment to educational reform and expansion.12 Latin America, for a variety of reasons, has been very late to recognize the importance of universal education to economic growth as well as to a more equal income distribution. In recent years, catch-up has begun to occur: the number of years of education available to the young has expanded, but still lags behind the impressive efforts in Asia. Repetition of beginning years of schooling continues, poor quality of teaching remains, and free public universities absorb significant budgetary resources. Past mistakes inevitably linger as individuals with limited literacy persist in the labor force. Reform is further complicated by its necessary continuity: one requires consistent policies over a generation, not over a single presidential term. So despite expenditures as a percentage of national product that approach 5 percent, few countries stand out positively. Comparable international test results confirm this regional backwardness.

Advance in coping with inequality requires continuing advance in the educational system. That is clear even from the extensive comparative report put together by the World Bank that correctly stresses other
factors. Education is hardly the only element influencing the distribution of income, but it frequently comes out as a principal one. Most relevantly, the quality of education is likely dramatically different for individuals whose parents are poor versus those who are wealthier. The appropriate conclusion is that ‘educational disparities account for an important share of Latin America’s high income-related inequality, but are not the only explanatory factor’.

A second area requiring attention is continued macroeconomic reform. In particular, the extraordinary regional difference in public finance seems to suggest a fundamental area of needed advance. Brazil, at one extreme, has revenues that exceed 38 percent of product, while Mexico, excluding oil receipts, barely enters into double digits. Neither level is likely to be efficient or effective. Even when large expenditures are made, governmental investment is small, and an inadequate infrastructure (physical, as well as individual, where health, nutrition, and so on weigh heavily) results. Countries, for example, have invested little in research and technology. Consequently, they have less capability to follow on closely the advances achieved elsewhere. Solving the inflation problem, as most countries have, does not mean the end of concern; instead, it is just the beginning.

Regional countries equally must expand their savings rates if they are again to achieve in the twenty-first century the per capita expansion accomplished in the 1960s and 1970s. Something of the order of 25 percent of national product is required, of which the foreign component should not exceed a small proportion. An inversion is needed. Historically, the private sector saved, allowing the public sector to invest. Now primary surpluses should become the rule, financing not only public capital formation, but permitting private firms to borrow at lower interest rates and for longer terms.

Trade advance should continue, on a more geographically as well as product-diversified basis. In the midst of the present rise in the terms of trade, and strong trade surpluses, there is greater enthusiasm about international trade within the region. Latin America continues to be a region where trade has been less significant than objective indicators suggest it should. Chile and Mexico are now the exceptions rather than the rule. The real test will come if, and when, the commodity boom begins to slacken. Latin America has had that experience before, and the inevitable result was to emphasize the internal market and strengthen protection against import competition, rather than to upgrade the quality of its exports and compete more effectively.

Whether an increasing commitment to the international market will persist, is the question. Globalization is seen in many countries as a hindrance rather than an opportunity. Last time, in the expansion of income
through 1980, that option was largely ignored. Now one hears increasing doubts about whether an open market should persist. Latin America needs the chance to enter, however belatedly, into the international marketplace. For it to choose an alternative path will again deter rather than enhance the region’s emergence.

A final, but important, subject is the need for institutional reform. This refers not merely to reform of the judiciary and the need for persistent and independent standards, but also to the area of regulation of the newly privatized activities. As natural monopolies, compounded by the reality of extensive foreign participation, these activities require supervision and reasonable certainty. Otherwise, investment required in infrastructure will continue to lag behind. Not only economic institutions are involved. Political reform is likewise needed in so many countries, where electoral rules are inadequate, and fundamental changes in structure are required.16

**Conclusion**

The countries of Latin America and the Caribbean find themselves challenged once again, just as they have been over the entire post-World War II period. This time, after having lost out to South Korea and Taiwan in the 1980s and 1990s, the task is much greater. China, India, South Africa and still other Asian countries are now the new competitors in world markets for trade and foreign investment. These are much larger challengers to cope with. And, over the last several years, they have been impressive performers.

In the face of this new challenge, some have advocated greater South–South exchange, as was the mantra in the 1960s and 1970s. Some see the present World Trade Organization (WTO) Doha Round as the place for large reductions in agricultural protection that will assure fairer trade and better opportunities for Latin American advance. Some have called for a resurgent state to replicate the advances achieved in the earlier era, a return to subsidies and import substitution, but this time fully committed to reduction of inequality in the distribution of income. Second-stage reform, as advocated here, is rejected in favor of a different approach.

That route of denial, attractive as it sometimes seems, runs a substantial risk of casting away the advances of almost a generation of reforms. International competition is inevitable, and increasing, in the present age of globalization, as even the developed nations have been discovering. Rather than try to deny that reality, the countries of the region must seek to confront it. Only through a process of internal productivity advance, fueled by advances in technology and greater domestic savings, will Latin America be able to compete, and thereby achieve greater continuing growth.

There is not much time left to choose.
Statistical appendix

There are now four comprehensive series of estimates of national income dating back to the nineteenth century, and encompassing the countries of Argentina, Brazil, Chile, Colombia, Mexico and Venezuela. Peru is the additional country incorporated in the Maddison estimates. Uruguay is also sometimes found. Although these series do differ with the particular base year selected, as well as sources utilized, which does explain some difference in absolute value relative to the United States or OECD countries, they are quite similar in their reported longer-term rates of growth.\(^\text{17}\) Table 65A.1 provides these results. Shorter-term movements in individual countries do differ, however, and sometimes importantly.

I will not enter here into the index number problems presented by such long-term comparisons, and the familiar Laspayres bias found as growth incorporates new sectors into production over time. These, and other methodological issues, are discussed extensively by Prados de la Escosura.\(^\text{18}\) But less time seems to have been spent in searching out and considering critically the national historical estimates that underlie the aggregates. For Brazil, for example, three different series seem to have been chosen by the authors. Ironically, the most recent estimates by Prados de la Escosura choose a series of Raymond Goldsmith, put together in the 1960s, well before recent research on historical Brazilian development flourished.

\[\text{Table 65A.1 Comparative per capita rates of growth}\]

\[
\begin{array}{cccc}
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LA 6</td>
<td>LA 6</td>
<td>LA 6</td>
<td>LA 7</td>
</tr>
<tr>
<td>1900–13</td>
<td>2.2</td>
<td>2.5</td>
<td>2.2 (2.6)</td>
<td>2.4</td>
</tr>
<tr>
<td>1913–29</td>
<td>1.0</td>
<td>1.5</td>
<td>2.4 (1.5)</td>
<td>1.6</td>
</tr>
<tr>
<td>1929–50</td>
<td>1.4</td>
<td>1.6(^\text{a})</td>
<td>1.8 (1.8)</td>
<td>1.5</td>
</tr>
<tr>
<td>1950–80</td>
<td>3.1</td>
<td>2.7(^\text{b})</td>
<td>2.5 (3.0)</td>
<td>2.8</td>
</tr>
</tbody>
</table>
\end{array}
\]

Notes:
LA 6 is Argentina, Brazil, Chile, Colombia, Mexico and Venezuela
LA 7 includes Peru
\(a\) 1929–45
\(b\) 1945–81

Sources:
Col (1) Prados de la Escosura (2004)
Col (2) Thorp (1998, Statistical Appendix), calculated from Total GDP minus population growth
Col (3) Hofman, (2000, p. 169)
Col (4) Maddison (2001)
I have chosen to use the Maddison estimates for Latin America – despite his use of the Goldsmith estimates for Brazil – principally because of their greater comparability with those for other regions. That database extends back to 1820 and incorporates information not only on the present developed countries but also on many developing nations in Africa, Asia and Southern and Eastern Europe. In all, there are 56 countries utilized, far larger than in any other source.

Notes

1. There has been a proliferation of new research covering Latin American economic history over the course of the last two decades. Much of this work is quantitative and a large part has been the product of Rosemary Thorp, who merits special credit. She authored *Progress, Poverty and Exclusion: an Economic History of Latin America in the Twentieth Century* for the Inter-American Development Bank in 1998, and co-edited three volumes of background papers in the series *An Economic History of Twentieth Century Latin America*, published by Palgrave in 2000. I have made much use of these volumes, without explicit citation. But additionally, there are the recent contributions of Leandro Prados de la Escosura, as well as that of Andre Hofman. And, of course, there has been much independent research done on a national level. In the brief ‘Statistical appendix’, I explain my choice of Angus Maddison’s national income estimates as the basis for long-term comparisons.

2. Lewis (1978).

3. Bulmer-Thomas (2003, p. 144), finds an R-squared of 0.82.


7. The classic exposition is found in CEPAL (1951).

8. This can be easily seen from the formula for revenue from the inflation tax: $p(M/P)$, the rate of inflation times the public holdings of money. Of course, as inflation occurs, desired holdings of money decline, which is why the process tends to become cumulative.

9. The Economic Bulletins issued by CEPAL at the beginning of the 1960s give evidence of this shift, as do the *Annual Economic Surveys*.

10. Ironically, many sometimes forget the degree to which the gains were associated with military governments, as well as substantial debt accumulation.

11. Nancy Birdsall has put together another 11 measures beyond the Washington Consensus, ten that are seen as required for the future. They stress the need for great social equity in the region. There is some overlap with my modest four; Birdsall et al. (2001). 

12. PREAL, *Quantity without Quality, 2006 Report Card*, provides extensive details on the reforms required, and the poor marks countries continue to receive for only partial commitment.

13. de Ferranti et al. (2003).


15. This is true of various gravity models using physical characteristics such as distance from markets as well as income levels. See Carillo-Tudela and Li (2004).


17. For some reason, Andre Hofman (2000) has chosen to present his aggregate results for per capita income, as well as for other measures, as a simple average of the individual country results, rather than weighting by the relevant magnitudes. This is appropriate were one interested exclusively in differences among countries, rather than also concerned with group, that is, Latin American, performance. In Table 65A.1, I have provided his original averages in parentheses, and recalculated the LA 6 totals.

References


Inter-American Development Bank (various years), Economic and Social Progress Report, Washington, DC: IDB.


Introduction
In 1960, economic development was the mandate and keenly sought-after province of a founding generation of African political leaders. Visions of economic progress were central to liberation rhetoric (Ake, 1996) and were widely embraced within the broader development community (for example, Karmarck, 1971). Yet development failed, and it failed uniquely. Over the next 40 years, as extreme poverty fell dramatically on a worldwide basis, poverty in sub-Saharan Africa (henceforth SSA or ‘Africa’) increased. At the turn of the millennium, nearly one in two Africans or 300 million in total consumed less than $1 a day, a proportion twice as high as the global rate and double the number prevalent in Africa in 1970. Home to 10 percent of the world’s population, the region now accommodates 30 percent of the world’s poor – who spend a quarter less than the Asian poor on their livelihood (World Bank, 2005). The African development challenge has become the global development challenge. How and why did it become so, and what can we expect, looking ahead?

We approach these questions through the lens of economic growth. The second section describes the African record from 1820 to the present, focusing on the period since 1960 and emphasizing demographic and other features that differentiate African growth patterns from those of other developing regions. In the third section we outline two main structural explanations for the African experience, based in turn on governance and resource endowments. We develop the former theme in a detailed exploration of the political economy of African growth. The fourth section documents the revival of growth that got under way in the mid-1990s; we interpret this revival as a lagged response to the economic and political reforms of the late 1980s and early 1990s. We argue that Africa now faces a window of opportunity, with politically stable countries facing the prospect of mutually reinforcing declines in fertility rates and increases in capital formation and growth. We conclude with a summary of lessons from the African experience, both for the region itself and for development economics.

The growth record
Any adequate account of Africa’s development experience must acknowledge the overall weakness of Africa’s growth record, its juxtaposition with
population explosion, and its variability over time and across countries. We begin with Africa’s long-run growth experience, drawing first on the painstaking historical research of Maddison (2001) and then on World Bank data for 100 developing countries since 1960. Our central concern here is the relative stagnation of economic growth in SSA in the period since 1950. This performance is not an outlier in historical terms: in Table 66.1, Africa’s per capita growth rate of 0.7 percent after 1950 matches that of the rest of the developing world over the previous century. But African populations missed out on the economic transformation that took place in the developing world – particularly in Asia – in the second half of the twentieth century. The result was that by the 1950s, African incomes, which had gained considerable ground in relative terms since 1913, had begun to diverge powerfully from incomes elsewhere in the developing world. By comparison with East Asia and Pacific, a shortfall of less than 50 percent in purchasing power parity (PPP)-adjusted terms around 1960 rose to well over 300 percent by the end of the century. The consequences of this growth failure are apparent in Table 66.2, which provides a snapshot comparison of human development, first at the outset of the 1960–2000 period and then at the end. With the exception of the primary enrollment rate, which was already high outside of Africa in the early 1960s, Africa fell further behind the rest of the developing world. Regress was not as severe for non-income measures as it was on income or poverty: by 2000 Africa exceeded the levels of primary enrollment, adult literacy and life expectancy that had prevailed elsewhere in 1960. But the failure to raise per capita incomes significantly had critically undermined these achievements.

At the sectoral level, slow growth has gone hand-in-hand with limited structural diversification. Traditional agriculture continues to absorb the majority of the labor force in many African countries, a feature no longer observed in any other region of the world (O’Connell and Ndulu, 2000). Irrigation is expensive and extremely sparse, with the result that African agriculture remains largely rain-fed and subject to periodic drought. Exports have tended to remain concentrated in a narrow band of primary commodities (Berthélemy and Soderling, 2001, 2002), including exploitation of mineral resources. Collier and O’Connell (2007) use global data to identify ‘resource-rich’ economies as those that exceed threshold values for the ratios of primary commodity rents (from energy, mineral and forest resources) to gross domestic product (GDP) and primary commodity exports to total exports on a sustained basis. Comparing SSA with other developing regions (excluding the Middle East and North Africa), they find that a stark difference already existed in 1960, with 12.5 percent of the SSA sample classified as resource-rich and only 7 percent of the non-SSA sample. This difference expanded over time, with another 16.7 percent of
Table 66.1 Long-run growth rates by developing region.

<table>
<thead>
<tr>
<th>Population</th>
<th>SSA</th>
<th>Total</th>
<th>LAC</th>
<th>ASIA</th>
<th>MENAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1820–70</td>
<td>0.3</td>
<td>0.2</td>
<td>1.3</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>1870–1913</td>
<td>0.7</td>
<td>0.6</td>
<td>1.6</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>1913–50</td>
<td>1.7</td>
<td>1.0</td>
<td>2.0</td>
<td>0.9</td>
<td>1.3</td>
</tr>
<tr>
<td>1950–2001</td>
<td>2.6</td>
<td>2.0</td>
<td>2.3</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>1950–60</td>
<td>2.2</td>
<td>2.1</td>
<td>2.8</td>
<td>2.0</td>
<td>2.6</td>
</tr>
<tr>
<td>1960–2001</td>
<td>2.7</td>
<td>2.0</td>
<td>2.2</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Peak rate*</td>
<td>3.00</td>
<td>2.52</td>
<td>2.87</td>
<td>2.52</td>
<td>3.34</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Real GDP (PPP-adjusted 1990 dollars)</th>
<th>SSA</th>
<th>Total</th>
<th>LAC</th>
<th>ASIA</th>
<th>MENAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1820–70</td>
<td>0.6</td>
<td>0.1</td>
<td>1.2</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1870–1913</td>
<td>1.1</td>
<td>1.3</td>
<td>3.5</td>
<td>1.1</td>
<td>1.6</td>
</tr>
<tr>
<td>1913–50</td>
<td>2.7</td>
<td>1.5</td>
<td>3.4</td>
<td>0.9</td>
<td>2.5</td>
</tr>
<tr>
<td>1950–2001</td>
<td>3.3</td>
<td>5.0</td>
<td>4.0</td>
<td>5.4</td>
<td>4.9</td>
</tr>
<tr>
<td>1950–60</td>
<td>4.1</td>
<td>5.6</td>
<td>5.1</td>
<td>5.8</td>
<td>5.6</td>
</tr>
<tr>
<td>1960–2001</td>
<td>3.1</td>
<td>4.9</td>
<td>3.7</td>
<td>5.3</td>
<td>4.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Real GDP per capita (PPP-adjusted 1990 dollars)</th>
<th>SSA</th>
<th>Total</th>
<th>LAC</th>
<th>ASIA</th>
<th>MENAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1820–70</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>1870–1913</td>
<td>0.4</td>
<td>0.7</td>
<td>1.8</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>1913–50</td>
<td>1.1</td>
<td>0.5</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>1950–2001</td>
<td>0.7</td>
<td>3.0</td>
<td>1.7</td>
<td>3.4</td>
<td>2.2</td>
</tr>
<tr>
<td>1950–60</td>
<td>1.9</td>
<td>3.4</td>
<td>2.3</td>
<td>3.7</td>
<td>2.9</td>
</tr>
<tr>
<td>1960–2001</td>
<td>0.4</td>
<td>2.9</td>
<td>1.5</td>
<td>3.4</td>
<td>2.0</td>
</tr>
</tbody>
</table>

| No. of countries | 53 | 90   | 27  | 42   | 21    |

Notes: SSA is Maddison’s ‘Africa’, excluding Algeria, Egypt, Morocco, and Tunisia, but including Libya (see below). LAC comprises ‘Latin America’ and includes the Caribbean. ASIA is Maddison’s ‘East Asia’. MENAT (Middle East, North Africa and Turkey) corresponds to ‘West Asia’ plus North Africa. Libya could not be separated out and we therefore include it here in SSA rather than in MENAT. Note that these growth rates are based on regional totals and are therefore dominated by the large countries in each region. More specifically, the population, real GDP and real GDP per capita growth rates correspond to population-, GDP- and (approximately) GDP-weighted individual-country growth rates. Note also that data before 1950 unavoidably contain major ‘guesstimates’. For example, the African growth rate for 1870–1950 is based on case study work on Algeria, Egypt, Ghana, Morocco, South Africa and Tunisia. Maddison applies the average growth rate of GDP per capita for this group of six (= 0.90) to all countries in SSA other than Ghana and South Africa. Our SSA aggregate for 1870–1913 combines this indirect estimate for 51 countries with Maddison’s direct estimates of 2.07 and 1.25 for Ghana and South Africa.

* Peak year and peak rate correspond to the earliest year after which all subsequent population growth rates are lower.

Source: Ndulu and O’Connell (2007), Table 1.4. Calculations are based on country and regional data from Maddison (2001).
<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Initial values</th>
<th>End-to-end annual growth rates</th>
<th>Ending values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1960 or earliest year before 1965, or as indicated</td>
<td>1970</td>
<td>1970</td>
</tr>
<tr>
<td>SSA</td>
<td>35</td>
<td>1278.1 (53.8)</td>
<td>55.8</td>
<td>41.1</td>
</tr>
<tr>
<td>OTHER</td>
<td>43</td>
<td>2591.5 (90.5)</td>
<td>26.5</td>
<td>53.2</td>
</tr>
<tr>
<td>DEV'ING</td>
<td>305</td>
<td>3338.4 (99.1)</td>
<td>17.4</td>
<td>56.4</td>
</tr>
<tr>
<td>LAC</td>
<td>22</td>
<td>934.4 (58.6)</td>
<td>55.5</td>
<td>45.3</td>
</tr>
<tr>
<td>SASIA</td>
<td>4</td>
<td>1833.1 (94)</td>
<td>20.4</td>
<td>50.6</td>
</tr>
<tr>
<td>MENAT</td>
<td>7</td>
<td>2402.5 (81.9)</td>
<td>42.3</td>
<td>51.7</td>
</tr>
</tbody>
</table>
Table 66.2  (continued)

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Initial values (1960 or earliest year before 1965, or as indicated)</th>
<th>End-to-end annual growth rates (earliest year before 1965 to latest year between 1995 and 2000)</th>
<th>Ending values (latest year between 1995 and 2000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDUST</td>
<td>22</td>
<td>8507.6</td>
<td>101.8</td>
<td>70.2</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>3433.3</td>
<td>80.4</td>
<td>38.1</td>
</tr>
<tr>
<td>SSA v SASIA</td>
<td></td>
<td>(136.8)</td>
<td>(91.8)</td>
<td>(100.5)</td>
</tr>
<tr>
<td>SSA v Other Dev</td>
<td>(49.3)</td>
<td>(59.4)</td>
<td>(210.6)</td>
<td>(77.3)</td>
</tr>
</tbody>
</table>

Notes: Except in the final 2 rows, the numbers in parentheses give the relevant developing-country mean as a percentage of the industrial-country mean. The final 2 rows show the SSA mean relative to the SASIA mean and the mean for all non-SSA developing regions. Regions: SSA = Sub-Saharan Africa; OTHER DEV'ING = Other Developing (LAC = Latin America and Caribbean, SASIA = South Asia, EAP = East Asia and Pacific, MENAT = Middle East, North Africa and Turkey); INDUST = Industrial countries.

Source: Ndulu and O’Connell (2007), using PWT6.1 and World Development Indicators.
the African sample acquiring resource-rich status by 1990, as compared with only 10.5 percent of the non-African. A consequence of Africa’s delayed structural transformation has been the continued vulnerability of its population to shocks to rainfall and world commodity markets.

Tables 66.1 and 66.2 document the demographic explosion that is a striking correlate of Africa’s economic stagnation over the post-1960 period (O’Connell and Ndulu, 2000; Lucas, 2003). With the exception of life expectancy rates, standard indicators of demographic pressure differed only modestly between Africa and the rest of the developing world in the early 1960s (Figure 66.1). But the demographic transition was already well underway in Latin America, where population growth rates had peaked in 1960 (Table 66.1). For the next 25 years, total fertility rates fell sharply outside of Africa while remaining virtually unchanged within Africa. Population growth rates therefore diverged sharply, and from the early 1970s through the remainder of the century, the population of SSA grew more rapidly than the non-African developing-country population had grown at its peak. The ratio of (overwhelmingly young) dependents to working-age population grew steadily, exceeding historical developing-country norms by 1970 and remaining above these through 2000. The fertility rate began to fall in Africa in the mid-1980s, suggesting entry into the final phase of the demographic transition. We will return to this observation, which is complicated by the huge impact of HIV/AIDS on life expectancies starting in the late 1980s.

The aggregate growth record conceals considerable variability, both over time and across countries. Table 66.1 indicates that a sharp deceleration of growth took place in Africa between the end of the colonial period (1950–60) and the remainder of the century. Within the latter period, a further distinction can be made (Figure 66.2), between the moderate growth rates of the 1960s and late 1990s and the deep contraction of 1974–94. African economies continued to grow in the 1960s. This growth was already weaker than that of other developing regions, however, and the global deceleration of the 1970s therefore took substantial portions of the continent into outright contraction. Between 1960 and 1994, nearly half of African countries with comparable data suffered per capita income losses exceeding 20 percent in constant domestic currency (Rodrik, 1998). The bulk of these losses occurred between 1974 and 1994, a period that began with a set of shocks to energy and tropical commodity markets (1974–79) and ended with a concentrated wave of African democratic reforms (1989–94). As indicated in Figure 66.2, the growth shortfall between 1974 and 1994 is much larger on a population-weighted basis than in the simple averages. This is partly a large-country phenomenon within Africa, but after 1980 it is driven primarily by the dramatic growth performance of
Source: Ndulu and O’Connell (2007), using World Development Indicators 2005. The figures show simple averages of country observations, for all countries with continuously available data.

Figure 66.1 Demographic pressure: SSA versus Other developing regions, 1960–2000

The averages discussed here also mask wide variation in the growth performance of individual countries. Many African countries have experienced limited episodes of robust growth. Nearly half of the African economies studied by Pritchett (1998), for example, exceeded a per capita income growth threshold of 1.5 percent through the mid-1970s. Ghana and more strongly Uganda have consistently exceeded a 2.5 percent threshold (roughly the long-run median for developing countries) since the mid-1980s. Botswana and Mauritius grew spectacularly; their long-run records compare favorably with those of the East Asian miracle economies. A diversity of outcomes also characterizes the period after 1994, during which 15 African countries have seen growth rates of total GDP in excess of 5 percent. Focusing on the most recent five-year period (1999–2004) and excluding the oil countries, median per capita growth rates in the fastest-growing, middle, and slowest-growing thirds of the African sample – each comprising 12 or 13 countries – were 2.8, 1.0 and –1.8 percent respectively.4
Outside of the mineral-exporting group, rapid growth after 1990 has also been associated with substantial diversification of production and exports. Finally, the weak contribution of increases in measured physical and human capital per worker to African growth is yet another significant feature. Weak investment effort is part of this story (see below): standard growth accounting exercises assign about half of the post-1960 growth shortfall relative to other developing regions to a shortfall in measured capital deepening per worker. But the data also suggest profound problems in translating investment effort into effectively utilized capital (Pritchett, 2000). In order to reconcile observed growth outcomes with measured capital inputs, one must conclude that the average productivity of African inputs per worker fell considerably over time, not just relative to productivity elsewhere but in absolute terms. Ndulu and O’Connell (2007), for example, find that the country-level cumulative changes in total factor productivity over 1960–2000 were as likely to be negative within Africa as positive (see also Hall and Jones, 1999). The correlation of physical capital accumulation with growth, moreover, is considerably lower within Africa than in the rest of the developing world, even over periods of a decade or longer. The latest example of this is Africa’s recovery starting in the mid-1990s, which was not accompanied by a commensurate boom in aggregate investment (Berthélemy and Soderling, 2001).

**Explaining poor growth performance in Africa**

Why did the African environment prove hostile to economic growth after 1950? Two broad lines of argument dominate the literature, based in turn on economic mismanagement and structural impediments to growth. With some imprecision we refer to these below as the ‘governance’ and ‘geography’ views. Demographic trends pose distinct structural challenges, and we treat these separately. A synthesis of these strands has yet to be achieved, though we make a tentative approach after reviewing the main arguments.

**The governance critique**

A critique of economic management has dominated the literature on African economic performance since the early 1980s. The central themes of what we will call the ‘governance critique’ were laid out in the World Bank’s 1981 *Accelerated Development in Sub-Saharan Africa* (the ‘Berg Report’, after its chief author Elliot Berg), Robert Bates’s classic 1981 *Markets and States in Tropical Africa* and, with a ten-year lag, Paul Collier’s 1991 *African Affairs* article on agencies of restraint.

In its initial and most influential form, the governance critique sought to explain what it characterized as excessive intervention by African governments in economic markets. Bates (1981) argued that African policy-makers...
had sacrificed both the agricultural sector and industrial efficiency in order to divert resources to favored interests. The 1980 Berg Report had located African economic stagnation in overtaxation of export agriculture, over-protection of import-competing industries, and oppressive state control of finance, industry and agricultural marketing. Bates (1981) argued that while these policies were economically inefficient, the ‘urban bias’ they displayed was rational from the point of view of political elites. Farmers would lose out, and their losses would exceed the gains to the political elites and to the urban interests more generally (the civil service, the military, and labor and capital in the formal private sector or state enterprise sector). But farmers faced deep collective action problems; while numerous, they were too poorly organized to constitute the ‘selectorate’ on which the government depended to retain power (Bates and Devarajan, 2001).

Governments would therefore penalize agriculture and support urban-based industrialization far beyond what could be justified by the correction of market failures. Moreover, they would do so using inefficient quantity-based instruments rather than price-based interventions (that is, quotas, exchange controls and marketing monopolies rather than tariffs and explicit export taxes), because the distribution of policy-generated rents was central to their political security. Growth would fail, but the protected urban electorate would be well served. Exceptions to the urban bias pattern could be explained, in the Bates analysis, by appealing to the rural political roots and business interests of founding political leaders in countries like Kenya, Malawi and Côte d’Ivoire.

The Berg–Bates contribution provided a description of African policy biases, a link from these to growth outcomes, and a grounding of these choices in the interests of African political elites. Each element of the argument provoked important voices of dissent or moderation. But the empirical content of the critique gave it substantial appeal to Africa’s donors and to economists seeking to understand the continent’s lagging growth performance. For donors, the governance critique provided a rationale for using conditional lending to push market-based reforms – a tendency undoubtedly strengthened by the ascendancy of conservative governments in the United States and Europe in the early 1980s. Donors could act as the agents of disenfranchised African populations, imposing conditionality on easily monitored policy reforms like trade liberalization and exchange rate unification. Since the existing policies penalized the poorer rural sector, their removal would simultaneously improve growth and distribution. For economists, the view that policies represented the largely autonomous choices by a self-interested political elite provided a causal interpretation of regression evidence linking measures of policy distortion with economic growth (for example, Sachs and Warner, 1995).
A generalization of the critique to misaligned interests
Subsequent contributions developed the governance critique beyond its initial application by restating it in terms of a conflict of interest between African policy-makers and African populations. The basic argument is that under authoritarian rule, development policy in a large number of African countries has tended to be captured by a narrow political elite operating under relatively weak institutional constraints. Characterized by weak legitimacy and tenuous bureaucratic control, autocratic regimes in Africa until the 1990s did not function as agents of the public interest, tending instead to sacrifice growth in favor of patronage-based redistributive politics or outright predation (see also Adam and O’Connell, 1999; Ndulu and O’Connell, 1999; Humphreys and Bates, 2001).

An anti-growth bias may become severe if policy-makers undervalue future interests relative to present. Policy-makers may discount the future excessively if they have little expectation of remaining in power or being held accountable. Using African data from the 1980s, McMillan (2001) finds that high rates of presidential turnover predict inefficiently high rates of export taxation in Africa, particularly for crops with large fixed inputs. Fosu (2002) documents the frequency of attempted coups in Africa and finds a strongly negative causal link to overall growth. In an extreme case of misaligned interests, the expected tenure of an authoritarian leader or oligarchy may therefore be inversely related to growth performance. Diamond (1977) appeals to global evidence, for example, to argue that economic success tends to bring democratization. If autocratic leaders internalize such a link, then development may represent a threat rather than an investment in future rents (Robinson, 1997). Political elites may then actively oppose development – as in some cases the colonial powers themselves did, fearing the contestability development might create in the economic and political spheres.

A combination of autocracy and tight economic controls during 1960–85 presented a particularly potent environment for patronage and predation. Using data on the type of African leadership Ndulu (2007) shows that 80 percent of autocratic regimes since independence imposed soft or hard controls (see Collier and O’Connell 2007) for a substantial portion of their tenure.

Economic controls were not new to Africa at the time of independence. Colonial administrations had embraced an activist, developmental role by the late 1940s, reflecting the suddenly temporary nature of their remaining trusteeship and the allure of state intervention following depression, wartime mobilization and the emergence of the Soviet Union as a great power. A mentality of market regulation was therefore in place well before the formal transfer of sovereignty in Africa, as were some of its key institutional mechanisms including monopoly export marketing boards,
exchange restrictions and economic plans (Fieldhouse, 1986). But the period from 1960 to 1975 saw a dramatic expansion of the regulatory presence of the African state. Although this sharp expansion was part of a global phenomenon and grounded in the global development paradigm of the day, it was propelled further by rent-seeking behavior.10

Late in the colonial period, the colonial powers had begun to introduce the rudiments of democratic self-government, and independence constitutions reflected the institutional structures of Western democracies, with contested multi-party elections, checks and balances, and substantial civil liberties. But the degree of de facto democracy deteriorated steadily following independence. As in the case of economic controls, authoritarian government was not an African innovation. Its growth impact in SSA, however, appears to have been far from beneficent, in contrast to the broad Asian experience with authoritarian rule during the same period (Alesina and Perotti, 1994). The governance critique asserts that in the African context, causality runs from government institutions to growth.

Limitations and extensions of the governance critique
While the governance critique had instinctive appeal both to Africa’s donors and to economists seeking to understand the continent’s lagging growth performance, it also raised serious conundrums.

Institutions Van de Walle (2001) argues that the unifying feature of African political economy is not the power of urban interests but rather the absence of coherent domestic political interest groups of any kind. In this view, domestic interests are everywhere too weak to restrain the behavior of the tiny political elite that holds or shares power in its own interest. To explain the persistent ‘choice’ of stagnation, he appeals to the low capability of African public bureaucracies and the dysfunctional influence of strategically motivated external donors. These conspired to undermine the medium-run coherence of virtually any policy initiative, he argues, while simultaneously protecting high-level African elites from personal responsibility for economic decline.

In an influential extension of the governance critique, Collier (1991) similarly appealed to an institutional vacuum that left too much discretion and too little accountability in the hands of political elites. Collier focused on agencies of restraint – public institutions designed to protect national assets (including privately held ones) from predation. Before 1960, this role was provided by colonial administrations operating under tight mandates of internal security, fiscal solvency, and openness to metropolitan trade and finance. Immediately following independence, the new counterparts to colonial institutions – export marketing boards, national central banks,
multiparty parliamentary systems, independent judiciaries – struck a balance between flexibility and restraint that reflected the conservative mandates of their colonial predecessors. But political leaders sought to consolidate power and could appeal to the need for ambitious development programs. By the mid-1970s, political leaders had systematically relieved national ‘agencies of restraint’ of their powers to restrain executive action (see also Collier, 1982; Bratton and van de Walle, 1997). Some palpable gains in policy flexibility were observed – Botswana, for example, left the Rand Monetary Area (RMA) and subsequently achieved lower inflation than the RMA countries. But in most cases such gains appear to have been overwhelmed by increased macroeconomic instability and deterioration in the protections afforded to private investment (points developed in detail in Collier and Pattillo, 1999).11

In the mid-1990s, the cross-country growth literature began to develop a broader institutional version of the governance critique, based on the tendency of African countries to cluster in the lower ranks of institutional performance measures that are correlated with growth on a global basis. The attentions of policy reformers meanwhile began to shift from conventional economic reforms to problems of improving public service delivery, reducing bureaucratic corruption and strengthening the rule of law. While economists still know relatively little about how durable improvements in public sector performance are achieved, three observations seem relevant to the African situation. First, when institutions are initially weak, the initiative of high-quality political leaders is critically important in determining how well or poorly existing institutions actually perform (Reinikka and Collier, 1999; Glaeser et al., 2004). Second, as suggested above (and as measured by shares in formal sector employment or total investment), African governments have tended to be large and overbearing rather than small and efficient. Third, institutions are known to display persistence: learning takes place and interests form around existing patterns of behavior (North, 1990). These observations suggest that much of the logic of the governance critique applies directly to the performance of public sector institutions over the 1960–2000 period. Wherever political leaders were unable to reconcile the benefits of a market-friendly institutional environment with their own priorities, institutional performance deteriorated and growth suffered.

Did Africa’s abrupt democratization between 1988 and 1994 improve the institutional ground of policy-making? Contested elections and a free press are among potentially important agencies of restraint cited by Collier (1991). Partly for this reason, our overall answer is a positive one.

Fractionalization, polarization and nation-building We next look at the impact on policy of the patterns of sub-national identity that existed at the
time of independence and that in many cases became the dominant mode of political mobilization and conflict. The countries of SSA came to political independence both later and more rapidly than those of other developing regions (Table 66.3). While only Ethiopia, Liberia and South Africa existed as independent states at the end of 1955, fully three-quarters of colonial Africa, representing the vast bulk of its population and GDP, had achieved political independence by 1966. In 1966 the average independent state in SSA had held sovereignty for fewer than ten years; its counterpart in the rest of the developing world had been independent for the better part of a century.

Colonial structures of political control were both arbitrary – with boundaries cutting across historical patterns of politics and trade – and effective. Their abrupt departure meant that the challenge of economic development was in many cases confounded from the outset with an acute problem of nation-building. Nigeria provides a telling example of the impact of \textit{ex ante} regional polarization on political and economic development. But similar patterns of internal polarization, often created or reinforced in the encounter with conquering European powers, existed throughout the continent in 1960. Azam (2007) emphasizes the salience of coastal–interior cleavages in West Africa, operating as in Nigeria on a North–South axis and tending to separate a nomadic, pastoralist Muslim interior from a more sedentary, educated, Christianized coast. In the Horn of Africa, both Sudan (Arab North, Christian and Animist South) and Ethiopia (federated with richer Eritrea after World War II, to guarantee Ethiopia’s access to the coast) have engaged in ethno-regionally based civil wars since the early 1960s. In Central Africa, Belgian favoritism towards

\begin{table}[h]
\centering
\caption{Dates of political independence, developing countries}
\begin{tabular}{llccccccc}
\hline
Region & \multicolumn{1}{c}{Proportion of countries politically independent} \\
 & \multicolumn{1}{c}{Region} & \multicolumn{1}{c}{\(n\)} & \(10\%\) & \(25\%\) & \(50\%\) & \(75\%\) & \(All\) \\
\hline
Other Developing & 66 & 1830 & 1830 & 1946 & 1961 & 1981 \\
of which: & & & & & & \\
LAC & 27 & 1818 & 1825 & 1840 & 1962 & 1981 \\
ASIA & 20 & 1816 & 1933 & 1948 & 1956 & 1975 \\
Total & 112 & 1822 & 1907 & 1960 & 1964 & 1993 \\
\hline
\end{tabular}
\end{table}

\textit{Source:} Gleditsch and Ward (1999) database, as compiled by Ndulu and O’Connell (2007), Table 1.2.
the Tutsi minority produced enduring cleavages that erupted in civil war in the 1990s. In South Africa, and in those portions of colonial Africa with large settler populations – including Kenya and Zimbabwe – race-based geographical discrimination was a matter of state policy; in these cases the pattern of *ex ante* polarization would eventually require determining the status of settler minorities.

While the salience of ethno-regional polarization was clear to political scientists in the early 1960s (for example, Carter, 1966), economists have only recently begun to come to grips with the implications of nation-building for African economic growth. Two approaches have been important. The first is due to Easterly and Levine (1997), who focused on fractionalization rather than on polarization *per se*. Easterly and Levine noted that the degree of ethno-linguistic fractionalization – measured by the probability that two randomly chosen individuals in a given country spoke a different first language – was extremely high in many African countries, by comparison with global norms. Moreover, on a global basis ethno-linguistically heterogeneous countries tended to grow more slowly, as a result of weaker public sector performance. Miguel (2004) reports a similar finding for Kenya and Tanzania, using data on local provision of public services. Collier (2000) finds, however, that the adverse impact of heterogeneity is strongly contingent on political institutions. In democracies, ethno-linguistic heterogeneity has no impact either on overall growth or on microeconomic efficiency (as measured by the economic return on World Bank projects), while in dictatorial regimes the adverse impact is strong.

Azam (1995, 2007) focuses directly on polarization, defining a polarized society as one like Nigeria’s, in which there are two or three large sub-national ethnic groups that dominate population and politics in separate regions. Azam argues that in a situation of *ex ante* ethno-regional polarization, regionally-based redistribution may be required to buy off the threat of armed conflict. The existence of such a risk is consistent with the global evidence of Collier and Hoeflfier (2004), who find the risk of civil war maximized under conditions of polarization: homogeneous societies have low exposure to civil war, but so do heterogeneous societies. In cases of *ex ante* polarization, then, the Azam analysis may force a reinterpretation of what is conventionally viewed, within the governance critique, as distortionary redistribution. If the absence of redistribution invites armed conflict and economic collapse, then a program that distorts efficiency relative to an irrelevant peaceful counterfactual may in fact be growth-promoting relative to the true counterfactual of civil war. This shifts the ground of the governance critique from redistribution per se to the instruments that are employed to achieve it. Political elites attempting to ‘buy the peace’ should be observed doing so transparently and credibly.
(perhaps via constitutional means), and with a minimum of distortion; and they should simultaneously employ instruments directly targeted at reducing polarization.

**Geography**

Starting in the mid-1990s, Jeffrey Sachs and Adrian Wood began to build an empirical case for the adverse influence of resource endowments and geography on African economic growth. Wood argued that in a world of capital mobility, comparative advantage was determined by endowments of immobile factors: primarily unskilled labor, human capital and natural resources. Africa’s rich endowment of natural resources relative to human capital implied a deep comparative advantage in the production and export of primary commodities (Wood and Berge, 1997). The failure of African countries to achieve competitiveness in manufacturing was therefore largely independent of trade policy or the quality of governance, although these factors may have affected the long-term evolution of factor endowments. Sachs argued that high transport costs and a hostile disease environment conspired to make capital accumulation and productivity growth much more expensive in Africa than elsewhere in the developing world (Sachs and Warner, 1995, 1997, 2001; Bloom and Sachs, 1998; Gallup and Sachs, 1999).

**Distance and landlockedness**

African populations are internally fragmented and isolated from world trade by unusually large land distances, unhealthy lowland coastlines, a sparse network of ocean-navigable rivers, and multiple political borders (Gallup and Sachs, 1998; Faye et al., 2004). With its 48 economies, the region has by far the highest density of countries per land area of any developing region; on average, each country shares a border with four neighbors (Ndulu, 2004). Nearly 40 percent of the African population lives in countries that are landlocked or virtually so. The unusual distance of African population concentrations from coastlines and ocean-navigable rivers seems to follow in part from the inland locations of water resources critical to agriculture, including the Great Lakes, major non-ocean-navigable rivers, and fertile rain-fed uplands.

The remoteness of African population concentrations may have severely limited the scope for Asian-style growth patterns based on proximity to global markets, scale economies and agglomeration. Africa’s relatively more sparse distribution of population, significantly low population density and relatively lower rate of urbanization raises significantly the transport intensity of its economic activities (C. Kessides, 2005). Furthermore, the lower population density and urbanization tend to increase the amount of infrastructure investment required to produce similar levels of income (Esfahani and Ramirez, 2003).
Unusually high internal transport costs accentuate this remoteness. Limão and Venables (2001) estimate that it costs nearly twice as much for the median African country to move a 40-foot container from a coastal port to its in-country destination, as it does for countries in other developing regions. These costs are particularly damaging for manufacturing, where the share of traded intermediate inputs is relatively large. They also penalize physical capital accumulation by raising the relative price of investment. Investment in Africa is unusually expensive in terms of local income, so that a given national saving rate delivers a lower increment to real capital accumulation in Africa than in other regions. The average relative price of investment goods for sub-Saharan Africa was 70 percent higher than for Organisation for Economic Co-operation and Development (OECD) countries or East Asia. Artadi and Sala-i-Martin (2003) find that the high relative price of investment goods reduces Africa’s predicted growth rate by 0.44 percent on an annual basis, holding saving effort constant.

Notwithstanding variations across countries in the region, for most African countries distance from their primary markets and the high transport intensities of their products (low value, high weight and sparsely produced) are major impediments for production and trade (Esfahani and Ramirez, 2003). Using a gravity model Limão and Venables (2001) estimated the elasticity of trade with respect to transport costs, and found it typically to be quite high at −3. Distance to key markets is an important impediment to trade as expected, but in their model poor infrastructure (measured by an index combining road, rail and telecom density) accounts for 40 percent of the predicted transport cost for coastal countries and up to 60 percent for landlocked countries. The median landlocked country has only 30 percent of the trade volume of a median coastal country. What is also striking from this study is that, holding activity levels and direct distances between trading partners constant, improving internal infrastructure within the landlocked country itself is as important as improving the infrastructure in the transit country.

Landlockedness adds a political dimension to remoteness. Transport costs now depend crucially on the infrastructure investments and pricing policies of coastal neighbors. These neighbors may also be important but unreliable hosts for the export of labor services. Sachs and Warner (1997) and others find that landlocked status reduces predicted growth by up to 1 percent per year on a global basis.

Disease burden Sachs and Warner (2001) and Masters and McMillan (2001) emphasize the high burden of human and animal disease in tropical climates and its impact on life expectancy, human capital formation, labor
force participation and economic growth. Ninety-two percent of SSA lies within the tropics, as compared to 60 percent for East Asia. Following Sachs and Warner (2001), Artadi and Sala-i-Martin estimate the foregone growth in Africa as a result of malaria prevalence at 1.25 percent per annum, a figure that surely reflects the influence of other highly correlated aspects of the health environment. Acemoglu et al. (2001) take a very different, institutions-based approach to linking disease burden with growth. They argue that the quality of contemporary institutions reflects the nature of the institutions introduced by European powers during the colonial period. Where the local disease environment was inhospitable, Europeans introduced extractive institutions, leaving a legacy of predation and violence that continues to undermine the rule of law and the security of property. Where the disease environment was favorable to European settlement, colonial regimes set up institutions conducive to long-term growth (see also Easterly and Levine, 2003).

Demography
In contrast to the experience of other regions, a fertility transition has not happened in Africa despite sharp reductions in infant mortality since the late colonial period and (until the HIV/AIDS epidemic starting in the 1990s) gradual improvements in life expectancy across the age distribution. The distinctive demographic features of African countries weigh unusually heavily on national saving and undermine the building up of the human capital needed for growth (Bloom and Sachs, 1998). Indeed, although we saw earlier that human development indicators have not diverged as strongly as income levels when comparing SSA averages with averages for other developing regions, these indicators have nonetheless deteriorated in relative terms. Enhancing human capacity by increasing the longevity of working life and improving skills and organizational effectiveness are important components of a strategy to close the growth differential with other regions.

Two distinct consequences of continued high fertility stand out. Firstly, the average population growth rate is at least one full percentage point above that for other developing regions. This increases the amount of national saving required to achieve any given increase in human and physical capital stocks per capita. It also increases the age dependency ratio, which reduces the per capita purchasing power associated with any given level of output per worker. High dependency ratios may also undermine the quality of human capital accumulation by spreading educational resources more thinly.

Secondly, until the early 1990s, rapid population growth produced not just a high but also a rising age dependency ratio. As we saw earlier, as Africa’s age dependency ratio gradually increased, the rest of the developing world
experienced a fertility transition that lowered population growth rates sharply and gradually reduced the ratio of dependents to working population over time.

In a regression-based counterfactual exercise, O’Connell and Ndulu (2000) estimate that Africa’s average growth is reduced by 0.85 percentage points relative to the sample mean, and by nearly 1.5 percentage points relative to East Asia as a consequence of its distinctive demographic patterns. This situation is made worse by the fact that HIV/AIDS has become epidemic and added to the burden particularly for survivors. HIV/AIDS patients in Africa account for 60 percent of the world’s people living with HIV/AIDS. This has a profound social and economic impact due to the large number of premature deaths of people in their prime age of employment and parenting.

A window of opportunity and sustained growth since the mid-1990s
Over the decade since 1995, 16 countries have had annual GDP growth of 5 percent or higher, by comparison with only five during the previous decade (Table 66.4). These countries account for 35 percent of the sub-Saharan Africa population. Some of the fastest-growing countries have also done relatively well in terms of poverty reduction, as demonstrated by a group of eight low-income African countries that grew at an average rate of 2.9 per capita per year and reduced poverty headcounts at an annual rate of 1.5 percentage points (World Bank, 2005). Excluding the oil-producing countries, income per capita in the fastest-growing one-third of African countries grew at a median rate of 2.8 percent over the five years 1999–2004. During the same period, the slowest-growing countries – predominantly those affected by conflict – saw their economies contract at a median rate of 1.8 percent. These host 21 percent of the region’s population. There has nonetheless been a striking decline in the number of countries posting negative growth rates of total GDP: four during the most recent period, down from 13 during the first half of the 1990s. In the middle of the growth distribution (accounting for some 13 percent of Africa’s population) the median growth rate was 1.0 percent per capita over 1999–2004.

These trends reflect important changes that are taking place across the continent. Policies and institutions are improving, peace and security is returning to the region, and African governments are increasingly taking control of their own economic destiny. Increased political participation and competition are giving Africans a greater stake in their own future. Demographic trends appear to have taken the first turn towards a transition that will reduce pressures on fiscal resources, encourage savings and support productivity growth. We elaborate below on each of these areas of progress.
### Table 66.4 Average GDP growth rates: sub-Saharan Africa

<table>
<thead>
<tr>
<th>Country</th>
<th>85/94</th>
<th>95/04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>-0.9</td>
<td>7.7</td>
</tr>
<tr>
<td>Benin</td>
<td>2.9</td>
<td>5.0</td>
</tr>
<tr>
<td>Botswana</td>
<td>8.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>3.6</td>
<td>4.7</td>
</tr>
<tr>
<td>Burundi</td>
<td>2.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Cameroon</td>
<td>-1.6</td>
<td>4.6</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>4.4</td>
<td>5.8</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>0.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Chad</td>
<td>4.2</td>
<td>7.4</td>
</tr>
<tr>
<td>Comoros</td>
<td>1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Congo, Dem. Rep.</td>
<td>-3.6</td>
<td>-0.6</td>
</tr>
<tr>
<td>Congo, Rep.</td>
<td>-0.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Côte d’Ivoire</td>
<td>1.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>2.9</td>
<td>22.3</td>
</tr>
<tr>
<td>Eritrea</td>
<td>-</td>
<td>2.3</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Gabon</td>
<td>1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Gambia, The</td>
<td>2.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Ghana</td>
<td>4.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Guinea</td>
<td>4.0</td>
<td>3.9</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>3.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Kenya</td>
<td>3.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Lesotho</td>
<td>5.0</td>
<td>3.3</td>
</tr>
<tr>
<td>Liberia</td>
<td>-18.7</td>
<td>16.6</td>
</tr>
<tr>
<td>Madagascar</td>
<td>1.2</td>
<td>2.9</td>
</tr>
<tr>
<td>Malawi</td>
<td>1.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Mali</td>
<td>1.7</td>
<td>5.9</td>
</tr>
<tr>
<td>Mauritania</td>
<td>3.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Mauritius</td>
<td>6.4</td>
<td>5.0</td>
</tr>
<tr>
<td>Mozambique</td>
<td>4.2</td>
<td>8.0</td>
</tr>
<tr>
<td>Namibia</td>
<td>3.5</td>
<td>3.4</td>
</tr>
<tr>
<td>Niger</td>
<td>2.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Rwanda</td>
<td>-4.3</td>
<td>10.9</td>
</tr>
<tr>
<td>Sao Tome and Príncipe</td>
<td>1.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Senegal</td>
<td>2.2</td>
<td>5.1</td>
</tr>
<tr>
<td>Seychelles</td>
<td>5.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>-1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Somalia</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.6</td>
<td>2.7</td>
</tr>
<tr>
<td>Sudan</td>
<td>3.6</td>
<td>6.2</td>
</tr>
<tr>
<td>Swaziland</td>
<td>6.6</td>
<td>2.9</td>
</tr>
</tbody>
</table>
A significant and durable improvement in the policy and institutional environment

During the last decade a large number of reforming African countries have re-established sustained macroeconomic stability, committed credibly to more open trade regimes, and consolidated market-based economic reforms that have improved the conditions for private sector involvement in the economy.

Sustained macroeconomic stability has returned to a large number of countries in the region, as evidenced by significantly lower inflation, narrower fiscal and external trade deficits, and a widespread move to Article VIII status in the International Monetary Fund (IMF), implying a commitment to currency convertibility for current account transactions – a status hardly conceivable only a decade earlier.

Consumer price inflation has persistently and sharply fallen within a decade, from 27 percent in 1995 to about 6 percent by 2004. This has resulted from a combination of significantly stronger fiscal controls across a wide range of countries and a shift by central banks towards a focus on price stability as the primary goal (Ndulu, 2004). In the median African country, government spending as a proportion of GDP fell sharply in the past decade, as it has in other developing countries in the world. During the 1990s, fiscal deficits among the 31 low-income African countries for which comparable data exist dropped from double digits to 5.2 percent before grants and only 2.5 percent after grants (World Bank, 2000). The residual deficit has been financed largely through non-inflationary sources. Where financing through external grants is assured, fiscal deficits (before grants) have tended to be higher, but without destabilizing the macroeconomic environment. At the turn of the century black market premia were extremely low across the continent, averaging just 4 percent outside of a few countries like Zimbabwe in acute political turmoil. Through unilateral trade reforms, African countries have also compressed both the tariff rates and categories with average tariff rates of 15 percent. In these respects the

Table 66.4  (continued)

<table>
<thead>
<tr>
<th>Country</th>
<th>85/94</th>
<th>95/04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>3.1</td>
<td>5.1</td>
</tr>
<tr>
<td>Togo</td>
<td>1.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Zambia</td>
<td>0.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>3.7</td>
<td>−1.5</td>
</tr>
</tbody>
</table>

Source: World Bank data.
continent now more resembles other developing regions, where reforms have been pursued in earnest for prolonged periods.

We also use here the Country Policy and Institutional Assessment (CPIA) data from the World Bank to show the extent of improvement in the policy and institutional environment since the mid-1990s. Annually the World Bank assesses the quality of borrowers’ policy and institutional performance in areas relevant to economic growth and poverty reduction. These assessments began in the late 1970s but the criteria of good performance have evolved over time. While in earlier years assessments focused mainly on macroeconomic policies, they now include other aspects such as institutional quality and governance. CPIA scores for individual countries have been found to be highly correlated with private-sector country ratings, such as the Institutional Investor, International Country Risk Guide (ICRG) and Euromoney ratings (Easterly, 1993). The CPIA scores range between 0 and 6 and are benchmarked globally. A score of 3.5 is broadly considered as a threshold of good performance. The average CPIA score for African countries rose sharply starting in the mid-1990s, and between 1997 and 2004 the number of countries scoring above the 3.5 threshold tripled, from 5 to 15 – the latter number accounting for nearly a third of all countries in the region.

Peace and security is spreading in the region

After protracted periods of conflict, peace and security has returned to many parts of the region. Southern Africa entered the millennium as a region of stability following the resolution of long-standing conflicts in Mozambique, Namibia, Zimbabwe and Angola (2002) and the transition to majority rule in South Africa. West Africa likewise has seen the end of conflicts in Chad, Sierra Leone and Liberia. In Eastern Africa and the Horn of Africa, the resolutions of conflicts in Burundi and more tentatively in Sudan and Somalia hold out the prospect of durable change. Africa’s progress is significant relative to other regions of the world, as reflected in global data on the incidence of civil war. The incidence of civil war shows a sharp increase in both absolute and relative terms in SSA in the early 1990s, a period that coincided with the wave of first democratization documented above. Since the mid-1990s, however, there has been a sharp decline in the proportion of countries under civil conflict in the region; and the proportion of African population in such countries has fallen even more sharply, dipping below the average of other developing regions at the turn of the century. This positive change is corroborated by a new global database on the incidence of violent conflicts (Gleditsch, 2004). Collier and Hoefler (2004), using this data, show that immediately after the end of the Cold War the incidence of wars declined. The number
of wars peaked in Africa in 1992 and since then the number of civil wars seems to be declining.

*Increased political participation gives Africans a greater stake in their own future, laying a stronger foundation for domestic accountability and policy restraint*

During 1990–94, SSA experienced a sharp wave of democratization (Bates, 2007). By the mid-1990s, this episode had fully made up for nearly three decades of absolute and relative deterioration, placing the region well above the norms prevalent in the rest of the developing world during 1960–85.

Political competition and participatory processes improved by more in Africa during the 1990s than in other regions. In 1982, only one-tenth of African countries and two-tenths of other developing countries had competitively elected executives. As late as 1991, Africa showed virtually no improvement, while other developing countries had doubled their figure to 40 percent. By 1995, however, the gap was nearly closed, despite continuing increases in other regions. In 2002, Africa was ahead of the other regions by about eight percentage points.

The political democratization drive in Africa has created space for peaceful regime changes, deeper debates about societal development visions, and greater respect for human rights. Although we observed earlier that the initial wave of democratization was associated with a sharp rise in civil conflict in the region, this situation appears to have reversed itself as democratic practices have taken hold. The democratization process has also raised the expectations from citizens throughout the region. Analysis of recent Afrobarometer surveys and the World Values Survey show that the majority of Africans believe democracy is good for the economy. They also prefer democratic political systems to authoritarian alternatives, as indicated in Figure 66.3. The African public expects democracy to deliver access to the basic necessities of life, including food, water, shelter, and also education. These ‘values’ surveys also show that Africans care about equity and public action to reduce poverty. They report discomfort with wide wealth differentials and a strong commitment to political equality (‘voting not a privilege of the better-educated’).

*African governments are pursuing collective action to improve the region's attractiveness and taking increasing control of their economic destiny*

There is now a strong revival in regional integration initiatives in Africa, with a change in focus from preoccupation with preferential trade arrangements to an approach that emphasizes market integration and promoting the region as an attractive investment destination for foreign and African
capital. The African Union and the New Partnership for Africa’s Development (NEPAD) have embraced the latter two objectives. These objectives can be met if Africa as a region can: (1) achieve a critical mass of countries with a policy environment friendly to capital accumulation and private business; (2) improve cross-country infrastructure links; (3) moderate the risks faced by domestic and foreign capital, and employ risk-mitigating instruments more effectively; and (4) strengthen as well as retain its pool of human skills.

As part of the effort to improve the region’s collective reputation and attractiveness, African governments are taking action to improve governance and connectivity under the African Union (AU) and the NEPAD initiatives. These initiatives are designed to: (1) push African countries to be assertive about ownership and to assume leadership and accountability for their development programs; (2) improve the reputation of the region through certification of good practices in governance for a critical mass of African countries under the African Peer Review mechanism; (3) increase regional connectivity to improve capacity to trade within the region and with the outside world through regional initiatives to scale up collaborative effort in improving infrastructure; and (4) enhance the capacity of a rationalized system of regional bodies to provide regional public goods – such as cross-country transportation and power-sharing networks, coordination in managing pandemics such as HIV/AIDS and malaria, and protection of regional commons such as the Nile river basin and the Great Lakes.


Figure 66.3 African Political Values
Implications for future development strategies

We have anchored the analysis of Africa’s development experience in the overall weakness of the region’s growth record, the juxtaposition of this record with a population explosion, and the variability of experience across countries and over time. If recent improvements are to be sustained and deepened, a combination of addressing the fundamental bottlenecks to scaling-up growth, and accelerating the demographic transition now slowly under way, is fundamental. The wide diversity of opportunities and constraints in the region rules out generalizations about the country-level growth strategies appropriate to these goals. Nevertheless there are a few broad lessons worth highlighting by way of concluding remarks.

The development constraints African countries face are neither static nor decisive. Geographical disadvantages and natural-resource dependence are not destiny, as their effects can be offset or ameliorated. Botswana, the fastest-growing economy in Africa (and among the fastest globally) since its independence in 1966, presents a striking example. It is landlocked and natural-resource dependent. Arguably, the strength of its state capacity, together with its being part of Southern Africa’s relatively effective infrastructure system, customs union and monetary area (for a long period), helped offset the negative effects of remoteness and served as a commitment instrument against rent-seeking.

The historical analysis showed clearly that Africa virtually missed the rapid development that has taken place in other developing regions in the closing four decades of the twentieth century and indeed the region can be considered the last frontier of the global development challenge. Being a late starter has its advantages and disadvantages. Access to knowledge from development experience and technological progress presents an opportunity to fast-track the development process and leapfrog. At the same time, Asia’s cumulative success presents a challenge to the competitiveness of late-starting Africa, as trade preferences are eroded and opportunities to learn before facing intense competition shrink. Among the resource-poor coastal economies, the incumbency of the successful Asian coastal economies has probably created a more challenging playing field for African export diversification starting in the 1990s than existed in earlier decades.16

We can categorize the constraints to growth discussed above into three groups: risks, transactions costs and capacity. Risk is largely associated with macroeconomic instability and absence of credibility to commit due to weaknesses in governance and institutions. The higher transactions costs in the region are largely associated with its unfriendly geography, climate and bureaucratic processes. Capacity constraints relate primarily to low human capital, partly associated with the late demographic transition and weak institutions.
There is no doubt that raising the level and efficiency of investment is critical if Africa is to close the gap in growth with other regions of the world. Getting the right policy environment in place and sustaining such an environment is a key signaling instrument for credibility to investors (both local and foreign). Indeed, given the central role of the modern state in defining the incentive environment for private economic activity, the failure to engender sustained growth must ultimately be traced to unsuccessful policy choices. Collier and O’Connell (2007) estimate the contribution of anti-growth policies to the growth differential between Africa and other regions (p. 24); taking 40 years of African growth experience as a whole and controlling for differences in the composition of opportunities, they conclude that policies inimicable to growth account for more than half of the overall growth differential with the rest of the developing world, or as much as 1.8 percentage points out of an overall (population-weighted) differential of 3.5 percent. This result is confirmed by a regression analysis in which they estimate the effect of anti-growth policies, controlling for shocks and differences in growth opportunities across countries. This magnitude of relative importance of policy mistakes in explaining the growth performance differential with other regions is corroborated by evidence from cross-country growth studies reviewed extensively in O’Connell and Ndulu (2000).17

Notwithstanding the importance of policy choices in explaining Africa’s growth differential with other developing regions, however, a substantial portion of the growth differential is accounted for by geographically based proxies for differential growth opportunities (Collier and O’Connell, 2007, pp. 81–8). Separating landlocked and resource-poor, coastal and resource-poor, and resource-rich countries, Collier and O’Connell find that while African countries tended to underperform in each category, nearly a third of Africa’s overall growth shortfall is associated with the unusually high share of African countries that are either landlocked and resource-scarce or resource-rich – both relatively low opportunity categories as compared to the coastal and resource-scarce group, based on the global evidence. The underperformance was most severe for Africa’s coastal resource-scarce economies (suggesting that missed opportunities dominate the story) and least severe for Africa’s land-locked resource-scarce economies (suggesting poor potential). Two sources of this differential are important – proneness to policy errors and the higher cost of doing development business, discussed above.

Proneness to policy mistakes is likely to be more important in resource-rich and coastal countries where the rent-seeking stakes are higher (from resource and trade rents), while higher transactions costs of economic activities from geographical disadvantages are likely to be more important
in landlocked countries. The convergence of resource wealth and ethno-regional fragmentation in many of the countries of SSA presents a particularly high risk of adoption of anti-growth policy regimes. The management of resource rents under conditions of ethnic diversity is one of the most important and distinct policy challenges for SSA.

A third challenge is dealing with the consequences of a dramatically delayed demographic transition in Africa compared to other regions. The fertility rate began to fall in Africa in the mid-1980s, suggesting entry into the final phase of the demographic transition. But unlike the experience in other regions, the pace of this transition is very slow; and the HIV/AIDS pandemic complicates the situation further. Although population changes are longer-term phenomena, as Srinivasan (1988) urges, it may be worthwhile to look at potential policy responses in the shorter term to help accelerate the demographic transition. These include education policy, population policy and gender equality to induce greater attention to the quality of children.

Much of the above discussion has employed *ceteris paribus* counterfactuals in order to isolate the individual effects on growth of policy mistakes, underprovision of public goods or weaknesses in human capital. If there are critical synergies across these factors, or if there are sharply increasing returns to individual state variables over some interval (as suggested by Azariadis and Drazen, 1990 and Berthélemy, 2005 in the case of human capital and Sachs et al., 2004 for both human capital and public infrastructures), then the achievement of rapid growth may require complementing policy improvement with a ‘big push’ to deal with poor governance, conflict and insecurity, and inadequate infrastructure. In terms of their ambition, contemporary versions of the big push are not novel; the record of Africa’s early decades is replete with the ambitions of visionary leaders who sought to engineer a sharp break from patterns of specialization and distribution inherited from the colonial era (Ndulu, 2007). What is different this time around, and provides a critically important window of opportunity for stakeholders, is that the core functions of market-friendly governance are widely understood to be among the binding constraints.

**Acknowledgment**

We thank Lopa Chakraborti for research assistance.

**Notes**

2. Here and throughout the chapter, we confine group totals to countries with continuously available data. This eliminates potentially misleading compositional effects within
groups, but it also means that group composition can differ by table or variable (for example, World Bank data are available for only 35 countries in SSA, while Maddison provides estimates for 52). For each variable (or table, as indicated), we report group totals for the full set of countries with continuously available data.

3. Nigeria, Ethiopia, Democratic Republic of the Congo and South Africa together account for 45 percent of African population among countries with available data; these economies contracted at an average rate of 1.8 percent over 1974–94.

4. We exclude oil countries, which grew at a median per capita rate of 3.4 percent. The upper, middle and lower thirds of the growth distribution comprise 35, 13 and 21 percent of Africa’s population, respectively.

5. A recent contribution in this line is van de Walle’s *The Politics of Permanent Crisis* (2001), discussed further below.

6. On within-Africa variation in agricultural policy and export taxation, see Mkandawire and Soludo (2000). On the weakness of within-country links from policy to long-term growth, see Easterly et al. (1993). On the ideological (as opposed to self-interested) motivations of African policy-makers, see Ndulu (2007), who notes the profound influence of Fabian socialism and dependency theory on founding African leaders. On the relative weakness of initial urban interests see Ndulu and O’Connell (1999), who observe that urban interests were nonetheless created by policy in some cases, and thereby came to undermine subsequent prospects for reform (see also the discussion of van de Walle, 2001 in the text).

7. Sahn (1996) argues cautiously, on the basis of calibrated general equilibrium models, that if market-based reforms had been fully implemented in the 1980s, their impact would indeed have been mildly progressive.

8. In McMillan’s analysis, policy-makers seek to maximize the revenue from taxing agricultural exports. They face a time-consistency problem, however: once farmers have sunk planting costs in the hope of receiving high producer prices, there is an incentive for policy-makers to pay very low producer prices that cover only the costs of harvesting. Evidence on crop- and country-specific export tax rates confirms that this incentive is strongest when policy-makers have urgent revenue needs and short planning horizons, and when the ratio of harvest to planting costs is small.

9. In a celebrated article, Acemoglu et al. (2001) argued that colonial regimes brought development-oriented institutions only where local health conditions supported the establishment of a large settler presence (see p. 507).

10. Migdal (1988) describes the development paradigm spanning this period as one in which the state is the ‘primum mobile’ of socio-economic progress. The idea of ‘developmentalism’ and the idea of state intervention were seen as inseparable, and policies and planning were seen as offering boundless possibilities for social engineering. It was taken for granted by multilateral and bilateral development agencies that the state had a pivotal role to play in transforming societies from backwardness to modernity (Ljunggren, 1993, pp. 7–8).

11. Collier argued that in the resulting situation of executive dominance, effective restraints in the areas of trade and monetary policy would have to mimic the reciprocal and supra-national structure of international trade agreements, where countries reciprocally commit to growth-promoting policies and to penalty structures capable of enforcing them. Donor conditionality, he argued, was ill-suited to fill the institutional vacuum: donors had their own constituencies and could not credibly threaten to terminate aid based on poor policy performance. As examples of partially successful supra-national arrangements he cited the maintenance of low inflation in the CFA countries (the 13 members of 2 monetary zones issuing respectively the West African CFA – Communauté financière d’Afrique – franc and the Central African CFA – Coopération financière en Afrique centrale – franc) and Rand Monetary Area, and the avoidance of highly distorted trade regimes by members of the Southern African Customs Union.

12. The Democratic Republic of Congo belongs in the latter category; perhaps also the Sudan, with its vast internal territory and limited access to its Red Sea coastline, and Ethiopia before the independence of Eritrea in 1994.
13. Limão and Venables (2001) estimate the median transport cost for a 40-foot container, from coastal port to destination (including transshipment), at $7600 for African countries. The comparable figures for Latin America and the Caribbean, East and South Asia, and the Middle East and North Africa are $4600, $3900, and $2100.

14. The 16 countries are Angola, Benin, Botswana, Cape Verde, Ethiopia, Mali, Mauritius, Mozambique, Rwanda, Senegal, Tanzania, Uganda, Chad, Equatorial Guinea, Liberia and Sudan (see Table 66.4).

15. These countries are Senegal, Mozambique, Burkina Faso, Cameroon, Uganda, Ghana and Cape Verde.


17. There is a wide range of other studies that corroborate this same conclusion. Ndulu (1998) reported the results of counterfactual simulations of Africa’s growth performance if conditions obtaining in East Asia were present in the region. Using results earlier obtained by Elbadawi and Ndulu (1995), Easterly and Levine (1997) and Elbadawi et al. (1997), a better policy environment as obtaining in East Asia would have raised growth by an additional 1.5–2.6 percent.

References


Berthélemy, Jean-Claude (2005), ‘Conférence du G8: Pour chaque dollar de dette annulée, autant sera supprimé dans l’aide’, Marché tropicaux et méditerranéens, 60 (3112), July: 6–8 [interview].


Robinson, James A. (1997), ‘When is a State Predatory?’ mimeo, Department of Economics, University of Southern California, Los Angeles, October.


Introduction

The growth and development performance of the Middle East and North Africa (MENA) region presents one of the major anomalies that current economics literature seeks to resolve, which is how to reconcile the existence of massive natural resources with the high unemployment, low growth and general underdevelopment of the region. In this debate, much attention is focused on the problems arising from: (1) state-oriented inward-looking economic policies; (2) lack of ‘integration’ with the world economy; (3) underdeveloped financial sectors and chilling investment climate; and (4) low levels of human capital development. In this chapter, we attempt to present a summarized yet more balanced and hopefully more insightful analysis of the growth and development experience of the countries in the region, with special attention given to the existing bottlenecks hindering future development prospects.

While discussing the MENA region as a whole we will divide the countries into five subgroups: (1) oil-rich labor-importing states (Bahrain, Kuwait, Oman, Libya, the United Arab Emirates, Qatar and Saudi Arabia); (2) oil-rich labor-abundant states (Algeria, the Islamic Republic of Iran, Iraq, Syria); (3) oil-poor labor-abundant NICs (Egypt, Morocco, Turkey); (4) oil-poor limited natural resource states (Israel, Tunisia, the West Bank and Gaza, Jordan, Lebanon); and (5) natural resource-poor states (Sudan, Yemen) (Richards and Waterbury, 1996). Although the inclusion of Turkey, Israel and Iran is controversial as the trajectory of the Arab and other Middle Eastern countries constitute a more appropriate whole, they share many commonalities as well. However, unless stated otherwise, the general statements will exclude Turkey and Israel.

The economic history of the MENA region is characterized by several cycles of growth and accumulation. In retrospect, the region formerly enjoyed higher levels of economic development and prosperity compared to its counterparts in Europe. While Istanbul with its 700,000 inhabitants in the sixteenth century was the largest city in the world, North Africa overall was much more urbanized than Europe (Paris with 125,000 inhabitants versus Cairo with 450,000 in around 1500) (Bairoch, 1997, pp. 517–37). However, in the last of these cycles, the region experienced a decline in its growth and development indicators starting from the early eighteenth century, with the
factors that precipitated this decline remaining a source of continuing debate. The current chapter will focus on the region’s most recent cycle, namely that after the 1950s following the gaining of political independence and control over natural resources by the countries in the region.

**Growth and development in the MENA**

From the late 1950s to the late 1970s (and in some cases, till the late 1980s and early 1990s) the economic structure of the region was characterized by an import-substituting industrialization (ISI) regime, the main features of which (as elsewhere) included strict quantitative controls on international trade, overvalued exchange rates and severe rationing in foreign exchange and credit markets.

Following the hikes in petroleum prices in the early 1970s, growth and development indicators in the MENA region improved rapidly. The sudden increase in investment and growth rates in the oil-exporting countries spread to the rest of the region through increases in worker remittances, and capital flows. In addition, gross capital formation jumped to exceptionally high rates, generating a locomotive effect on growth rates and overall standards of living. On the financial front, considerable amounts of financial savings were accumulated abroad, resulting in the famous expansion of the Eurodollar market through the recycling of petrodollars.

In contrast, the downside of the above picture has been the high level of volatility of gross domestic product (GDP) growth since the 1970s: the average volatility of GDP growth in the region as a whole has been twice that of the developing-country average, and twice more volatile in the oil-rich economies than the rest of the region (Abed and Davoodi, 2003; Hirata et al., 2004, pp. 62–3).

The single most important determinant of growth in the MENA (where fuel products account for about half of the region’s GDP and around 90 percent of total exports in the oil-rich countries) has been the fluctuations in international oil prices. In addition to growth volatility, as a result of high dependence on oil revenues, fiscal policy in the oil-rich countries is also volatile and procyclical. Likewise, the oil-poor labor-abundant countries are also oil price-sensitive because a large part of their economies is dependent on worker remittances as well as on development aid and tourism revenues from the oil-rich labor-poor countries. The non-oil-producing sectors, on the other hand, suffer from the ‘Dutch disease’ where the continuous flow of large oil revenues result in an appreciation of the real exchange rate, making it less competitive.

In large part due to the collapse in oil prices during the 1980s and 1990s the growth rates in the region experienced marked declines. Low growth rates failed to provide the rapidly expanding labor force with sufficient...
employment opportunities and led to a deterioration of living standards and a rise in poverty rates. As a result, growth performance and GDP per worker and total factor productivity growth rates in the MENA region since the early 1980s has been near zero and negative, closer to Latin America and lagging far behind East Asia. Furthermore, despite substantial improvements since the gaining of political independence, the region lags behind both East Asia and Latin America in the UN Human Development Index (that is, in adult illiteracy rate, life expectancy at birth, and so on) (Bosworth and Collins, 2003; UNDP, 2002).

Nevertheless, there is considerable heterogeneity in performance across countries within the region. While the per capita incomes of the oil-producing countries declined at an average annual rate of –0.79 percent between 1980 and 2000, those in the non-oil-producing countries increased by around 2 percent over the same period. Looking at the oil boom years we get a similar picture where non-oil-producing countries grew almost twice as much as oil-producing ones. What are the reasons behind this diverse yet overall poor performance? We will turn to this question in the coming sections.

State, institutions and development

Despite the diversity in state structures, resource endowments and economic performance a characteristic shared virtually by all countries in the region (including Turkey and Israel) is the dominant role played by the public sector in the development process (Richards and Waterbury, 1996).

The majority of the states also experimented with the usual sequencing in terms of development models, going through an ISI period (accompanied by land reform) which would be disrupted (usually following an economic and/or political crisis, such as Egypt in 1967, 1974; Turkey in 1980) and be replaced by an outward-oriented development model where the role of the state is attempted to be downsized through domestic and external liberalization programs and public sector restructuring.

The first ISI attempts in the region were launched by Turkey in its first five-year plan of 1934, emphasizing the establishment of state enterprises in textiles, primary commodities and minerals, ceramic and glass, paper, chemicals and cement, and iron and steel as well as state banks for financing these enterprises. The Turkish model would provide a guide for the other MENA countries and be replicated throughout the region (Richards and Waterbury, 1996). In the oil-rich labor-abundant countries, the availability of massive oil rents, or what are termed ‘soft-budget constraints’, allowed the large countries to launch ambitious ISI projects. This led to a proliferation of capital-intensive (often turnkey) industries producing protected intermediate and final products for the domestic market. However, tariff
protection and credit access was often granted wholesale and no technological upgrading or other performance measures were required as was the case in East Asian countries.

During this period, despite the presence of a generally hostile attitude by the state bureaucracy, the private sector in many countries benefited largely from intermediate products supplied by the state enterprises at discounted prices or from other subsidies in the form of cheap credits or foreign exchange. One of the characteristics of the ISI era was that during this period the accumulation process was highly dependent on politics rather than markets. The political and economic environment thus created opportunities for wide-ranging rent-seeking behavior within the business community, as businesses competed for the special set of incentives (subsidized credit and foreign exchange, import licenses, and so on) provided by the state. The pre-liberalization ISI era thus gave rise to a narrow distributional coalition between the state bureaucracy and the business community.

Regarding financial and banking sector development, the region continues to suffer from the lack of an efficient banking system with long-term credit availability for private investment projects (excluding Israel). Furthermore, the use of state banks for political rent distribution in the form of distributing cheap credits on non-economic grounds manifests itself in the accounts of large ‘duty losses’ of these banks (OECD, 2001; Mitchell, 1999, pp. 29–30). In the case of capital market deepening, the money markets are mostly dominated by short-term government securities, while capital markets in private securities remain underdeveloped throughout the region.

Moreover, the tax system of the countries in the region has been characterized by the inability or unwillingness of policy-makers to implement an efficient and fair tax scheme which, in addition to resulting in a narrow tax base and high tax evasion, further contributes to the unequal distribution of the tax burden on low-income groups. As a result in virtually all countries in the region the business environment for private firms with no political ties with the state bureaucracy is not encouraging. Informal tax traps are common and lack of clear-cut and consistent tax laws result in incorrect assessments of tax obligations (for example EIU, 2004).

On the other hand, following independence the survival of these mostly autocratic regimes required distribution of economic rents to a wider group of supporters. Indeed, what is common in the region is that the continuous flow of revenues (mostly from oil rents) has helped postpone economic and political reforms, since the region as a whole (excluding Turkey and to some extent Israel) could manage to avoid the balance-of-payments crises that other developing countries faced at the last stage of their ISI experience.
Natural resources and development
After gaining full independence and national sovereignty in the post-colonial era, oil-producing MENA countries (which account for about three-quarters of the world’s proven crude oil reserves and 35 percent of global oil production) reaped the benefits of increasing oil revenues. Once national governments secured control over their oil production and pricing, oil revenues started to flow in cascades. For example, in the case of Saudi Arabia, crude oil revenues increased from around $10.4 million in 1946 to $104.2 billion in 1980 (Owen and Pamuk, 1998, p. 210).

However, this development has led to a Dutch disease with destructive effects on non-oil industrial sectors while retarding economic diversification and growth (Sachs and Warner, 2001). Appreciating domestic currency resulting from large oil-related foreign exchange inflows created an unsuitable environment for the development of domestic industries by making non-oil exports less competitive. Furthermore, spending on massive construction projects further turned the terms of trade against manufacturing. Another major reason for the misalignment is the pegged or fixed exchange rate regimes adopted in the region as a whole (excluding Turkey) (World Bank, 2003, p. 110). Nabli and Veganzones-Varoudakis (2002) argued that MENA countries experienced overvaluation of more than 20 percent a year in their real exchange rates from the mid-1970s to 1999. They also suggest that the exchange rate policy explains losses in competitiveness and in manufactured exports in the region as a whole where real exchange rate overvaluation has decreased the ratio of manufactured goods to GDP by 18 percent a year.

On the other hand, in contrast to the abundance of oil resources, with 5 percent of the world’s population, the MENA countries have only 1 percent of the world’s renewable fresh water. According to the World Bank the region’s per capita supply ‘stands at only one-third of its 1960 level, and water availability is expected to halve over the next 25 years if the present pattern of use continues’ (World Bank, 2004, p. 4). Water shortage means that in addition to the strain of providing clean water to a rapidly increasing population, the countries are also increasingly dependent on food imports. Moreover, conflicts over water distribution and sharing have been exacerbated due to a lack of adequate regional conflict resolution mechanisms.

Trade and development
Historically, the MENA region was a thriving center of trade both originating within the region and as a crossroads for trade routes between Europe, East Asia and southern Africa. However, the shift in the balance of power between the MENA and Europe over the seventeenth and eighteenth centuries and Europe’s subsequent industrialization instituted a new
pattern of trade, that of manufactures exports from Europe in return for
primary products and raw materials, and led to the subsequent decline and
decimation of existing manufactures and crafts production that the region
had enjoyed. During this period, any attempt by the region to industrialize
was forcefully prevented (most notably by Britain and France), such as the
This not only significantly shifted the pattern of production and trade, but
also served to disrupt intra-regional trade in agriculture and manufactured
goods, which had expanded under the consolidation of the region under
the Ottoman rule (Owen, 1993).

More recently, the fortunes of the region (excluding Turkey and Israel)
in the post-World War II period have been dependent on two types of trade.
The first is the inter-regional export of fuels and other primary products
(for example natural gas, iron phosphates), which during oil price booms
reached almost 50 percent of GDP in the oil-exporting countries, ‘with 35
to 40 percent of GDP “spillover” effects for the region as a whole’ (Shafik,
1998). The spillover was mainly due to the second, intra-regional trade in
labor, which has been a vehicle of transmitting the rents throughout the
region, reaching as high as 20 percent of GDP for some countries such as
Jordan and around 5–10 percent of GDP or higher for several countries
such as Egypt, Syria, Morocco, Jordan and Tunisia (Galal, 2000).

Both the oil-rich labor-poor and labor-abundant countries have fuel
exports that made up around 85 percent of their total exports as of 2000.
The oil-poor small states, on the other hand, have successfully diversified
their exports whereby manufactures made up around 75 percent of total
exports in 2000. Furthermore, Israel and Turkey have highly diversified
exports compared to the rest of the region, with Israel emerging as a world
leader in high-technology exports.

Although regional integration or ‘Pan-Arab unity’ has been a prominent
topic in the region, intra-regional trade in the MENA has never exceeded
8 percent of exports and is the lowest of any region in the world (Galal,
2000). The lack of diversified production structures has undoubtedly been
a hindrance for intra-regional trade: gravity model estimations reveal that
MENA countries trade about a third less than otherwise identical countries
(Rose, 2002). 5

In order to reverse this trend, the Arab Free Trade Agreement has been
established, with 18 countries signing the agreement in 1997. Furthermore,
as a sign of expanding regionalism, in addition to intra-Arab treaties,
several MENA countries have signed bilateral association agreements with
the EU, with others to follow (Fawzy, 2003).

The uncertainty in gains from regional integration is a risk for regimes
that are wary of engaging in potentially destabilizing reforms. Moreover,
the availability of windfall rents has allowed the regimes to appease domestic constituencies, and prevented the formation of coalitions pressuring integration or other deep structural transformations (Carkoglu et al., 1998).

**Labor markets and human capital in the MENA**

Regarding demographical challenges, the MENA region has the second-highest population growth rate in the world after sub-Saharan Africa, exacerbating labor market problems. Although the rate has been decreasing in recent years, average annual growth in the labor force is still expected to be 3.4 percent a year in 2000–2010, which is twice that of other developing countries, with adverse effects on per capita incomes (World Bank, 2003, p. 19).

The presence of a disproportionately high share of the young in the population (under-30-year olds constitute almost two-thirds of the population on average), low growth rates and lack of skill development has resulted in high unemployment throughout the region.

It is estimated that 16 MENA countries that represent 60 percent of the regional population need to provide 47 million new jobs between 2002 and 2012 just to keep up with the increasing labor supply (Keller and Nabli, 2002). As a result, the unemployment (and underemployment) rate is quite high in the region, and despite underestimated official figures stands at around 15 percent in the Arab countries (UNDP, 2002). Furthermore, since 1981 the labor force has grown faster than population growth and can be expected to increase further with increasing female participation rates.

In the case of oil-poor countries, another problem lies in their dependence for job growth on the oil-producing countries. As of 1997, for example, foreigners in Kuwait held 99 percent of private sector and 42 percent of public sector jobs. The ratio for the foreign to domestic workforce is 90 percent in the UAE, 83 percent in Qatar and 69 percent in Saudi Arabia (McMurray, 1999, p. 19).

Regarding human capital, following political independence the MENA countries faced a daunting task to educate their population, with adult illiteracy of 70 percent in Syria and 85 percent in Algeria, Iraq and Libya around independence (El-Ghonemy, 1998). The colonial powers had established parallel systems of education and the systemic discrimination in education left the majority of the population, especially in rural areas, with dilapidated and low-quality public schools while the expatriates, the urban elites and sectors friendly to colonial powers enjoyed high-quality educational establishments (El-Ghonemy, 1998).

Since the post-independence period MENA countries have invested a high proportion of their GDP towards education and health, and have
made remarkable gains on both counts. Average illiteracy rate dropped from 60 percent in 1980 to about 43 percent in the mid-1990s, while enrollment at all levels went up from 31 million to 56 million during the same period (UNDP, 2003). However there is still widespread illiteracy among youth and adults and even higher rates among women and the rural poor. A side-effect of the industrialization attempts by the MENA countries was an allocation of resources towards secondary and higher education, which typically have lower social rates of return than primary education. The result has been the oddity of unemployed highly educated workers, while having large numbers of illiterate adults and youth (Richards and Waterbury, 1996).

**International conflicts and socio-political instability**

The region has been plagued with ongoing conflicts since the eighteenth century, starting with the decline and the following collapse of the Ottoman Empire and the erection of colonial regimes. Having borders drawn by the colonial powers based on politics rather than historical, cultural or ethnic backgrounds or social consensus led to subsequent ethnic and religious civil conflicts (for a detailed list of these conflicts, see for example Elbadawi, 2005, pp. 306–7).

In addition, since 1948 the Middle East has witnessed: four wars between Israel and several of its Arab neighbors; three wars with Western countries; the full occupation of Iraq and Palestine and the partial occupation of Egypt, Lebanon and Syria; extended periods of economic sanctions on Syria, Iraq, Sudan and Libya; and several *coup d’état* instigated from within and outside the region. The Iran–Iraq war alone left around 1 million dead and 2.5 million refugees, with an estimated cost of $200 billion. These conflicts have had a direct impact on state structure and overall trajectory of development.

The artificial mapping of the region with sovereign borders overlapping with different ethnic and religious groups further fed into the authoritarian state structure, thanks to the excuse that the survival of the unity of the country is dependent on the suppression of popular demands by different groups.

The majority of publications on socio-political risk and the investment and growth relationship find a negative correlation between these variables. Venieris and Gupta (1986), Alesina and Perotti (1996) and others find an inverse relationship between political instability and growth or investment, or savings rate. In addition, Asteriou and Price (2001) found that socio-political instability not only negatively affects the growth rate but also increases its volatility. Similarly, Rodrik finds a significant negative relationship between external shocks and growth in countries where there
are latent domestic social conflicts and poor conflict management institutions, as in the case of the MENA (Rodrik, 1998). The idea behind the above research is that socio-political unrest and instability disrupts market activities and investment decisions by increasing uncertainty and risk while directing limited resources to non-productive security-related expenditure.

As a result (or on the pretext) of non-stop civil or military conflicts the existing regimes have devoted a sizable portion of their budgets to military spending. The average military expenditure to GDP ratio in the region is 6.6 between 1990 and 2004 with a maximum of 21.8 in Kuwait and minimum of 1.8 in Tunisia. Comparatively, the averages were 1.4, 0.5 and 2.5 in Argentina, Mexico, and Malaysia for the same period (SIPRI, 2005). Such military spending creates a substantial potential for peace dividend in the region. However, for the peace dividend to materialize, the peace must be ‘real and durable, and perceived as such’ (Fischer et al., 1993).

**Economic reform in the MENA**

Despite the presence of a general consensus among policy-makers and economists on the need for reform, the question regarding which path to follow remains unanswered. Several countries in the region have embarked on structural adjustment programs (SAPs) under the guidance of the International Monetary Fund (IMF) and the World Bank. The reforms included standard policy packages by the twin institutions such as fiscal reform (introducing value-added taxes – VATs; eliminating state subsidies; increasing transparency in public expenditures), liberalizing trade and capital accounts, and shifting to more flexible foreign exchange regimes. Despite differences, the countries that have enjoyed higher rates of growth since the early 1990s have been those that implemented reform programs (that is, Egypt, Jordan, Morocco, Tunisia) (Hirata et al., 2004).

Nevertheless, it is difficult to make a generalized statement on the success of the SAPs in the MENA due to credits and debt reliefs extended to certain reforming countries based on political considerations during the adjustment period (for example Egypt, for its support to the first Gulf War) (Gray, 1998).

Despite the implementation of comprehensive trade and financial liberalization programs including tariff reductions, privatization, tax breaks and eased restrictions on foreign ownership, as well as establishment of free trade zones and other incentives to encourage foreign direct investment (FDI), capital flows to the region remain minimal. The region’s share of FDI fell to 0.7 percent in 2000 from 2.5 percent in 1980 (Hirata et al., 2004). In addition, most of the capital flowing into the region appears to be in short-term funds. In the case of Turkey, capital account liberalization has
exposed it to the uncertainties and instabilities associated with short-term capital flows, which have demonstrated themselves in three major crises in 1994, 2000 and 2001. Furthermore, as shown in the case of banking sector crises and subsequent cost of bank defaults resulting from endemic corruption as well as rent-seeking promoting IMF engineered policies (such as 100 percent state insurance on private bank deposits in Turkey), the countries in the region need a major restructuring in their financial systems.

On the other hand, income inequality and poverty rates have increased since the implementation of reform policies (Ali and Elbadawi, 2002; Fergany, 1998). The region had previously enjoyed the lowest incidence of poverty and income inequality of any region in the developing world (Adams and Page, 2003). As a result, only 5.6 percent of the population in the region lived on a less than the $1 a day benchmark compared with 14.7 percent in East Asia and 28.8 percent in Latin America (Shafik, 1995). Adams and Page (2003) pointed out two statistically and economically significant reasons for this: international remittances, and public sector employment and the welfare state. The policy of public sector employment and subsidized public services and pricing to generate popular support for the survival of the political regimes appears to be the common element in the region. However, the slow-down in growth as well as neoliberal reforms, which have scaled back the role of the state, have reversed the trend of lowered inequality (Ali and Elbadawi, 2002; Fergany, 1998).

**Conclusion and policy suggestions**

Contrary to the view that finds ‘little reason for gradualism’, as in World Bank (2003, p. 7), there may arise significant socio-political costs to a ‘big bang’ approach in the region. In a majority of MENA countries, certain sectors and groups of people (that is, peasantry, civil servants, organized labor) will stand as absolute losers from the reform programs, at least in the short run. Furthermore, the economic and political failures of the past have created an unstable environment pregnant with socio-political fault-lines, which are further exposed by slow growth rates, increasing unemployment, and increasing income inequality and poverty among different income groups and different regions.

The worsening economic performance has radicalized the divide between urban and rural, secular and Islamist, and ethnic identity groups, and these politicized fault-lines have, in turn, been accompanied by increasing authoritarian governance in the region (Lubeck, 1998, p. 299).

As a result, increasing hegemony of neoliberal policies along with economic liberalization and deregulation of markets may have the opposite effect on political liberalization and consolidation of democracy in the
region, by further deepening such divisions through increasing economic insecurity and social dislocation among the public.

The experiences of MENA countries suggest that historically determined institutional characteristics and the political environment of a country are of crucial importance in determining both the nature of the adjustment process and subsequent economic performance. Given that the market-led, outward-oriented reform programs have not produced the anticipated results so far, there is a continuing debate among economists about the underlying reasons. This chapter follows the line that developing countries share common structural problems in their institutional settings, and that policies that are designed to liberalize their economy (and political and civil life) may also generate serious instabilities without necessarily eliminating the previously existing ones. The existence of strong state hegemony in the form of military, legislative and economic institutions with a lack of clear-cut lines between private and public spheres resulted in a lack of democratic accountability and transparency during the design and implementation of reform packages. In addition, previously formed rent-seeking coalitions have prevented the implementation of a comprehensive reform program designed according to the needs of the countries in the region.

As a result, instead of removing the state, the neoliberal reform programs helped the state become instrumental in distributing rents to a new group of rentiers that make their living from financial rents (Mitchell, 1999, p. 30; Yeldan, 2001; Demir, 2004, 2005). Hence the state(s) ‘now subsidizes financiers instead of factories, speculators instead of schools’ (Mitchell, 1999, p. 31).

The recent experience of MENA countries suggests the state and the institutional infrastructure need to be reformed before embarking on reform programs that may undermine the legitimacy of the state structure and lead to socio-political instabilities. In other words, sequencing of reforms is a must both for the sustainability of the reforms and for their further deepening. The future of the region in terms of economic and political outcomes depends on the following:

1. Institutional reform in the form of judicial, legal, administrative and prudential regulation including the rule of law should be established. In addition, rent-seeking groups need to be controlled if any economic reform is to be successful.
2. Providing social safety nets for the disadvantaged and the losers during transition.
3. Political liberalization including reforming the state and making it democratically accountable rather than populist.
4. Privatization of the ownership or the management of State Economic Enterprises (SEEs) and public banks are needed to avoid corruption, rent-seeking and subsidized credit distribution to a few wealth groups based on political considerations.

5. Boom–bust cycles need to be stabilized especially in the case of resource-rich countries, which are dependent on the changes in the oil markets. As also argued by the World Bank (2003, p. 10), the countries need to: establish rules that shield fiscal spending from fluctuations in oil revenues; create deposit accounts for oil revenues to be set aside for future generations; and avoid misalignment in exchange rates. This may have solved the ‘resource curse problem’ in the oil-rich countries by offering an alternative to investing revenues in non-profitable and non-competitive domestic investment projects. This may also pave the way to avoid currency appreciation and support competitive domestic sectors. Also, this may provide an outlet for intergenerational resource distribution for future generations, given the limited supply of oil reserves.

Notes
1. Israel is considered an industrialized country and in that sense is in a separate category to other MENA countries.

2. Our classification here is slightly different from Richards and Waterbury and is meant to highlight export structure, intra-regional labor migration, patterns of industrialization and dependence on oil revenues. For example, the proven oil reserves of Syria are negligible; however it was highly dependent on oil exports for revenues during the ISI period and until today as fuel exports made up 76 percent of total exports in year 2000.

3. Some recent scholars attempted to explain the lagged performance in the region with the legacy of Islam. Kuran (2004) for example, blames the Islamic waqf or trusts that locked capital into a dysfunctional institution, Islamic inheritance law which dispersed inheritance among multiple heirs, and the individualism of Islamic law as preventing capital accumulation à la Europe. However, the lagging performance of the region vis-à-vis Europe came far too recently on a historical scale to be pinned on the influence of religious (or cultural) institutions. Moreover, as Inalcik (1969) emphasized, Islamic society and law ‘shaped themselves from the very first in accordance with the ideas and aims of a rising merchant class’ (Inalcik, 1969, p. 101). Finally, any attempt to explain the decline in economic performance of the region after the eighteenth century with the religious and cultural factors or institutions should also be able to explain how the same institutions could create the opposite results prior to that date.

4. In Turkey, unlike others, the state assumed a direct role in creating and supporting the development of a national business class; Bugra (1994).

5. The colonial legacy on regionalism warrants greater attention. As Ventura-Dias (1989) argues, colonial powers promoted intra-regional trade in Asian countries which allowed ‘permanent marketing channels to be established’ in contrast with both MENA and Latin America, where colonial intervention disrupted intra-regional trade.

6. The only country in the region with a definite plan to limit the harmful effects of the oil curse is Kuwait, which preferred to utilize its oil revenues on investments abroad (Owen and Pamuk, 1998, p. 216).
References


Economic Intelligence Unit (EIU) (2004), Country Finance Egypt, London: Economic Intelligence Unit.


Many economists believe that China today is largely a market economy. Yingyi Qian (2003), a leading economist on China, observes: ‘In the last 22 years of the 20th century, China transformed itself from a poor, centrally planned economy to a lower-middle-income, emerging market economy.’ Barry Naughton, another leading economist on China, echoes this view in his recent textbook (Naughton, 2007). This chapter counters this claim with evidence that shows that China today is far from a market economy – defined as one predominantly based on private ownership.

This chapter first reviews China’s growth experience since 1978. I will note briefly the enormous gains China has made, but the emphasis here is to highlight aspects of Chinese performance that are less well known, such as the increasingly investment-driven growth, the slowdown of productivity growth, and some notable microeconomic inefficiencies (as compared with, say, India). Social performance also deteriorated in the 1990s.

I then turn to ask the question, ‘Just how capitalist is China?’ The answer, surprisingly, after nearly 30 years of reforms, is not at all clear. There is evidence that China today is a commanding-heights economy similar to some of the most statist economies of the 1970s. I use data on fixed-asset investments by what I call the registered domestic private sector to show that in the 1980s the domestic private sector developed vigorously but in the 1990s the pace of liberalization and denationalization slowed down. The final section concludes with some broad implications of this analysis.

**China’s growth experience**

China’s economic success is indisputable. Its gross domestic product (GDP) growth has led the world in the growth table. Between 1978 and 2004, according to the World Bank, real GDP growth per annum averaged 9.73 percent, the fastest in the world. Poverty has fallen dramatically since the onset of the reforms in the late 1970s. By one estimate, the overall poverty level in China – measured as the percentage share of the population living under the poverty line – declined from 53 percent in 1981 to only 7.97 percent in 2001 (Ravallion and Chen, 2004). In the 1970s, the Chinese economy was plagued by shortages; today, it is often blamed as the source of the worldwide deflation as the economy has become a powerful export engine.
I do not repeat these well-known success indicators. Rather, this is a focused treatment of those topics that have received less attention in the academic literature. These include the non-economic foundation of China’s rapid growth, the heterogeneous growth record during the reform era, and the divergence between the GDP indicators and other development indicators. The purpose of the discussion is to highlight the complexities involved in trying to understand China’s growth experience.

The social and political foundations of growth
Economists assign a heavy weight to the reforms when explaining China’s economic success. This is not wrong, but it is important to point out that cross-country evidence on the linkage between policy reforms and economic growth is not nearly as clear-cut. Yet, it is uncontroversial that China’s impressive growth clearly followed the policy changes introduced in the late 1970s. Even if we acknowledge the importance of the reforms, the appropriate way to frame the discussion is to ask whether the policy changes introduced in the late 1970s interacted with some of the favorable pre-existing conditions in the country and whether it was an interaction effect, rather than just the policy changes alone, that spurred China’s growth.

In this context, it is useful to compare China with India. When examining the details of India’s growth record, Hausmann et al. (2004) correctly note that India’s growth rate began to pick up relative to the historic benchmark in the 1980s and that this acceleration of growth coincided with a number of modest policy changes. In addition, they note that China’s growth also followed seemingly modest policy changes in the 1980s. If we accept this characterization of the respective policy changes in China and India, we are still left to explain why China’s growth was so much stronger than India’s growth, as indicated in Table 68.1. For the period on which Hausmann et al. focus, that is, from 1978 to 1990, the average annual GDP growth in China was 9.28 percent, almost twice India’s growth rate of 5.01 percent.

This differential suggests the importance of identifying some initial differences between China and India that might have contributed to their divergent growth performances in the 1980s. The most important initial difference in favor of China has to do with the social conditions. For complex reasons, during the orthodox socialist period (1949–78) China invested heavily in health and education, especially in the rural areas. For example, as early as 1965 the life expectancy of Chinese women at birth was 55 years, compared with only 44 years in the case of Indian women. In the 1980s, infant mortality in China was substantially lower than infant mortality in India. (In the mid-1980s, China’s infant mortality rate was 54 per
1000, compared with 122 per 1000 in India.) Furthermore, China’s primary education enrollment ratio was far higher than that of India as early as 1975. China also had a more equal initial distribution of income as compared with that in India.4

Panel (A) of Table 68.1 gives the average annual growth percentages of real GDP for China and India. Panel (B) gives the average annual growth percentages of real GDP per capita. Comparing the data in Panel (A) with the data in Panel (B) reveals an interesting pattern: Although China’s aggregate GDP performance was better than that of India, its per capita GDP performance was even better. For the period from 1978 to 2004 as a whole, China’s GDP growth rate was 1.81 times that of India (9.37 percent compared with 5.37 percent). However, the ratio of Chinese GDP per capita growth to that of India was 2.51 for this period (8.44 percent compared with 3.37 percent), substantially exceeding the GDP growth ratio of 1.81 of the two countries. It is important to understand the source of this differential.

One hypothesis centers on population control. China has a draconian population control program, but for political reasons India cannot replicate this aspect of China’s ‘development strategy’. Because both countries have a chronic surplus of labor, it is plausible to argue that the higher GDP growth per capita in China – far in excess of its aggregate GDP growth – is

<table>
<thead>
<tr>
<th>Year Period</th>
<th>China</th>
<th>India</th>
<th>Ratio of China to India</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978–2004</td>
<td>9.73</td>
<td>5.37</td>
<td>1.81</td>
</tr>
<tr>
<td>1978–90</td>
<td>9.28</td>
<td>5.01</td>
<td>1.85</td>
</tr>
<tr>
<td>1991–97</td>
<td>11.53</td>
<td>5.43</td>
<td>2.12</td>
</tr>
<tr>
<td>1998–2004</td>
<td>8.76</td>
<td>5.97</td>
<td>1.47</td>
</tr>
</tbody>
</table>

Table 68.1 GDP growth and GDP per capita growth, China and India, 1978–2004 (%)

<table>
<thead>
<tr>
<th>Year Period</th>
<th>China</th>
<th>India</th>
<th>Ratio of China to India</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978–2004</td>
<td>8.44</td>
<td>3.37</td>
<td>2.51</td>
</tr>
<tr>
<td>1978–90</td>
<td>7.73</td>
<td>2.78</td>
<td>2.78</td>
</tr>
<tr>
<td>1991–97</td>
<td>10.26</td>
<td>3.53</td>
<td>2.91</td>
</tr>
<tr>
<td>1998–2004</td>
<td>7.95</td>
<td>4.29</td>
<td>1.85</td>
</tr>
</tbody>
</table>

Notes: GDP data are calculated on an exchange rate basis.

Source: Data are downloaded from World Development Indicators at http://devdata.worldbank.org/dataonline/.
partially due to the coercive capacity of the state. Many economists neglect this fact when they tout China’s supposedly superior record of economic reforms; in reality it is the political, not the economic, management that matters.

**Heterogeneous growth experience**

It is interesting to explore the substantial heterogeneity in China’s growth experience during the long time-span since reforms (1978). I deal with three respects: the importance of investments for China’s recent growth performance, productivity development and social performance.

Let us revisit Table 68.1, which breaks down the reform era into three sub-periods. GDP growth was fastest pace during the 1991–97 period, averaging 11.5 percent per year. The period from 1978 to 1990 came second, at an annual 9.28 percent on average. The most recent period from 1998 to 2004 turned out to be the least impressive, averaging 8.76 percent per year.

These fluctuations in the GDP growth coincided with huge changes in the investment levels. In the 1980s, the gross fixed capital formation as a percentage ratio to GDP averaged around 30 percent. During the recessionary years of 1989 and 1990, this ratio declined to 25 percent and then it surged to 35 percent in 1993. Starting in 1997, the ratio increased sharply, reaching 40 percent in 2004 and then 48 percent in 2005, another new high.5

So there was a change in the drivers of growth over time. Of the three sub-periods presented in Table 68.1, China’s GDP performance is the least impressive during the 1998–2004 period, but this is also a period when China was investing at its highest level. Thus, at 8.76 percent per year, although China was still leading the world in GDP growth, it was achieving this performance at a substantially higher level of investments than when it was growing faster in the 1980s and the early 1990s.

Is China’s latest growth spurt as sustainable as the one it experienced in the 1980s and the early 1990s? The answer depends on the productivity of the latest investment surge. If the high levels of investments are accompanied by or directly lead to technological progress, then an investment-driven growth pattern can be sustainable. Research on Western economies has shown a sustained, long-term positive correlation between productivity growth and capital deepening (Wolf, 1991).

The evidence, however, suggests that the latest investment growth did not lead to improved productivity. An exhaustive survey of the various studies on China’s total factor productivity (TFP) reveals one consistent pattern: TFP performance declined beginning in the late 1990s relative to earlier periods. The key findings are summarized in Table 68.2. These studies differ on the TFP estimates but they converge on trend developments: TFP growth during the last period, that is, in the late 1990s or early 2000s, was
Table 68.2  Estimates of the annual TFP growth in the Chinese economy (%)  

<table>
<thead>
<tr>
<th>Sources of estimates</th>
<th>Level of data</th>
<th>Reference periods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1980s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>First half</td>
</tr>
</tbody>
</table>

Note: a  Kuijs and Wang (2005) estimate TFP growth to be 3 percent during the 1978–2004 period and 2.7 percent during the 1993–2004 period. I have calculated TFP growth to be 3.74 percent during the 1978–93 period on the basis of their estimates.

Sources: See the table for citation information.
considerably more moderate than TFP growth in the 1980s and the early 1990s. For example, Zheng and Hu (2004) report that TFP grew annually by 3.26 percent between 1978 and 1995, but during the 1995–2001 period TFP growth virtually disappeared (0.32 percent). Focusing only on Chinese industry, Ren and Sun (2006) report a reduction in TFP growth of a similar magnitude.6

Although during the reform era as a whole China made substantial progress in eradicating poverty, progress was uneven across both space and time. The largest gains in poverty reduction occurred in the first five years of the 1980s. According to Ravallion and Chen (2004), between 1980 and 1985 poverty declined drastically. In the rural areas, the incidence of poverty – measured by the headcount of those living below the poverty line – declined from 75.7 percent to only 22.7 percent. Income distribution improved in the early 1980s, as indicated by a reduction of the Gini coefficient.

Since then, the pace of poverty reduction has been considerably more measured and there have been episodic setbacks. In 1998 rural poverty was at 11.6 percent, and in 2000 it rose to 13 percent and in 2001 to 12.5 percent. Income distribution also deteriorated. In 2001 China had a Gini coefficient of 39.45, compared with 27.98 in 1980. This remarkable speed of reversal, according to two experts, is ‘almost unheard of in the developing world’ (Khan and Riskin, 2001).

The retrogressions in the late 1990s warrant special scrutiny. First, productivity indicators for this period deteriorated as well, so there was no economic and social trade-off. Second, according to the data provided by Ravallion and Chen (2004), the poverty level increased in the late 1980s as well. But the circumstances of the late 1980s differed substantially from those of the late 1990s. The surge in poverty in the late 1980s can be easily explained by the macroeconomic shocks – the Chinese economy went into a severe contraction in the late 1980s. It is more difficult to explain the increase in poverty during the economic boom period of the late 1990s.

The adverse social developments in the late 1990s are a sign that Chinese growth may have acquired an inherent anti-poor bias. The magnitude of the effect is substantial. Consider the increase in poverty incidence from 11.4 percent in 1999 to 12.96 percent in 2000. In percentage terms this increase may not be striking, but because China has a huge rural population, such a seemingly small rise in poverty incidence in fact corresponds to 11.24 million rural residents newly thrust into poverty, equivalent to the entire population of Greece. Other – far less known – indicators are also telling. For example, the World Bank has documented that during the reform era China has underperformed – both against other countries such as India and against its own economic potentials – in terms of reducing
infant mortality. The World Bank has also reported that China is one of only seven countries in the world to have a higher infant mortality rate among girls than among boys.\(^7\)

The most recent evidence is even more alarming. The World Bank has just reported that the income of China’s poorest 10 percent of the population declined by 2.4 percent between 2001 and 2003 (McGregor, 2006). This is the first documented evidence that a large number of Chinese people – about 130 million people – has actually experienced an absolute reduction in their living standards. The issue is no longer one of ‘relative deprivation’, about which economists tend not to be as concerned, but one of ‘absolute deprivation’. If this trend continues, it will have serious implications for the prospects for China’s growth and political stability.

**Creating output vis-à-vis creating value**

Bai et al. (1997) draw the distinction between the technical capabilities on the part of the SOEs to produce outputs and their economic capabilities to create value. State-owned enterprises (SOEs) can produce a massive quantity of a product, but the product itself may lack demand. In this case, there is a divergence between technical and economic measures of efficiency.

We can apply the same reasoning to the Chinese economy as a whole. It is well known that India’s GDP growth has lagged that of China for much of the last two decades of the 1980s and 1990s. But a little-known fact is that the Indian economy is able to create more value and wealth for a given unit of GDP than is the Chinese economy. It is intriguing to note that India has a higher manufacturing value added per worker than China. The value added per worker in manufacturing was 2885 dollars per year during the 1995–99 period for China, but 3118 dollars per year during the same period for India. In fact, the value added in manufacturing declined between the mid-1980s and the mid-1990s in the case of China but increased in the case of India.\(^8\)

An International Monetary Fund (IMF) paper shows that India ‘overly’ specialized in highly skilled industries at the expense of low-skilled industries.\(^9\) But even if the two countries differed in their initial starting points, it still begs the question why the manufacturing value added declined over time in China. The aforementioned IMF paper shows that between 1981 and 1996 China’s share of output in skill-intensive industries was not only lower than that of India, but it was also declining over time. The declining value added in Chinese production suggests that Chinese firms were not climbing up the value chain during a period of massive boom.

**Just how capitalist is China?**

The conceptual framework explaining China’s transition to a market economy is gradualism – the idea that the reforms are an endogenous
process whereby ‘unhooking a single key connection can cause the entire fabric to unravel’ (Naughton, 1996, p. 311). A critical empirical benchmark is the size of the private sector. China started out with a very small private sector, but due to the increasingly supportive policy environment, the private sector grew and overtook the previously dominant state sector.

I do not question the logic of a gradualist approach nor the empirical basis for applying a gradualist approach to China in the 1980s. The findings of a deterioration of TFP performance, worsening social performance and the increasingly investment-driven nature of the high growth since the mid-1990s raise a question whether gradualism is still an accurate empirical characterization of China today. Apart from the fact that it sheds light on the sustainability of China’s growth, the TFP performance can be a proxy measure of the reforms. In this regard, the across-the-board finding that TFP growth began to deteriorate in the late 1990s is significant. Did the reforms stall in the 1990s?

I examine a key benchmark in the gradualist interpretation of China’s reforms – private sector development. I ask, ‘After nearly 30 years of transition, just how capitalist is the Chinese economy?’ Surprisingly, the answer is not straightforward. I distinguish between an output-based measure and an input-based measure of the size of the private sector and I show that there is a huge discrepancy between these two measures. First, while the output-based measure shows the size of China’s private sector to be quite large, the input-based measure gives rise to a far smaller estimate. Second, while the output-based measure shows a continuously rising private sector, the input-based measure indicates severe setbacks – and even retrogressions – in private sector development in the 1990s. The latter finding directly contradicts the predictions of the gradualist framework.

In the following paragraphs, I will first describe the conceptual and methodological problems underlying many of the output-based measures of the Chinese private sector. I will then present detailed statistics based on a critical input – fixed-asset investment – on the evolution of the registered private sector in China. This measure shows a robust development of the private sector in the 1980s and a sharp contraction in the 1990s.

The registered private sector firms, either individual businesses (getihu) or privately operated enterprises (siying qiye), refer to newly established private businesses registered as such with the government. For sure, this is a narrow and conservative measure of the private sector but, as I will explain, an examination of the development of this sector provides an important insight into the course of the economic reforms in the 1980s and 1990s.
Output-based measures
The most frequent measure of private sector development used by economists is the share of the private sector in production. By this measure, China’s private sector has made huge strides. For example, Naughton (2007, p. 300) shows that the domestic private sector accounted for 19 percent of industrial output value in 1996, up from zero in 1978. This type of data is often used to support the argument that the policy environment improved steadily for the domestic private sector.

An output-based measure incorporates two very different effects. One is the ‘policy effect’: the increase in the private sector share that resulted from a more favorable policy environment. But this measure also incorporates what might be called an ‘efficiency effect’. The private firms are more efficient than the SOEs and, therefore, even given a very narrow business space, they can outcompete the SOEs. This suggests that the ratio of the private to the state sector can rise without any improvement in the policy environment for private sector firms and with rising inefficiencies of the SOEs. Thus, this measure tells us as much about China’s policy environment as about the huge inefficiencies of the state sector, and we cannot distinguish which of the two dynamics is driving this ratio. As an illustration, in 1985 the industrial output of the private sector was about 2.9 percent that of the state sector; by 1997, this ratio had risen to 70.2 percent. Even if the argument is correct that the policy environment improved between 1985 and 1997, it would be highly misleading to conclude that the policy environment facing the private and state sectors converged at about 70.2 percent in 1997.

China economists use a broader measure than the registered private sector to showcase China’s transition success. The Chinese style of reforms has spawned a variety of hybrid and highly ambiguous ownership forms, such as SOEs with some private revenue rights, collective firms controlled at the local levels, and private–state joint-ownership firms. It is difficult to sort out who actually controls these myriad firms. The most careful analysis to decompose the ownership of Chinese firms has been carried out by two economists at the Organisation for Economic Co-operation and Development (OECD) (Dougherty and Herd, 2005). For their estimation, they use a detailed industrial firm data set maintained by the National Statistical Bureau (NSB).

One feature of the NSB dataset makes this estimation possible: the data set identifies the controlling shareholder of the firm. The OECD economists then use the shareholding structure information to generate estimates of the size of the Chinese private economy. In the NSB industrial data sets, the shareholders are classified among the following categories: (1) state (direct or indirect); (2) collective (that is, local governments); (3) individuals;
(4) domestic legal persons; and (5) foreign companies. The OECD economists make the crucial assumption that individuals, domestic legal persons and foreign companies comprise the private ownership. They conclude that the private economy accounted for 52.3 percent of industrial value added in 2003, compared with 27.9 percent in 1998.

The most serious problem with the OECD study is the assumption that domestic legal-person shareholders are private. Legal-person shareholding refers to cross-shareholding by firms. The term implies a type of keiretsu arrangement whereby firms own one another’s stocks. The difference with Japan, however, is that in China much of the legal-person share capital originates in the state sector, that is, SOEs establishing or holding significant equity stakes in other firms. The subsidiaries of the SOEs, on account of their final ownership, should be classified as SOEs. However, the OECD calculation classifies the entire output of these firms as ‘private’. As an example, the OECD methodology would classify Shanghai Automotive Industry Corporation (SAIC) as a private firm. SAIC is a quintessential SOE but its largest shareholder is a legal-person shareholder, an investment company of the Shanghai government.

But let us take at face value the claim that the Chinese private sector – inclusive of the foreign firms – is producing 52.3 percent of industrial production and let us place that claim in perspective. Here a comparison with India in the 1970s is revealing. India at that time was at the apex of its commanding heights after Indira Gandhi had nationalized all major banks, significantly expanded the scope of the ‘License Raj’, and created numerous barriers for the private sector. But even at the height of the ‘License Raj’, the importance of the Indian private sector far exceeded the level of the Chinese private sector in 2003. One estimate puts the share of private sector firms in total manufacturing GDP at 93 percent in the early 1960s and at 69 percent in 1983–84. The share of fixed-asset investments of the private sector was around 58 percent, a ratio that is several multiples of the Chinese level today. Thus, even a generous accounting of the current size of the Chinese private sector puts China roughly in the same league as some of the world’s most statist economies of the 1970s.

An input-based measure of the domestic private sector
I focus only on the registered domestic private sector firms. This definition covers newly established private businesses (exclusive of foreign firms). This is a narrow definition of the private sector and, by construction, it understates its true size. The usefulness of this measure is that it is a test of a dynamic claim in the gradualist framework – that China chose the politically and economically prudent path to encourage the entry and the organic growth of new private businesses over time. Comparing this measure across...
different time periods provides a way to assess the claim that the size of newly established private businesses has grown over time.

The input we focus on is capital allocated for fixed-asset investments. Fixed-asset investments, compared with similar activities in a market economy and with other economic activities in the Chinese economy, are heavily controlled by the government. All investment projects above a fairly low threshold require government scrutiny and approval. For this reason, fixed-asset investments are a superior indicator of the ownership policies of the state as compared with the output measure, because they are not subject to the confounding influences of efficiency differentials between the state and the private sectors.

Table 68.3 provides the fixed-asset investment data on four ownership types: SOEs, collective firms, the ‘individual economy’, and firms of other ownership. (The Chinese statistical system records fixed-asset investments by the registered private sector under the ‘individual economy’.) These four categories are exhaustive and mutually exclusive and thus their totals add up to 100 under Column (6).

One striking pattern emerges from this table. The investment share of the individual economy in the 1990s was actually smaller than it was in the 1980s. In the first six years of the 1980s, between 1980 and 1985, the ‘individual economy’ accounted for 20.7 percent of the total fixed-asset investments. This share climbed slightly in the second half of the 1980s, to an average of 21.9 percent between 1986 and 1990. In contrast, during the 1991–95 period, the ‘individual economy’ share declined to 13.2 percent and during the 1996–2000 period it was 13.9 percent. Panel (B) of Table 68.3 provides annual data for selected years. In 1993, the ‘individual economy’ only accounted for 11.9 percent of total fixed-asset investments, a full 10 percent drop from that prevailing in the second half of the 1980s (at 21.9 percent). After 1993, this ratio climbed slowly to 15 percent in 2002 and then fell back to 14.2 percent in 2004, just one percentage point higher than that at the very onset of the reforms in 1980.

Because our measure only covers fixed-asset investment activities in the registered private sector, and because of the possibility that the individual economy category may not fully record activities by the established private firms, we must return to the question of whether this measure is too narrow. In particular, the ‘other’ ownership category exploded from effectively zero in the second half of the 1980s to 11 percent in the 1991–95 period and then to 18.7 percent during the 1996–2000 period. If the ‘other’ ownership category encompasses mainly private sector firms, then the ‘true’ investment share of the private sector should be the sum of the individual economy and the ‘other’ ownership. That would put the investment share of the private sector during the 2001–2003 period at 43.2
percent (14.4 + 28.8), roughly double the level of the early 1980s at 20.7 percent.

But assigning all the firms in the ‘other’ ownership category to the private sector is a massive overstatement. The ‘other’ ownership category encompasses four types of firms: (1) joint-ownership firms; (2) shareholding firms; (3) foreign-invested enterprises (FIEs); and (4) unclassified firms. Shareholding firms and FIEs dominate this category of firms. These two types of firms accounted for 95.6 percent of the fixed-asset investments in this category during the 1996–2000 period and 97.2 percent during the 2001–03 period.

### Table 68.3 Ownership composition of fixed-asset investment (%)

<table>
<thead>
<tr>
<th>Year</th>
<th>(1) SOEs</th>
<th>(2) Collective firms</th>
<th>(3) Individual economy</th>
<th>(4) Of individual economy: urban only</th>
<th>(5) ‘Other’ ownership</th>
<th>(6) Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel (A): Period data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980–85</td>
<td>66.7</td>
<td>12.7</td>
<td>20.7</td>
<td>1.6</td>
<td>0.0*</td>
<td>100.0</td>
</tr>
<tr>
<td>1986–90</td>
<td>64.8</td>
<td>13.4</td>
<td>21.9</td>
<td>2.9</td>
<td>0.0*</td>
<td>100.0</td>
</tr>
<tr>
<td>1991–95</td>
<td>59.0</td>
<td>16.3</td>
<td>13.2</td>
<td>2.7</td>
<td>11.0</td>
<td>100.0</td>
</tr>
<tr>
<td>1996–2000**</td>
<td>52.5</td>
<td>15.0</td>
<td>13.9</td>
<td>4.1</td>
<td>18.7</td>
<td>100.0</td>
</tr>
<tr>
<td>2001–03</td>
<td>42.7</td>
<td>14.1</td>
<td>14.4</td>
<td>7.6</td>
<td>28.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Panel (B): Annual data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>61.5</td>
<td>17.9</td>
<td>11.9</td>
<td>2.7</td>
<td>8.8</td>
<td>100.0</td>
</tr>
<tr>
<td>1997</td>
<td>52.5</td>
<td>15.4</td>
<td>13.8</td>
<td>3.0</td>
<td>18.3</td>
<td>100.0</td>
</tr>
<tr>
<td>2000</td>
<td>50.1</td>
<td>14.6</td>
<td>14.3</td>
<td>5.5</td>
<td>21.0</td>
<td>100.0</td>
</tr>
<tr>
<td>2001</td>
<td>47.3</td>
<td>14.2</td>
<td>14.6</td>
<td>6.6</td>
<td>23.9</td>
<td>100.0</td>
</tr>
<tr>
<td>2002</td>
<td>43.4</td>
<td>13.8</td>
<td>15.0</td>
<td>7.8</td>
<td>27.9</td>
<td>100.0</td>
</tr>
<tr>
<td>2003</td>
<td>39.0</td>
<td>14.4</td>
<td>13.9</td>
<td>8.1</td>
<td>32.7</td>
<td>100.0</td>
</tr>
<tr>
<td>2004</td>
<td>14.1</td>
<td>14.2</td>
<td></td>
<td></td>
<td></td>
<td>100.0</td>
</tr>
</tbody>
</table>

Notes:

*: Constructed as zero since this category did not exist prior to the 1991–95 period.

**: In 1997 the government changed the investment reporting and approval procedure. The investment reporting threshold was revised from 50,000 yuan to 500,000 yuan, but this change only applied to SOEs and urban collective firms. The effect of this change is that the published amount of fixed asset investments in the state and urban collective sectors is smaller than the actual amount. For 1996, the government published both the revised and unrevised data. In the unrevised data, the SOEs invested 1205.6 billion yuan in fixed assets and the collective firms invested 366 billion yuan. In the revised data, the SOEs invested 1200.6 billion yuan and the collective sector invested 365.2 billion yuan. This is about a 0.4 percent and 0.2 percent difference, respectively.

Sources: Based on various sources on fixed asset investments compiled by the NSB. See the text for a detailed explanation.
A critical empirical issue is whether shareholding firms are private. During the 2001–03 period, they accounted for 68.2 percent of the fixed-asset investments in the ‘other’ ownership category. The largest shareholding firms are SOEs that have issued shares on the stock market; and only 6.97 percent of the shareholding firms were private Initial Public Offerings (IPOs) between 1990 and 2003. These shareholding firms are firmly in the hands of the state although they have some private revenue rights. According to a detailed study of over 600 firms on the Shanghai Stock Exchange and the Shenzhen Stock Exchange, the three main groups of shareholders – the government, legal-persons and private individual investors – each controlled about 30 percent of the outstanding shares (Xu and Wang, 1997). But the control rights are far less dispersed. According to the same study cited above, on average individual shareholders controlled only 0.3 percent of the board seats of those firms, whereas the government retained 50 percent of the board seats and state-owned institutions controlled the remainder.

Another test of the gradualist claim is to compare firms that have clear, straightforward ownership rights at the extreme ends of the ownership spectrum. One useful indicator is the ratio of fixed-asset investments in the registered private sector relative to the state sector. If, as commonly alleged, ownership biases against the private sector declined over time, it must be the case that the ownership biases against registered private sector firms relative to the treatment of the explicit state sector firms must have declined. A decrease in the ownership biases should be associated with a rising ratio; an increase in the ownership biases should be associated with a declining ratio.

The ratio in fact declined over time. Table 68.4 presents data on fixed-asset investment in the registered private sector as a ratio of fixed-asset investment in the state sector, collective sector and firms of ‘other’ ownership under Columns (1a), (2a) and (3a). In the 1980s, the fixed-asset investments undertaken by the registered private sector in both urban and rural areas already amounted to about one-third of the fixed-asset investments in the state sector. The ratio of the individual economy to SOEs, under Column (1a), was 0.31 during the 1980–85 period and 0.34 during the 1986–90 period. But this ratio declined sharply between 1991 and 1995, to only 0.22. Between 1996 and 2000, the ratio rose moderately, to 0.27. Between 2001 and 2003, despite a period of rapid growth and economic reforms widely perceived as bold, the ratio of fixed-asset investment by purely private to state firms only managed to recover to the level prevailing at the very onset of the reform era. For much of the 1990s there is no evidence that the ownership biases abated by this measure. In fact, the evidence suggests the opposite. This pattern holds when comparing the
individual economy with the collective sector and with firms in the ‘other’ ownership category.

How do we reconcile the above findings with the widespread view that China’s private sector developed rapidly in the 1990s? Part of the answer is provided in Tables 68.3 and 68.4 where it is clear that the investment share of the urban private sector rose dramatically in the 1990s. Thus, the boom in the private sector occurred in the easily observable urban sector. But it is important to stress that private sector development in China was overwhelmingly rural in origin. In the 1990s, the private sector in the rural areas faced increasing difficulties and because of their sheer weight the problems in the rural areas weighed down the total size of the private sector.21

<table>
<thead>
<tr>
<th>Year</th>
<th>(1) Individual economy/ SOE ratios</th>
<th>(2) Individual economy/Collective firm ratio</th>
<th>(3) Individual economy/’other’ ownership ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1a) Urban &amp; rural</td>
<td>(1b) Urban only</td>
<td>(2a) Urban &amp; rural</td>
</tr>
<tr>
<td>Panel (A): Period data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980–85</td>
<td>0.31</td>
<td>0.024</td>
<td>1.64</td>
</tr>
<tr>
<td>1986–90</td>
<td>0.34</td>
<td>0.045</td>
<td>1.64</td>
</tr>
<tr>
<td>1991–95</td>
<td>0.22</td>
<td>0.045</td>
<td>0.80</td>
</tr>
<tr>
<td>1996–2000</td>
<td>0.27</td>
<td>0.078</td>
<td>0.93</td>
</tr>
<tr>
<td>2001–03</td>
<td>0.34</td>
<td>0.18</td>
<td>1.02</td>
</tr>
<tr>
<td>Panel (B): Annual data</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>0.19</td>
<td>0.044</td>
<td>0.66</td>
</tr>
<tr>
<td>1997</td>
<td>0.26</td>
<td>0.056</td>
<td>0.89</td>
</tr>
<tr>
<td>2000</td>
<td>0.29</td>
<td>0.11</td>
<td>0.98</td>
</tr>
<tr>
<td>2001</td>
<td>0.31</td>
<td>0.14</td>
<td>1.03</td>
</tr>
<tr>
<td>2002</td>
<td>0.35</td>
<td>0.18</td>
<td>1.09</td>
</tr>
<tr>
<td>2003</td>
<td>0.36</td>
<td>0.21</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Sources: Based on various sources on fixed asset investments compiled by the NSB. See the text for a detailed explanation.
Conclusion
According to a famous formulation, the Chinese reform has been ‘Pareto-optimal’ in that it has created winners without creating losers (Lau et al., 2000). It is time to reassess this claim about the Chinese reforms. The decline in income of 130 million Chinese in recent years and the massive forcible land seizures are at direct odds with the view that the Chinese reforms have been ‘Pareto-optimal’. A substantial portion of the Chinese population may have lost absolutely.

It is also time to reassess a central claim in the gradualist framework – that the Chinese reforms have continuously deepened over time and have succeeded in establishing a market economy. Many economists marvel at China’s speed of transition but, as I have shown, by an investment measure China may have retreated in establishing a market economy in the 1990s, and by an output measure the Chinese economy today is less private than the Indian economy in the early 1980s. Let us also keep in mind that in 2008 China will celebrate the thirtieth anniversary of its reform program (1978–2008). This is one year longer than the duration of the orthodox phase of central planning in China (1949–78). While some argue that China has chosen a different reform strategy as compared with other transitional economies (Stiglitz, 1999), a more plausible hypothesis is that the Chinese may have a different goal: they have reformed to preserve socialism, not to institute capitalism.

Notes
1. I thank Professors Jaime Ros and Amitava Dutt for comments on an earlier draft and Nancy Hearst for editorial assistance. The usual caveats apply. The empirical component of this chapter is based mainly on my book (see Huang, 2008).
2. For a comprehensive account of the Chinese economy, see Naughton (2007).
3. The GDP data are calculated on a foreign exchange basis and the source of the data is the World Development Indicators, compiled by the World Bank (available at http://devdata.worldbank.org/dataonline).
4. The data on social development in China and India can be accessed in the World Development Indicators (available at http://devdata.worldbank.org/dataonline).
5. The data on fixed-asset investments are reported in the China Statistical Yearbook, various years. For the latest data, see National Statistical Bureau (2006).
6. One exception to the findings reported here is Wang and Meng (2001), who report that TFP growth averaged 7.3 percent during the 1992–97 period but only 2.5 percent during the 1978–91 period. However, the authors themselves dismiss this finding as ‘a statistical error’ since they cannot locate the sources of this dramatic acceleration of TFP growth.
7. It is not known whether this is because of female infanticide. If it is, it is useful to determine whether female infanticide increased or decreased in the 1990s. Demographers believe that economic hard times tend to be associated with a higher rate of female infanticide. See the findings by the World Bank in research notes on China’s health sector in World Bank (2005a) and World Bank (2005b).
9. The output measures here refer to the ratios of output in labor-intensive (skill-intensive) industries to the output in less labor-intensive (less skill-intensive) industries. High (low) labor-intensive industries are those industries above (below) the median value of labor.
intensity. The skill-intensity measure is similarly derived. For details, see Kochhar et al. (2006).

On the logic of gradualism, see Roland (2000). Naughton (1996) provides a convincing explanation of the reforms in the 1980s based on a gradualist framework. For a rejoinder of the gradualist approach to China, see Woo (1999).

On TFP in China, see Chen et al. (1988), Borensztein and Ostry (1996) and Hu and Khan (1997). Groves et al. (1994) link specific reform measures to some aspects of the operating improvements of the SOEs. The disagreements over TFP center around the relative importance of productivity improvement vis-à-vis factor accumulation to explain China’s growth, and evidence on whether the state sector also experienced productivity growth. See the debates between Jefferson et al. (1992) and Woo and Fan (1994).

The other problem is that the study treats domestic private sector firms and foreign-invested enterprises (FIEs) as a single homogenous category. This treatment does not recognize that China has favored foreign firms at the expense of domestic private sector firms. Thus, the estimate implicitly incorporates a substitution effect between FIEs and domestic private sector firms.

An analogy would be those firms owned and controlled by Temasek, the holding and investment arm of the Singaporean government. Whether Temasek behaves as if it is a private firm is a separate question, but, from an accounting point of view, because Temasek itself is state-owned, the firms controlled by Temasek ought to be classified as state-owned as well.

One huge difference separates China today from India of the late 1970s – the role of foreign direct investment. The government of Indira Gandhi severely restricted FDI whereas China today welcomes FDI. It should be stressed that the OECD’s estimate of China’s private sector at 52.3 percent is inclusive of foreign firms and a comparison of only domestic private firms would be even more unfavorable to China.

The data on fixed-asset investments used in this section come mainly from a series of NSB publications specifically devoted to covering fixed-asset investment activities. We have checked the data in these specialized publications with those published in the annual China Statistical Yearbooks. In comparison with the Chinese data on output, the Chinese data on fixed-asset investments are remarkably consistent across a number of publications. The data used in the text come from NSB (1987), NSB (1989b; National Statistical Bureau 1989a), NSB (1991), NSB (1992), NSB (1997) and NSB (1999). The data for some of the later years are from NSB (2003).

On the extent of state controls of fixed-asset investments, see Rawski (2001b).

The individual economy includes households, that is, self-employment proprietorships.

Another type of shareholding firm – known as a shareholding cooperative – is genuinely private. These were converted from township and village enterprises (TVEs) or small SOEs and are typically majority-owned by their employees. But they are very small. As of 2002, the shareholding cooperatives accounted for only 2.89 percent of China’s industrial output by value, as compared with 11.7 percent for the privately-operated enterprises (siying gaiye). Therefore, this is not a serious downward bias. It should be noted that the NSB no longer uses the ‘individual economy’ in its data series on industrial output, although it still uses the ‘individual economy’ category for its fixed-asset investment reporting. The 11.7 percent quoted in the text refers only to siying gaiye and presumably does not include industrial getihu. See NSB (2003).

I explore this topic in detail (Huang, 2008).

One issue that clouds an assessment of Chinese performance is whether the Chinese growth rate is as truly impressive as the official figures suggest. Young (2000) and Rawski (2001a) raise questions about the veracity of the Chinese data. This is a complicated issue that I will not deal with here, except to note that better research is needed to reconcile the well-documented microeconomic inefficiencies in the Chinese economy with the apparent macroeconomic success.
References


World Bank (2005a), ‘China’s Progress Towards the Health MDGs’, Briefing Note No. 2, March.

World Bank (2005b), ‘Rural health insurance: Rising to the challenge’, Briefing Note No. 6, May.


Similarities among South Asian economies

The countries of the South Asian region differ in size, resource endowment, the specificities of class configurations and the nature of the ruling regimes. Nonetheless they share certain common structural characteristics: high degrees of inequality of asset ownership, especially in Pakistan and India; the presence of substantial underemployment; a strong dualism between organized and unorganized sectors, especially in manufacturing, which sometimes (but not always) translates into the dualism between large-scale and small-scale economic activities; the continuing significance of agriculture as a major employer; the recent emergence of service activities as the largest incremental employers; and the involvement of the dominant share of the workforce in what are essentially low-productivity activities, often in the form of self-employment.

There is also an apparent synchronicity of policies and processes across the region, despite very differing social and political pressures. All the economies of the region had import-substituting industrialization strategies and substantial state regulation over economic activity for the first few decades after independence. From the 1980s onwards, all of them moved, in varying degrees, to a strategy of development based on export-orientation, internal deregulation, trade liberalization and privatization. The process started in Sri Lanka, as the Jayawardene government in 1977 moved towards liberalization and dismantling of the earlier universal food security system. Thereafter, especially from the early 1990s, all the governments in the South Asian region introduced policies of internal and external liberalization and privatization (Mahmud, 2000).

There was reduction in state control in terms of administered prices, regulation of economic activity and direct responsibility for a range of goods and services. Along with internal deregulation there was trade liberalization, entailing shifts from quantitative restrictions to tariffs and sharp reductions in the average rate of tariff protection. Financial liberalization involved reductions in directed credit (especially to agriculture and small industries), freeing of interest rate ceilings and other measures which raised the cost of borrowing for governments, peasants and petty producers. There was privatization of state assets, often in controversial circumstances.
All the economies moved towards ‘market-determined’ exchange rates, liberalization of current account transactions, and some degree of capital account liberalization, such as easing rules for foreign direct investment, allowing non-residents to hold domestic financial assets and making it easier for domestic firms to access foreign commercial borrowing. The implications of such external liberalization were very different in traditionally aid-dependent economies such as Pakistan and Bangladesh compared to India or Sri Lanka: in the latter, net capital inflows kept real exchange rates at levels that generated current account deficits, while in the former capital inflows were a substitute for aid.

In fiscal policy, most countries experienced some degree of ‘rationalization’ (a euphemism for reduction) of direct and indirect tax rates. This was associated with declining tax–GDP (gross domestic product) ratios in several cases, as tax buoyancy failed to meet the optimistic expectations that had justified tax rate cuts. In particular, the cuts in import tariffs (and the associated cuts in domestic duties required to establish ‘level playing fields’) involved lower aggregate collections relative to GDP in most of these countries, except Bangladesh. Attempts to reduce fiscal deficits typically involved cutting back public productive investment and social expenditure, reducing subsidies to farmers and increasing user charges for public services and utilities. Ironically, fiscal deficits in most cases did not fall relative to GDP, as the largest increases in expenditure came about in interest payments, partly due to the burden of past debt and partly because of the increased costs of public sector borrowing. In addition to falls in public investment as a proportion of GDP, there was also a resource crunch at regional and lower levels of government. This tended to reduce per capita spending in important areas such as basic infrastructure development, health and education (as in India), and affected the viability and legitimacy of local government institutions (as in Nepal and Pakistan).

Results of the transition from dirigisme to neoliberalism

The results of this process had similar outcomes in most of these economies, despite their very different initial conditions. On the positive side, growth rates on average have increased and this has generally been associated with greater macroeconomic stability in terms of lower inflation and avoidance of balance-of-payments difficulties. There has been some increase in private investment in all of these countries due to the immediate effects of liberalization and increased export orientation. However, aggregate investment rates have not increased much except in Bangladesh and Sri Lanka, and very recently in India.

However, income inequalities have increased in all the economies of the region. Growing economic inequalities are evident between rural and
urban residents; between households in various size-classes of expenditure; between sub-regions and provinces within countries. This has been associated with increased social and political tensions in the region, which have often been expressed not so much in direct demands for redressal of income imbalances, but in terms of other ethnic, social, cultural or regional demands.

There has generally been a deceleration of employment growth, compared to the *dirigiste* period. This has occurred despite an improvement, or at least the same trend level, of growth in aggregate economic activity. In general, employment has not kept pace with the increase in population, resulting in higher rates of unemployment and underemployment, and also in declining labour force participation (which is not fully explained by increased involvement in education). The quality of employment also appears to have deteriorated, with declines in regular work and increases in either casual contracts or self-employment in adverse conditions. Wage shares of income have typically declined; and real wage rates have stagnated or declined.

The relative decline of manufacturing, especially in the small-scale sector, and the stagnation or decline of manufacturing employment, is marked across the region, with the exception of the garments industries in Bangladesh and Sri Lanka. Agriculture and/or services appear to have become residual refuge sectors for workers who cannot find productive employment in industry. Across the region, there appears to have been relatively little link between rates of aggregate economic growth and employment generation.

The decline of institutional credit has been a major factor affecting the viability of agriculture and small-scale industrial development in most countries of the region. This has become a particularly severe problem in the recent past, when trade liberalization in the context of stagnant or declining world trade prices for South Asian crops has put additional competitive pressure on farmers, and contributed to an agrarian crisis across all South Asian countries.

Standard indicators of human development have improved on the whole, but the rate of improvement is much lower than desirable, and some indicators have even worsened in some cases. Literacy rates and primary enrolment have improved across the region, but the progress is uneven, with some pockets of backwardness according to region and social group; and average levels remain low everywhere except Sri Lanka. Human development indicators have improved faster only in Bangladesh, where public expenditure in social sectors has been high. Nutritional improvements have been slow or minimal, and per capita calorie consumption has fallen even in supposedly fast-growing countries like India. Infant mortality rates
have actually increased in recent years in Pakistan and in some states of India.

Notwithstanding some of these outcomes, the transition from dirigisme to neoliberalism has been seen by many as a vindication of the latter, especially in the context of India where the retreat from dirigisme has been followed by a fairly creditable export performance in some spheres and a higher rate of economic growth than most of the developing world since the mid-1990s (Joshi and Little, 1996). However, such sanguineness about neoliberalism is unwarranted. This is evident from the analytical reconstruction, within a political economy context, of India’s development story, which we attempt below, and from our brief sketches of trends in other South Asian economies that follow.

India

The economic policy regime erected in the 1950s in India had its roots in the freedom struggle. The economy had been dominated by metropolitan capital and metropolitan commodities before independence. Freedom meant freedom from this domination; and this could not be ensured without giving the state in independent India a major role in building infrastructure, expanding and strengthening the productive base of the economy, setting up new financial institutions and regulating and coordinating economic activity. This was necessary for building capitalism itself, although some saw it as a means of transition to socialism. State capitalism and state intervention were essential instruments for the development of a relatively autonomous Indian capitalism, displacing metropolitan capital from the pre-eminent position it had occupied in the colonial economy.

This intervention however occurred within a certain context. Despite talk of land reform, of providing ‘land-to-the-tiller’, and curbing the concentration of economic power, little was done to attack or redress asset and income inequality. The worst forms of absentee landlordism were done away with, but concentration of landownership remained largely intact. And while some monopolistic practices were curbed, asset concentration in the industrial sector was never really challenged. India’s monopolists were thus able to use state intervention as a device to consolidate and expand their positions.

The persistence of asset and income inequality imposed a constraint on the expansion of the market for mass consumption goods in the country. The absence of any radical land redistribution meant that the domestic market, especially for manufactured goods, remained socially narrowly based. It also meant that the growth of agricultural output, though far greater than in the colonial period (where the last half-century had witnessed virtual stagnation), remained well below potential. For the country
as a whole, the benefits of such agricultural growth as did occur were largely confined to a relatively narrow stratum of landlords-turned-capitalists and sections of rich peasants who had improved their economic status. Meanwhile, industrial growth was not sufficiently employment-generating to create large increases in demand from this source.

In this context, continuous growth in government spending became crucial for imparting a dynamic to the system. Import-substitution policies provided domestic capitalists with a large initial market for manufactures, which the government sought to expand through its current and capital expenditures. At the same time the state supported the domestic capitalist class through its infrastructure investment and through the creation of a number of specialized institutions for financing private investment. While this strategy succeeded in overcoming the long stagnation in agriculture, established a range of basic and capital goods industries, and achieved on average a rate of industrial growth in excess of 7 per cent during the 1950s and the early 1960s, by the mid-1960s it was clearly running out of steam. Not only was the initial stimulus offered by import substitution exhausted, but the capacity of the state to continue to provide the stimulus to growth was undermined by its inability to raise adequate resources.

This reflected the contradictory nature of the roles the State was required to fulfil (Patnaik, 1994). While increasing government expenditure was necessary to keep the domestic market expanding, the state also became the most important instrument for what Marx had called ‘primary accumulation’ by the domestic bourgeoisie, which was reflected in the fiscal crisis of the state. This implied that the government had either to cut back the tempo of its investment, or to maintain this tempo through increased borrowing with inflationary consequences, since the dirigiste economy was not a demand-constrained one.

The state’s inability to impose a measure of discipline (essential for viable capitalist functioning) on the capitalists, which made tax evasion rampant and contributed to the fiscal crisis, also made the attempts at regulation through licensing and other instruments quite ineffective. The imbalance between the pattern of domestic production envisaged in the plans, and the pattern of demand emanating from the relatively affluent sections who were the main beneficiaries of growth, gave rise to unutilized capacity through the illicit diversion of resources towards unplanned end-uses. The consequent ‘irrationality’ reflected inter alia in an increase in the capital–output ratio, compounded the problem arising from the increasing fiscal constraints upon the state.

This basic contradiction of the dirigiste regime, namely that it could sustain the tempo of development only at the cost of accentuating inflation
which undermined its own social stability, became apparent after the mid-
1960s. After 15 years of rapid industrial expansion until the mid-1960s,
industrial growth slowed over the next decade. This was mainly because of
a cutback in public investment in the face of strong inflationary pressures,
accentuated by the oil price hike of 1973. Public investment recovered
somewhat after the mid-1970s, by putting a check on inflation, first by
turning the terms of trade against agriculture and then, after the mid-
1980s, by going in for external borrowings. But such borrowings made the
economy vulnerable to capital flight. One such episode, triggered by the
payments difficulties arising from the Gulf War in 1991, led to the explicit
introduction of a neoliberal economic regime.

Growth had already picked up from the mid-1980s. The annual rate of
growth of real GDP rose from an average of over 3.5 per cent per annum
in the three decades after independence, to between 5 and 6 per cent in the
1980s and 1990s. Rather than import liberalization per se, the growth of the
1980s was related to a strong revival of agriculture through the spread of
the ‘Green Revolution’ to Eastern India. It also owed much to higher gov-
ernment spending, financed increasingly by external commercial borrow-
ing in the late 1980s. While domestic demand was kept up by higher
government spending, import liberalization (especially of capital goods
and components for manufacturing) reduced inflationary pressures and
imparted an impetus to final good production. Inflation was also kept
under control by lower relative prices in agriculture.

The inter-sectoral terms of trade for agriculture in the 1980s and early
1990s remained low compared to the early 1970s; but, for almost a decade
prior to the mid-1990s, Indian agriculturists were protected from the far
more adverse international movement of terms of trade against primary
products. Relative prices for agriculture worsened again only in the late
1990s, when trade liberalization exposed farmers to declining world prices.
This marked the onset of the period of agrarian crisis and increase in
peasant suicides.

While the internal contradictions of the dirigiste regime were bringing it
to a dead-end, major international changes also undermined dirigisme. The
easier access to private capital flows negated the basic assumption of a
binding foreign exchange constraint. More fundamentally, the process of
globalization of finance sought to break down all barriers to the free flow
of finance capital. It has been argued that the economic reforms after 1991
reflected an acceptance of the validity of the neoliberal economic paradigm
(Bhagwati, 1993). It is more correct to locate the policy shift in the totality
of circumstances produced by the interplay between the changing external
context and the accentuating domestic contradictions within the earlier
regime.
These contradictions encompassed other elements too (Chandrasekhar and Ghosh, 2004). A comparatively narrow social segment had provided the main source of growth in domestic demand under *dirigisme*. This social segment, eager to emulate the lifestyles and consumption patterns of the metropolitan centre, was not satisfied with having more domestically produced goods; its demand was increasingly for the new goods produced in the metropolitan centres, which could not be locally produced using only indigenous resources and technology. This imbalance increased over time because of further innovations in the metropolitan economies, creating a powerful and growing pressure from the more affluent groups for a dismantling of controls.

The emergence of newer strata, through a process of proliferation and diversification of the Indian capitalist class during the years of *dirigisme*, was another contributory factor. New capitalists operating outside the traditional bases of existing monopolistic groups, such as trade, finance, services of various kinds, had come into being, and operations abroad by non-resident Indian groups had become significant. These new entrants sought to diversify into manufacturing, and therefore welcomed deregulation. Because of access to newer technology, they were less averse to import competition. The established big businesses too which were, to start with, beneficiaries of state controls of various kinds, began to chafe against these controls at a certain stage in their search for new avenues for investment and production.

Support for economic liberalization also came from other quarters: new businessmen involved in what were essentially ‘parallel market’ transactions; a section of the top bureaucracy; the large and politically powerful urban middle classes; the more prosperous farmers whose real incomes increased in the 1980s. The technological and media revolutions imparted a significant impetus to the international demonstration effect, further fuelling neoliberal and consumerist demands. The ‘Non-Resident Indian (NRI) phenomenon’, whereby a significant number of people from the Indian elites and middle classes actually became resident abroad, also contributed to demands for opening up the economy.

The neoliberal reform strategy did not have much impact on growth to start with. The growth rate during the decade of the 1990s was scarcely any higher than during the 1980s; in the material commodity-producing sectors, agriculture and industry taken together, it was lower. Agricultural growth in particular decelerated dramatically during the 1990s, which became the first decade since independence to witness a decline in food-grain output per capita. So great however was the compression in domestic demand, especially in rural India where government expenditure was sharply reduced, that per capita foodgrain absorption fell even more
dramatically, resulting in the accumulation of 63 million tonnes of foodgrain stocks by mid-2002, of which 41 million tonnes were excess stocks. Per capita absorption of foodgrains for the country as a whole fell from the post-war triennial peak of 178 kg per annum during 1988–91 to 157 kg by 2001–04, which was the level in ‘British India’ on the eve of World War II. The excess stocks were largely dumped on the international market, where they were bought up to be used as animal feed for the rich countries. This was because neoliberal orthodoxy was hostile to the use of foodstocks on employment schemes, lest it enlarge the fiscal deficit (even though such deficit entailed no inflationary consequences in a demand-constrained system, and would not even raise the government’s net indebtedness much, as the foodstocks were with the government-owned Food Corporation of India).

GDP growth accelerated only after 2001, reaching more than 8 per cent between 2003–04 and 2005–06, mainly because of certain ‘newer’ services and in some export-oriented manufacturing of garments and chemicals. This period however also saw an even sharper divergence between agricultural and non-agricultural growth. Foodgrain production showed a zero trend, so that notwithstanding fiscally caused demand compression in the countryside, demand-pull inflation resurfaced by 2006. This was worsened by the government’s policy of stock reduction and winding down of public procurement operations in foodgrains.

Does liberalization per se account for India’s more successful recent export performance? If we leave aside ‘gems and jewellery’ where India was a successful exporter even before ‘liberalization’, and garments where India and her low-wage neighbours have been traditionally powerful, the one new area of strength that has emerged recently is information technology (IT)-related services and business process outsourcing (BPO). Here, the existence of a substantial educated manpower, whether with high skills as in the case of software exports, or with low skills as in the case of BPO, has been an important contributory factor. But the credit for this must go to the earlier dirigiste regime which defied conventional wisdom in setting up institutions of higher education, including several of excellent quality. Likewise, even the recent surge in exports of cement, steel and construction material are forays into the world market of industries that were set up and strengthened during the dirigiste period. All these avenues of export success would have eluded the Indian economy had it adopted neoliberal policies from the beginning, and not broken decisively out of the inherited colonial pattern of international division of labour.

Indeed, this was precisely the idea of P.C. Mahalanobis, the architect, along with Jawaharlal Nehru, of the dirigiste strategy in India. Mahalanobis assumed a closed economy in his 1950s plan model (inspired
by Soviet experience) because of his export pessimism on the basis of the then existing production structure. However, he strongly believed that by breaking out of this production structure through protectionism and state intervention, and imparting divergence and depth to it, the economy would be able to emerge as a more successful exporter at a later date. Recent Indian experience appears to vindicate him more than his neoliberal critics.

The recent acceleration in growth however cannot be explained by export performance. True, the export performance, especially in the service sector, has boosted domestic incomes and consumption to an extent at the upper end of the spectrum. But much of this boost is self-propelling, giving a lift to ‘animal spirits’ à la Keynes, which explains the recent increase in investment ratio. Underlying this boom are: a consumption splurge by the upper income groups, driven by the international demonstration effect of metropolitan living standards; a construction-cum-real estate boom, led by an enormous appreciation in land values; and an explosion in corporate profits. Two of the most palpable features of the contemporary Indian economy fall into place here: the increase in the unemployment rate even in the midst of this unprecedented boom, since structural-cum-technological change is in the direction of high-productivity sectors; and the rampant drive to displace peasants from land in the name of ‘infrastructure projects’, ‘Special Economic Zones’ and such like, all of which camouflage land speculation.

This growth has been accompanied by significant increases in inequality – both across regions of India and within regions across different economic and social categories. There is a widening gap between incomes in agriculture and non-agriculture: the ratio of per-worker domestic product in non-agriculture to that in agriculture which was about 2 in the 1950s has increased to well over 4 in the early 2000s. The wage share of national income has also fallen sharply and the wage share of value added in organized manufacturing declined to only around 10 per cent in 2004. Consumption surveys show absolute declines in the income and consumption of a substantial share of the population. During the last two decades of the twentieth century, the urban top 20 per cent of households experienced historically unprecedented increases in per capita consumption in real terms, while the per capita consumption of the bottom 40 per cent of the rural population actually declined.

Agriculture, which continues to employ the bulk of the workforce, has been afflicted by an intense and prolonged crisis. Its problems are closely related to more open trade combined with domestic deflationary policies. From the mid-1990s, the financial viability of cultivation has been eroded, owing not only to longer-term problems such as declining soil fertility, changing weather conditions and excessive dependence on depleted
groundwater, but also to a combination of sharply rising input prices and volatile and declining output prices. Indian farmers were encouraged to shift to cash crops, and exposed to international competition from the highly subsidized crop production in the North, even as they were facing reduced support from central and state governments in the form of inadequate availability of institutional credit, decline in public extension services, insufficient regulation of some inputs and reduced subsidies for other inputs such as fertilizers, reduction in timely crop price support, and poor storage and marketing facilities.

Poor employment generation remains a critical issue. Agricultural employment has fallen, due to both agricultural stagnation and technological and cropping pattern changes that reduced labour demand in agriculture. Such changes were also hastened by the growth of landlessness (as peasant cultivation became less viable because of rising input costs and falling or stagnant crop prices) accompanied by a shift to commercial cultivation with hired labour. In urban India, manufacturing is increasingly characterized by more capital-intensive techniques, and therefore declining employment elasticity of production. Even the ‘newer’ and more dynamic services such as IT-enabled activities that have increased their share of output still remain minuscule in employment terms. All IT-related activities currently employ less than 0.2 per cent of the total workforce. Recent employment increases have been mainly in the form of subsistence self-employment in low-value services, despite the economic boom.

Until the late 1980s, there was a secular trend towards declining poverty. Subsequently, while official estimates show a decline in poverty, nutritional indicators suggest the contrary. The coverage and quality of public services has worsened, which has had particular impact upon the condition of women and girl children (Dreze and Sen, 1994). The persistence of illiteracy, especially among females; the inability to ensure even primary education to all children and high drop-out rates over successive years of schooling; the poor indicators of health and the recent stagnation of infant and maternal mortality rates; the absence of proper sanitation for a large proportion of the population – all these provide an indication of the current state of the development project in India.

The external sector, however, has displayed a degree of overall stability in the balance of payments and a relative absence of the boom-and-bust cycles that marked some other emerging markets. To some extent this reflects the relatively limited extent of capital account liberalization over much of the period, and the fact that the Indian economy was not really ‘chosen’ to be a favourite of international financial markets until 2002. The greatest stability to the balance of payments was imparted by the substantial inflows of workers’ remittances from temporary migrant workers in the
Gulf and other regions, which has amounted to more than all forms of capital inflow put together. Since 2004 there has been a sharp increase in portfolio capital flows and external borrowing, but FDI remains relatively small.

The alternative to neoliberalism in India consists of policies to ensure the viability and sustainability of agriculture, and greater emphasis on public expenditure with high direct and indirect effects on employment generation, especially in infrastructure, health and education. This requires higher resource mobilization from the rich. It is also necessary to counter some of the adverse effects of trade liberalization on employment, apart from more directly addressing the basic structural issues of asset and income inequality and the persistence of low-productivity employment. Of course, such a policy shift requires political will and therefore a change in political configurations.

South Asian scenario
The political economy of the transition from dirigisme to neoliberalism in other South Asian countries, while differing in detail from that of India, was roughly similar: it represented everywhere the abandonment of the domestic bourgeoisie’s quest for a relatively autonomous trajectory of capitalist development, and the pursuit of an alternative trajectory with much closer integration between domestic and metropolitan capital, between finance and industry, and between the local and the global. It therefore meant a growing hiatus between the bourgeoisie and the urban upper income groups on the one side and the bulk of workers (facing insecure employment), peasants (facing agrarian crisis), petty producers and small businessmen (facing closures), and agricultural labourers (facing shrinking employment) on the other.

Pakistan
Despite a respectable per capita growth of around 5 per cent per annum over the second half of the twentieth century, Pakistan systematically underperformed on most social and political indicators, including education, health, sanitation, fertility, gender equality and political instability. In general, output growth has been associated with very low employment growth, at the trend rate of only 2 per cent per annum for the long period 1960–99. Employment growth worsened after the imposition of an International Monetary Fund (IMF) structural adjustment programme in 1987–88. In the 1990s, economic growth reduced on average and also became much more volatile. This was associated with historically low rates of investment, as private investment failed to revive or compensate for the decline in public spending. The investment–GDP ratio declined from 17.3
per cent in 1998–89 to 14.7 per cent in 2000–01, largely due to the collapse in public investment. Industrial growth rates almost halved from 8.2 per cent to 4.8 per cent per annum. Further, the earlier success at reducing poverty was reversed in the 1990s, as the percentage of households living in absolute poverty increased from 21.4 per cent in 1990–91 to 40.1 per cent in 2000–01 (Hussain, 2004).

The initial years after Pakistan’s third military coup witnessed a worsening of the macroeconomic situation, with increasing poverty and unemployment, falling real wages and worsening income distribution. However, recent geopolitics has impacted in some positive ways upon Pakistan's economy, mainly because of the willingness of the Musharraf regime to be a key ally of the USA. This caused the waiver or rescheduling of more than one-third of Pakistan's external debt, an increase in foreign aid flowing back to Pakistan, and the reinstating of export quotas in textiles and garments. Pakistani workers abroad have contributed to the massive recent increase in remittances, to as much as 14 per cent of GDP. However, since the domestic investment rate is still below the savings rate, the inflow of aid and remittances has not really contributed to economic activity, but is simply stored as foreign exchange reserves.

**Bangladesh**

Bangladesh shows a different and somewhat more optimistic economic trajectory than other South Asian countries. Since independence in 1971, there has been a slow but accelerating improvement in living standards. GDP grew at an average rate of 3.7 per annum in the 1980s, 4.8 per cent in the 1990s, and 5 per cent in the 2000s. Per capita income has grown even faster as population growth has slowed down in the recent years. With the fertility rate reduced from 6.3 children in 1975 to 3.3 in the mid-1990s, the rate of population growth has fallen from over 3 per cent to less than 1.5 per cent in a space of three decades. Per capita income growth, consequently, has doubled from about 1.6 per cent in the 1980s to over 3 per cent in the following decade and a half (Muqtada et al., 2002). This has been associated with a reasonable degree of macroeconomic stability, with inflation rates falling from over 10 per cent in the 1980s to just over 5 per cent in the subsequent 15 years, and moderate falls in the fiscal deficit and trade deficit ratios to GDP. Strong export performance in the garments sector and a steep increase in workers’ remittances from abroad have played an important role in improving both external and internal balances of the economy, as the inflow of foreign aid has experienced a secular decline.

The rate of investment in Bangladesh increased slightly from 17 per cent of GDP in the 1980s to close to 20 per cent in the subsequent 15 years, with
both public and private sectors contributing to this increase. A major factor was the increase in public expenditure over this period, despite declining foreign aid. This was partly because of the rising tax–GDP ratio, reflecting higher indirect tax revenue after the imposition of a uniform value-added tax. The share of social sectors, such as education and health, increased from 15 per cent of the growing total budgetary expenditure to over 20 per cent. The share of physical infrastructure also increased. In consequence, the headcount poverty ratio declined from 71 per cent in 1973–74 to 40 per cent in 2000, with the 1990s witnessing a more rapid decrease. Human development indicators also improved: the decline in infant mortality in Bangladesh, from 95 (per 100 live births) in 1972 to 30 (per 100 live births) in 2000, was among the fastest in the developing world. In spite of these achievements, the basic development challenges remain, requiring continued government intervention to ensure income diversification and improvements in living standards in the future.

**Sri Lanka**

Sri Lanka has been often hailed for its high achievements in human development, despite limited growth. Since independence in 1948, Sri Lanka has registered an average annual growth rate of 3–4 per cent while achieving adult literacy rates of 92 per cent and life expectancy of more than 72 years by 2001. Nevertheless, Sri Lanka remains a low-income, food-deficit country with more than 34 per cent of households below the poverty line in 2001.

In the 1970s, poverty alleviation strategies included free or highly subsidized social and economic services, such as education and health care. There were food subsidies as well as producer subsidies intended to help agricultural producers. These created the highest levels of human development in the region. However, since 1977, Sri Lanka pursued an economic liberalization policy. This was initially accompanied by increased public investments financed by external borrowing, but this could not be sustained (Dunham and Jayasuriya, 2000). From the mid-1980s, the government pursued a macroeconomic stabilization course along with mounting military expenditures in the civil war. From the early 1990s, deregulation and liberalization policies were supposed to bring about economic growth, but success here has been confined to the garments industry. Employment has stagnated, real wages have fallen throughout the 1977–2001 period, and most of the population continues to be employed in commodity production activities as farmers or production workers, and possesses limited education. Remittances (largely from women workers abroad) have been crucial in stabilizing the balance of payments and providing buffer incomes for the poor.
Nepal
A landlocked least-developed country, Nepal has major infrastructure gaps and much of it is physically inaccessible. There are marked regional differences between the Terai, hill and mountain areas, with decreasing infrastructure and increasing poverty as one moves up. Agriculture dominates the production structure, and rural people dominate the poor. The problems of backwardness are compounded by a long and relatively open border with India, which creates a high degree of dependence upon the Indian economy and also a vulnerability to India’s macroeconomic policies. The extremely adverse material conditions have generated extreme political movements, which in turn have generated much political instability.

There was a systematic decline in public investment from around 25 per cent of GDP in the late 1970s, to 18 per cent in the late 1980s, to only 11 per cent in the 1990s. This was associated with low and falling rates of private investment. Neoliberal reforms from the early 1990s have been associated with stagnation in agricultural incomes. The manufacturing sector has focused on exports, of garments, carpets and pashminas; there has been little development of manufacturing for the home market. Import penetration has prevented the development of infant industries and led to closure of small units and low employment generation in manufacturing. Financial sector reforms have been associated with reduced flow of institutional credit to agriculture and small enterprises, and fewer possibilities of subsidies for providing micro-credit to the poor through government channels. Open unemployment rates are low, but rates of underemployment are very high, estimated to be around 43 per cent of the labour force. However, recent political changes in Nepal give rise to guarded optimism about the future, and the possibility of economic strategies that will focus more on livelihoods of the poor and productive employment generation in a diversified economy.

Conclusion
Dirigisme outside India was always more vulnerable, since the countries concerned were small in size. Paradoxically however, unlike in Latin America, no effort was made during the entire dirigiste period for closer economic integration between the South Asian countries. On the contrary, competition between the South Asian countries, such as between Sri Lanka and India in tea, and between Pakistan and India in jute, was intense and yielded suboptimal outcomes, compared to what cooperation even after an initial phase of competition could have achieved.

South Asia’s quest for economic development has been accompanied by a parallel quest for building modern nation-states encompassing multiple religious, linguistic, ethnic and regional groups. The difficulties of this
process, exemplified by Tamil secessionism in Sri Lanka, by the break-up of erstwhile Pakistan, by the periodic eruptions of secessionist movements in India, and by a host of conflicts dotting the entire region at any time, have been compounded by the consequences of the development trajectories pursued. In a dirigiste regime the distribution of the ‘gains’ from development across the proto-bourgeoisies belonging to different ethnic, linguistic and such other groups is determined by the nature of the state, by who has greater control over the state apparatus. In a neoliberal regime there exists a spontaneous tendency towards increasing divergence starting from certain initial differences. The resulting discontent under neoliberalism allows scope for self-serving intervention by powerful elements of metropolitan capital backed by metropolitan states, which creates a tendency towards a fracturing of the nation. The growing inequalities become particularly significant in this context: since any growing hiatus necessarily has some regional or linguistic or religious dimension, it can become a means of fragmentation of the nation. Dirigisme in South Asia did not always succeed in keeping these contradictions in check. Neoliberalism is further accentuating these contradictions. The development of humane societies in this region still has a long way to go.

Note
1. The ‘poor’, officially, are those with consumption below the official poverty line. This is determined by applying price-index adjustment to the observed cost of that 1973–74 consumption basket at which a certain number of calories were just accessed.

References
The image of rapid growth throughout East Asia following Japan has implied a more coherent region than actually exists, and a corresponding tendency to see economic progress in the region as similar in origin and nature, leading to regional terms such as the ‘Far East’, ‘Asia-Pacific’ and ‘Pacific Asia’ preceding others such as ‘East Asia’, ‘yen bloc’, ‘flying geese’, ‘tigers’, ‘mini-dragons’ and so on. On the other hand, rather amazingly, the World Bank’s (1993) influential *East Asian Miracle* volume hardly considered the significance of geography or location. For the Bank, it did not seem to matter that the eight high-performing Asian economies (HPAEs) – incredibly, with the benefit of hindsight, excluding China – were all from the same part of the world and were not only contiguous but often also linked by trade, finance and investment relations.

The next section starts with a critical review of the World Bank’s (1993) influential study of the East Asian economic miracle. The following section emphasizes the diversity of East Asian experiences and the significance of recognizing this diversity for drawing appropriate lessons. The chapter then considers implications of the East Asian financial crises of 1997–98, before the final section addresses some issues in drawing lessons from the East Asian experience.

**Interpretations of the ‘East Asian miracle’**

There are at least three major competing explanations for the rapid growth and industrialization of several East Asian economies, often referred to as the ‘East Asian miracle’. The dominant version in the 1970s and 1980s presented sustained rapid growth in the East Asian region as essentially due to unfettered market forces (Little et al., 1970). The obvious policy implication was to liberalize, as well as to open up or globalize. In the late 1980s, this was challenged by an almost opposite, *dirigiste* position, which emphasized the leading role of the ‘developmental state’ (Johnson, 1982; White, 1988; Amsden, 1989; Wade, 1990). The World Bank’s 1993 study seemed to offer an intermediate view by acknowledging the role of the state, while emphasizing the importance of the market.
The World Bank’s (1993) East Asian Miracle identified at least seven types of state interventions. It approved of the first four, deemed functional interventions, namely: (1) ensuring macroeconomic discipline and macro-economic balances; (2) providing physical and social infrastructure; (3) providing good governance more generally; and (4) raising savings and thus investment rates. Functional interventions were said to compensate for market failures, and hence, were deemed acceptable, if not desirable, and less distortive of markets. Three types of market-distortive strategic interventions were considered in the areas of finance: directed (that is, subsidized) credit, trade policy and industrial policy.

The interventions were not just market-conforming, but also played important proactive roles which have been more than simply market-augmenting, as suggested by the World Bank’s analysis. Of the more controversial strategic interventions, the Bank conceded that financial interventions had been important and successful in East Asia, particularly in North-East Asia, that is, in Japan, Korea and Taiwan, but considered other strategic interventions to be failures. However, the Bank maintained that other developing-country governments were not capable of successfully pursuing the types of policies North-East Asians had successfully implemented because their state capabilities were almost unique and virtually non-replicable.

The Bank volume’s evaluation of the role and record of Japan’s Ministry of International Trade and Industry (MITI) and its counterparts elsewhere in the region is more predictable, arguing that government interventions have been trade-distortive and generally unsuccessful in East Asia, although export promotion in particular comes in for much praise. The Bank disapproves import protection and fails to see the connection with export promotion. The Japanese, South Korean and Taiwanese governments implemented import substituting industrialization (ISI) policies from the 1950s, but also pursued export-promotion soon after to ensure that their industries quickly become internationally competitive by requiring a rapid transition from import substitution to export orientation, that is, ‘effective protection conditional on export promotion’ (EPconEP). EPconEP is quite different from just import substitution or export promotion for example in export-processing zones (EPZs). Both import substitution and export promotion (for example subsidies) are trade distortions. It is also misleading to suggest that EPconEP is tantamount to ‘simulated free trade’ because the ‘bad’ import-protection distortion is negated by the ‘good’ export-promotion distortion. EPconEP allows industries and firms to enjoy rents from import protection to develop new industrial and technological capabilities in order to become internationally competitive in terms of both cost and quality.
Infant industries have often been provided with effective protection conditional on export promotion, which has had the effect of forcing firms and industries quickly to become internationally competitive. By giving firms protection for certain periods, depending on the product, and by also requiring that they begin exporting certain shares of output within specified periods, discipline was imposed on the firms in return for the temporary trade protection they enjoyed. Such policies forced firms to reduce their unit production costs as quickly as possible, for example by trying to achieve greater economies of scale and accelerating progress up learning curves. Requiring exports has also meant that producers have had to achieve international quality standards quickly, which imposed pressures to progress technologically in terms of product design and quality as well as technological processes. With strict discipline imposed, but also some flexibility in enforcement, many firms managed rapidly to achieve international competitiveness.

**Regional diversity**

Some important differences among the East Asian miracle economies suggest that not all East Asian economies have been proceeding inexorably in the same direction in a similar manner. The East Asian experiences are far from constituting a single model. Some of the major differences in East Asia are themselves very instructive. In the case of the role of foreign direct investment (FDI), tremendous contrasts are found. In the case of Singapore, FDI has constituted about a quarter of gross domestic capital formation and about 15 per cent in Malaysia. At the other end of the spectrum, in the case of Japan and Korea, the percentage has long been below 2 per cent. Some of the other countries fall between these two extremes, with few near the mean for developing countries of around 5 to 6 per cent. Those most successful in developing industrial capacities and capabilities in East Asia – namely Japan, South Korea and Taiwan – have hardly depended on FDI.

FDI’s far greater importance in South-East Asia has been due to a variety of reasons, which have not been entirely economic. One reason for the major role of FDI in Singapore and Malaysia is political. After Singapore seceded from Malaysia in 1965, the regime decided to attract FDI in massive quantities to ensure its own survival, so that the major foreign powers would quickly develop a stake in the survival of the Singapore regime. Subsequently, this FDI preference has been justified in terms of improving access to foreign markets and technology.

Malaysia has long had ethnic rivalries and an ethnic affirmative action policy. Some policy-makers tried to limit ethnic Chinese control of the economy by encouraging foreign direct investment. Again, one finds a
political motivation for the important role of FDI in Malaysia. In other words, political considerations have been a very important reason for attracting, even privileging foreign investment in Singapore and Malaysia.

There has also been considerable diversity in the role and performance of public investments, including state-owned enterprises (SOEs), in East Asia. In South Korea, Japan and, of course, Hong Kong, SOEs are hardly important today, but historically, state-owned enterprises were important in Japan before World War II, and in South Korea until more recently. Conversely, however, one finds that state-owned enterprises have been extremely important in Singapore and Taiwan. Again, this is partly explained by political factors, but there are also economic considerations. And very importantly, the performance of these SOEs has also been quite impressive.

Industrial and technology policies in East Asia have also been quite diverse. One extreme, of course, is Hong Kong, where there has been relatively little industrial policy, although more than most opponents of industrial policy care to admit, especially in recent years. It is far more detailed and sophisticated in Japan and Korea at the other end of the spectrum. In Korea, industrial policy is largely oriented towards large firms, whereas in Taiwan, much more emphasis is given to medium-sized and relatively smaller enterprises.

Industrial policies in the region have also had different orientations, emphases and instruments. For example, trade policy has been very important in almost all economies in the region except Hong Kong and Singapore, while financial policy has been important in all the countries, including Singapore, but again, with the exception of Hong Kong before the 1998 crisis. Since Hong Kong’s reversion to China in mid-1997, there have been many indications of greater government interventions in the territory, presumably in line with its new status and China’s envisaged role for the deindustrialized financial centre.

The World Bank recommended that the rest of the developing world emulate South-East Asia, not North-East Asia because of important differences between them. These misleading claims require us to recognize the far more impressive achievements and superior economic performance of the first-tier East Asian newly industrialized economies (NIEs) (including Singapore), compared to the second-tier South-East Asian NIEs. The World Bank (1993) argued that the South-East Asian high-performing economies were the preferable model for emulation by other countries seeking late development.

According to Yoshihara (1988), the South-East Asian economies have been characterized by ‘ersatz capitalism’ because of the compromised and inferior role of their states, their maltreatment of ethnic Chinese and their failure to develop better technological capabilities. Jomo et al. (1997)
criticized the World Bank’s claims, suggesting various problems associated with the growth experiences of the South-East Asian economies praised by the Bank. In any case, the currency and financial crises of 1997–98 radically transformed international opinion about the East Asian models, with earlier praise quickly transformed into condemnation (Jomo, 1998).

The Malaysia, Indonesia and Thailand (MIT) economies as second-tier or second-generation East Asian newly industrializing countries (NICs) share some common characteristics with Singapore, which is also in the region. However, they are not only far less advanced in developmental terms, but also quite different from the city-state’s heavy reliance on trade and financial services besides manufacturing. Essentially, the MIT economies have had somewhat different, even ersatz developmental states and industrial policies, compared to the first-generation East Asian newly industrialized economies (NICs). Although Singapore too has pursued industrial policy, it has used fewer trade policy instruments and has been far more reliant on foreign direct investment compared to the other East Asian NIEs. Though Singapore, like Hong Kong, has eschewed trade policy instruments, it has used state-owned enterprises (SOEs) – usually referred to as government-linked corporations (GLCs) in the island republic – more than any other East Asian economy, and perhaps any other economy in the world in the early twenty-first century.

Most importantly, the South-East Asian high-growth economies (including Singapore) have relied much more heavily on foreign direct investment (FDI) to develop most of their internationally competitive industrial capabilities. Trade policy instruments in the region have been less well formulated and implemented, with rather mixed consequences, but have nonetheless been part of the region’s industrial policy story. Generally, government interventions in the region have been influenced by a variety of considerations besides economic development and late industrialization. Consequently, industrial policy has also varied in nature, quality and effectiveness. Yet, the economies in the region would not have achieved as much as they did without selective government interventions, including industrial policy.

Growth performance has been superior in North-East Asia over the long term despite the much greater resource wealth of South-East Asia. Over the period studied by the Bank, that is, from the 1960s until the early 1990s, growth in the former averaged about 8 per cent, compared to about 6 per cent for the latter. A 2 per cent difference, compounded over a period of a quarter-century or more, adds up to a lot. Very importantly, except in Hong Kong (due to immigration from China) and perhaps Singapore, population growth has been much lower in the former compared to the latter. Immigration into Hong Kong and Singapore involves a very high
proportion of the labour force, thus raising the average labour utilization rate. Political factors have also ensured far more equitable distribution of economic welfare than would otherwise have been the case in the first-tier NIEs, whereas such considerations have been less influential in the second-tier South–East Asian NICs despite Malaysia’s ethnic ‘social contract’ and Indonesia’s rural development efforts to achieve political stability.

Improvements in per capita income and economic welfare have been much more significant in North–East Asia, compared to South–East Asia (with the exception of Singapore), despite the greater resource wealth of the latter. Income inequalities have also been far less in North–East Asia, although there is some evidence of rapid recent increases in inequality. In other words, what South–East Asia has achieved has been less impressive in some critical ways. Drawing from this contrast, some people now argue that resource wealth is not a blessing, but a curse, in so far as it may have postponed the imperative to industrialize.

The North–East Asian NIEs have generally had much more sophisticated and effective industrial policy compared to South–East Asia’s NICs. This accounts, in no small way, for the very important differences in industrial and technological capabilities between North–East Asia and South–East Asia. Also, South–East Asian industrialization is still driven by FDI, whereas North–East Asian industrialization is primarily an indigenous phenomenon.

Japan and the first-generation NIEs began to industrialize in the very specific economic and political conditions of the post-World War II Golden Age and Cold War. North–East Asia grew rapidly in the immediate post-war period under a ‘security umbrella’ provided by the Americans, including aid and privileges no longer available to others. Besides providing generous aid, the Americans were anxious for them to ‘succeed’ economically in order to be showcased as attractive alternatives to their neighbours under communist rule or influence. Hence, the Americans were quite happy to tolerate trade, finance, investment, intellectual property and other policies violating neoliberal economic norms that they are now strongly opposed to. Such conditions are simply no longer available to others, and hence, their experiences are more difficult to emulate. To discourage other developing countries from trying to emulate the first-generation East Asian NIEs, it is now often argued that their state capabilities are culturally unique and impossible for others to emulate.

The Guomindang government of Taiwan was the same regime driven out of mainland China by the communists because of its incompetence and corruption. One could say the same of the Rhee regime in Korea in the 1950s. Japan has hardly been scandal-free in recent years, and most observers trace recent abuses to the nature of Japan’s post-World War II
political economy. The superior policy-making and implementation capabilities of the North–East Asian decision-makers was, at least until the 1997–98 financial crisis, widely acknowledged, but this in itself does not prove that policy-makers were thoroughly competent and incorruptible.

Some also claim that East Asia cannot be emulated owing to its very different initial conditions. Such differences are real, but often exaggerated. There is no doubt that Japan as well as the first-tier East Asian NIEs have also been distinguished by much higher levels of educational achievement. However, the level of literacy in Korea in 1950 was lower than the literacy rate in contemporary Ethiopia, which has one of the lowest literacy rates in Africa today. The educational achievements of contemporary South Koreans reflect the tremendous human resource investments in East Asia in the post-World War II period.

Some fortuitous circumstances must also be considered. Japan, South Korea and Taiwan all implemented relatively virtuous American-sponsored land reforms soon after the end of World War II. There was also significant redistribution of other non-land assets in Japan, most notably, of the pre-war and war-time zaibatsu industrial conglomerates. Much of the motivation for such redistributive reforms was, of course, anti-communist, that is, to undermine and minimize support for the communists by those desiring asset redistribution.

In contrast to South–East Asia, more egalitarian asset redistribution in Japan, South Korea, Taiwan and China have also been important. The Americans were not uninfluenced by the left, partly because of the nature of the wartime anti-Axis alliance and the nature of the most influential scholarship available. During the post-World War II American occupation of Japan, it was widely presumed that the zaibatsu ‘military industrial complex’ had been responsible for the militarization of pre-war Japan. The American occupation forcibly broke up family control of the zaibatsu to create a unique, corporatist ‘stakeholder’ economy. Assets were sold to employees, and to local communities, thus developing worker and community stakes in the companies.

Hence, the peculiarly Japanese economy was created by deliberately redistributive policies with unique outcomes. The ‘human relations’ school of industrial relations influenced the post-World War II development of guaranteed life-long employment and the seniority wage system, which have effectively strengthened a strong employee commitment to the fate of their firm.

From miracle to debacle
Although East Asian economic performance before 1998 was debated, the East Asian debacle of 1997–98 was not anticipated, partly because it was
not principally due to a failure of the real economy despite various recognized economic weaknesses (Jomo et al., 1997). The financial crisis from mid-1997 was precipitated by an eventually successful currency attack on the Thai baht, overvalued after the strengthening of the US dollar after mid-1995. The crisis was greatly exacerbated by herd-like panicky withdrawals from the entire East Asian region, inducing currency and stock market collapses (Jomo, 1998). Those who control financial assets usually enjoy disproportionate policy influence in most contemporary economies, especially in ‘emerging markets’. The greater role of foreign capital in South-East Asia subordinated domestic industrial capital in the region, allowing finance capital, both domestic and foreign, to become more influential in the region, thus rendering it more economically vulnerable (Jomo, 1998).

Finance capital in the region had developed complex symbiotic relations with politically influential rentiers, dubbed ‘cronies’ in the aftermath of the 1997–98 crisis. Although threatened by the full implications of international financial liberalization, East Asian financial interests were quick to identify and secure new possibilities for capturing rents from arbitrage as well as other opportunities offered by gradual international financial integration. Thus, foreign dominance of South-East Asian industrialization facilitated the ascendance of financial interests and politically influential rentiers.

Such increasingly influential alliances were primarily responsible for promoting financial liberalization in the region, both externally and internally. However, insofar as the interests of domestic financial capital did not entirely coincide with international finance capital, international financial liberalization was necessarily partial. The processes were necessarily uneven, reflecting the variety of interests involved and their varying strengths in different parts of the region. Such flows were also desired to finance current account deficits in both countries, principally due to service account deficits (mainly for imported financial services as well as investment income payments abroad) and growing imports for consumption and output of non-tradeables, mainly in the property (real estate) sector. There is little evidence that such capital inflows contributed significantly to accelerating economic growth, especially of the tradeable sectors. Instead, they often contributed to asset price bubbles, consumption binges and ‘over-investments’.

**Challenges**

There are important lessons to be drawn from East Asia, but clearly, there is no single model as such. For other reasons as well, it does not make much sense to emulate any particular economy in East Asia. Most other
developing countries would find it impossible to do so even if they wanted to. Nevertheless, some important lessons can be drawn from the East Asian experiences. Such lessons are best drawn from careful analysis rather than broad-brushed generalizations about a rather diverse region.

Economic liberalization, including globalization, since the 1980s has fundamentally changed the environment and conditions for selective industrial policy and, hence, for aspiring developmental states. Most importantly, economic liberalization – at both national and international levels – has seriously constrained the scope for government policy interventions, especially selective industrial promotion efforts. This is true of both the international and domestic policy environments, where policy conditionalities and World Trade Organization (WTO) membership obligations have radically reduced the scope for national economic policy initiatives.

The mid-1980s onwards has seen widespread, sweeping and rapid opening up of trade, investment, finance and other flows. Very often, such liberalization has been externally imposed by the Bretton Woods institutions as part of conditions imposed to secure access to emergency credit during the debt crisis of the 1980s, and more recently, in the wake of the financial crises since the mid-1990s. Various policy packages for (price) stabilization in the short term or for structural adjustment in the medium term have involved such conditionalities. The new intellectual and policy environment from the 1980s – the so-called ‘Washington Consensus’ – has promoted such policy reforms.

Such policy changes as well as limited government capabilities have meant little preparation in terms of a proactive strategy to anticipate and cope with the new international competition. Few industrial policy instruments of the past are viable or feasible options today, including many tools used successfully in post-World War II East Asia. Many, if not most of, the main industrial policy tools still available are already intensively used by most advanced industrial economies.

These policies are probably necessary, but certainly not sufficient for stimulating and sustaining economic growth and structural change for developing countries’ ‘catch-up’. Special policies are urgently needed to prevent such economies – already at a historical disadvantage in various respects – from falling further behind, if not to begin to close the gap with the industrially more developed economies of the North as well as the industrial economies that have emerged in recent decades, that is, during the last third of the twentieth century.

As a region, East Asia has led other developing-country regions in terms of economic performance, growth rate, increased exports as well as technological progress. Lall (2003) notes the great divergence between East
Asian countries with and without selective industrial policy, and finds the latter (mainly in South–East Asia) far more economically vulnerable.

Industrial development in the new circumstances clearly requires international competitiveness, and such competitiveness is increasingly defined in manufacturing, related services and institutions, and not simply in terms of wage costs or exchange rate competitiveness, as important as these may be. Inability to compete effectively implies being bypassed, and likely, stagnation at the lower end of the technological and income ladder.

Appropriate industrial policy will require selective interventions as well as effective coordination among firms, clusters and factor markets, which should presumably be consistent with a clear and coherent ‘vision’ of the future as well as the ‘road-map’ towards policy goals (Lall, 2003). For this purpose, there are still many useful lessons to be drawn from the varied experiences of the more successful East Asian NIEs and China, as well as the more modest and flawed achievements of the South–East Asian NICs.

Notes
1. This chapter is based on considerable earlier work. I am grateful to all who have provided me with critical feedback on this work. Needless to say, however, no one else bears responsibility for this version.
2. After the East Asian crisis, even the IMF seemed to back off from its previous advocacy of financial market liberalization.

References
Introduction

‘Transition’ is a designation that is widely used for the transformation of state socialist countries, characterized by state-controlled economies and political dictatorship, into democratic market economies. Transition started to receive intensive attention when a number of countries freed themselves in 1989 from either particularly the Soviet control (think of Poland, Hungary, Czechoslovakia, East Germany and the Baltic states) or mainly domestic tyranny (such as Romania and Albania). Transition was extended in 1992 to the former Soviet Union (FSU) states, also dubbed the Commonwealth of Independent States (CIS), after the disintegration of the Soviet Union in December 1991.

We also will consider countries such as China and Vietnam as transition economies, though there is a reason to exclude them. In China and Vietnam political power remained in the hands of the leaders of the communist parties. In contrast, the member countries of the FSU and the Central and Eastern European (CEE) countries have undergone a dramatically rapid democratization process – with some exceptions, such as Belarus, Turkmenistan and Uzbekistan. If a high speed of political turmoil were a determining feature of transition, countries such as China and Vietnam would stay out. The differences between the outcomes in, on the one hand, the countries in CEE and of the FSU and, on the other hand, China are striking. From the countries of the first group, only three countries were able to recover from the ‘transition depression’ of gross domestic product (GDP) after ten years. However, China did not experience a decline in GDP in any year from 1978 onwards – China’s transition period – while it had a high and stable annual GDP growth figure in the order of 8 percent on average. It is, of course, quite interesting and informative to analyze the reasons for such economic divergences. For that goal, we will not employ a rapid political transformation as a distinctive feature of transition, so that we will focus on economic transition.

At the start of transition in Europe, around 1990, knowledge about the preferred materialization of the transition process was partly borrowed from the experience obtained from the process of economic growth in
developing countries. Now, more than 15 years later, there is ample information about the transition process to consider the reverse question: What are the lessons of this transition from practice for developing countries that are eager to give additional spurts to growth? The search for these lessons is the aim of this chapter. The chapter consists of three parts. It starts with a short history of political developments, mainly in Russia, the motor of state socialism after World War II. The next part describes the stylized facts of the transition process in the countries in CEE and of the FSU, and China as well. The following part discusses the lessons for developing countries that can be drawn from the transition process.

A short political history

After the October Revolution of 1917, the Bolsheviks found all political power in an internationally isolated and underdeveloped country rested in their hands. They considered it their main task to industrialize the country at full speed, using internally raised investment funds. It was for this purpose that the Soviet planning system was created (Knaack, 1996).

The Soviet Union experienced a ‘golden age’ in the 1950s. The country grew rapidly, propelled by increases of capital, labor and raw materials. Economic growth directly benefited consumers as their diet and housing improved apace (Schroeder, 1992). Space flights and Nobel prizes symbolized the achievements of Soviet science.

However, from the 1960s onwards, the Soviet economy settled on a slower growth path. According to Ellman and Kontorovich (1992, pp. 10–12), there are three explanations for this slowdown of economic growth. First, a loss of control of the economy. If an economy becomes more complex, coordination from above becomes increasingly difficult. Second, a reduction in the growth rates of both the volume and the productivity of production factors. One could not, for example, increase the participation rate of women any more, and also stocks of natural resources were running out. Third, a weakening of the ‘entrepreneurial spirit’. In a command economy, pressures from above provide the main source of dynamics in an economy. Brezhnev’s policy of ‘stability of the cadres’ represented a codification of the process of slackening the pressures from above.

Declining growth rates alone cannot explain the collapse of the Soviet Union at the end of the 1980s. According to Kornai (1992), the command system was still able to guarantee the population a decent way of life. What caused the crisis was its weak economic performance relative to that of the USA and other Organisation for Economic Co-operation and Development (OECD) countries. The dynamic efficiency argument was the raison d’être of the Soviet Union. Consequently, the relatively poor growth figures of the 1980s threatened the political legitimacy of the whole system.
Economic reform under Gorbachev must be understood as part of his effort to revive and modernize the Soviet economy. The results of the reform process were disappointing, mainly because Gorbachev’s policies were not feasible (Hewitt et al., 1987). But his policies also had unintended consequences. This happened especially for his glasnost policy intended to unmask bureaucrats sabotaging the reform process.1

On 1 November 1989, the Berlin Wall fell. From that moment on in CEE one country after another unlinked itself from the Soviet dominance and started a process of conversion from a centrally governed economy to a market economy. On 30 June 1990, the two parts of Germany were officially reunited and the German economic, monetary and social union was created. In Czechoslovakia, the Velvet Revolution ended the ruling of the Communist Party in November 1989. The new government of Prime Minister Klaus introduced a series of measures as from 1 January 1991, aimed at the integration of the Czechoslovakian economy into the world economy. The Soviet Union was also contaminated with this spirit of the time. After the breakdown of its economic system the new rulers strove for a rapid transition towards a new system characterized by market relations, private ownership, and a liberal democracy. After the abortive coup d’état of August 1991, the Soviet Union disintegrated and Yeltsin became President of the new Republic of Russia. On 2 January 1992, the Gaidar administration introduced a number of market reforms, which inflicted an enormous shock to the Russian economy.

Stylized facts of the transition process

Central and Eastern Europe and the former Soviet Union
The processes of change in the countries of CEE were not based on a blueprint showing how a formerly communist country ought to be restructured as a capitalist country. According to the prevailing opinion, that was not necessary. Due to the German unification, East Germany took the West German legislation and rules over at one stroke and Eastern German enterprises were privatized at a quick pace. Other countries had the possibility to copy a large part of the existing and tried and tested recipe as applied earlier by, for example, Spain at its accession to the European Union and developing countries at their integration into the world economy. It was the recipe developed by the International Monetary Fund (IMF) and the World Bank (Taylor, 1993) and is often dubbed the Washington Consensus.

The Washington Consensus comprises the ‘Holy Trinity’ of stabilization, liberalization and privatization (Rutland, 1999). This Consensus initially was advice for a reform agenda for the Latin American countries at the end of the 1980s to adapt their policies and institutions.2 Applied to the former
communist countries at the time, this trinity would read as follows. ‘Stabilization’ refers to the need to reduce inflation, both open and suppressed, to create a currency with a stable external value, and establish balance-of-payments equilibrium. To achieve these goals, a standard package of fiscal and monetary measures was recommended, together with price liberalization so as to eliminate suppressed inflation. ‘Liberalization’ means the freeing of enterprises and individuals of the old planning institutions. The initial expectation was that the disappearance of these planning institutions would create space for the new market institutions to arise spontaneously. ‘Privatization’ was considered crucial for both political and economic reasons. Politically it would create powerful interest groups with a stake in transition and, for that reason, strongly opposed to the old system. In addition, it would secure active support from the West. It was thought, moreover, that only privately owned enterprises could operate successfully in a market economy. The collapse of the communist regimes in CEE after 1989 and the dissolution of the Soviet Union in 1991, spread optimism about the chances of rapid economic growth.

The transition process appears to have several characteristics, both intended and unintended (Campos and Coricelli, 2002; Ellman, 2005; Knaack and Jager, 2007). First, all countries of the former Eastern bloc faced a dramatic fall in output. The individual country patterns of recession and recovery cover the years 1989–97. They have largely been of the L-curve shape (sharp fall, followed by slow recovery), rather than the initially hoped-for J-curve type (small fall, followed by fast growth). Figure 71.1 displays the growth rates of real GDP in the years of transition for the distinct country groups. The strongly negative GDP growth rates for each group in the first years of transition indicates that real GDP levels should exhibit L-curves. The fall of investment expenditures, especially for inventories and housing, was even larger than the fall in real GDP. Defense expenditures on equipment and materials have declined sharply as a proportion of GDP, especially in the countries of the FSU. Consequently, during the period of falling output, private consumption has declined only moderately relative to production in most transition economies.

Second, a large inter-country variation in the adjustment patterns is observable, with large differences in both depth and length of the L-curve. According to Figure 71.1, for the whole group of CEE countries the average growth rate already became positive again in 1993, after a depth of −9 percent for the sub-group EU8 (the first eight CEE countries that joined the European Union, that is, Poland, Hungary, the Czech Republic, Slovakia, Slovenia and the three Baltic States) and −14 percent for the countries in the sub-group SEE (or South-Eastern Europe), both in 1991. The CIS, however, only experienced for the first time since 1989 positive economic growth in
1997. The low-income CIS countries had their depth in economic growth in 1992, with a decline in GDP of over 20 percent, and the middle-income CIS countries in 1994 with a negative growth rate of 15 percent. The variation in the cumulative fall of real GDP per country in the beginning of the transition process has been large. It ranges from moderate for Poland (6 percent, in two consecutive years of output decline) to high for Russia (40 percent, in seven years of consecutive years of output decline) and extremely high (over 60 percent fall) for Armenia (63 percent, in four consecutive years), Moldova (63 percent, seven years) and Georgia (78 percent, five years).

Third, like output, the level of foreign trade in transition economies has followed a pattern of decline and recovery. In the countries of Eastern Europe, foreign trade declined by 62 percent in the period 1990–93 and then rose to 71 percent of 1990’s export level in 1998. As for trade redirection, the share of the Western countries in the export of the CEE countries has increased tremendously. It nearly doubled from 35.8 percent in 1992–93 to 67.5 percent in 1998–99 (World Bank, 2002, p. 7). For the CIS countries this share did not change at all: it remained at 28 percent. On the other hand, the

**Figure 71.1 GDP growth rates for four regions in CEE and the FSU, 1990–2003**

form. CMEA trade collapsed. Appropriate exchange rate changes to promote competitiveness proved to be a useful help in redirecting and promoting foreign trade. For example, the Czech Republic was successful in redirecting its trade to the West by devaluing its currency by 50 percent. In contrast, East Germany could not benefit from devaluation at all. Instead, it suffered from a managed real revaluation. Given the one-to-one substitution of the West mark for the East mark, in practice the competitiveness of East Germany badly worsened, creating mass unemployment.

Fourth, the sectoral composition of GDP in current prices changed during transition in favor of services and away from manufacturing. Largely this was due to the relatively larger contraction of manufacturing during recessions. However, productivity gains in the manufacturing sector and, therefore, declining relative prices of manufacturing products during economic recoveries also contributed to this sectoral change.

Fifth, during the transition the participation rates in the labor markets changed rapidly. Under communism, the participation rates were high, in particular for women. During the transition there was a significant increase of flows out of the labor force, especially for women and older persons near to retirement age. At the same time, officially registered unemployment rates increased much faster in Eastern Europe relative to the FSU countries. Labor also moved in great numbers from the state sector to the private sector. In contrast, geographical labor mobility remained very low.

Sixth, the production fall in the formal sector led to sharp increases of the official unemployment rates. Across the transition countries, the employment rates differed widely. The outcome for a specific country strongly depended on whether or not the loss of employment in the old enterprises was compensated by the creation of jobs in new (de novo) enterprises. In transition countries where restoring sustained growth has proved relatively illusive, new enterprises account for a low share of employment and value added – both between 10 and 20 percent (World Bank, 2002, p. xxv). According to the World Bank, in practice the transition process proves only to get momentum when the share of medium- and small-scale enterprises in the national employment is more than 40 percent. The CEE countries reached this percentage in 1996. In this respect, the countries of the FSU stayed behind. In these countries the growth of small-scale enterprises stagnated completely; here the share of small enterprises did not rise above 20 percent. Both the government and the Mafia seem to be responsible for that, as they viewed the new enterprises as cash cows instead of centers of new economic initiatives that have to be stimulated.

Seventh, a major result of the transition has been the emergence of a large informal sector. Partly, this concerns new enterprises, which wish to escape the attention of bureaucrats and tax authorities. Partly, it concerns
enterprises that already existed under the old system. These enterprises undertake substantial volumes of activity ‘off the books’ and make extensive use of barter trade. In addition, there has been a widespread informalization of the labor market. This takes the form of, for instance, unilateral determination of wages and conditions by the management of the firms, regardless of laws about these matters. Other examples are employers that disregard paying wages on time, and employees that do not receive maternity leave, are not protected from dangerous working conditions and are not represented by trade unions.

Eighth, a dramatic result of the transition has been a huge increase in relative and absolute poverty, as well as income inequality. According to the World Bank (2002, p. 8), between 1990 and 1998 the population living on less than US$1 a day in the CEE and CIS regions increased from 1.5 to 5.1 percent. It was the only area in the world for which this share increased in that period (in the poorest area worldwide, that is, sub-Saharan Africa, the share remained stable at 47 percent). The norm of US$1 is in effect not appropriate for the transition region as the costs of living are higher there than in other regions. Think, for example, of the cost of heating. Adapting the poverty line for region-specific costs, the World Bank estimates that between 1987–88 and 1993–95, the share of the poor in the total population in the FSU and CEE rose from 3 percent to about 25 percent, and in number of persons from 7 million to 89 million. This means that the transition process pushed more than one-fifth of the population below the poverty line. A group of the population that was particularly badly affected was children. Pensioners also suffered heavily. As a survival strategy, the population in all countries involved put much time and energy into cultivating food in the many allotments. For example, in 1996 this line of food production was estimated to yield 43 percent of Russia’s total food production. The deterioration of the living conditions led also to demographic changes. There has been a striking increase in mortality, concentrated among adult men in Russia and Ukraine. This increased the gender gap in life expectancy (in Russia women have a life expectancy that is 13 years longer than men). There was also a dramatic decline in birth rate and an increase in emigration. The combined effects of the current trends in mortality, birth rate and migration are that the population in many countries in CEE and the FSU declines.

Poverty increased not just because of the fall in output, but also due to greater income inequality in all European and FSU transition countries. Table 71.1 contains these developments over time in the years 1987–98, expressed through the Gini coefficients. Without any exception, the countries included in the table had a higher Gini coefficient, and thus a more unequal income distribution, at the end of that period. The CIS exhibited,
Table 71.1 Income inequality during transition

<table>
<thead>
<tr>
<th>Countries</th>
<th>Gini coefficient of income per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSB</td>
<td>0.23</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>0.23</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.36</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.19</td>
</tr>
<tr>
<td>Estonia</td>
<td>0.24</td>
</tr>
<tr>
<td>Hungary</td>
<td>0.21</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.24</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.23</td>
</tr>
<tr>
<td>Poland</td>
<td>0.28</td>
</tr>
<tr>
<td>Romania</td>
<td>0.23</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.22</td>
</tr>
<tr>
<td>CISa</td>
<td>0.28</td>
</tr>
<tr>
<td>Armenia</td>
<td>0.27</td>
</tr>
<tr>
<td>Belarus</td>
<td>0.23</td>
</tr>
<tr>
<td>Georgia</td>
<td>0.29</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>0.30</td>
</tr>
<tr>
<td>Kyrgyz Republic</td>
<td>0.31</td>
</tr>
<tr>
<td>Moldova</td>
<td>0.27</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>0.26</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>0.28</td>
</tr>
<tr>
<td>Turkmenistan</td>
<td>0.28</td>
</tr>
<tr>
<td>Ukraine</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Notes:
— Not available.
a. Median of countries with data.
CSB is the acronym of Central and South-East European countries and the Baltic States.


on average, a much larger increase than the CSB (Central and South-East European countries and the Baltic States). The Gini coefficient of Hungary hardly increased, whereas Armenia showed the largest increase. The picture of changes in the Gini coefficient in the years after 1998 (until 2003) is divergent: for example, Poland and Romania experienced further increases in the coefficient, for Hungary there is no change, while Russia and Armenia exhibit substantial declines in the coefficient (see, World Bank, 2005, p. 15).

Ninth, growth of crime and the widespread criminalization of society has been a significant feature of transition. This has been particularly
marked in the FSU. In some countries, kleptocracy has been an important part of the political system. In these countries, there are also close links between the criminal, political and business worlds. This is connected with the inability of the state to perform even its night watchman functions. Hence, the supply of property protection and rule enforcement are privatized, that is, taken over by criminal organizations.

Tenth, during transition the old institutions collapsed, creating an institutional vacuum (Schmieding, 1993; Knaack, 1996). As the transition experience has advanced, it has become clearer that the disparity between the rapid progress in liberalization and privatization and the slow development of institutions that support markets and private enterprises directly and negatively affects overall economic performance.

China
The reform processes in the countries of Eastern Europe and the former Soviet Union can be better evaluated when we compare them with the reform process in China. It is interesting to compare, for example, Russia and China, because there are so many similarities between them as economies in transition and yet the outcomes of the transition process were so different. As said before, the fall in GDP in Russia in the period 1991–94 was more than 40 percent, while in the period 1978–2002 the annual growth rate in China was 9.4 percent. During this period in China, the per capita income of rural and urban population, measured at constant prices, increased 5.3 and 4.7 times, respectively (Lin, 2004). How can we explain these differences?

There are many similarities between Russia and China as economies in transition (Buck et al., 2000). These include the enormous geographical scale, abundance of cheap labor, and large potential markets to attract foreigners. In addition, both countries inherited similar economic and political ideologies from their Stalinist and Maoist periods, with a common emphasis on Party control, high defense budgets, large industrial enterprises, heavy bureaucratic and tariff protection against manufactured imports, and subsidized public services.

Of course, there were important differences too. The main difference was the economic structure. On the eve of the start of the transition process (in Russia in 1991 and in China in 1978) China was largely an underdeveloped and rural country with 71 percent of the workforce active in agriculture, despite heavy industrialization efforts in the 1950s and 1960s. Russia was an industrialized country with only 13 percent of the population working in agriculture. China’s financial system was also relatively underdeveloped.

To some extent, these differences in initial conditions and structural characteristics can explain the divergence of transition policies. Since
agriculture was the biggest sector, accounting for 37 percent of output and 71 percent of employment, agriculture was the biggest Chinese economic problem in 1978. It explains why the transition process started in agriculture and gradually spread to industry. When China started its transition in 1978, initially the government did not question the feasibility of the old system. Its attempt was simply to improve the system by giving agents in collective farms and state enterprises some autonomy so that a closer link between personal rewards and individual efforts could be established. Agricultural prices were raised substantially, mandatory quotas for sown area and output eliminated, and compulsory procurement quotas reduced, with the sale of above-quota output on free markets and increased possibilities for so-called sideline activities. Finally, the commune system gave place to the family responsibility system, accounting for 94 percent of the peasant households in 1984. In response to these reforms, per capita grain production rose by one-sixth between 1978 and 1985, the per capita output of cotton rose by two-thirds, and that of oilseeds, sugar and tobacco doubled (Balassa, 1987, p. 411).

With respect to industry, the Chinese government adopted a dual-track policy, liberalizing the entries to the labor-intensive sectors, but also creating conditions to address the viability issue of those firms in the capital-intensive high-priority sectors. The enterprises that were most able to take advantage of the entry option were collective enterprises sponsored by local governments, particularly in rural areas. These township and village enterprises had started already in the Maoist period, but expanded rapidly after 1978. They already produced 33 percent of total industrial output in 1991. Private firms then produced 11 percent, implying that the share of the state sector reduced to a mere 56 percent of total production (Naughton, 1994, Table 1). These developments in industry had two effects. The intensified competition resulted in the disappearance of monopoly rents and the occurrence of a division of labor between the state and non-state sectors. The state sector increasingly consists of large firms in heavy industry, while the non-state sector consists of medium-sized and small firms in light industry.

The success of the reforms in agriculture and industry had a positive effect on the macroeconomic stability of the economy. Formal rationing was in place for more than 20 consumer goods in 1978, accounting for more than 50 percent of consumer expenditures. Shortages of even the basic commodities were common. Shortages of consumer goods disappeared quickly, even though price controls remained in place in China. These policies had two effects: there was a significant increase in living standards and, therefore, an increase of popular support for the reform measures. Due to high saving rates of the Chinese population, there was no collapse of
investment in China, but a dramatic change in the way these investments were financed. Savings by government and by state enterprises has dropped sharply, while savings by private business and households has increased sharply. By the late 1980s, households were saving 15 percent of their income, compared with 2 percent before 1978.

The dynamics of the non-state enterprises exerted a heavy pressure on the state enterprises and triggered a state policy of managerial reforms. These reforms had some success. For example, during the 1980s the output of the state enterprises grew 7.7 percent annually – though other ownership forms of production exhibited even more rapid growth rates. Total factor productivity also grew in the state sector, although less than in the non-state sector. As a result, the state sector is still less competitive than the non-state sector. The relatively low export levels in the state sector compared with the export levels of comparable firms in the West are a clue to that (Buck et al., 2000, p. 393). The continued government subsidization and protectionism of the state enterprises through low-interest loans and monopolistic practices is both cause and consequence of this disappointing competitiveness. The government hopes to eliminate these distortions in the near future.

China also put into effect a policy of opening up to the outside world, but the scope of the opening to the outside world was only expanded gradually. The first step was to set up special economic zones in the south near Hong Kong and Macao, as laboratories for market reforms. Enterprises in these zones had the right to retain most of the foreign exchange they earned and were more flexible in firing workers. Furthermore, foreign investors in these zones could repatriate profits and own land. After the successful performance of these zones, a number of coastal cities were opened, followed by economic areas along some rivers, such as the Yangtze River, and along the borders, and finally capital cities of the various provinces were turned into open cities. In the first years, it was especially the ‘overseas Chinese’, including Taiwanese, who made use of the new possibilities. Later China could welcome a huge influx of foreign direct investment (FDI), also compared, for example, with Russia. In the period 1989–95, Russia attracted FDI to the amount of $3.9 billion, or $1.1 per capita, in 1995. In the same period, China received $121.7 billion of FDI, equivalent to $18.2 per capita (Buck et al., 2000, p. 384). Virtually all FDI was in the form of joint ventures. Joint ventures were made possible, although the government limited foreign ownership and control of businesses. In the late 1970s, foreign involvement was limited to 35 percent of any venture. This was raised to 49 percent in 1985, while from 1988 onwards majority foreign ownership is allowed, though still subject to state approval. The joint ventures were crucial for the success of the Chinese reform process, because the foreign
companies delivered the technological knowledge for the Chinese partners to produce goods that could compete on the world market.

An important difference between China and the CEE and the FSU is that in China the Communist Party kept control, while in the other transition countries the power of the Communist Party dissolved in favor of a democratic political system. The main argument of the Chinese leadership is that the continued rule of the Communist Party guaranteed social stability, which is conducive for the success of the economic reform. Leaders occupied with competing for power would create uncertainty about whether or not the economic reforms would be continued. This is not to say that there were no political reforms in China. Actually, economic reform often contains some elements of political reform. In the case of China, there was devolution of power from the center to the provinces. This empowerment of the regions has created what is now the major driving force behind economic reform (Woo, 1994, p. 289).

Lessons to be learned
For most economists and politicians the depth and duration of the depression which accompanied the transition process in CEE and the FSU came as a surprise. The initial idea was that the removal of the overwhelming apparatus of political control over economic activity could only imply additional prosperity in the medium term. The previous system was characterized by a myriad of distortions, and the removal of most of them would lead to a vigorous impulse to output. This optimism was not a monopoly of neoclassical economists. Well-known heterodox economists, like Janos Kornai, also held this view (Campos and Coricelli, 2002). This raises the question of how to explain this anomaly.

Economic stabilization
As mentioned before, in order to suppress open and hidden inflation, most transition countries adopted a standard package of restrictive fiscal and monetary policy (sometimes supported by exchange rate and income policy). Experience has shown that, despite widespread initial skepticism, such packages – if persisted in – are successful in reducing macroeconomic instability (Ellman, 2005). However, the time for the package to work may be far longer than anticipated. For example, Poland implemented the stabilization package at the beginning of 1990, but only nine years later inflation fell below 10 percent. Moreover, this restrictive macroeconomic policy may contribute to a sharp decline in output and welfare, as happened in Russia after the unsuccessful attempt at shock therapy by Prime Minister Gaidar. As a result, Kornai (1994) argued – writing on the Hungarian situation where inflation was about 20 percent – that growth must be the main
economic objective when the danger of accelerating inflation does not exist anymore. Stiglitz (1998) strengthens this view, summarizing the evidence that only high – and not moderate – inflation is costly: When countries cross the threshold of 40 percent annual inflation in an upward direction, they fall into a high-inflation, low-growth trap. Below that level, there is little evidence that inflation is costly. Recent research even suggests that low levels of inflation may improve economic performance relative to what it would have been with zero inflation.

**Optimal sequencing**

Economic stabilization is a prominent part of the discussion on the optimal order of reforms. The transition process of the 1990s has intensified and widened this discussion, which previously was concentrated on the reform process in developing countries. It concerns reform on three levels of aggregation. On the highest level of aggregation, it is about the optimal order over time of economic liberalization, economic stabilization, privatization and the creation of supporting institutions, necessary for a smooth working of the markets. On a lower level of aggregation, the optimal-sequencing discussion focuses the order over time of the different parts of economic liberalization, in combination with economic stabilization. Here one distinguishes the liberalization of domestic goods and labor markets, international trade, domestic financial markets, international capital flows, and the foreign exchange market. On the lowest level of aggregation, the focal point is the optimal order of domestic sectoral reform: agriculture prior to industrialization, or perhaps the other way around?

A very useful empirical analysis of the transition lessons of the optimal order of the highest aggregation level is Beyer (2001). His data set consists for each of 14 CEE countries of the months in which they switched to the new regime, the majority of their prices were liberalized, their most substantial attempt for stabilization was undertaken, and a new or adjusted constitution was adopted (as an indicator of a country’s new institutional system). By using groups of countries with a similar order of reform it appears that the sequence over time of stabilization, privatization and liberalization is significantly the best sequence in terms of GDP level six years after the system switch. Beyer dubs this sequence the graduality approach. Slovenia and Hungary have followed this adjustment path. These two countries started the reform process with constitution-building. If instead liberalization took place together with stabilization at the start of the reform process, Beyer considers the adjustment process would be a ‘big bang’ approach. The countries that belong to this group (Albania, Bulgaria, the Czech Republic, Poland and Slovakia) display a worse development of GDP relative to the gradualists, but a significantly better outcome than the
transition countries that liberalize and/or privatize before stabilization. Early privatization appears to work badly, as Belarus, Romania, Russia and the Ukraine witness. However, Estonia and Lithuania are counter-examples.

The previous paragraph gives the important clue for optimal sequencing at the lower level of aggregation: that stabilization should in any case not come after liberalization. This gives a strong support to the earlier literature on optimal sequencing for developing countries, which concluded that stabilization should be carried out first, followed by liberalization. The standard outcome of that literature with respect to the optimal sequence within liberalization is: domestic goods and labor markets, international trade simultaneously with unifying the exchange rate and realizing the equilibrium value of that rate, domestic financial markets, and finally, freeing international capital flows from border restrictions.

The transition gives some support to this optimal order, though only of an anecdotal character. The anecdotes that follow have to do with the space for the exchange rate to find its equilibrium value in time. In former East Germany, the unification with West Germany implied a ‘big bang’ liberalization and the introduction of a unified, but highly overvalued currency from former East Germany’s viewpoint. The politically motivated choice of a one-to-one conversion of the East mark into the West mark resulted in a huge deterioration of competitiveness of former East Germany and a concomitant disaster with respect to its GDP, creating mass unemployment. This outcome gives support to the earlier-mentioned optimal liberalization order, which requires that the exchange rate is able to reach its equilibrium value in the process of trade opening.

The collapse of the CMEA trade soon after the start of transition and the resulting loss of jobs in the big state enterprises had to be counterbalanced by an increase of exports to mainly the West and the creation of new jobs by small and medium-sized enterprises. The Czech Republic, for example, was successful in both respects. The large devaluation of its currency resulted in a strong swing of its foreign trade to the West and the process of ‘small’ privatization contributed to strong employment growth in the private sector. Obviously, the Czech Republic also profited from its geographical position and the possibilities of the tourist industry, especially in Prague.

Like East Germany, Russia has opted for, ultimately, a less successful road. As with the Czech Republic, it devaluated its currency in 1992, but made the mistake to choose a more or less fixed exchange rate of the rouble against the dollar under conditions of high internal inflation. The cause was that Russia had already liberalized international capital movements before the economy was sufficiently stabilized. The potential instability was the lax policy stance on fiscal deficits. Large capital inflows initially allowed
the government to finance a continuing fiscal deficit at relatively low interest rates. This can be considered the so-called good equilibrium (see Gros and Steinherr, 2004, p. 243). However, in the meantime, for Russia the debt-to-GDP ratio increased and so did the country’s debt service burden. This development gradually undermined the country’s credibility. Reinforcing simultaneous developments were a growing overvaluation of the rouble, in response to the large capital inflow, and inflationary pressure, due to capital inflow as long as the central bank stabilized the exchange rate. Both reinforcing developments usually worsen the country’s current account over time – in a gradual, but inevitable way. The real appreciation of the rouble did not lead so much to a deterioration of the trade balance, given the strong export potential of the gas and oil reserves. But this appreciation resulted in the crowding-out of the Russian industrial production in line with ‘Dutch disease’ features. Industry became more and more expensive and lost its possibilities to export. The industrial loss of sales became still more severe because the real rouble appreciation stimulated the Russian consumers to opt for cheaper foreign consumer goods. In this state the country was ripe for a shift in expectations leading to the so-called bad equilibrium (see Gros and Steinherr, 2004, p. 243). Given the openness of the capital account, Russia was exposed to sudden withdrawals of foreign capital. This fear became reality during the summer of 1998, after which the rouble collapsed and a severe economic crisis occurred.

With respect to the third level of aggregation and the concomitant optimal sequencing of sectoral reform, the different reform roads of China and Russia are informative. From the success of the economic reform process in China, some economists concluded that the Chinese road of agriculture first was also applicable to Russia. By starting the reform in industry, Russia was unable to gain the credibility that probably would have come if it had started the reform process in agriculture. This position is debatable. The situation in China and Russia was quite different. In China, the agricultural sector was the biggest sector, accounting for 37 percent of the output and 71 percent of employment. In Russia, the agricultural sector employed only 13 percent of the labor force and generated 18 percent of gross national product (GNP). Moreover, relative to China the Russian agriculture was much more mechanized. It is much easier to assign property rights to the individual plots that farmers have been working on than to assign property rights to the capital equipment that workers have been using jointly.

However, the argument gains weight when we include in agriculture the activities of the townships and village enterprises. In the 1990s they already produced 33 percent of industrial output. Including the private sector, the enormous dynamism of the non-state sector had a positive effect on the
supply of consumer goods and the living standards of the population, and therefore created popular support for the reform measures. A rapid growth of the non-state sector can also absorb the unemployed in the state sector. This happened also in the Czech Republic. In the period 1989–1995, about 25 percent of the workers in the state sector left that sector. They could easily find new jobs in the new private sector, especially in the new service sector. Especially in the Prague area, the new private sector showed amazing growth. A recent report of the World Bank (2002) confirms the importance of the stimulation of the non-state sector in an early phase of the reform process. According to this report, a key for economic growth in transition countries is the shift from capital-intensive to labour-intensive enterprises. The last group consists overwhelmingly in small enterprises (with a maximum of 50 workers). According to the World Bank, the transition gets momentum when the share of medium- and small-scale enterprises in the national employment is more than 40 percent. China fulfils this condition, as well as to a lesser extent some CEE countries.

Compared with China and the Czech Republic, the position of the small and medium-sized enterprises in Russia is delicate. In the years 1995–97 employment in Russian businesses with a maximum of 50 employees fell officially by 50 percent. Disappointing productivity growth cannot be an explanation. On the contrary, these firms often had good performance (Commander et al., 1996, Chapter 8). The true explanation is the ‘grasping hand’ of the Russian government and the Mafia. They compelled small enterprises to pay highly unpredictable taxes and regular payments to their ‘protectors’. This explains why the transparency of government behavior and a reduction of risks in the business environment are important determinants of the success of the reform process.

Institutions
After the fall of the Berlin Wall in 1989, in principle all CEE countries followed a liberalization process directed at the breaking down of the planning systems. It was expected that markets would arise spontaneously as soon as the old planning bureaucrats disappeared. In other words, the policy-makers expected that a process of ‘organic growth’ would create the political and economic institutions necessary for the smooth functioning of a market economy. Obviously they trusted that the fundamental propensities of human nature to ‘truck, barter and exchange one thing for another’, as postulated by Adam Smith, were not foregone during the decades of communist rule (Knaack, 1999, p. 357).

However, they did not take into account the fact that proper functioning markets require an institutional infrastructure and that it takes a lot of time before the new institutional system and the persons who have to work in
those markets are adapted to the new circumstances. A clearly delineated system of property rights; a regulatory apparatus curbing the worst forms of fraud, anti-competitive behavior and moral hazard; a moderately cohesive society exhibiting trust and social cooperation, the rule of law and clean government – these are the social arrangements that economists usually take for granted, but which were absent in the transition economies (Rodrik, 2000). Not only must new institutions be created, but they must also prove their value during a time-consuming process of trial and error. Each economic transition process is fundamentally an incremental process, during which the country constantly experiments with new forms and finally keeps that form which is satisfactory. In this way, the existing institutional structure actually improves.

The abolition of the old planning system in one stroke without the construction of new institutions of a market economy has irrevocably led to an institutional vacuum. That vacuum has many forms (Knaack, 1999, p. 363). The old rules lose their value, but the enterprises have not yet learned how to behave in the new situation. Further, the information structure of the old system disappeared, while the new market signals were not yet fully developed. For the enterprises it was difficult to find new customers, and when they finally succeeded it was difficult to assess their creditworthiness. As a result, the enterprises operate in an environment characterized by an extreme uncertainty.

It is obvious that the institutional vacuum must be filled. In the CEE countries and the countries of the FSU this happened in different ways. In the case of East Germany, the country took over in one stroke the institutions of West Germany. The Czech Republic profited heavily from the neighborhood of the large markets of Germany and Austria, and the country also learned quickly from the international trade relations. Moreover, from 1995 onwards the Czech Republic gradually adopted the _acquis communautaire_, the legal structure of the European Union. It must be stressed that imported blueprints do not do their work instantaneously. The main reason is that blueprints are highly incomplete. Much of the knowledge to operate with the blueprints has not been written down and has to be learned. However, blueprints give a direction for knowledge acquisition.

Russia did not have these advantages. For a big country, it has a surprisingly huge international trade. However, nearly all export is energy and raw materials. Actually, these are the features of a small and open developing country. With regard to the possibility of the import of institutions it only had to fulfil the requirements of the IMF when it borrowed some money. Consequently, much more than the other small CEE countries, Russia had to fill the institutional vacuum on its own terms. Given the fact that the
creation of new institutions is a time-consuming process, one can understand that it fell back on its old routines and that, given the weakness of the state, organizations such as the Mafia also filled the vacuum.

Compared with the CEE and FSU countries, the problem of the institutional vacuum in China was less acute. First, we have to remember that after the reforms of 1978 the overwhelming majority of the economic relations in China was still shaped and guided by the official planning system. Only in a very gradual way was there a shift from the planning system to more market relations. Consequently, firms and individuals had enough time to adapt to the new circumstances and to learn the rules of a market economy. Second, insofar as the enterprises in the economic zones had to obey immediately the rules of the market, they could profit from the experiences, knowledge and economic networks of the so-called ‘overseas Chinese’, businessmen especially from Hong Kong, who invested heavily in the zones. Third, sales in the big cities of agricultural surpluses and light industry products of the village enterprises need relatively little organization and structure. Permission from the local authorities to sell on a street corner or a square is sometimes enough. This also explains why, for example, in the Czech Republic it took so little time to start small enterprises in the service sector.

Political reform

There is an intensive debate about the relationship between political regime type and economic performance. Based on the experience in a handful of economies in East and South-East Asia, which (until recently at least) registered the world’s highest growth rates, under authoritarian regimes, one could conclude that economic development requires a strong hand from above. To embark on self-sustained growth, deep economic reforms are often needed, which cannot be undertaken in the messy pull and push of fragile democratic politics. The main argument is that economic reform necessarily imposes costs on some segments of society, and that political openness would provide the losers with the opportunity to form coalitions to stop the reform. An example of this occurred in 1992 in Russia when the apparatchik Chernomyrdin replaced the reformist Gaidar. The first new acts were to squeeze the thousands of small shops that had appeared since January 1992 and to extend cheap credits to the industries under the Ministry of Oil and Gas that Chernomyrdin had headed (Woo, 1994, p. 288). A strong and committed leadership can also push economic reforms against the interests of some interest groups. For example, Buck et al. (2000) describe that the Chinese Communist Party stimulated joint ventures with foreign partners against possible dissent of insiders of state enterprises excluded from the deals. The central authorities stimulated
foreign investors to select from the state enterprises the best physical and human assets to form joint enterprises, usually geographically separated from the unreformed parts of the state enterprise left behind. Normally, the incumbent manager and workers repel any outside investor, unless they are prepared to give employment guarantees for all branches of the enterprise, including those involved with social provisions.

This line of thought met a lot of criticism. Rodrik (2000), for example acknowledges that in effect the Asian countries have prospered under authoritarianism, but that many more have seen their economies deteriorate – think of Zaire, Uganda or Haiti. Moreover, some of the most successful economic reforms in South America were implemented under newly elected democratic governments – witness the stabilizations in Bolivia (1985), Argentina (1991) and Brazil (1994). Moreover, the transitions in the democratic European countries were more successful than the transitions in the authoritarian FSU countries, like Belarus, Turkmenistan and Uzbekistan. It is obvious that we cannot subtract from these examples a clear-cut answer about the relationship of political and economic reform. But it is evident that the reform process is helped when a strong government is able and willing to create the necessary market institutions and resist the interest, especially of the insiders in the state enterprises.

Conclusions
After the fall of the Berlin Wall in 1989, one country after another in Central and Eastern Europe freed itself from Soviet domination and started a transition process from a centrally planned economy into a market economy. In this transition process, they followed the recipe from the IMF and the World Bank, developed earlier for developing countries, mainly in Latin America. In all transition countries in Central and Eastern Europe, the results were rather disappointing. In 1999, only three of all these transition countries surpassed the 1989 national income levels. The new countries that belonged to the former Soviet empire underwent an even more severe income fall. The national income reductions were much more profound than initially expected. When we compare these figures with China’s experience, the difference is striking. China started its reform process in 1978, and for many years had double-digit positive growth figures without any intermediate fall.

The length and depth of the recession in most countries can be explained by the fact that the reform process was based on an incomplete theory about the functioning of a market economy. The policy-makers recognized too late the precondition for the functioning of a market economy, namely an institutional infrastructure, and the dynamics of the reform process, namely that it takes time before the new institutional infrastructure and the
persons who have to work in it are adapted to each other and the new situation. The collapse of the old planning institutions placed the enterprises in an extremely uncertain situation, in which it was difficult to find new customers and to decipher how trustworthy they were. From this perspective the length and depth of the depression depended on the time it cost to build new institutions, for example the new private property rights, and the time for the market players to adapt to them.

Our study also reveals that the negative aspects of transition can be compensated for, and in the Chinese case even more than fully compensated for if the authorities allow structural flexibility. This takes two forms. First, the speed with which the de novo enterprises can expand is important for the success of the transition process. This change is observable in the Czech Republic and especially in China. According to the World Bank the transition gets momentum if the share of medium-sized and small-scale enterprises in national employment is more than 40 percent. This condition is fulfilled in China. In Russia, to the contrary, the de novo enterprises were unable to expand. The government could not protect the new enterprises against the negative practices of the Mafia and the already existing big enterprises. Behaviour of the government itself, such as an erratic tax burden, also was counterproductive. Second, the loss of the export market that the COMECON was before the regime switch had to be offset by an increase of exports to mainly the West. The Czech Republic was particularly successful in this respect. The strong devaluation of its currency resulted in a strong swing of foreign trade to the West. China’s export possibilities to the West also profited strongly from an undervalued domestic currency, the yuan. In contrast, Russia did badly in the 1990s due to an overvalued rouble, leading to a crowding-out of the traditional industry.

Besides these institutional lessons, which are also applicable to developing countries, the transition process in Eastern European countries and the newly independent countries that arose from the former Soviet Union also produced some lessons about the order of reform measures. Institutional adjustments and economic stabilization, both early in the reform process, prove to have positive effects on a rapid restoration of the pre-transition national income level. Late stabilization, in contrast, is devastating in this respect. A late adjustment over time of the exchange rate system towards more flexible – and thus less rigid, disequilibrium – exchange rates appeared to be economically costly in the transition countries: witness the negative experience of former East Germany and Russia, and the positive experience of the Czech Republic. This is an implicit support for the optimal sequence of liberalization steps as developed before the transition experience of the 1990s.

Summing up, the success of a transition process, and thus a development policy in general, not only depends on the building of a viable market
sector. It also depends on the existence of a strong government that is able and willing to create the necessary market institutions, fight the vested interests, and formulate an economic policy that aims at an immediate and thorough economic stabilization. Privatization and full liberalization can come later. As soon as domestic markets function, a rapid opening of international trade and stimulus of the international trade relations, including the introduction of a unified exchange rate which closely approaches its equilibrium value, should be focal points of economic policy.

Notes
1. For example, latent nationalism was fuelled by new publications about the Chernobyl catastrophe and the contents of the Molotov–Ribbentrop pact.
3. CMEA is the group of countries that belonged to the Council for Mutual Economic Assistance. The latter was the body that was supposed to govern trade among Soviet-bloc nations.

References


Index

‘1-2-3’ model, see dependent economy model

Acemoglu, D. 266–70, 385–6
Afghanistan 101
Africa 24, 458–9, 463
poverty in 220
skilled worker migration from 122
Africa, sub-Saharan 103, 125, 152, 459, 463, 490–516
brain drain from 119
child labor in 254
children’s health in 250–51
democratization in 512
demographic change in 495, 507–8
development strategies 514–16
economic structure 491–5
gender gaps in education in 237
geographical disadvantages 505–7
growth experience 490–98
interpretations of poor growth performance 498–508
poverty in 220, 490, 508
African Development Bank 157
aggregate demand 557–8, 560–61
agricultural productivity and gender inequality 239
agriculture 37, 397–8, 400
in China 588, 593
effects of war on 430
foreign aid and 108
gender issues and 228, 229, 230, 231
in India 557–8, 562
in Russia 593
in South Asia 556
AIDS 154, 221, 224, 225, 251, 383, 435, 507–8
Aitken, B. 142
AK model of growth 4
Amsden, A. 168
Arab Free Trade Agreement 527
Asia 73, 75, 76, 78
gender inequality in 236
Asian Development Bank 157
Asian financial crisis 72, 74–6, 84–5, 463, see also financial crisis in East Asia.
asymmetric information 342
in international capital markets 88
in technology markets 138, 140
international migration and 124–5
balance of payments 151, 451–2
of India 563–4
Baland, J-M. 394–406
Balasubramanyam, V. 22–3, 48–58, 143–4
Bangladesh 556, 565–6
Bank for International Settlements 66
bank lending international 71, 79–88
banks 79–88, 89–91, 385, 453, 525
Bardhan, P. K. 381–93
Barro, R. J. 22, 195, 413, 420
Bates, R. 270, 498–9
Bauer, P. 98
Bebbington, A. 414
behavioral economics 389
Benabou, R. 201
Berg, E. 498–9
Beyrer, J. 591
Bhagwati, J. 5–6, 51, 170
Birdsall, N. 193–212, 461
Blecker, R. 170
Bloom, H. 24, 26
Bloom, D. 126
bonds
GDP-indexed 65
local currency-denominated 65–6
Borensztein, E. 144
Borjas, G. 125
Boserup, E. 225–6
Brady plan 481
brain drain, 119–22, 132–3, 171
Branstetter, L. 145
Braunstein, E. 242–3
Brazil 72, 154, 156
ethnic/racial inequality in 264
Bretton Woods institutions 150–53, 159
Bruton, H. 30–47
buffer stocks 27
bureaucracy 52
Burgstaller, A. 170
Burundi 265
Canada
ethnic/racial inequality in 264–5
capabilities 312, 418, 420
capital account liberalization 531
in India 563
capital accumulation 32, 35, 166–7, 349–50, 458–9
capital flows see international capital flows
capital markets 382, 384, 525
capital mobility 295
capital requirements 68
Caribbean
skilled worker migration from 121
CEPAL see ECLAC
centrally planned economies,
lessons from 282–4
Chang, H-J. 367
Chen, S. 219
child labor 253–4, 255–6, 259–60
children
and development 250–61
effect of female education on 238
Chin, J. 173
China 54, 78, 165, 457, 463, 536–50,
579, 587–90
comparison to India 537–8
FDI in 51–3
gender issues in 234–5
growth experience 536–42
income distribution in 189–90, 541
investment levels 539
legal system in 387–8
population control in 538–9
poverty in 220, 536, 541
private sector participation 543–9
state owned enterprises in 542,
544–5, 548
total factor productivity 539–41
Chiswick, B. 124
civil war 265, 423–41, 511 see also war
class 390–91
classical economists 165
Coase theorem 342, 385
coefficient of variation 182–3
collective action 398–9
Collier, P. 270, 426, 427, 498, 501–2
Colombia 427
colonies 31, 167, 267–9, 529
commodity agreements 158
Commonwealth of Independent States
(CIS) 579 see also transition economies
comparative advantage 16, 21–2, 31,
153, 167
Compensatory Financing Facility
(CFF) 151
Computable General Equilibrium
(CGEl Models 347–8, 354
conditionality
for foreign aid 111–13
IMF 151–2
contagion, financial 60, 81, 88
contract enforcement 203, 411
contracts, informal 388
convergence, of income across
countries 169, 172
corporate governance 387
corruption 199–200, 369–77, 386
costs of 371–2
in East Asia 574–5
foreign aid and 107, 113
and growth 290, 372
levels and types of 370–71
measurement of 369–70
policies towards 372–6
and size of government 374
cost-benefit analysis 32, 351–3
Cote d’Ivoire 425
credit markets 198–9, 384, 389
crime 203
in transition economies 586–7
cross-country growth regressions 10–12
culture 407–22
definition of 409
cycles, effect of international capital
flows on 90–94
Dahi, O. 522–35
Damon, A. 250–61
Darity, W.A. 262–77
data problems 215–16, 217–18, 229–30
ternational migration and 120
<table>
<thead>
<tr>
<th>Term</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>De Soto, H.</td>
<td>411</td>
</tr>
<tr>
<td>Deardorff, A.</td>
<td>173</td>
</tr>
<tr>
<td>debt crisis</td>
<td>88, 151, 454</td>
</tr>
<tr>
<td>debt relief</td>
<td>101</td>
</tr>
<tr>
<td>deflation</td>
<td>297</td>
</tr>
<tr>
<td>Deininger, K.</td>
<td>208</td>
</tr>
<tr>
<td>Demir, F.</td>
<td>522–35</td>
</tr>
<tr>
<td>democracy</td>
<td>408, 412, 531–2</td>
</tr>
<tr>
<td>and war</td>
<td>427–8</td>
</tr>
<tr>
<td>dependency theory</td>
<td>167, 478</td>
</tr>
<tr>
<td>dependent economy model</td>
<td>322–4, 331–2</td>
</tr>
<tr>
<td>Development Assistant Committee (DAC)</td>
<td>99</td>
</tr>
<tr>
<td>discrimination</td>
<td></td>
</tr>
<tr>
<td>against women</td>
<td>231</td>
</tr>
<tr>
<td>economic</td>
<td>197</td>
</tr>
<tr>
<td>ethnic/racial</td>
<td>264</td>
</tr>
<tr>
<td>distortions</td>
<td>5–6</td>
</tr>
<tr>
<td>market</td>
<td>32–3, 37–8</td>
</tr>
<tr>
<td>policy-induced</td>
<td>6</td>
</tr>
<tr>
<td>divergence, in growth</td>
<td>455–6</td>
</tr>
<tr>
<td>division of labor</td>
<td></td>
</tr>
<tr>
<td>international</td>
<td>287</td>
</tr>
<tr>
<td>Doha development round</td>
<td>151, 153, 154, 155</td>
</tr>
<tr>
<td>dollarization</td>
<td>298</td>
</tr>
<tr>
<td>Douglas, M.</td>
<td>417</td>
</tr>
<tr>
<td>dual economy</td>
<td>36</td>
</tr>
<tr>
<td>Dutch disease</td>
<td>43, 322, 333, 523, 526</td>
</tr>
<tr>
<td>foreign aid and 108</td>
<td></td>
</tr>
<tr>
<td>remittances and 132</td>
<td></td>
</tr>
<tr>
<td>Dutt, A.</td>
<td>22, 163–76</td>
</tr>
<tr>
<td>East Asia</td>
<td>48, 203–7, 459, 569–78</td>
</tr>
<tr>
<td>asset redistribution in 575</td>
<td></td>
</tr>
<tr>
<td>diversity of experiences</td>
<td>571–5</td>
</tr>
<tr>
<td>foreign direct investment in 571–2</td>
<td></td>
</tr>
<tr>
<td>industrial policy in 572–3</td>
<td></td>
</tr>
<tr>
<td>interpretations of growth experience</td>
<td>569</td>
</tr>
<tr>
<td>lessons from 576–7</td>
<td></td>
</tr>
<tr>
<td>trade policy in 572–3</td>
<td></td>
</tr>
<tr>
<td>Easterlin, R.</td>
<td>194</td>
</tr>
<tr>
<td>Easterly, W.</td>
<td>98, 201–2, 264, 269, 270, 411</td>
</tr>
<tr>
<td>Economic Commission for Latin America, (ECLAC)</td>
<td>158, 474–6, 478</td>
</tr>
<tr>
<td>education</td>
<td>199, 209, 217</td>
</tr>
<tr>
<td>of children</td>
<td>251–3, 256–8</td>
</tr>
<tr>
<td>effects of war</td>
<td>435</td>
</tr>
<tr>
<td>inequality of 206</td>
<td></td>
</tr>
<tr>
<td>in Latin America</td>
<td>484–5</td>
</tr>
<tr>
<td>in MENA 528–9</td>
<td></td>
</tr>
<tr>
<td>policy 224, 438</td>
<td></td>
</tr>
<tr>
<td>policy, for children 259</td>
<td></td>
</tr>
<tr>
<td>primary 207</td>
<td></td>
</tr>
<tr>
<td>in South Asia</td>
<td>556</td>
</tr>
<tr>
<td>university 206</td>
<td></td>
</tr>
<tr>
<td>efficiency 55, 342–3</td>
<td></td>
</tr>
<tr>
<td>of private firms and state-owned enterprises compared 361–4, 366–7</td>
<td></td>
</tr>
<tr>
<td>Elson, D.</td>
<td>227–49</td>
</tr>
<tr>
<td>enclaves 460–61</td>
<td></td>
</tr>
<tr>
<td>Engel’s law</td>
<td>20</td>
</tr>
<tr>
<td>Engerman, S.</td>
<td>201, 385</td>
</tr>
<tr>
<td>England 381</td>
<td></td>
</tr>
<tr>
<td>entitlements, individual</td>
<td>430–32, 434–5</td>
</tr>
<tr>
<td>entrepreneurs 359</td>
<td></td>
</tr>
<tr>
<td>environment 173–4</td>
<td></td>
</tr>
<tr>
<td>environment, and war</td>
<td>428, 429</td>
</tr>
<tr>
<td>environmental Kuznets curve 173</td>
<td></td>
</tr>
<tr>
<td>environmental planning models 349</td>
<td></td>
</tr>
<tr>
<td>equity 342–3, 360</td>
<td></td>
</tr>
<tr>
<td>Ethiopia 24</td>
<td></td>
</tr>
<tr>
<td>ethnic diversity 270–75</td>
<td></td>
</tr>
<tr>
<td>ethnic fractionalization</td>
<td>504</td>
</tr>
<tr>
<td>ethnic/racial conflict 264–6, 275</td>
<td></td>
</tr>
<tr>
<td>ethnicity 262–77, 423–4, 531</td>
<td></td>
</tr>
<tr>
<td>ethno-linguistic fractionalization (ELF) index 270–74</td>
<td></td>
</tr>
<tr>
<td>Europe, Central and Eastern 579, 581</td>
<td></td>
</tr>
<tr>
<td>see also transition economies</td>
<td></td>
</tr>
<tr>
<td>evolutionary economics 405–6</td>
<td></td>
</tr>
<tr>
<td>exchange rate 206, 243, 245, 526</td>
<td></td>
</tr>
<tr>
<td>devaluation of 335</td>
<td></td>
</tr>
<tr>
<td>as a nominal anchor 297–8</td>
<td></td>
</tr>
<tr>
<td>policy 40, 42, 293, 295–6, 526</td>
<td></td>
</tr>
<tr>
<td>stability 296</td>
<td></td>
</tr>
<tr>
<td>targets 301, 336</td>
<td></td>
</tr>
<tr>
<td>exchange rate regime</td>
<td></td>
</tr>
<tr>
<td>fixed 296–8</td>
<td></td>
</tr>
<tr>
<td>floating 296–8</td>
<td></td>
</tr>
<tr>
<td>intermediate 296–8</td>
<td></td>
</tr>
<tr>
<td>export processing zones</td>
<td></td>
</tr>
<tr>
<td>in East Asia 570</td>
<td></td>
</tr>
</tbody>
</table>
export promotion 10, 21, 50–51, 144
in East Asia 570
export-oriented industrialization
women and 232
exports 33, 35, 39, 40, 54
effects of war on 430
and gender inequality 239
of India 561
manufacturing 449–50, 459
technology transfer and 137
externalities 130, 167, 168, 169, 342,
384, 395, 396, 399, 400–401, 447
from skilled emigration 132–3
factor price equalization 168
families,
bargaining power of women in 232,
235
migration decision by 126
famines 418
Fel’dman model 350
female education
effect on children 238
effect on fertility 237
effect on growth 236–7
effect on population growth 237
fertility 237
finance 388
financial crisis 71, 72, 74–5, 454, 531
in East Asia 62, 575–6
in Mexico 72, 74–6, 82–3
in Russia 62
financial liberalization 88, 89, 90, 91–4,
452–3, 458, 463
in East Asia 576
financial markets 452–3
women and 240–41
financial volatility 60, 71, 89–90
effects on growth 59–60
Findlay, R. 166
fiscal deficit 312–13
relation with growth 313
in Latin America 477, 481–2
in Sub-Saharan Africa 510
fiscal policy 91, 305–17
characteristics in developing
countries 305–11
and gender equality 241–3, 244
in India 558
and privatization 360–61
procyclical 311
in South Asia 555, 557
fisheries 402–3
Fishlow, A. 469–89
FitzGerald, V. 319–40
food security 438
foreign aid 98–118, 430, 436–8
amounts of 100
definition of 99
donor motives for 105
donors of 101, 104
foreign policy and 104
recipients 101–3
tyling of 104–5, 110
types of 99, 110
foreign debt 170
foreign direct investment 48–58, 61, 64,
170, 172, 448, 451
in China 589
determinants of 49–53
in East Asia 571–2
effect of 53–6
effects of war on 433
export oriented 143
and gender equality 244
import substituting 143
in Latin America 476–7
tax incentives for 54
and technology transfers 138–46
foreign exchange 54, 106, 447–8
effects of war on 430
Foster-Greer-Thorbecke index 215
fragility, economic 77–8
France, legal system in 387
Friedman, B. 410
Friedman, M. 98, 390
Fukuyama, F. 413–14
gap models 106, 350
gender,
and development 227–49
distribution, effect of growth and
development on 227–36
distribution, effect on growth and
development 236–41
inequality, in India 563
General Agreement on Trade and
Tariffs (GATT) 150
genocide 265–6
Ghosh, J. 554–68
Gibson, B.  341–55
Gini coefficient 179, 182–3, 195–7, 208, 463
Glass, A.  137–49, 173
Glewwe, P.  250–61
global warming 174
globalization 71, 119, 234, 288, 445–6, 473, 485–6, 577
Goulet, D.  417–19
governance 134, 388
  and foreign aid 113
  and growth 290
government 39, 343, 436
  failure 201, 286, 343
  policy intervention 39
  revenue 359
  foreign aid and 107–8
  in transition economies 596
government expenditure 209
  effects of war on 433–4
  reform 316
Gray, C.W.  369–80
greed and grievance, and war 426, 429
Griffith-Jones, S.  59–70
Grilli, E.  27
gross domestic product
  -indexed bonds 65
Grossman, G.  173
growth 23–4, 30, 33, 37, 88, 165, 203–7, 447, 450, 454–6,
  effect of brain drain on 133–4
  effect of ethnic diversity on 270
  effect of foreign aid on 105–10
  effect of income distribution on 194–201
  effect of international capital flows
    on 90–94
  effect on ethnic/racial inequality 262
  effect on gender inequality 235–6
  effect on well-being of children 254–8
  effect of war on 432
  in India 560–61
  in MENA 524, 529–30
  models 349–50
  in South Asia 555
  of transition economies 582
  relation with life expectancy 223–4
  relation with literacy 223–4
  relation with poverty 223–4
Hagen, E.  412
Hamilton, A.  168
happiness 194, 207–8
Harrison, A.  142
Harrison, L.  411, 412, 416
Harrod-Domar model 32, 36, 106, 321
Hatton, T.  125–6
Heal, G.  342, 349
health 199, 217–18, 383
  of children 250–51, 256–8
  effects of war on 435
  foreign aid and 106, 107
  impact of brain drain 132
  in MENA 528–9
  policy 224, 257, 438
  policy, for children 258
  in South Asia 556–7
  in transition economies 585
Heckscher, E.  31
Heckscher-Ohlin-Samuelson model
  129–30, 167, 168
Heinz, J.  242–3
Helpman, E.  173
HIV 221, 224, 225, 435, 507–8
Hoksbergen, R.  407–22
Homer-Dixon, T.  428
Hong Kong 463
horizontal inequality, see inequality, horizontal
Huang, Y.  52–3, 536–53
human capital 55–6, 133, 144, 171, 270
  effect of war on 432
  in MENA 528
Human Development Index 218, 418
  for Bangladesh 566
  ethnic/racial inequality in 263–4
Hymer, S.  49–50
identity 409, 416
immigration policy 127
imperfect information 389
imperialism 167
import substitution 10, 21, 32, 33, 36,
  42, 43, 50–51, 144, 288, 523, 524–5
  in East Asia 570
  in India 558
  in Latin America 477–8
income, measurement of 215–16
income distribution 193–212, 364, 398,
  see also income inequality
effect of migration on 129
income inequality
aggregation problems in measuring 181–5
in China 541
comparisons across countries 184–5
comparisons over time 185
determinants of 185–8
in different regions 183–4
and economic growth 188–90
in Latin America 484–5
measurement of 179–91
policies towards 190
increasing returns to scale 130, 167, 168, 169, 171, 386, 403–4, 446–7
India 156, 203, 415, 557–64
FDI in 51–3
gender issues in 230, 234–5
information technology sector in 133, 561, 563
international migration from 120
legal system in 387–8
uneven pattern of poverty reduction in 221
Indonesia 139, 425
industrial policy 168, 460
and gender equality 244
industrialization 188, 475–6
industrial sector, gender issues and 228
industry, in India 558
inefficiency 33, 395, 396
inequality 194–201, 399, 557, see also
income, income distribution between women 236
in distribution of education 199–200
effect on growth 194–201
ethnic/racial 262–4
horizontal 208–9, 425, 426, 428
in India 562
international 164–5, 456–7
international migration and 124
in landownership 199
in MENA 531
North–South 30
in South Asia 555–6
in transition economies 585–6
vertical 427
inflation 295, 301, 306
effects of war on 433
in Latin America 477, 481
in Sub-Saharan Africa 510
inflation targeting 294, 298–9, 325–30
and central bank independence 300
experience in developing countries 299–301
prerequisites 299–300
informal employment 232, 240
informal sector 37, 433–4, 437
in transition economies 584–5
information technology 133
in India 561, 563
infrastructure, effects of war on 432
innovation 134, 172, 385
input–output models 344–5
institutional economics 381, 385
new 394
institutions 41–2, 194, 201, 202, 266–9, 385–6
dysfunctional 385–6
in transition economies 587, 594–6
insurance markets 198, 389
intellectual property rights 138, 145–6, 172, 382–4
Inter-American Development Bank 157
International Bank for Reconstruction and Development (IBRD) 104, 150–52, 157, see also World Bank
international capital flows 55, 59–68, 71–118, 169–70, 328, 448–9, 452–3, 463, see also foreign direct investment
boom–bust cycles of 59–61, 64
effects of war on 430, 432–3
main characteristics of 60–64
policies towards 64–8
regulation of 66–7, 77, 88, 94, 156, 295
to India 564
international capital markets 88
International Development Agency (IDA) 151
International Development Association (IDA) 104
international institutions 150–62
international migration 119–36, 171
effects of 128–33
growth in 120
returns to 123–4
<table>
<thead>
<tr>
<th>Index</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Monetary Fund (IMF)</td>
<td>53,</td>
</tr>
<tr>
<td>66, 111–12, 114, 151, 153,</td>
<td>155–</td>
</tr>
<tr>
<td>7, 159, 243, 288, 319, 530–31, 581</td>
<td>159</td>
</tr>
<tr>
<td>Basic Financial Programming Framework 320–21</td>
<td></td>
</tr>
<tr>
<td>international production chains 451, 460</td>
<td></td>
</tr>
<tr>
<td>international trade 3–13, 31–48, 167–9, 287, 448</td>
<td></td>
</tr>
<tr>
<td>effects on income distribution 7</td>
<td></td>
</tr>
<tr>
<td>effects on growth 4, 8–12</td>
<td></td>
</tr>
<tr>
<td>effects of war on 433</td>
<td></td>
</tr>
<tr>
<td>gains from 4–8, 16, 167</td>
<td></td>
</tr>
<tr>
<td>liberalization 11, 38–9, 172, 449–52, 482, 510–11</td>
<td></td>
</tr>
<tr>
<td>liberalization, and gender equality 243</td>
<td></td>
</tr>
<tr>
<td>mainstream view of 3–4</td>
<td></td>
</tr>
<tr>
<td>in MENA 526–8</td>
<td></td>
</tr>
<tr>
<td>policy 21–2, 30–47, 168, 225, 460</td>
<td></td>
</tr>
<tr>
<td>of transition economies 583–4</td>
<td></td>
</tr>
<tr>
<td>International Trade Organization 150</td>
<td></td>
</tr>
<tr>
<td>investment 35, 36, 197, 451, 458</td>
<td></td>
</tr>
<tr>
<td>effects of war on 432</td>
<td></td>
</tr>
<tr>
<td>foreign aid and 106, 108</td>
<td></td>
</tr>
<tr>
<td>public 458–9</td>
<td></td>
</tr>
<tr>
<td>Iraq 101</td>
<td></td>
</tr>
<tr>
<td>Islam 533</td>
<td></td>
</tr>
<tr>
<td>Jager, H. 579–600</td>
<td></td>
</tr>
<tr>
<td>Jha, R. 305–18</td>
<td></td>
</tr>
<tr>
<td>joint ventures 144</td>
<td></td>
</tr>
<tr>
<td>Jomo K. S. 569–78</td>
<td></td>
</tr>
<tr>
<td>Kaldor, N. 207</td>
<td></td>
</tr>
<tr>
<td>Kaminsky, G. 71–97</td>
<td></td>
</tr>
<tr>
<td>Keen, D. 426</td>
<td></td>
</tr>
<tr>
<td>Kerala 222, 224</td>
<td></td>
</tr>
<tr>
<td>Keynes, J. M. 27, 150</td>
<td></td>
</tr>
<tr>
<td>Kindleberger, C. P. 54</td>
<td></td>
</tr>
<tr>
<td>Knaack, R. 579–600</td>
<td></td>
</tr>
<tr>
<td>knowledge,</td>
<td></td>
</tr>
<tr>
<td>codified 36</td>
<td></td>
</tr>
<tr>
<td>tacit 41, 43, 44</td>
<td></td>
</tr>
<tr>
<td>Korea, South 21, 39, 40</td>
<td></td>
</tr>
<tr>
<td>Kornai, J. 580</td>
<td></td>
</tr>
<tr>
<td>Korten, D. 416</td>
<td></td>
</tr>
<tr>
<td>Kozul-Wright, R. 445–68</td>
<td></td>
</tr>
<tr>
<td>Krugman, P. 168, 171–2</td>
<td></td>
</tr>
<tr>
<td>Kuran, T. 533</td>
<td></td>
</tr>
<tr>
<td>Kuznets curve 188–90</td>
<td></td>
</tr>
<tr>
<td>Kuznets, S. 188, 193</td>
<td></td>
</tr>
<tr>
<td>labor markets 123–4</td>
<td></td>
</tr>
<tr>
<td>effects of war on 430</td>
<td></td>
</tr>
<tr>
<td>in MENA 528–9</td>
<td></td>
</tr>
<tr>
<td>labor, skilled 55</td>
<td></td>
</tr>
<tr>
<td>Lall, S. 56</td>
<td></td>
</tr>
<tr>
<td>Lamb, J. 420</td>
<td></td>
</tr>
<tr>
<td>land 394, 397, 402</td>
<td></td>
</tr>
<tr>
<td>markets, women and 231</td>
<td></td>
</tr>
<tr>
<td>reform 201, 386, 557</td>
<td></td>
</tr>
<tr>
<td>rights, women and 230–31</td>
<td></td>
</tr>
<tr>
<td>titles 382</td>
<td></td>
</tr>
<tr>
<td>Landes, D. 413</td>
<td></td>
</tr>
<tr>
<td>Latin America 38, 203–7, 459, 469–86</td>
<td></td>
</tr>
<tr>
<td>capital flows to 72, 73, 75, 76, 78</td>
<td></td>
</tr>
<tr>
<td>gender inequality in 236</td>
<td></td>
</tr>
<tr>
<td>gender issues in 230</td>
<td></td>
</tr>
<tr>
<td>growth experience 470–83</td>
<td></td>
</tr>
<tr>
<td>law</td>
<td></td>
</tr>
<tr>
<td>and development 381–93</td>
<td></td>
</tr>
<tr>
<td>civil 387</td>
<td></td>
</tr>
<tr>
<td>commercial 388</td>
<td></td>
</tr>
<tr>
<td>common 387</td>
<td></td>
</tr>
<tr>
<td>contract 381</td>
<td></td>
</tr>
<tr>
<td>rule of 381–2, 386, 391</td>
<td></td>
</tr>
<tr>
<td>learning by doing 167, 168, 172, 384</td>
<td></td>
</tr>
<tr>
<td>Leontief model 345</td>
<td></td>
</tr>
<tr>
<td>Leontief, W. 344</td>
<td></td>
</tr>
<tr>
<td>Levine, R. 264, 270</td>
<td></td>
</tr>
<tr>
<td>Lewis, W. A. 36, 166, 227–8, 275, 286</td>
<td></td>
</tr>
<tr>
<td>liberalization 445, 581</td>
<td></td>
</tr>
<tr>
<td>in India 560–61</td>
<td></td>
</tr>
<tr>
<td>in MENA 531–2</td>
<td></td>
</tr>
<tr>
<td>in South Asia 545–55</td>
<td></td>
</tr>
<tr>
<td>licensing, technology 137, 139, 144</td>
<td></td>
</tr>
<tr>
<td>life expectancy 219–24</td>
<td></td>
</tr>
<tr>
<td>linear programming models 346–7</td>
<td></td>
</tr>
<tr>
<td>linkages 188–9</td>
<td></td>
</tr>
<tr>
<td>liquidity 77</td>
<td></td>
</tr>
<tr>
<td>List, F. 160</td>
<td></td>
</tr>
<tr>
<td>literacy rate 219–20, 223–4</td>
<td></td>
</tr>
<tr>
<td>loan-pushing 170</td>
<td></td>
</tr>
<tr>
<td>lock-in 167</td>
<td></td>
</tr>
<tr>
<td>Lorenz curve 182, 184, 463</td>
<td></td>
</tr>
<tr>
<td>Lucas, R. 171</td>
<td></td>
</tr>
<tr>
<td>lump sum transfers 6–7</td>
<td></td>
</tr>
<tr>
<td>Lutz, M. 23–4</td>
<td></td>
</tr>
</tbody>
</table>
macroeconomic policy 225
and gender distribution 241–9
Mahalanobis model 351
Mahalonobis, P. C. 561–2
Malaysia 54, 78–9
Mali 24
Malloch, T. 414
Mansfield, E. 140
market 41, 342, 354, 394, 401, 408–9, 446–8
failures 198–9, 285–6, 342
imperfections, capital 126–7
role in development 281–91
structure 20
and technological progress 282
in transition economies 594
Marshall Plan 150, 157–8
Masson, P. 293–304
Mattoo, A. 153, 154, 155
McCleary, R. 413, 420
McClelland, D. 411–12
McNamara, R. 151
Mexico 72, 400
financial crisis 72, 74–6, 82–3
migration from 120, 130
microfinance 240
middle class 201–2
Middle East and North Africa
(MENA) 522–35
migration, determinants of 122
Milanovic, B. 165
military expenditure 434, 530
Millennium Development Goals 157, 237
Mishra, P. 130
modernization theory 407, 410, 416
monetary policy 91, 293–303
and development 293–4
and domestic nominal anchors
298–9
and gender equality 241–3
independence 294–5
see also stabilization policy
Montalvo, J. 273–4
moral hazard, in technology markets
140
Moran, T. 143
Morgan, W. 27
Morley, S. 179–92
mortality, child 107
mortality, infant 250–51, 255–6, 425, 435
mortality rate
adult 435
child (under 5) 219–20, 250–52, 255–7
multinational firms, see transnational
corporations
mutual funds 72–9, 95
Myrdal, G. 417
natural disasters 105
natural resources 394, 395, 396, 397, 400–402, 404, 450, 461, 526
war and 426, 428
Ndulu, B. 490–521
need-for-achievement 411
neoclassical approach 167, 168
Nepal 425, 567
new growth theory 22, 167, 168, 171, 350
New International Economic Order
(NIEO) 159
newly-industrialized countries (NICs)
174, 460
Nigeria 101
ethnic/racial inequality in 264
non-government organizations
(NGOs) 99, 416, 434, 437
Norback, P.-J. 139
norms, social 202, 389
North Africa 552–3
North America 201
North, D. 371
North–South
inequality 164–5
models 21, 145, 165–7
relations 160, 163–76
trade 167–9
Nurkse, R. 286
nutrition
effects of war on 434–5
in South Asia 556
nutritional status, of children 257
O’Connell, S. 490–521
official development assistance 99
Ohlin, B. 31
oil 523, 526
Okedihi, T. 274
Olson, M. 399, 411
open access 394, 395, 398
openness 38–9, 450, 460
growth and 462
Organization for Petroleum Exporting Countries (OPEC) 158–9
outsourcing 561
outward-oriented development 40, 524

Pack, H. 141
Pakistan 564–5
patents 383, 391
Patnaik, P. 554–68
pharmaceuticals 383
planning 341–55, 594–5, 598
economy-wide 342–4
in India 561–2
models 341, 344–50
see also centrally planned economies
Platteau, J-P. 394–406
polarization 273
policies
and effects of foreign aid 108
to promote child welfare 257–60
towards economies in conflict 437–8
policy 343, see also under education,
fiscal, health, industrial, monetary
and trade policy
planning models and 353–4
political constraints, on privatization 365–6
political economy
of India 560
of migration policy 127
of South Asia 564
political instability 200, 529
political reform, in transition economies 596–7
political voice 217–18
politically relevant ethnic groups index 271–2
pollution 173
population 237, 428
portfolio flows, international 71
Posner, D. 271–2
poverty 213–26, 235, 417–18
absolute 216
in China 536, 541
foreign aid for reducing 104
gap 214

in India 563
line 214–17, 568
measurement of 214–15
in MENA 531
one dollar a day 216–17, 221
rate 219
relative 216
in sub-Saharan Africa 490, 508
targeted programs 325
in transition economies 585
two dollars a day 217, 221–2
and violent conflict 427
war and 428
Prebisch, R. 18, 21, 27, 158, 475, 478
Prebisch-Singer thesis, 18, 20–22, 25–6, 27, 158
primary commodities, exporters of 25
income inelasticity of 20
price inelasticity of 20
principal agent problem, foreign aid and 111, 113–14
Pritchett, L. 165
privatization 208, 356–9, 364–7, 400, 401, 403, 533, 581
in Latin America 482
reasons for 360–64
in transition economies 592
productivity 446
growth 34, 141, 459
project appraisal 351–2
property rights 38, 201, 202, 267, 381, 389, 395, 401, 404, 405, 411
property, common 389, 396, 400, 401
protection conditional on export promotion 570–71
import 38, 153, 201
import in OECD countries 153
psychological factors 410–12
public goods 282, 399
global 174
public investment 206
public sector 524
purchasing power parity 216
Putnam, R. 202, 264, 269, 413–14
Puttermann, L. 281–92
Quah, D. 164
Quesnay, F. 344
race 262–77
Radelet, S. 98–118
Raffler, K. 150–62
Ramchandran, V. 140
Randomized experiments 258–9
Ravallion, M. 220
Rayment, P. 445
reforms,
in China 589–90
in MENA 530–32
sequencing of, in transition economies 591–4
regional planning models 350
Reinhart, C. 88
religion 412–14, 417, 420, 423–4
remittances,
migrant 119, 127, 130–32, 523
migrant, to South Asia 563–4, 565
rent-seeking 33, 201, 202, 264–5, 525, 532
Research and Development 55, 138, 382–3
reverse engineering, 55, 137
Reynal-Querol, M. 273–4
Ricardo, D. 15–17, 21, 31
risk
   indicators 77–8
   political 529
Rivera-Batiz, F. 119–36
Rodríguez, F. 3–15
Rodrik, D. 152, 200–201, 382
Romeo, A. 140
Rostow, W.W. 286
Roy model 124
Russia 594
   financial crisis 62, 74–6
Rwanda 265
Saavedra-Rivano, N. 170
Sachs, J. 98
Saggi, K. 137–49, 173
Sala-i-Martin, X. 164
Samuelson, P. 168
sanctions, economic 438, 529
Sapsford, D. 16–29
Sarkar, P. 19
saving 197
   foreign aid and 106, 107, 108
   private 324
relation with growth 313
school attendance 217
school enrollment
   gender gap in 229
   primary 217, 219, 251–3, 435
   secondary 253, 255
Schumpeter, J. 407–8
Seguino, S. 236, 238
self-employment 230, 240
services, in India 563
shadow price 346, 351–2
Sierra Leone 225
Singapore 54, 463
Singer, H. 18–21, 23–4, 27, 170
Sjaastad, L. 122
skilled workers 119, 120–22, 124, 130, 133, 168, 171
slavery 167, 267, 270
Smarzynska, B. 140
Smith, A. 16, 165, 207, 381, 394
Social Accounting Matrices 347
social capital 202–3, 413–15
   effects of war on 430
social conflict 200
social diversity index 274
social inclusion 217–18
Sokoloff, K. 201, 385
Solow growth model 43
South Africa 154
South America 201
South Asia 459, 554–68
   child labor in 254
   children's health in 251
   poverty in 200
South East Asia 460
Soviet Union 341, 580–81
speculative attack 297, 302
spillovers
   technology 55–6, 141–3, 145, 452, 462
   wage 142
spiritual capital 414
Spraos, J. 27
Squire, L. 207, 208, 213–26
Sri Lanka 222, 224, 566
Srinivasan, T. N. 5–6
stabilization 285, 581
   effect on women's employment 233
   new Keynesian approach to 330
policy 319–37  
in transition economies 590–91  
Stark, O. 126  
state 427  
capture 386  
as coordinator of development 287  
developmental 208, 385, 462, 524  
in India 557, 562  
intervention 390  
in MENA 531–3  
role in development 281–91  
state-owned enterprises 356, 359–60, 524–5  
in China 542, 544–5, 548, 589  
in East Asia 572–3  
reasons for establishing 359–60  
Stern, N. 98, 152  
Stewart, F. 423–41  
Stiglitz, J. 98  
stock markets 77, 89, 90, 95, 365, 384  
effect of financial liberalization on, 81  
Stolper-Samuelson theorem 7  
Streeten, P. 56–7  
structural adjustment 112, 152, 288, 319–37, 356, 458  
effect on women's employment 233  
and income distribution 335–6  
in MENA 530–32  
structuralist CGE models 347–8  
subjective well-being 194, 207–9  
Subramanian, A. 153, 154, 155  
subsidies  
on FDI 50, 54  
in MENA 532  
to Northern farmers 154  
surplus labor 286, 330  
Taiwan 39, 40  
Tanzania 203  
tariffs, import 32, 50  
tax buoyancy 310  
tax-GDP ratio 306  
tax reform 313–16  
Taylor, J. E. 126  
Taylor, L. 166  
teacher absenteeism 259  
technological capability 172  
technological change 20–21, 56, 134, 403, 447–8  
technology 53, 385  
diffusion 145  
transfer 36, 53–5, 137–49, 171–2, 384, 449, 462  
transfer, foreign aid and 106  
terms of trade  
agriculture-industry 559  
between primary products and manufactured goods 17–20  
international 16–29, 20–21, 451, 461  
net barter 17, 26  
North–South 18–19, 31, 170  
volatility of 23–4  
Thompson, E. P. 391  
Throsby, D. 411, 418  
total factor productivity 141, 142, 384  
growth of 33, 206  
township and village enterprises 588, 593–4  
trade, see international trade  
trade, free 39  
Trade-related Intellectual Property Rights (TRIPs) 154, 383  
Trade-related Investment Measures (TRIMs) 154  
transaction costs 389, 394, 397, 401–3, 405  
transition economies 579–600  
capital flows to 73, 76–8  
Triplett, R. E. 262–77  
trust 202  
Turkey 524–5, 530–31  
underemployment 233, 242  
unemployment 233, 242, 459  
in MENA 528  
in South Asia 556  
in transition economies 584  
uneven development 22, 166–8, 169, 171  
United Nations 157  
United Nations Conference on Trade and Development (UNCTAD) 158–9, 453  
United Nations Development Programme (UNDP) 158, 218, 263
unpaid work, women and 230, 231,
232, 233, 234
urban bias 499
Uruguay round 154, 155, 172
USA 154
legal system in 387
migration to 130
Value at Risk models 60–61
values 409, 412–13, 418, 419
Venezuela, technology spillovers in 142
Vietnam, well-being of children in 257
violence 203
violent conflict 423–41
causes of 423–9
consequences of 429–37
in South Asia 567–8
Voluntary Export Restrictions 154
von Neumann model 345
wage inequality, gender 232
Walrasian CGE models 347
war 225, 423–41, 529–30, see also
violent conflict
foreign aid and 107
Washington Consensus 37, 39–40, 42,
44, 483–4, 577, 581–2
water, in MENA 526
Weber, M. 386, 407
Weil, D. 413
Weingast, B. 381
Wilber, C. 407–22
Williams, E. 266–70
Williamson, J., 44
Williamson, J. G. 125–6
Williamson, O. 415–16
women, missing 234–5
Women-in-Development literature 228,
229, 231
Wood, A. 168
working time, for men and women
234
World Bank 27, 38, 104, 109, 111–14,
158, 216, 243, 319–21, 414, 457–8,
530, see also International Bank
for Reconstruction and
Development (IBRD)
interpretation of East Asian
miracles 569–70, 572
Revised Minimum Standard Model
321–5
World Food Program 158
World Health Organization 383
World Trade Organization (WTO) 11,
48, 150, 153–6, 158, 159, 172–3,
383
Yang, M. 27
# Contents

## PART VI INTERNATIONAL ISSUES

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>International trade and development</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td><em>Francisco Rodríguez</em></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Terms of trade and economic development</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td><em>David Sapsford</em></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Trade policy and development</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td><em>Henry J. Bruton</em></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Foreign direct investment</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td><em>V.N. Balasubramanyam</em></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Private capital flows and development</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td><em>Stephanie Griffith-Jones</em></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>International capital flows to emerging economies: short- and long-run effects</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td><em>Graciela L. Kaminsky</em></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Foreign aid</td>
<td>98</td>
</tr>
<tr>
<td></td>
<td><em>Steven Radelet</em></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>International migration and the brain drain</td>
<td>119</td>
</tr>
<tr>
<td></td>
<td><em>Francisco L. Rivera-Batiz</em></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>International technology transfer: the role of foreign direct investment</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td><em>Amy Jocelyn Glass and Kamal Saggi</em></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>International institutions and development</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td><em>Kunibert Raffer</em></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>North–South issues</td>
<td>163</td>
</tr>
<tr>
<td></td>
<td><em>Amitava Krishna Dutt</em></td>
<td></td>
</tr>
</tbody>
</table>

## PART VII DISTRIBUTION AND POVERTY

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>Measures of the distribution of income and their interpretation in developing countries</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td><em>Samuel A. Morley</em></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>Income distribution: effects on growth and development</td>
<td>193</td>
</tr>
<tr>
<td></td>
<td><em>Nancy Birdsall</em></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Measuring progress in reducing poverty</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td><em>Lyn Squire</em></td>
<td></td>
</tr>
<tr>
<td>Page</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>50</td>
<td>Gender issues in development</td>
<td>Diane Elson</td>
</tr>
<tr>
<td>51</td>
<td>Children and development</td>
<td>Paul Glewwe and Amy Damon</td>
</tr>
<tr>
<td>52</td>
<td>Ethnicity and economic development</td>
<td>William A. Darity Jr and Russell E. Triplett</td>
</tr>
<tr>
<td>53</td>
<td>The role of the state and markets in development</td>
<td>Louis Putterman</td>
</tr>
<tr>
<td>54</td>
<td>Monetary policy</td>
<td>Paul R. Masson</td>
</tr>
<tr>
<td>55</td>
<td>Fiscal policy</td>
<td>Raghbendra Jha</td>
</tr>
<tr>
<td>56</td>
<td>Stabilization policy and structural adjustment</td>
<td>Valpy FitzGerald</td>
</tr>
<tr>
<td>57</td>
<td>Economic planning in developing economies</td>
<td>Bill Gibson</td>
</tr>
<tr>
<td>58</td>
<td>State-owned enterprises and privatization</td>
<td>Anindya Sen</td>
</tr>
<tr>
<td>59</td>
<td>Corruption and development</td>
<td>Cheryl W. Gray</td>
</tr>
<tr>
<td>60</td>
<td>Law and development</td>
<td>Pranab Bardhan</td>
</tr>
<tr>
<td>61</td>
<td>Institutions, property rights and development</td>
<td>Jean-Philippe Platteau and Jean-Marie Baland</td>
</tr>
<tr>
<td>62</td>
<td>Culture and development</td>
<td>Roland Hoksbergen and Charles K. Wilber</td>
</tr>
<tr>
<td>63</td>
<td>The economics of war: causes and consequences</td>
<td>Frances Stewart and Graham Brown</td>
</tr>
<tr>
<td>64</td>
<td>Economic development in the Third World: an international perspective</td>
<td>Richard Kozul-Wright and Paul Rayment</td>
</tr>
<tr>
<td>65</td>
<td>Latin America and the Caribbean</td>
<td>Albert Fishlow</td>
</tr>
<tr>
<td>66</td>
<td>Sub-Saharan Africa</td>
<td>Benno J. Ndulu and Stephen A. O’Connell</td>
</tr>
</tbody>
</table>

PART VIII  THE STATE, INSTITUTIONS AND DEVELOPMENT

PART IX  ECONOMIC DEVELOPMENT EXPERIENCE
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
<th>Authors/Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>67</td>
<td>The Middle East and North Africa</td>
<td>Omar S. Dahi and Firat Demir</td>
</tr>
<tr>
<td>68</td>
<td>China</td>
<td>Yasheng Huang</td>
</tr>
<tr>
<td>69</td>
<td>South Asia</td>
<td>Prabhat Patnaik and Jayati Ghosh</td>
</tr>
<tr>
<td>70</td>
<td>The East Asian newly industrializing countries</td>
<td>Jomo K.S.</td>
</tr>
<tr>
<td>71</td>
<td>Transition economies: lessons for development</td>
<td>Ruud Knaack and Henk Jager</td>
</tr>
</tbody>
</table>

Index: 601
PART VI

INTERNATIONAL ISSUES
In July of 2007 more than 1000 economists – including four Nobel prize winners – signed a letter to Congress petitioning it not to impose tariffs on China. The phrasing of the petition is revealing of the extent to which many economists believe in the beneficial effects of free trade:

As economists, we understand the vital and beneficial role that free trade plays in the world economy. Conversely, we believe that barriers to free trade destroy wealth and benefit no one in the long run . . . There is no foundation in economics that supports punitive tariffs.1

This view of trade is widely shared by the overwhelming majority of economists with a neoclassical training. In a recent survey of graduate students at the six top economics departments in the USA, only 7 percent of respondents disagreed with the statement ‘tariffs and quotas reduce general economic welfare’ (Colander, 2005, p. 189) Of all the policy statements presented in the survey, trade policy was the one that commanded the broadest consensus among students. As Alan Blinder recently commented, in the economics profession ‘anyone who says anything even obliquely that sounds hostile to free trade is treated as an apostate’.2

This consensus is not due to the absence of theoretical results showing the existence of situations in which greater trade can decrease welfare. Students of international trade are commonly exposed to examples of optimal tariffs, infant industries and strategic trade policy, all of which can serve to justify intervention in international trade. Rather, the consensus in the profession emerges from the vision that these examples are not relevant in most real-world circumstances and that the dangers from attempting to carry out activist trade policy far outweigh its potential benefits. As one of the architects of strategic trade theory has himself written:

The gains from intervention are limited by uncertainty about appropriate policies, by entry that dissipates the gains, and by the general equilibrium effects that insure that promoting one sector diverts resources from others . . . It is possible, then, both to believe that comparative advantage is an incomplete model of trade and to believe that free trade is nevertheless the right policy. In fact, this is the position taken by most of the new trade theorists themselves. (Krugman, 1987, p. 143)
Deciding whether to accept this conventional wisdom or to question it is a key step in the framing of a country’s development strategy. The purpose of this chapter is to analyze critically the basis for the pro-trade consensus, and to lay out clearly the arguments of supporters and detractors of activist trade policy. Such an exercise requires a dual discussion of the theoretical literature and the empirical evidence, which I present in the next two sections. As I will argue, a careful examination of these two literatures provides grounds for the raising of serious questions about the desirability of outward-oriented trade strategies for development.

What the theory says
In its simplest version, the gains-from-trade theorem (Kemp, 1962; Samuelson, 1962) states that in the absence of distortions and when lump-sum transfers are feasible, all individuals in an economy can be made better off from an appropriate combination of full liberalization and compensating transfers. This result follows from the basic Ricardian insight that it will be more efficient for countries to shift production from goods for which their opportunity cost of production is high to goods for which it is low.

The static intuition can be readily extended to an intertemporal model of growth, as shown by Smith (1979). In such models, the efficiency effects of trade will show up in higher steady-state levels of consumption and welfare. This extension can generate considerable confusion as openness will not be associated with higher long-run growth rates. This is more than anything a consequence of the fact that the long-run growth rate in a Ramsey-style model of economic growth is determined by the rate of technological progress and unaffected by any other variables. However, openness does generate higher steady-state levels of income and higher growth rates on the transition to these new steady states.

The intuition readily carries forward to endogenous growth models. In an AK model of growth, static efficiency losses generated by distortions turn up as declines in the level of efficiency captured by the productivity term in the production function. Trade policy thus has an unambiguously negative effect on the growth rate. In more complex endogenous growth models that seek to endogenize productivity as a consequence of decisions to innovate, the public goods nature of knowledge introduces a distortion. The existence of this distortion opens up the possibility of second-best effects in which trade policy can potentially enhance welfare, a possibility that we discuss in more detail below. A full discussion of alternative models of this type is given by Grossman and Helpman (1991).

In all of these renderings, the gains-from-trade theorem is little more than an extension of the first and second fundamental theorems of welfare economics. The possibility of trading shifts the consumption possibilities
frontier outwards, and the competitive equilibrium allows us to attain any point on that consumption possibilities frontier. The possibility of lump-sum transfers allows us to redistribute the gains from trade in such a way that all individuals can be made better off.

The theorem is thus open to the same caveats as the first and second welfare theorems. In the first place, it is vulnerable to the existence of distortions. One of the earliest recognized distortions in the trade literature is the possibility of the home country being able to affect international prices. In this case the world economy becomes a monopolist in the world market and perfectly competitive levels of production are no longer optimal. In the second place, the result requires the existence of non-distortionary taxes and subsidies that allow the redistribution of the gains from trade. In the absence of these instruments, some individuals may – and generally will – be made worse off by greater openness.

Obviously, in the real world distortions do exist and lump-sum taxes do not. Most of the academic literature’s defense of the gains-from-trade principle as a useful guide to policy action comes from the interpretation of results that appear to indicate that it will generally be suboptimal to deal with distortions through trade policy, and that reasonable approximations to lump-sum transfers exist.

Let us first discuss the issue of distortions. In a set of classic articles, Bhagwati (1971) and Bhagwati and Srinivasan (1969) showed that in the presence of distortions, trade tariffs or taxes will generally not be the optimal way to address the distortions or objectives that are commonly taken as rationales for activist trade policy. The exception is the case in which the distortion is actually generated by the existence of monopoly power in trade, in which the first-best policy is a tariff. But in the case in which there is an externality that leads to underproduction of a given good, the optimal policy is to subsidize the production of that good. A tariff is suboptimal because it is actually a combination of a production subsidy and a consumption tax, and there is generally no reason to expect that one would simultaneously want to increase production and reduce consumption of any particular good. These seminal contributions are thus generally taken as the demonstration that trade policy is a misguided way to address most of the problems generated by the existence of distortions.

The Bhagwati–Srinivasan contributions reflect a vision of policy-making in which the first-best equilibrium can be attained and thus serves as a useful policy benchmark. The ideal policy will be the one that accurately identifies all existing distortions and introduces an optimal intervention to eliminate the effects of each distortion. The existence of a government with the capabilities to carry out this complex exercise is presumed. One way of understanding this presumption is as a reflection of the view that the
majority of distortions are policy-induced. In this case, the policy prescription is clear: a laissez-faire elimination of all government-induced distortions could lead the economy sufficiently close to its first-best optimum to take advantage of the full gains from trade.3

While such a vision of the world is certainly coherent, it is also quite reasonable to hold to an alternative vision of the world as completely ridden with distortions. In this world, interacting complex processes such as innovation, knowledge networks, geographical clusters, public goods and international market power combine to create an economy in which distortions are a fact of everyday life. The task of identifying all such distortions and crafting interventions to address each one of them is beyond the realm of rationally effective policy-making. Even if one could identify all existing distortions, the design of interventions to eliminate some of them may be out of the sphere of possible policy actions by the government because of institutional or political constraints.

In such a world, policy-making should not try to replicate the first-best equilibrium. Rather, it makes sense to think of policy reforms as taking place in a setting of radical Knightian uncertainty, where the expected effects of removing a policy distortion can only be deduced from local experimentation. Hausmann et al. (2004) have recently proposed such an approach to policy reforms. Rather than attempting to eliminate all distortions at once, they suggest that reformers should concentrate on the reforms that have the greatest expected pay-off, given that other distortions are in place. They suggest an empirical method to infer whether certain distortions are in effect binding constraints on growth, and thus to identify whether altering them will lead to greater growth. The resulting method of policy-making is akin to the use of non-linear programming algorithms to search for local maxima, which do so through the search for incremental improvements rather than by the explicit calculation of a global solution.

One of the consequences of taking seriously such a world is that it turns a common free trade argument on its head. Advocates of trade often argue that even though some level of intervention in trade may be optimal theoretically to address trade-induced externalities, identifying such interventions in a context of considerable uncertainty would be beyond the capacities of most governments (Krugman, 1987). However, if we view the real world as ridden by trade and non-trade induced distortions, the same dose of realism that leads us to conclude that the government cannot address all of these distortions also leads us to recognize that we are unavoidably in a second-best world, in which the incremental effects of trade policy on welfare could well be positive.

Let us now turn to the issue of lump-sum transfers. Non-distortionary taxes require conditioning on characteristics that individuals cannot
change. It is doubtful whether such taxes actually exist or not in real life, and even if they do it would be extremely difficult to design them to redistribute the gains from trade—it as the observable characteristics that are out of the control of individuals are unlikely to correlate perfectly or even reasonably well with the benefit or loss from trade that these individuals experience.

The existence of compensation to losers is particularly relevant precisely because trade theory predicts stark effects on income distribution from trade openness. In the absence of compensation, the Stolper–Samuelson theorem predicts that the real return to a country’s scarce factor will decrease with greater openness. In this case developing-country unskilled workers would actually benefit from greater openness, so that trade may be distributively beneficial for poor countries with an abundance of unskilled labor. However, the factor endowments model of trade appears to have little empirical support (see Feenstra, 2004, Chapter 2), so that this may not be the most appropriate theory with which to think about the distributive effects of trade. Alternative theoretical frameworks can produce different predictions concerning income distribution. For example, a set of recent theoretical and empirical contributions (Rodrik, 1997; Reddy and Dube, 2000; Ortega and Rodríguez, 2006) have argued that trade can diminish the bargaining power of unions and thus lead to a decline in labor shares.

In the case of compensation, the literature’s optimism comes from the belief that while lump-sum taxes do not exist, reasonable approximations can be constructed in practice to carry out the necessary compensations to losers. Indeed, it has been shown that lump-sum taxes are not even theoretically necessary in this respect: factor and income taxes will suffice to enact the desired redistribution (Dixit and Norman, 1980, pp. 79–80). This result appears intuitive even if the real-world setting is much more complex than that of our models: it is generally possible to identify—at least ex post—the key groups that gain and lose from trade openness and to design transfer programs to redistribute the gains.

The key question regarding compensation is not whether it is feasible in a technical and operational sense to design and implement such compensation, as it almost certainly is, but whether this compensation is likely to take place in practice. There are a number of political economy reasons why one may expect that such compensation is unlikely to take place. One is that while it may be optimal for the gainers to promise to compensate the losers before the reforms are carried out, such promises are likely to be time-inconsistent, particularly if their gains are protected by some degree of irreversibility in trade reforms. While the manipulation of activist trade policies by interest groups has received considerable attention in the literature (for example, Grossman and Helpman, 1994) and has often been adduced as a reason for
the superiority of simple rules such as free trade, the political economy of compensation arrangements has received much less attention. An early exception can be found in the work of political scientists such as Ronald Rogowski (1987), who argued that increased trade alters the political landscape, making owners of abundant resources much more powerful and assertive, and thus much less likely to accept demands for compensation.4

In sum, whether one considers the free trade case a reasonable one or not on theoretical grounds depends on whether one considers two basic assumptions of the gains from trade theorem reasonable. The first one is that the extent of distortions is sufficiently restricted so that the government can identify them and deal with them through policies designed on the basis of the theory of optimal intervention. The second one is that redistributive policies can and will be implemented to compensate the losers from trade liberalization, particularly when these are the most disadvantaged groups in society.

The decision of whether these two assumptions are reasonable or not on purely theoretical grounds is far from clear-cut. What should be clear is that a critical vision, which is based on skepticism about the appropriateness of the first-best model as a guide for policy and the likelihood of implementation of appropriate redistributive strategies, cannot be deemed insensible on a purely theoretical basis. The belief in the optimality of free trade should thus be based on the belief that the empirical evidence decisively points in favor of a beneficial effect of trade on growth. We turn now to that issue.

**What the data say**

Broadly speaking, the empirical literature that has studied the effect of openness on growth has taken one of two vantage points. The first one is to analyze the correlation between openness and growth in data sets that cover a large section of developing and developed countries, in the tradition of cross-country growth empirics initiated by Robert Barro (1991). The second one is to concentrate on country or region-level analytical case studies of economic growth. Both literatures have been appealed to by proponents and detractors of trade-oriented development strategies. In what follows, I will attempt to shed some light on the reasons behind these differences in interpretation.

The empirical literature on openness and growth is voluminous indeed. I will not attempt to provide a full survey of the main contributions (the interested reader may consult Rodriguez, 2008). Broadly speaking, however, a number of findings appear to emerge from this literature.

First, there is no strong unconditional or conditional correlation between economic growth and a number of direct measures of trade policy,
such as weighted or unweighted tariffs, import quotas or other non-tariff barriers. This point was first made by Rodríguez and Rodrik (2001) and generated some surprise in the literature. It has since been confirmed by, among others, DeJong and Ripoll (2006), who argue that there may be a non-linear relationship where the effect of tariffs on growth depends on the initial level of a country's income and may be positive or negative. In Rodríguez (2007), I have also shown evidence in favor of a non-linear effect, although I have argued that the precise form of that effect may be difficult to discern.

Second, there appears to be a reasonably strong correlation between growth or productivity and the ratio of trade to gross domestic product (GDP), especially when the latter is measured in prices of a constant base year (Dollar and Kraay, 2002; Alcalá and Ciccone, 2003). Some attempts have been made to discern whether this correlation actually embodies a causal relationship. The most well-known attempt, formulated by Frankel and Romer (1999), consists in using instrumental variables estimates of the effect of trade volumes on growth where the latter is instrumented with its geographic determinants as derived from the estimation of gravity equations. These results are controversial – as Rodríguez and Rodrik (2001) and Irwin and Tervio (2000) have shown, they are not robust to controlling for the direct effect of geographical variables on income or productivity. Other attempts to discern causality using alternative methods to instrumental variables do not confirm the existence of causal effect (Rigobón and Rodrik, 2005).

A drawback of using the trade to GDP ratio as an indicator of openness is that it may capture many non-policy-induced changes in trade openness which are largely irrelevant if one is preoccupied with designing a developing country’s trade strategy. Natural resource booms, the emergence of new export sectors, changes in other countries’ trade policies, and changes in foreign aid can all have an effect on the trade to GDP ratio without necessarily having an obvious link to trade policy. In sum, the key problem of the trade–GDP ratio is that it is an indicator of results and not of policy actions. To take just one example, if the infant-industry argument for protection were correct, initial levels of trade protection would lead to the development of productive, competitive domestic industries that would later on be capable of competing internationally. Tariffs would be associated with higher growth, but so would exports. A correlation between trade volumes and growth may thus not be very informative about the desirability of activist trade policies.

Some authors have tried to produce compound measures of trade policy that capture the different ways in which an economy can be closed to international trade. According to these authors, one would not expect to observe
a simple correlation between simple measures of trade policy such as tariffs and economic growth because countries can use many policy devices to impose trade protection, of which only one is import tariffs. The most famous of these measures was provided by Sachs and Warner (1995) and recently updated by Wacziarg and Welch (2003). What these indices actually measure is very controversial. Rodríguez and Rodrik (2001) argue that the Sachs and Warner variable’s effect on growth was purely driven by two subcomponents of the index – black market premia and export marketing boards – which are not obviously linked to trade policy. For example, they argue that the effect of export marketing boards on growth in the Sachs–Warner study comes from the fact that the variable was taken from a 1994 World Bank study called *Adjustment in Africa* that covered only 29 African economies undergoing adjustment programs during the 1980s, leading to the exclusion of non-African or African non-adjusting economies from the sample and strongly biasing the results in favor of a trade–growth correlation. Rodríguez (2008) levies similar criticisms at the Wacziarg and Welch (2003) exercise.

In recent years, there has been growing skepticism of the possibility of establishing strong conclusions regarding causal growth effects using the cross-country regression framework. A growing consensus appears to have emerged around the belief that the problems of causality, robustness and specification are simply too pervasive and difficult to solve in the context of highly aggregated cross-national empirical data. This skepticism has led authors such as Bhagwati and Srinivasan (2001) to discount the aggregate growth evidence altogether, and to call for concentrating exclusively on the evidence from case studies. While these criticisms should be taken seriously, it is important to note that even if one takes the cross-country evidence at face value, accepting the framework without questioning, it does not appear to lend the strongest of supports to the pro-trade view. As in the case of the theoretical literature, it appears to be open to multiple interpretations, some of which are consistent with the view that protection is not unequivocally harmful for growth.

Country-level studies of openness and growth are also open to multiple interpretations. Bhagwati and Srinivasan (2001) cite the Organisation for Economic Co-operation and Development (OECD) and National Bureau of Economic Research (NBER) studies of more than a dozen major developing countries carried out in the 1960s and 1970s, which uncovered key differences between the constraints on economic performance in countries that pursued import substitution strategies and those that pursued export promotion. A revised interpretation of this view was given by the World Bank’s 1993 study *The East Asian Miracle*. Broadly speaking, the key argument of this study was that the openness to trade and reliance on market
forces of East Asian economies played a fundamental role in making possible their sustained growth acceleration.

The World Bank’s characterization of the high-growing East Asian tigers as economies that followed a strategy of free trade has, however, been strongly questioned by several authors. Some of these criticisms were collected in a 1994 volume published by the Overseas Development Council (Fishlow et al., 1994) in which Dani Rodrik, Robert Wade and Stephen Haggard disputed the key findings of the World Bank study. In Robert Wade’s words, ‘the [World Bank’s] report uses standards of inference so elastic that practically anything could be confirmed’ (2003, p. xix)

One of the key points of dispute concerns whether East Asia can adequately be characterized as a region that followed a non-activist trade policy. The World Bank study had concluded that East Asia’s relative prices were closer to international averages than those of other regions, supporting the contention that its international trade was relatively undistorted. Wade pointed out that this is only true when one uses an unweighted average that includes the island economies of Hong Kong and Singapore, where price distortions were necessarily negligible. In contrast, during the 1976–85 period, relative prices in Japan, South Korea and Taiwan deviated more from international prices than those of countries which are generally perceived to have had strong records of intervention, such as India, Pakistan, Brazil, Mexico and Venezuela in the period 1976–85. Similarly, Alice Amsden’s (1992) in-depth study of South Korea’s industrialization contends that the success of its industrial policies was largely due to an active intervention in the determination of relative prices, a strategy that she labels ‘getting relative prices wrong’.

During the 1990s, the set of liberalization experiences that could be the subject of in-depth studies expanded dramatically. Between 1990 and 2002, the average tariff rate in the world went down from 10.5 percent to 6 percent, and the ratio of imports plus exports to GDP rose from 75.2 percent to 86.8 percent (World Bank, 2005a). In 1990, the General Agreement on Tariffs and Trade (GATT) had been signed by 96 countries: between 1990 and 2005, 65 countries joined it either as the GATT or in its most recent incarnation as the WTO.

While the result of these liberalization experiences has not yet been fully analyzed, what is clear is that aggressive trade liberalization proved to be very far from a necessary condition for a growth take-off. Some of the most aggressive liberalizers of this period were former communist economies such as Mongolia, Ukraine and Moldova, which suffered some of the deepest growth collapses in post-World War II history. But openness did not only fail to pay off in the former Soviet Bloc. With the exception of Cuba, the evidence suggests that virtually all Latin American
economies moved in a direction of greater trade liberalization during the 1990s. Yet the region’s growth performance during the post-reform period has been disappointing to say the least, with per-worker GDP and total factor productivity growing respectively at annual rates of only 0.1 percent and 0.2 percent between 1990 and 2002 (Ocampo, 2004). The region is said to have entered an era of ‘reform fatigue’ (Lora et al., 2004) in which voters are increasingly willing to vote for political platforms to roll back reforms.

In sum, neither cross-national empirical studies nor country-level case studies seem to give strong support to the idea that openness is unequivocally good for growth. A reading of the evidence in support of activist trade strategies is certainly possible and indeed has been carried out by reputable mainstream economists. These conclusions mirror our interpretation of the theoretical literature, which can also be interpreted as supporting a case for intervention in trade policy.

**Concluding comments**

One way to explain the apparent divorce between the favorable view that the majority of economists have about free trade and the lessons given by the empirical and theoretical literature is by thinking about free trade as one of the components of our discipline’s ‘hard core’ (in the sense of Lakatos, 1976), a set of beliefs and methodological assumptions that are not considered the appropriate subject of empirical tests. Since these core beliefs are never tested without auxiliary assumptions, any failure to explain the evidence can be handled by altering the assumptions but not the core belief. As a senior faculty member once quipped after seeing a presentation of my work, ‘if the data does not say that trade is good for growth, then the data must be wrong’.

It is not easy for a discipline to abandon or even begin to question a hard-core belief, but neither is it impossible. To take one example, during the 1990s the assumption of rationality has made the transition from a hard-core belief to an auxiliary hypothesis that is not even taken very seriously most of the time. This change has opened up a burgeoning new area of research in behavioral economics which has transformed our understanding of individual economic behavior.

Signs that this may be starting to happen in the study of the relationship between trade and development are beginning to appear. In 2005, the World Bank published a comprehensive assessment of the experience of the 1990s with economic reforms (World Bank, 2005b). The sobering assessment of this disappointing period recognizes that the results of economic reforms were far below what its proponents had expected and rejects the one-size-fits-all approach to reform that the institution espoused during
the greater part of the period in question. On the concrete matter of trade policy, the report concludes that: ‘[w]hile trade reforms can help accelerate integration in the world economy and strengthen an effective growth strategy, they cannot ensure its success’, and ‘the distributive effects of trade liberalization are diverse, and not always pro-poor’ (pp. 131–2). On the fairness of the world trading system, it states that ‘global markets are the most hostile to the products produced by the world’s poor’. As Dani Rodrik wrote in his review of this volume: ‘occasionally, the reader has to remind himself that the book he is holding in his hands is not some radical manifesto, but a report prepared by the seat of orthodoxy in the universe of development policy’ (Rodrik, 2006, pp. 974–5).

A reconsideration of the role of openness in countries’ development strategies would fundamentally alter the nature of the debate on generating and sustaining growth. Whether this occurs will probably depend not only on the internal dynamics of academia, but also on the extent to which outside reality exerts pressure for such a change. Political discontent with the experience of the 1990s is undoubtedly a key reason for the World Bank’s reappraisal of the reform experience. In the same way, the results of the current reassertion of state involvement in much of the developing world are likely to influence deeply the direction that development research will take in the future. Perhaps, to turn Keynes on his head, economists are nothing more than the slaves of long-defunct practical men.

Notes
1. At the moment of writing (August 2007), the process of signature collection is not yet finished and thus the letter has not been published. The text of the letter was obtained through personal communication with Andy Roth of the Club for Growth.
3. Obviously, even if only a few distortions remain there is no theoretical presumption that the resulting equilibrium will be better than the pre-laissez-faire equilibrium. It can be argued that at this point it becomes feasible to target the remaining distortions through optimal interventions. Alternatively, it can be argued that theoretical models are always approximations of the real world, and that a distortion-free model should be a reasonable approximation to a reality in which there is a reduced number of distortions.
4. For a recent exploration of this issue, see Davidson et al. (2006).

References
Amsden, Alice (1992), Asia’s next Giant: South Korea and Late Industrialization, Oxford: Oxford University Press.


Introduction

Why do countries trade with one another? What determines the terms on which trade between countries is conducted in the world marketplace? These two questions are perhaps the most fundamental to be considered in any discussion of international economic relations, be it trade between developed and developing countries or trade amongst countries in either the developing or the developed world. These questions are of especial importance in the context of economic development since if there are ‘gains from trade’ to be earned, the distribution of such gains between trading partners (especially when we are thinking of trade between developed or industrialized countries on the one hand, and developing nations on the other) carries important implications not only for living standards and economic welfare within the participating countries, but also for the continued willingness of developing and newly emerging economies to engage in processes, such as those initiated under the auspices of the World Trade Organization, designed to bring about further reductions in barriers which impede the process of international trade.

The classical economists, most notably Adam Smith (1723–90) and David Ricardo (1772–1823) (see in particular, Smith, 1776 and Ricardo, 1817), initially addressed the first of these questions and their, respective, analyses of absolute and comparative advantage as the basis for trade are widely recognized. Indeed, to this day the Ricardian model, despite its various (over)simplifying assumptions (including that of a two-country, two-good, one-factor world) still occupies centre stage in the economic theory of international trade. Despite its undoubted logical elegance, a major limitation of the Ricardian analysis is encountered when we move to the second of these two questions, because the model leaves us analytically with a range of indeterminacy, somewhere within which the terms of trade (defined in the usual way as the ratio of the price of a country’s exports to the price of its imports) must lie if trade is to offer benefits (in terms of increased output or consumption with unchanged resource endowments) to at least one of the trading partners. This indeterminacy is, in practical terms, more than a mere theoretical loose end, especially when we move
outside of the static world of Ricardo’s model and begin to seek answers to important real-world questions such as who gains most from trade and to what, if any, extent the pattern of winners and losers has changed over time with the evolution that has occurred in both the structure and performance of the world economy. Such questions are of particular importance in the context of the changes that have occurred over the long run in the terms on which trade between developed and developing nations has been conducted, especially since this carries important implications for the pace and nature of economic development in the Third World.

Who gains from trade?
While the elegance of Ricardo’s analysis and its correctness within the confines of its own assumptions can not be faulted it does, as noted, beg a question that is of central importance in the context of the trade that takes place between countries of the developed or industrialized world, on the one hand, and those of the developing or Third World on the other. While the analysis demonstrates quite clearly the potential benefits to trading partners from engaging in international trade in the world marketplace, it has nothing whatsoever to say about the division of these potential gains between them. Suppose that the two countries comprising the Ricardian world are, say, the USA and China, then it follows from the model that if relative prices in the world marketplace (the so called net barter terms of trade) were equal to US internal relative prices then China would effectively appropriate all of the gains from trade for herself, whereas at the opposite end of the spectrum, the USA would scoop all of the gains if Chinese relative prices prevailed.

In order to focus ideas let us consider trade between the countries of the developed or industrialized world and those of the developing world and, for simplicity, assume that the former produce manufactured goods while the latter produce primary commodities. The fact that Ricardo’s analysis did not shed any light on the issue of how the potential gains from trade are shared out in practice did not seem to constitute a problem in the minds of classical economists since in a related context Ricardo, like Smith before him, had argued that as an inevitable consequence of the twin forces of diminishing returns in the production of primary commodities from a fixed stock of land (including mineral resources) as population increased, and the downward pressures on production costs in manufacturing generated by the moderating influences of surplus population and urbanization upon wages, the price of primary products would rise over the long run in relation to the price of manufactured goods, thereby giving rise to an upward drift in the net barter terms of trade between primary commodities and manufactured goods. On the above assumptions this movement
will translate into an improvement in the terms of trade of developing countries vis-à-vis the developed countries. On the basis of this argument there was little, if any, reason to be concerned about the plight of developing countries in the context of their trading relations with the industrialized world since it predicted that over the long run the terms of trade would shift, as a consequence of the workings of the invisible hand of the market, steadily in their favour, with the result that they would in due course enjoy an increasing share of the potential gains from trade. Indeed, casual analysis of the views expressed by many politicians and international policy-makers over at least the last century would seem to suggest an underlying acceptance of this particular prediction of classical economics (Sapsford and Singer, 1998).

The evidence
In the early 1950s, the classical prediction of a secular improvement in the terms of trade of primary commodity-dependent developing countries was challenged by both Prebisch (1950 [1962]) and Singer (1950).3 Both argued forcefully that in direct contravention of the then still prevailing classical orthodoxy, the terms of trade faced by primary producers had actually as a matter of statistical fact been historically subject to, and could be expected to continue to be subject to, a declining as distinct from an improving trend. Both analyses therefore implied that contrary to the classical view, developing countries were actually obtaining a falling proportion of the potential gains from their trade with the countries of the developed world (Sapsford et al., 1992). The statistical techniques available for revealing and estimating trends in the middle of the twentieth century were by today’s standards extremely rudimentary. However, it is relevant to notice that some 50-plus years later the conclusion of a downward secular movement remains strongly intact, despite a rapid evolution in the sophistication of the time-series methods to which terms-of-trade data have been subjected.4 Figure 37.1 provides a time plot covering the period 1900 to 1999 of the log of most commonly analysed terms-of-trade series, together with the estimated least-squares trend line which provides clear evidence of the presence of a steady downward trend over the last century.5

From Singer I to Singer II
When the Prebisch–Singer thesis was launched in 1950 the world was, perhaps, a simpler place economically than it is today. In particular, the fact that the industrialized countries of the North specialized heavily in the production and subsequent export of manufactured goods, while those in the South concentrated intensively in the production or extraction and export of primary commodities, meant that the Prebisch–Singer focus upon the
terms of trade between manufactured and primary goods carried direct implications for the terms of trade of less-developed countries (LDCs) vis-à-vis developed countries. As the post-1950 decades unfolded, and many of the LDCs of the South moved increasingly down a path of industrialization, Singer (1975) argued that terms-of-trade issues continued to be of vital importance to LDCs despite their greater reliance on manufactured exports. In this context he distinguished between what he called ‘Singer I’ and ‘Singer II’; with Singer I referring to terms of trade between different commodities (manufactures versus primaries) and Singer II referring to terms of trade between different countries (less-developed economies, the ‘South’, versus developed economies, the ‘North’).

As we shall see in the following section there has been a range of alternative explanations put forward in the literature for the observed secular deterioration in the terms of trade. However, we will also see that these explanations relate at least as much to the characteristics and structures of the countries themselves as to the commodities they trade. The distinction between Singer I and Singer II is an important one because it alerts us to the possibility that potentially serious terms-of-trade problems for LDCs can continue to persist despite their increasing tendency to switch towards manufacturing exports. Various pieces of empirical evidence support this concern. For example, econometric evidence reported by Sarkar and Singer (1991) suggested that for a large sample of less-developed countries over the period 1979 to 1987 the terms of trade between the manufactures exported by them and those imported by them from the developed economies was subject to a declining trend. A similar finding has also been
reported by Lutz (1999). Some new evidence relating to the experiences of a sample of the world’s very poorest countries is reported later in this chapter. In an overview of the performance of the Prebisch–Singer thesis over its first half-century, Singer (1987) presented some statistical evidence to suggest that the terms-of-trade deterioration experienced by LDCs since the mid-1950s can be attributed, in approximately equal proportions, to three causes: first, the deterioration in the price of primary commodities in relation to manufactures (as emphasized in Singer I); second, a more rapid deterioration in the prices of the primary commodities produced by developing countries than in the prices of those produced by developed countries; and finally, a fall in the price of the manufactures exported by developing countries relative to those exported to them by developed countries (as emphasized by Singer II).

Explaining the downward trend
A number of theoretical explanations have been put forward in the literature to account for the observed downward trend in the terms of trade of developing relative to developed countries, and these can be conveniently summarized under the following four headings:

- Differing price elasticities of demand for primary commodities and manufactured goods (with the inelastic nature of the former resulting in a tendency for increases in the conditions of commodity supply to be felt more strongly in price decreases than in quantity increases).
- Differing rates of growth in the demands for primary commodities and manufactured goods (with the demand for primaries expanding less rapidly than the demand for manufactures due to their lower income elasticity of demand – especially so in the case of agricultural commodities due to the operation of Engel’s Law – plus the development of synthetic substitutes and the occurrence in manufacturing of technical progress of the raw material saving sort).
- Technological superiority (the argument being that the price of manufactured goods rise relative to those of primaries because they embody both a so-called Schumpeterian rent element for innovation, plus an element of monopolistic profit arising from the monopoly power of multinational producers).
- Asymmetries in market structure (the argument here is that differences in market structure – with primary commodities typically being produced and sold under competitive conditions, while manufacturing in industrialized countries is often characterized by a high degree of monopoly by organized labour and monopoly producers – mean that while technical progress in the production of primary
commodities results in lower prices, technical progress in manufacturing leads to increased factor incomes as opposed to lower prices).

In his 1950 paper Singer placed particular emphasis on points listed under the first two bullet points, while Prebisch (1950 [1962]) specially emphasized those covered under the third and final bullets. A number of recent authors have developed what have become known as North–South models, which further develop the points grouped under bullets three and four. See, for example, Sarkar (1994) and Dutt (1998).

Some policy implications
Although space constraints do not allow us to discuss in any detail the policy implications of the observed worsening trend in the terms on which trade is conducted in the world marketplace between primary commodity-producing countries and manufacturing countries, it is nevertheless important to note that the Prebisch–Singer hypothesis is sometimes advanced as one argument in favour of development and trade policies of the import-substituting industrialization (IS) as opposed to export promotion (EP) variety. A number of early enthusiasts of the Prebisch–Singer thesis recommended the adoption of such a policy stance. However, the policy issues here are not clear-cut and the fact, already mentioned, that all four of the above explanations relate as much, if not more, to the characteristics of different types of countries as to the characteristics of different types of traded goods highlights the need to devise and implement policies that address differences and imbalances of the former as well as the latter sort (Singer, 1987). Although a number of recent analysts (for example Krueger, 1997) have drawn an association between below-average growth performance and the adoption of IS as distinct from EP policies, the real-world situation, as Singer frequently pointed out, was rather more complex when one looks beneath the surface, in that countries such as Korea (held up in some circles – including Washington, DC – as a beacon to demonstrate the superiority of outwardly oriented trade policies) appear in reality to have achieved their above-average rate of growth by adopting a subtle dynamically evolving policy mix involving a combination of IS policies in certain sectors of the economy simultaneously accompanied by EP policies in other sectors. As Singer was also quick to point out (for example Sapsford and Singer, 1998), Ricardian comparative advantage is seldom, if indeed ever, exogenously given in the manner assumed in the simple Ricardian model, still so central in trade analysis. Thus while it is logically correct that in the static Ricardian world primary commodity-exporting countries can still potentially gain from trade by specializing in those sectors in which they possess (static) comparative advantage, what is relevant in reality is the
fact that many comparative advantages are actually either consciously engineered or acquired through learning by doing and increasing returns to scale. Accordingly, it has been argued that the Prebisch–Singer thesis provides one possible case for industrialization based on (a limited period of) infant industry protection. In development economics, as in prospecting, all that glitters is not gold.

Although outside of the scope of the current chapter, it is important to recognize that the terms-of-trade issue carries a range of important implications for issues other than the distribution of the gains from trade. These include questions relating to the level, and pattern, of Southern or LDC growth and the unevenness (or otherwise) of international development patterns. Some brief comments are, however, in order. As far as the implications of terms of trade for Southern growth are concerned, there now seems to be something of a consensus in the so-called ‘New’ (or Endogenous) Growth Theory literature that the terms of trade do matter when it comes to explaining inter-country differences in growth performance. In particular, in a study of the determinants of growth across countries Barro (1997) estimates the (partial) elasticity of real output with respect to the country-specific net barter terms of trade to be 0.137, with a standard error of 0.3 – implying an effect which is significantly different from zero. The long-run trends in the terms of trade, and the forces that drive these, also carry some important implications for our understanding of the observed patterns of international development. The literature here is extensive and is well covered by Dutt (1990, 1998). For the purpose of the current discussion it is relevant to notice that different causes of terms-of-trade deterioration carry with them different implications regarding the evenness of development. In cases where the predominant factor driving terms-of-trade deterioration is the income elasticity of demand, we would expect to observe uneven development. However, in cases where its roots lie in the pace of technological progress in the South then, especially in situations where there are inflows of foreign direct investment into the South accompanied by the possibility of significant spillovers of various sorts to the domestic economy, there is the potential for deteriorating terms of trade to go hand in hand with more even development (Balasubramanyam et al., 1996). In a nutshell: the terms of trade matter, in terms of both the speed of LDC growth and its evenness.

Some recent issues: country-specific evidence and terms-of-trade volatility
Terms-of-trade issues continue to attract the attentions of researchers with interests in a variety of trade-related fields and in this section we consider two of these. The first relates to the strength and periodicity of terms-of-trade volatility, while the second relates, in the spirit of Singer II, to the experiences of particular economies.
Trend versus volatility in the terms of trade

Inspection of the time plot in Figure 37.1 suggests that while the terms of trade declined secularly over the course of the last century they were, at the same time, subject to appreciable variability about that trend. In a statistical analysis of the time-series properties of the Grilli–Yang data (as plotted in Figure 37.1) Sapsford and Balasubramanyam (1999) produced evidence that suggested that the extent of volatility exhibited by the terms of trade about its trend increased over the twentieth century. This increased volatility was particularly marked after 1972 and appeared to have been accompanied by an acceleration in the trend rate of decline of the terms of trade. Taken together these results yielded a depressing picture for the primary commodity-dependent countries of the developing world for they indicated that since 1973 such countries have experienced a marked increase in the trend rate of deterioration in their terms of trade (from about 1.6 per cent per annum to around 4 per cent), accompanied by a marked increase in their volatility estimated to be to the order of 50 per cent. This clearly represents a doubly unpleasant state of affairs.

In a related study, Singer and Lutz (1994) report cross-country panel data estimates which seem to indicate that the magnitude of both the downward trend in the terms of trade and the extent of volatility about this trend exerted significant downward effects on gross national product (GNP) growth in a sample of some 79 non-oil-producing countries. In this study the authors argue that terms-of-trade volatility exerts a detrimental influence on growth performance via a number of channels, including: the increased risks which it imposes on investment (which they see as reducing the incentives to invest and thereby the realized volume of investment); its potential to disrupt development (including structural adjustment) plans, to increase price instability, to destabilize domestic incomes and to distort the structure of domestic prices, including the price of traded relative to non-traded goods.

The arguments developed and empirical results reported in these two studies would seem to imply that the trend and volatility in the terms of trade should not be treated as separable issues in the manner of classical time-series analysis. Instead these two issues are best seen from the analytical standpoint as twin pillars of the same underlying problem faced by LDCs: their heavy dependence on primary commodities (or more recently upon components and parts of manufactured goods) as a source of export revenue. The empirical work reported in these papers focused attention on both the long-run and cross-country evidence regarding the trend and volatility in the terms of trade seen as related dimensions of the same underlying problem. Amongst other things the results reported by Sapsford and Balasubramanyam (1999), as noted above, indicated the
occurrence after 1972 of a major increase in the trend rate of deterioration of the terms of trade and, moreover, that this was accompanied by a marked increase in volatility. This is the true ‘double whammy’ in which both of the twin pillars of the commodity problem seem to have turned simultaneously against these countries. However, the evidence reported by Singer and Lutz (1994) suggests that the bad news does not end there, in the sense that both the trend rate of terms-of-trade decline and the extent of terms-of-trade volatility would appear to exert a subsequent significant downward influence upon the rate of economic growth; a ‘triple whammy’.8

Country-specific evidence
A number of studies in the recent literature have moved away from the continued analysis of the aggregate data,9 preferring instead to focus attention on country-specific evidence, especially that relating to low-income primary commodity-exporting countries. It is a regrettable fact of current economic life that it is the very poorest economies in the world who are the most dependent on either a single or a small number of primary commodities for the vast majority of their export earnings. In an UNCTAD study, Sapsford (2001) noted that according to World Bank 1996 data, all but two of the world’s poorest 20 economies were located in Africa and of these, some 13 were dependent on either a single primary commodity or a small number of commodities for in excess of 90 per cent of their export earnings. Two examples are as follows: Mali (with an estimated annual per capita income of US$260 and life expectancy at birth of 50 years) earned 99.8 per cent of its export revenue in 1996 from cotton, while Ethiopia (the then poorest economy in the world, with an estimated per capita income of only US$110 per annum, and life expectancy of only 43 years)10 depended on coffee for 99.8 per cent of its export earnings. Clearly, for these seriously poor economies both trends and volatilities in the terms on which they undertake their trade in these particular primary commodities with the outside world are of vital importance. One might say that they are quite literally a matter of life and death.

Sapsford (2001) provided a detailed statistical analysis of the terms-of-trade experiences of this sample of the world’s poorest economies over the period 1960 to 1998. Using a simplified version of the structural econometric model proposed in Bloch and Sapsford (2000), this study undertook a structural stability analysis of country-specific terms of trade. Notice that unlike cruder previous approaches to trend estimation, this approach controlled for the influence of fluctuations in the level of production in the industrialized world on country-specific terms of trade. Although a detailed discussion of the nature and implication of the results revealed by
this study is beyond the scope of the present short chapter\textsuperscript{11} it is relevant to notice that the main results can be summarized as follows:

- Of the world’s 15 poorest commodity-exporting countries, all but two experienced a significant change in the trend rate of growth of their terms of trade during the period 1960–93.
- In nine of these cases, the change in the trend occurred between 1972 and 1982.
- Nineteen out of 28 reported trend estimates are negative; only three of the reported trend estimates are positive.
- In nine out of the 13 countries where there is a trend shift, the pattern shows a worsening of the situation in respect of terms of trade.
- In six out of the 13 countries where there is a trend shift, the pattern shows an increase in the volatility of the terms of trade.

Taken together these results seem to indicate that over the post-1960 period many of the very poorest commodity-exporting LDCs in the world have indeed been subject to Prebisch–Singer effects on their terms of trade, effects which have exerted a continuous downward pressure on economic and export growth of a magnitude sufficient to more than offset the upward effects which they might have experienced as a result of the positive influence of expanding output in the industrialized countries.

Some concluding remarks

The terms-of-trade is a topic that continues to attract the attentions of researchers in the fields of international trade and development economics: scarcely a week passes without yet another addition to the literature. Within the confines of this short chapter it has only been possible to scratch the surface of the many and varied issues involved. Particular emphasis was placed in the first half of the chapter upon the famous Prebisch–Singer hypothesis, the empirical evidence relating to the hypothesis and its policy implications, especially as these relate to countries situated at the lower end of the per capita income distribution. In the second half of the chapter we focused attention on some recent analyses that have extended the earlier approaches to encompass the effects, from a country-specific perspective, of terms-of-trade volatility upon economic growth performance and the relationship between terms-of-trade trend and volatility.

As we have seen, the Prebisch–Singer hypothesis that there is a long-term decline in the price of primary commodities relative to the price of manufactures has historically been an object of controversy, although given the overwhelming weight of empirical evidence in its favour, it seems to have become widely accepted in the majority of circles since the 1990s.\textsuperscript{12} Most
tests of the hypothesis use time-series models to estimate trend growth rates in selected relative prices. The focus of concern has typically been the net barter terms of trade between producers of primary products (equated with developing countries) and producers of manufactures (equated with industrialized countries). A new approach which has been developed recently is to construct a structural model which seeks to identify the various different factors which impinge on the prices of manufactured goods and primary commodities (Bloch and Sapsford, 1997, 2000). Applying this approach, it has been found that the overall trend identified in the time-series models is the net effect of a number of separate divergent influences. On the one hand, there are Prebisch and Singer effects that exert a downward pressure on the commodity terms of trade. These effects arise because of differences in market structure (markets for primary products are more perfectly competitive) and differences in the factor bias of technical change (technical change in manufactures is assumed to save raw material inputs and labour). On the other hand, rising output in the industrialized countries can have an offsetting effect, as primary products used in manufacturing activity experience rising prices when the level of manufacturing activity increases.

The final section of this chapter discussed some country-specific results that were obtained by applying this basic approach. These results seemed to suggest that many of the very poorest economies in the world – which are also, as it happens, heavily (if not totally) dependent on either a single primary commodity or a very small number of them for their export revenue – have been subject over at least the last four decades of the twentieth century not only to the downward pressures of the sort emphasized by both Prebisch and Singer but also to the additional pressures generated by increasing terms-of-trade volatility.

Some two and a half centuries have elapsed since the classical economists first proffered their prediction that the tide would inevitably turn, over the long run, in favour of the LDCs. To offer what is perhaps logically the only admissible defence – namely, that some two and a half centuries is too short a time period for the classical mechanisms to fully work themselves out – seems implausible not only to trained economists but more importantly to those citizens of the world’s very poorest economies who seem condemned to remain forever at the very margins of survival.

Notes
1. This chapter is dedicated to the memory of Hans Singer (1910–2006), friend, collaborator and inspiration.
2. For brevity I refer hereafter to the net barter terms of trade between primary commodities and manufactured goods (that is, to the ratio of the price of primary commodities to the price of manufactured goods) as simply their terms of trade.
3. Although conventionally referred to in the literature as the Prebisch–Singer hypothesis recent evidence reported by Toye and Toye (2003) suggests that it should correctly be termed the Singer hypothesis since their detailed archival research indicates that Prebisch’s interest in the topic stemmed directly from his receipt of Singer’s then unpublished UN manuscript on the subject.

4. Being in contradiction with the then prevailing orthodoxy, these papers attracted criticism on a number of (primarily statistical) grounds. However, detailed scrutiny of the literature seems to indicate that almost six decades after its initial launch the empirical validity of the declining-trend hypothesis first put forward by Prebisch and Singer has become pretty much universally accepted. See Sapsford and Chen (1998) for a review of the second wave of statistical studies that appeared during the late 1980s and the 1990s, and Spraos (1983) and Sapsford (1985) for reviews of the earlier criticisms. This longevity is all the more remarkable when one recognizes the wide array of statistical techniques to which the hypothesis has been subjected. There are few, if any, hypotheses in economics that can claim to have stood the test of time so well as this one. See Spraos (1980, 1983), Sapsford and Singer (1998), Sapsford and Chen (1999) and Bloch and Sapsford (2000) for detailed reviews.

5. The data series employed here were compiled by Grilli and Yang (1988). Dr Stephan Pfaffenzeller of the University of Liverpool compiled the post-1986 data according to the same definitions. This is the terms-of-trade series most commonly analysed in the literature, being compiled as the World Bank’s index of the prices of some 24 internationally traded non-fuel primary products deflated by the UN’s index of the unit values of manufactured exports from industrialized countries.


7. The early twenty-first-century experience of China may well turn out to offer an important lesson here regarding the distinction between what is logically correct and what is relevant.

8. Although outside of the scope of the present chapter it is important to notice that the existence and indeed persistence of volatility around the trend carries with it a potential policy implication regarding the possible merits of instituting some form of stabilization scheme (for example buffer stocks, compensatory financing mechanisms and the like) designed to at least smooth out, if not remove, fluctuations about the trend. The literature here is extensive and goes back in time beyond Keynes, although the latter’s views as expressed at the Bretton Woods Conference are of particular importance. For surveys of the major issues involved see Winters and Sapsford (1990), Maizels (1992) and Sapsford and Morgan (1994).

9. Or, perhaps more correctly, have chosen to move away from the continued reanalysis of the Grilli–Yang (1988) data set.

10. To put these life expectancy figures in perspective it should be noted that the corresponding figures for the USA, the UK and Japan are 76, 77 and 80 years respectively.

11. See UNCTAD (2002) for a summary. Notice that in this study (following Singer and Lutz, 1994, Sapsford and Balasubramanyam, 1999, and others) the standard error of estimate about the estimated regression plane was used as the sub-period measure of terms-of-trade volatility.

12. It is now the case that at least some of the international agencies involved in the world trading system have come to accept that primary commodity producers in developing countries do face real and significant uncertainties and risks regarding the prices that they will actually receive for their products when they come to the world market. At the time of writing in 2006, a task force set up under the auspices of the World Bank is investigating a range of possible ‘market-based’ approaches (including the formation of futures markets) for dealing with the price risks faced by primary commodity producers in developing countries. As pointed out by Morgan (2001) these approaches appear to represent an attempt to confront price risk by modifying the financial environment within which primary producers in less-developed countries operate. However, it remains to be seen whether such approaches will prove any more, or less, successful than the various policies which have preceded them.
References


38 Trade policy and development

Henry J. Bruton

Introduction
The countries of Western Europe, northern North America, and Australia and New Zealand (the North) began to achieve increasing per capita gross domestic product (GDP) in the first part of the nineteenth century. Growth, so measured, while not uninterrupted, became sustained enough that one may say that the routine functioning of these economies produced increasing per capita GDP. Growth became, in effect, built in. As a consequence of 150–200 years of this fairly routine growth, the countries of the North are now very rich, at least in terms of GDP per capita. The Great Question of Development Economics is: Why have not all countries been able to establish economies in which growth is built in to their routine operations? The broad policy question is therefore how to bring about modifications in these non-growing economies (the South) in such a way that growth becomes routine for them too. The theoretical question is to explain why and how one group of countries did grow while, another, much larger group, has not.1 In the best of all worlds, this explanatory story would be well grounded in fundamental economic principles, supported by convincing empirical evidence, and lead to policies that can be implemented and, when implemented, bring about the achievement of the objective stated above.

The purpose of this chapter is to study the way that foreign trade enters the development quest just defined. The approach is to tell a story of history and learning, of learning through time as evidence and theoretical insight accumulated. Such a historical, learning approach seems appropriate because of the changing views and continuing controversy, and because no widely held model or theory or policy position has become conventional wisdom.

The beginning
In the late 1940s and early 1950s when ‘development economics’ emerged as a distinct field of inquiry within the broad discipline of academic economics, empirical evidence and explanatory hypotheses about the process of development were relatively primitive. Yet it was widely accepted that the huge differences in per capita GDP between the North and South demanded action, demanded that the world ‘do something’.2 Trade policy
was, from the beginning, a major element both in terms of explanation and in terms of policy.

**Import substitution: argument and practice**

The earliest role of trade as an explanation of the failure of the South to grow rested on arguments and assumptions about the way that the ‘structure’ of production in a country affected its capacity to grow. Structure of production refers mainly to the composition of output of an economy, especially to the distinction between manufacturing on the one hand and agriculture and mining on the other. The economies of the South had, it was widely argued, become locked into agriculture and mining, largely due to foreign trade based (at least to some extent) on assumptions about comparative advantage. While this locked-in state may have resulted in the maximization of current production, the argument continued, it did prevent these economies from growing over time. There were several detailed arguments and assumptions on which these conclusions depended.

The earliest specific argument was that the terms of trade had consistently moved against the South countries. Several reasons were offered as to why this had been and was happening. Productivity growth in manufacturing was thought to be greater than in agriculture and mining, but this higher productivity growth was matched or more than matched by rising wages in the North. The rising wages, combined with the widespread monopoly power of producers in the North, prevented manufacturing prices from falling as productivity grew. While in the South, productivity grew more slowly and an abundant supply of labor prevented wage rates (and prices) from rising. Added to these considerations was the assumption that demand for food (and agricultural products in general) and minerals was inelastic with respect to income in world markets. As output of these commodities increased, their prices relative to those of manufactures tended to fall over time as per capita incomes rose in the North.

The South could not shift out of agriculture and minerals into manufactures because its technological capacity was such that it could not compete with the producers of the North as long as free trade prevailed. This disadvantage with respect to technological capacity prevailed despite the fact that the South had access to the same array of equipment and ideas that most of the countries of the North had – mainly imported from Britain.

These arguments disputed any role that conventional trade theory – comparative advantage as formulated by Ricardo and by Heckscher–Ohlin – could play in a development context. Comparative advantage, it was often noted, is a static notion that assumes technical knowledge is the same in all countries, constant returns to scale everywhere, and that equilibrium states always prevail, that is, no second-best issues arise. It thus had nothing to
communicate with respect to formulating a policy aimed at producing marked changes in the structure of an economy. Added to these considerations were references to colonialism and to the manner and extent to which it had drained resources from the poor to the rich countries.

These explanations of the failure of the South to grow had obvious policy implications, especially for trade policy: abandon free trade, and behind protection change the structure of the economy, that is, industrialize. The idea was simple. Products – especially consumer durables – now imported would be subjected to prohibitive tariffs or banned completely and local firms would emerge to take advantage of the already existing demand. These arguments and their policy implications became identified as the import-substitution (IS) approach to development.

Growth theory available at the time was the Harrod–Domar model in which physical capital formation was the basic source of growth. The main (often only) requirement to achieve growth (and changed structure) was a rate of capital formation sufficiently high, given the productivity of capital, to achieve a target rate of growth of output.\(^6\) Since virtually all capital goods were imported from the North, many countries sought to subsidize capital formation by maintaining an overvalued exchange rate. This practice, along with low, often negative, real interest rates, did induce relatively satisfactory rates of investment. At the same time such policies made it particularly difficult for new firms to enter export markets, they penalized employment and resulted in a great deal of underutilized capacity in the newly created firms.\(^7\) Physical capital, deemed the major lack in the South, was frequently found to be unused and poorly maintained. The importation of capital goods along with the inability to export new products created balance-of-payments problems which were often met by further exchange controls or more widespread use of tariffs and quotas.\(^8\)

Distortions were added to by the common practice of many countries of establishing individual tariffs to whatever level was deemed necessary to enable a given activity to come into existence. The result was a wide range of tariff rates and an even wider range of effective rates of protection.

In these highly distorted economies, market prices were quite misleading as signals of social costs and social returns. Considerable attention was given to benefit–cost analysis of large-scale projects using shadow prices, rather than market prices. This was difficult to do with any confidence, and even where it was done, faced the problem that costs and prices paid and received were those prevailing in the market, so that profitability of firms in shadow prices did not assure profitability in market prices which somehow had to be achieved.\(^9\) The use of shadow prices did not, in any way, offset the distortions that the policies to implement the shift in structure imposed on the economies.\(^10\)
It also became clear that the distortions, especially those that affected access to foreign exchange and domestic investment permits, resulted in a substantial allocation of resources to rent-seeking, that is, to seeking permits and licenses to access underpriced resources, rather than seeking investment projects, new technical knowledge and focusing on their own business. There is little doubt that such activities were costly in many ways and dampened the search for ways to make the new firms increasingly productive.\textsuperscript{11}

The general approach to development just described, despite violating many of the more established principles of textbook economics, was widely supported. The World Bank, for example, in its 1979 \textit{World Development Report}, noted (pp. 67–8) that the policy had had important positive effects on entrepreneurial and technological capacity in many developing countries, and had induced the growth of a manufacturing sector behind high levels of protection. This bank report noted the inefficiencies described above, but with much more restraint than was the case a decade or so later. There were other favorable reports and very few explicit criticisms expressed in the literature of the day.

\textit{Some consequences}

Tables 38.1 and 38.2 show the broad outlines of the post-World War II years when IS was widely followed. The 1950s were dominated by the recoveries from the dislocations of the war, but by the 1960s the effects of IS in many countries became paramount. During the 1960–73 period both labor productivity and total factor productivity growth were impressive over a wide range of countries. Most countries also experienced export growth that dispelled doubts about the ability of developing countries to export. An index from Little et al. (1970, p. 245) shows that manufactured exports from all developing countries increased from a base of 100 in 1953 to 283 in 1965. Capital formation also took place at impressive rates in most countries as the protection of domestic activities created many new opportunities. Even agriculture, despite being heavily penalized, grew at respectable rates in most – not all – countries pursuing IS.

In addition life expectancy increased almost everywhere, infrastructure was improved, and literacy increased. Clearly things got better in the 1950s and 1960s as a consequence of the IS strategy.\textsuperscript{12} By the end of the 1970s it appeared that IS was a great success and a sure guide to continued growth.

As the tables show however, difficulties appeared as the 1970s wore on. Total factor productivity growth (TFPG) slowed markedly, and many cases of negative TFPG appeared, a sure sign of increasing distortions and coordination failures. Falling growth rates also contributed to negative productivity growth. Export growth slowed and turned negative in many cases.
Table 38.1 Rates of growth of productivity

<table>
<thead>
<tr>
<th>Country</th>
<th>1960/73 GDP/L</th>
<th>TFP</th>
<th>1973/84 GDP/L</th>
<th>TFP</th>
<th>1984/94 GDP/L</th>
<th>TFP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>East Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>China</td>
<td>2.2</td>
<td>1.4</td>
<td>4.3</td>
<td>0.5</td>
<td>6.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Indonesia</td>
<td>2.5</td>
<td>1.1</td>
<td>4.3</td>
<td>0.5</td>
<td>3.7</td>
<td>0.9</td>
</tr>
<tr>
<td>S. Korea</td>
<td>5.6</td>
<td>1.4</td>
<td>5.3</td>
<td>1.1</td>
<td>6.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Malaysia</td>
<td>4.0</td>
<td>1.0</td>
<td>3.6</td>
<td>0.4</td>
<td>3.8</td>
<td>1.4</td>
</tr>
<tr>
<td>Philippines</td>
<td>2.5</td>
<td>0.7</td>
<td>1.2</td>
<td>−1.3</td>
<td>−0.3</td>
<td>−0.9</td>
</tr>
<tr>
<td>Thailand</td>
<td>4.8</td>
<td>1.4</td>
<td>3.6</td>
<td>1.1</td>
<td>6.9</td>
<td>3.3</td>
</tr>
<tr>
<td>Taiwan</td>
<td>6.8</td>
<td>2.2</td>
<td>4.9</td>
<td>0.9</td>
<td>5.6</td>
<td>2.8</td>
</tr>
<tr>
<td><strong>Latin America</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argentina</td>
<td>2.6</td>
<td>0.2</td>
<td>0.4</td>
<td>−1.0</td>
<td>1.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Brazil</td>
<td>4.4</td>
<td>2.9</td>
<td>1.0</td>
<td>−0.8</td>
<td>0.5</td>
<td>−0.2</td>
</tr>
<tr>
<td>Chile</td>
<td>1.6</td>
<td>0.7</td>
<td>−0.6</td>
<td>0.7</td>
<td>4.7</td>
<td>3.7</td>
</tr>
<tr>
<td>Colombia</td>
<td>2.9</td>
<td>1.9</td>
<td>1.2</td>
<td>0.0</td>
<td>1.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Ecuador</td>
<td>4.4</td>
<td>3.3</td>
<td>1.7</td>
<td>−0.5</td>
<td>0.0</td>
<td>−0.1</td>
</tr>
<tr>
<td>Mexico</td>
<td>3.8</td>
<td>1.6</td>
<td>0.7</td>
<td>−0.8</td>
<td>−1.1</td>
<td>−1.8</td>
</tr>
<tr>
<td>Venezuela</td>
<td>1.2</td>
<td>0.9</td>
<td>−3.1</td>
<td>4.3</td>
<td>−0.6</td>
<td>−0.4</td>
</tr>
<tr>
<td><strong>Middle East</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egypt</td>
<td>4.7</td>
<td>2.3</td>
<td>0.5</td>
<td>−2.2</td>
<td>−1.1</td>
<td>−1.5</td>
</tr>
<tr>
<td>Iran</td>
<td>6.1</td>
<td>2.4</td>
<td>−2.9</td>
<td>−5.7</td>
<td>−2.2</td>
<td>−2.2</td>
</tr>
<tr>
<td>Jordan</td>
<td>2.1</td>
<td>−0.9</td>
<td>6.7</td>
<td>2.3</td>
<td>1.2</td>
<td>−2.9</td>
</tr>
<tr>
<td><strong>South Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bangladesh</td>
<td>0.0</td>
<td>−0.6</td>
<td>2.5</td>
<td>1.8</td>
<td>1.1</td>
<td>0.7</td>
</tr>
<tr>
<td>India</td>
<td>1.8</td>
<td>0.1</td>
<td>2.4</td>
<td>1.0</td>
<td>3.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Sri Lanka</td>
<td>2.1</td>
<td>1.0</td>
<td>3.2</td>
<td>0.7</td>
<td>2.7</td>
<td>1.0</td>
</tr>
<tr>
<td>Pakistan</td>
<td>3.9</td>
<td>0.2</td>
<td>2.8</td>
<td>2.0</td>
<td>2.7</td>
<td>1.5</td>
</tr>
<tr>
<td><strong>Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>2.2</td>
<td>0.2</td>
<td>0.0</td>
<td>−0.9</td>
<td>−0.2</td>
<td>−1.6</td>
</tr>
<tr>
<td>Ghana</td>
<td>0.9</td>
<td>−1.0</td>
<td>−3.2</td>
<td>−3.2</td>
<td>1.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Kenya</td>
<td>3.4</td>
<td>3.4</td>
<td>0.4</td>
<td>−0.1</td>
<td>0.1</td>
<td>0.4</td>
</tr>
<tr>
<td>Nigeria</td>
<td>1.2</td>
<td>−0.9</td>
<td>−2.3</td>
<td>−4.6</td>
<td>1.3</td>
<td>2.0</td>
</tr>
<tr>
<td>Uganda</td>
<td>0.7</td>
<td>−0.3</td>
<td>−2.9</td>
<td>−3.0</td>
<td>1.3</td>
<td>1.1</td>
</tr>
<tr>
<td>Tanzania</td>
<td>3.0</td>
<td>2.2</td>
<td>−1.1</td>
<td>−1.7</td>
<td>1.0</td>
<td>0.6</td>
</tr>
<tr>
<td>S. Africa</td>
<td>2.3</td>
<td>0.9</td>
<td>1.0</td>
<td>−0.3</td>
<td>−2.0</td>
<td>−1.8</td>
</tr>
<tr>
<td>Zambia</td>
<td>1.0</td>
<td>0.2</td>
<td>−2.3</td>
<td>−1.9</td>
<td>−2.5</td>
<td>−1.1</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>2.9</td>
<td>2.7</td>
<td>−0.8</td>
<td>−1.3</td>
<td>0.2</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Capital formation in the poorer countries declined and, while remaining strong in a number of middle-income countries, declined in others as investment opportunities dried up. The slowing down of capital formation and productivity growth also dampened an already weak demand for labor,
and unemployment became an increasingly evident issue. Similarly almost all countries had trouble maintaining both internal and external stability.13

Two things were clear to most observers by the early 1970s: the first was that for well over a decade IS had been a genuine success, and the second was that, as practiced, it was not the strategy that could create an economy in which growth was built in to the routine operation of the economy.

Several aspects of the strategy became widely appreciated. That the assumption of fixed production coefficients of Harrod–Domar was the source of many difficulties especially became recognized.14 The assumption of fixed production coefficients was a principal reason for the failure of the high rates of investment to create the demand for labor that would match available supply as well as for the widespread underutilization. The assumption rested largely on the belief, widely held, that the technology of the North was so dominant that it made any other technology, any other factor combination, technologically inefficient, that is, inefficient at any set of factor prices. So in effect firms in the South had no choice of technique or machine if they were to compete in the world economy.

This last assumption rested on another, larger, more encompassing notion. The prevailing idea of development at this time was simply to duplicate the North. This idea is given full exposition by Arthur Lewis in a very influential article published in the early 1950s.15 Lewis divides the economy of a South country into a large, very poor, traditional sector and a small capitalist sector. Investment was to take place in the latter sector and was to be accomplished by importing physical capital and technology from the North. As investment continued, the capitalist sector would grow relative to the traditional sector until it encompassed the whole economy and the South would be as the North. Development was in effect imported, not indigenous, and capacity to import became a crucial constraint.16

The dual economy model was thus a story of displacement, not an explanation of how a non-growing traditional economy changed itself into a growing economy. This pervasive idea, development as replication of the North, had several consequences. It detracted attention away from the development notion itself, that is, to achieve the metamorphosis of a traditional, non-changing economy into a growing one. A theory of development is concerned with explaining how this traditional economy can be turned into a growing one in such a way that those characteristics that, in effect, define the society – its history, its values, its ethos, its very meaning – are not violated. Fixed production coefficients, the idea that all technical and organizational knowledge from the North could be codified and immediately utilized as in the North, the achievement of the North’s product mix, all contributed to the dampening, even preventing, of any domestic efforts to adapt, to seek and to learn, and to recognize the importance of
building onto and from and within existing institutions and within the boundaries defined by the ideas of the good life of the population. It is evident, from this statement, that the role of trade (and other foreign activities) is vastly important, but subtle and complex.17

Finally, the strategy as pursued not only neglected, but in general penalized agriculture. Agriculture was often taxed and more often price controls on foodstuffs were imposed to keep their prices (and therefore wages) low in urban areas as yet another means of subsidizing new manufactures. The neglect of agriculture also encouraged a more rapid exodus from rural to urban areas, the result of which was increased unemployment in the cities and the emergence of ‘informal sectors’ alongside the new manufactures. Agriculture in most countries of the South was the largest sector in terms of both production and employment. To penalize this sector meant that the major sector of the economy was being penalized. In a country where new exports were very slow in emerging, a weak agricultural sector meant that the growth of domestic demand for new, non-agricultural products was severely dampened, and this in turn impeded the learning-by-doing process in the new firms. That agriculture could be safely penalized was a by-product of the ‘structural change’ argument for import substitution: the other side of creating a new manufacturing sector was the killing off of agriculture. Where exports of the new activities were not possible and agriculture was penalized, then importing food grains often became necessary and was costly. The late 1950s in India were a prime example of this sort of phenomena.

That this set of policies did not result in the creation of a built-in growth process does not now appear surprising. Despite the burst of early growth in the 1950s and 1960s noted above, and the improvement in a number of welfare measures, problems began to emerge that convinced most observers that the import-substitution process was in no sense sustainable. The details of the story just summarized vary markedly among the countries of the South, but the broad picture seems generally applicable. For Africa less so, perhaps, than in most other places, and more so in India, Pakistan and most Latin American countries.18

The fall of IS and the rise of openness
While recognition that the IS policies had created an unsustainable situation became widespread, alternative strategies did not convince many policy-makers in the South. That distortions were ubiquitous and were penalizing the economies was appreciated by most observers. Economists still had only rough ideas of how to make an economy grow, but of course distortions and optimal allocation of resources were textbook stuff. It is not surprising therefore that the new strategy proposed concentrated
attention on eliminating the sources of these distortions: the ad hoc tariffs, internal controls of many kinds, overvalued exchange rate, soft budget constraint, public ownership of many firms, price controls and inflationary fiscal policies. The catchwords became ‘market friendly’, ‘privatizing’, ‘macro stability’ and ‘openness’. ‘Outward orientation’ and ‘Washington Consensus’ replaced IS as the summary term. This new approach originated largely in the North and the pressure to liberalize came from the North, the World Bank, the United States Agency for International Development (USAID) and Northern academic economists, hence the appellation Washington Consensus. Reluctance of South countries to move sharply toward liberalization was partly a matter of the immediate economic costs (especially the almost inevitable increased unemployment) to withdrawing protection from recently established activities that could not survive without protection, and partly a matter of genuine doubt that the newly pushed theory would in fact have the desired effect. In addition, there were numerous people who profited from IS, and who, therefore, opposed its abandonment.

There was no new ‘theory’ of development that led to this different strategy. The liberalization package as looked upon by the South as simply a return to pre-World War I international arrangements. In particular the idea that if property rights are clear and in place, if the price mechanism follows the textbook stories, if macrostability is maintained, and the investment rate is at least 15 percent, the economy will grow with full employment, was not a widely accepted view outside the Washington Consensus group. It was too easy to see that some countries – for example India, Korea, Vietnam, China and Botswana – were growing well by violating many of the conditions laid down as essential by the Washington Consensus. Similarly other countries – especially in Latin America – had followed the rule rather closely and performed much less well.19

The period from 1870 to 1914 is especially illuminating. During this period capital and trade moved easily to almost all parts of the world, government interventions in the economy were modest, migrations were large and widespread, foreign exchange markets were fairly stable due to the widespread commitment to the gold standard and to the British pound, the international capital markets were effective, price stability was common, and there were significant social and political changes throughout the world. In addition, there were marked improvements in transportation and communication around the world.20 These circumstances, however, did not produce industrialization in the South, and did not produce productivity growth, nor the growth of factory employment in the South.21 While there was some growth in the South during this interval, it is correct to say that the North grew faster and routine growth became even
more firmly established. This experience was rarely confronted by the advocates of openness.22

Doubts about the Washington Consensus
Distortions of course mean (by definition) that the economy operates within its production possibilities frontier with the given technology that is available. So eliminating the distortions would be expected to result in a once-over increase in output; but the real question was what produced growth after this one-shot effect has been absorbed.

The main evidence leading to the outward-oriented position was the success of Korea and Taiwan. Taiwan had begun to grow rapidly from the early 1950s and Korea from the early 1960s. The most obvious feature of these impressive performances was the high growth rate of exports, often non-traditional products and services that had come into existence behind the protection and the subsidies that IS had provided. Similarly it was widely believed initially that the two economies were largely market driven with very little government intervention. Both countries made marked policy changes in this direction in the 1960s. These changes did reduce distortions, but they also included other government policies that subsidized exports and capital formation. In particular it became clear that both countries maintained an undervalued exchange rate over much of the time after 1960. There were other policies and arrangements that impeded imports. Liberalization as practiced by these two countries did not mean ‘free’ trade. It also became understood that both countries had accumulated a great deal of collective learning during the Japanese occupancy of the first part of the twentieth century.23 Korea also learned greatly from the presence of United States armed forces engineers in their country. This collective learning had produced a labor force much more experienced and skilled than in most of the other countries trying the IS strategy. This last item is especially relevant in understanding how and why production in the two countries responded so well and so quickly to the incentives and subsidies offered by their governments.

There was one other characteristic of government policy-making that is highly relevant to this story. Both governments recognized that good policy-making required trial and error and hence willingness to change policies and try different approaches. The idea was not minimal government, but rather a searching, learning government.24 This searching for the right policies was necessary simply because it was recognized that there can be no right policy that can be arrived at in any other way.

The Washington Consensus became widely advocated and strongly pushed by international agencies and key academic figures. To repeat the point made above: that IS, as practiced, had run into a dead end was, in
general, appreciated, but the Washington Consensus seemed simply to go back to an approach to development that had also failed as evidenced in particular by the experience of the 1870–1914 period. Research in the 1970s and 1980s on Korea and Taiwan convinced almost all observers that their success was not due to a minimal government or a market-friendly approach, and that these economies were much less open than was thought to be the case in the 1960s and 1970s.25

There have been, from early on, efforts to examine empirically the effects of IS and outward-orientation policies on growth of GDP, on employment and productivity, and on macrostability. A survey of this literature is beyond the scope of this chapter, but a few remarks may give the flavor of the current status of this work.26 There are two general approaches: case studies of individual countries and cross-country regression analyses involving numerous countries. There are difficulties at all steps along the way in both approaches: definitions and measurement of inward- and outward-looking, quality of available data, the appropriate theoretical formulations, distinguishing the role of trade policy from that of macro stability, education and other possible factors, and many more. Cross-country regressions, once so widely used, are increasingly recognized as subject to so many difficulties that their results are essentially meaningless.27

Three issues of direct relevance to trade policy may be noted briefly. The first is the role of exports. Do countries grow well because they export or do they export because they grow well? At the aggregate level the evidence might well support the former position, but at the firm or sector level, the evidence generally supports the opposite view. The micro data appear somewhat more convincing, that is, firms must find their basic inducements to search and learn that lead to productivity growth within their own indigenous environment, and thereby become able to export. It does seem clear that once they begin to export, then they can gain additional knowledge from that activity. Secondly, evidence supports the view that exporting does not have much effect on productivity growth of traditional activities of the developing countries.28

Thirdly the empirical (and other) evidence supports the view that the exchange rate is a crucial policy variable. The overvalued exchange rate was a major, perhaps the major, reason for the failure of the IS approach. Recent empirical work and case studies show convincingly the strength of an undervalued exchange rate as an instrument that encourages import replacements and the search for foreign markets as well as the search for increased productivity and higher-quality output.29

While it is important that empirical and historical research continue, it now seems likely that for the foreseeable future dispute between the two approaches cannot be resolved by evidence of the conventional sort applied
in the usual way. The most illuminating approach is simply case studies of individual countries to ascertain what has gone on in these countries. Generalizations across countries and through time, supported by rigorous theoretical formulations and by equally rigorous econometric results, do not now appear possible, and certainly not possible to the extent that specific policy formulations emerge that are applicable to all countries. The last section addresses briefly the consequences of this position.

The upshot of it all
An important reason why the role of trade in development is difficult to isolate is that the development process itself is not yet fully understood. There are two main aspects of development that both IS and the Washington Consensus seemed not to have appreciated in their early formulation. The first was the assumption, implicit in most of the literature of both strategies, that technical and other knowledge are public goods, available to everyone and equally productive in all places. Research of the 1980s and 1990s has shown this assumption to be terribly misinformed. Productivity levels among firms in the same activity vary widely within a country and across countries, as do rates growth of productivity, profit rates, capital–labor ratios, product quality, and so on. These great and abiding differences are explained by the fact that much of knowledge cannot be codified and put into manuals, but is tacit and can be accumulated only by producing and is, therefore, unique to the firm where it is accumulated. That this is the case is due to the great differences in entrepreneurship, in the ingenuity of the shop floor people, and in the extent of commitment to searching and learning. This finding has had a fundamental impact on development theory and practice: since sustained growth requires sustained increases in productivity, which, in turn, requires continuing increases in knowledge, a growing economy must be one in which firms are, as a matter of routine, actively engaged in searching for new knowledge. Policy incentives must be geared to this objective, different from the incentives aimed at achieving an ‘optimal’ allocation of given resources and given technology. Trade policy, in particular, is greatly affected by the notion of tacit knowledge and its source.

The second important research finding has to do with institutions: institutions as norms and rules of the game. As such they have direct and significant effects on the response of economic agents to market incentives. Economic actors are rule-followers as well as profit-seekers, and the rules emerge from and in turn create institutions. So whether a policy will have the desired effect depends on the institutional environment within which it is applied, and any change is sure to be resisted to some extent. In particular it means that the usual market signals can be expected to
have diverse results in different countries and at different times in the same country.

Given the two preceding points, it is evident that the textbook notions of ‘optimal’ allocation of resources or some other ideal or maximum state are basically meaningless: technology is always changing, that is, its change is part of the firm’s activity, and institutions mean that the textbook ideal – for example perfect competition – cannot be defined independently of these institutions. To understand a given economy means to gain some insight into how it works and then to show that its workings can be improved upon according to some practical criteria. This is quite different from the more usual objective of trying to create an economy that follows a general strategy, for example perfect competition, Pareto optimality, and so on.

Incentives are to be aimed at creating inducements to search and learn. Import-substitution strategies created investment opportunities, but also dampened any inducements to search and learn, to create tacit knowledge and thereby continue to grow. The Washington Consensus assumed that simply eliminating distortions by virtue of an unfettered market and a ‘realistic’ exchange rate would produce growth. Both were and are inadequate, because they misunderstand tacit knowledge and the role of institutions.

The final question is what kind of a trade policy will induce firms to engage in the searching and learning process that does produce growth and improved quality of output. Foreign trade policy should aim at making it very profitable to export and to replace imports. Such an approach is different from that usually practiced of subsidizing inputs – capital, loanable funds, land, fertilizer, and so on – as noted earlier.

There are several trade policies available that reflect this point of view. The most evident is an undervalued exchange rate, that is, one that results in the accumulation of foreign exchange. Such an exchange rate policy has been pursued by China, Taiwan, South Korea, Brazil and Japan at times to considerable advantage. Foreign aid takes the form of guaranteeing favorable prices for a developing country’s exports, or prices that rise as the exporting firm increases its productivity or tax relief based on increased export earnings. A policy that rewards firms that increase employment over time with a given capital stock has had favorable effects on both employment and productivity growth.

The general conclusion as to the present state of play of the role of trade may be summed up in this way: we have learned over recent decades that simple notions such as IS and outward orientation lead nowhere. Trade policies – including exchange rate policy – must emerge from a clear vision of how development takes place. That vision must explicitly include knowledge accumulation and application and the recognition that much knowledge is necessarily tacit. It must also recognize that institutions are so
fundamental and so history-and culture-dependent that generalizations across countries are risky indeed. Thus a country, in determining its foreign trade policy, must study how its economy in fact works.35

Notes
1. There is some evidence that inequality among nations increased between 1500 and 1800. See, for example, Lindert and Williamson (2003).
2. President Harry Truman, in his inaugural address in early 1949, committed the United States to supply technical and financial aid to the countries of Africa, Asia, and Latin America.
3. This argument was first developed in Raul Prebisch (1950) and by Hans Singer (1950). A later and slightly different version is Lewis (1978).
4. A formal model that demonstrates that free trade can be harmful to long-run growth in those countries that are locked into sectors where productivity is low and whose indigenous technological change is slow or completely absent, is Rodriguez and Rodrik (2001).
5. See the papers by Gregory Clark and Robert Feenstra (2003) and by Clark (1987).
6. The notion of an optimal rate of saving was often introduced. If this optimal rate of domestic saving was not high enough to produce the target rate of growth of output, then foreign aid or borrowing could supplement domestic saving to the extent necessary to reach the target.
7. Gordon C. Winston (1974) is one of the earliest analyses of the underutilization of capital issues.
8. Countries that had long relied heavily on one or two primary exports often experienced the ‘Dutch disease’ even before the beginning of import-substitution policies. The overvaluation pursued as a policy to encourage capital formation often exacerbated an already misleading exchange rate regime.
10. The study by Little and Mirrlees (1974) is a fine review of the state and role of cost-benefit analysis at the time.
11. Anne O. Krueger (1974) first drew the attention of the profession to this important cost of the distortions.
12. African countries shared less in this success than in countries in other parts of the world. This failure had little to do with IS as few African countries made a substantial commitment to the strategy.
13. Although the term ‘import substitution’ seems to imply that imports should be reduced relative to GDP, this was rarely the case. The strategy proved quite import-intensive. See Carlos Diaz-Alejandro (1965).
14. Robert Solow’s neoclassical growth model (Solow, 1956) was motivated largely by the recognition that the fixed-coefficient assumption severely penalized employment growth as well as impeding adjustment of production techniques to the factor supply situation in general.
15. See W. Arthur Lewis (1954) and, somewhat later, Fei and Ranis (1964) for elaborations of the labor surplus models of growth.
16. Chenery and Strout (1966) developed a model in which capacity to import was a constraint on growth along with the capacity to save.
17. The role of ‘tacit knowledge’, in contrast to codified knowledge, is a crucial part of the story and will be discussed later.
18. There were many studies that explored the import substitution experience in specific countries and across several countries. Little et al. (1970) and the six country studies that accompanied it were among the first that attracted attention. Two studies by Balassa (1971) and Balassa et al. (1982) were also important in spreading understanding about the problems of IS, as was Krueger (1978). There are many others.
20. Some calculations show that the correlation between domestic saving and domestic investment was lower in this period than in recent years when capital markets are assumed to work extra well. With ‘perfect’ international capital markets there should be no or very little relationship between domestic saving and investment.

21. Elaboration of this point is in Lewis (1978) and Dowrick and DeLong (2003). See also Maddison (1970).

22. The story is similar if one begins in 1820 except that from 1820 to 1870 there were more tariffs and other interferences with the flow of trade and other forms of market interventions.

23. See Kohli (1994) for a good discussion of the impact of the Japanese occupancy on the creation of human capital in the two Koreas and Japan.

24. Gustav Ranis (in Winrock International Institute for Agricultural Development and USAID, 1991, pp. 128–9) points out that: . . . ‘what happened in Taiwan was not Mandarins sitting around saying is this what we have to do now. There was a lot of bumbling and stumbling and going back and forth.’ See also Biggs et al. (1995).

25. The World Bank study (1993) acknowledged this picture in general, but argued strongly that Korea and Taiwan were distinctive in several ways and their experiences and policy packages could not be replicated in other countries. See also Stiglitz (2000).


27. Jesus Felipe (1999), Durlauf (2000) and Brock and Durlauf (2001) all have especially helpful discussions of cross-country regressions and growth accounting in general.

28. Sectoral and firm studies have become plentiful in recent years, and results can always be questioned. Michael Hobday (1995) is a thorough study of the relationship between exporting and learning. Roberts and Tybout (1996), Aw et al. (1998), Temple (1999), Tybout (2000) and Bernard and Jensen (2001) all have helpful discussions and much data. Westphal (1990) is a particularly good discussion of the role of exports.

29. See Bruton (1997, Chapter 8) and Bruton (1998) for a discussion of the undervalued exchange rate as an instrument of development policy.


31. There are many studies available that confirm this state of things. See the sources cited in notes 28 and 30.

32. Tacit knowledge has been discussed by many people. Nelson and Winter (1982) were perhaps the first to employ it in a strategic way in economic analysis. More recently good discussions are found in papers by Chandler et al. (1998) and Chandler (1992). See also Helleiner (1992).

33. John Williamson (2003, p. 324) lists ten ‘points’ that are included in the Washington Consensus and that will (presumably) produce growth. The list does not include any item that acts directly on growth, but essentially defines the conditions for a perfectly competitive economy with no inflation and a government role limited to conventional post office, defense, infrastructure, and so on. See also Krueger (2000).

34. The differences among firms means that it is a dubious business to speak of a country’s comparative advantage. Some firms in an industry export, some do not. For a country to seek to produce those goods and services in which it has a comparative advantage is to seek that which is not there. See Hausmann and Rodrik (2002) for an interesting empirical study of the determination of activities in a country.

35. This position is increasingly recognized even by economists who differ sharply in other ways. See almost any of the recent writings of Dani Rodrik, especially Rodrik (1999, 2000) and Stiglitz (2000) on the one hand and Srinivasan and Bhagwati (2001) on the other. This point is also discussed in Bruton (1998).

References


39 Foreign direct investment

V.N. Balasubramanyam

Introduction
The vast literature on foreign direct investment (FDI) in developing countries may seem disproportionate to the volume of FDI they harbour. Developing countries as a group accounted for 25 per cent of the total stock of FDI of $8.9 trillion in the world economy at the end of the year 2004, most of which, around 70 per cent, was accounted for by 11 developing countries (Tables 39.1 and 39.2). One reason for the intense interest in FDI is the nature of the beast, which appears to evoke both admiration and opposition in equal measure; admiration because of the unrivalled ability of the multinational enterprises (MNEs), the main purveyors of FDI, to transfer technology and know-how across borders; opposition because the MNE is first and foremost a profit-maximizing entity. In the recent past attitudes towards FDI have turned from a mixture of suspicion and admiration towards one of unqualified admiration, shown by the eagerness of most developing countries to attract FDI. This change in attitude towards FDI on the part of host developing countries is influenced by a number of factors: a steep reduction in alternative sources of finance in the wake of the debt crisis, the collapse of the Soviet Union and with it a waning of ideological opposition to capitalism and its institutions, the demonstrable success of the East Asian countries based in part on FDI, and growth in knowledge and understanding of the nature and operations of FDI on the part of the host countries.

The issues that have surfaced in the recent literature on FDI reflect these changes. These have to do with the specific factors which figure prominently in the choice of locales for investment by foreign firms, and the factors which promote effective transfer of technology to the host countries and maximize the benefits they can expect from FDI. In addition, there is the suggestion mooted principally by the EU and Japan that FDI should be on the agenda of the World Trade Organization (WTO) on a par with trade in goods and services.

This chapter reviews these and other issues in the literature on FDI. The next section discusses the determinants of FDI, drawing on the theoretical and empirical literature. The subsequent section discusses the impact of FDI on growth and development in host developing countries. The final section draws some conclusions.
Determination

Stephen Hymer (1976) initiated the discussion on the determinants of FDI with the thesis that firms go abroad to maximize the rents inherent in the advantages they own. These advantages range from the possession of a

<table>
<thead>
<tr>
<th>Table 39.1 Stock of foreign direct investment: 1990–2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>World</td>
</tr>
<tr>
<td>$ US trillion</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
<tr>
<td>Developed countries</td>
</tr>
<tr>
<td>$ US trillion</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
<tr>
<td>Developing countries</td>
</tr>
<tr>
<td>$ US trillion</td>
</tr>
<tr>
<td>Percentage</td>
</tr>
</tbody>
</table>

Source: UNCTAD (various issues).

<table>
<thead>
<tr>
<th>Table 39.2 Stock of inward foreign direct investment in developing countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000–01</td>
</tr>
<tr>
<td>US $ million</td>
</tr>
<tr>
<td>All developing economies</td>
</tr>
<tr>
<td>Argentina</td>
</tr>
<tr>
<td>Brazil</td>
</tr>
<tr>
<td>Mexico</td>
</tr>
<tr>
<td>Africa</td>
</tr>
<tr>
<td>China</td>
</tr>
<tr>
<td>China, Hong Kong SAR</td>
</tr>
<tr>
<td>India</td>
</tr>
<tr>
<td>Korea, People's Republic of</td>
</tr>
<tr>
<td>Malaysia</td>
</tr>
<tr>
<td>Singapore</td>
</tr>
<tr>
<td>Thailand</td>
</tr>
<tr>
<td>Total of 11 countries</td>
</tr>
</tbody>
</table>

Source: UNCTAD (various issues).

Determinants

Stephen Hymer (1976) initiated the discussion on the determinants of FDI with the thesis that firms go abroad to maximize the rents inherent in the advantages they own. These advantages range from the possession of a
brand name to complex processes and product technologies. Hymer’s work set in train a number of studies on the foreign investment decision process of foreign firms: John Dunning (1973, 1981), Buckley and Casson (1991), and Markusen (2004) who encapsulated the key factors in the foreign investment decision process of firms as the ‘OLI paradigm’. Simply put, firms must possess ownership advantages (O), location advantages (L) and must be capable of internalizing operations (I) if they are to invest abroad. Internalization here refers to the ability of the firm to set up its own suppliers of materials and equipment (backward linkages) and marketing operations (forward linkages). Such internalization is necessary to overcome imperfections in the market which may result in a number of problems including the loss of ownership advantages through imitation by others, hold-up of operations by suppliers of materials and equipment, and the problems associated with decision-making with imperfect information (Williamson, 1981).

The concern of the host countries is with location advantages (L): how best to lure firms, which possess ownership advantages and are able to internalize, to their locale and benefit from the technology and know-how they bring along and the jobs they create. There is now much received wisdom on the sort of L factors which will attract FDI, the result of years of research and the experience of host countries with FDI. It is now well known that artificial incentives such as tax concessions and tax holidays offered to foreign firms do not always attract foreign investors; such incentives are not a substitute for the sort of economic environment foreign firms seek. Countries endowed with the sort of raw materials the foreign firms seek, those blessed with cheap but productive labour (low efficiency wages), and economies which enjoy macroeconomic stability and pursue stable economic policies tend to attract relatively large volumes of FDI. In the absence of these fundamentals, tax concessions and various sorts of subsidies are of little attraction to foreign firms.

In the same vein, it is suggested that policy-induced incentives such as tariffs on imports and subsidies to exports do not lure foreign firms. In fact a neutral trade policy which favours neither production for exports nor the production of import substitutes for the home market tends not only to attract large volumes of FDI but also promotes efficient utilization of FDI. The import-substituting domestic market-oriented strategy (IS strategy) is characterized by tariffs and quotas on imports, which on average outweigh the average level of subsidies given to exports. In the case of the export-oriented strategy (EP strategy) the opposite is the case. These sorts of incentives, which bias production in favour of either exports or the domestic markets, distort allocation of resources in the economy. They are also artificial and uncertain incentives. They are artificial in the sense that they do
not reflect the true market-dictated comparative advantage of various sectors in the economy; they are uncertain in the sense that they are subject to the vagaries of government policy. Foreign firms in general are not attracted by such artificial incentives and those that do respond and operate in the distorted environment may reap benefits but they tend to be transient. A neutral strategy favours neither the export markets nor the domestic markets; resource allocation in such an environment is dictated by market forces. Foreign firms seek such an environment, which allows for the full play of the comparative advantage they possess. There is robust statistical support for the proposition enunciated by Bhagwati (1978) that a neutral strategy attracts relatively large volumes of FDI and also promotes its efficacy (Balasubramanyam and Salisu, 1991; Balasubramanyam et al., 1996).

The recent experience of India and China with FDI, though, does not seem to conform to received wisdom on the determinants of FDI. They both possess most if not all of the location advantages foreign firms seek. Both countries are endowed with relatively cheap labour, both have liberalized their trade and FDI regimes to a considerable extent, they both possess large domestic markets and they have posted impressive growth rates in recent years. But they differ markedly in the volume of FDI they have attracted. China attracts ten times more FDI than India does: in recent years the annual average inflows into China have averaged around $50 billion compared with the $4 billion that India attracts (Figure 39.1 and

---

**Figure 39.1** Inflows of foreign direct investment in China and India: 1980–2004 (US million)

*Source: UNCTAD*
Table 39.3  Inflows of Foreign Direct Investment in China and India ($ US million)

<table>
<thead>
<tr>
<th>Year</th>
<th>China</th>
<th>India</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>57</td>
<td>79</td>
</tr>
<tr>
<td>1981</td>
<td>265</td>
<td>92</td>
</tr>
<tr>
<td>1982</td>
<td>430</td>
<td>72</td>
</tr>
<tr>
<td>1983</td>
<td>916</td>
<td>6</td>
</tr>
<tr>
<td>1984</td>
<td>1,419</td>
<td>19</td>
</tr>
<tr>
<td>1985</td>
<td>1,956</td>
<td>106</td>
</tr>
<tr>
<td>1986</td>
<td>2,244</td>
<td>118</td>
</tr>
<tr>
<td>1987</td>
<td>2,314</td>
<td>212</td>
</tr>
<tr>
<td>1988</td>
<td>3,194</td>
<td>91</td>
</tr>
<tr>
<td>1989</td>
<td>3,393</td>
<td>252</td>
</tr>
<tr>
<td>1990</td>
<td>3,487</td>
<td>237</td>
</tr>
<tr>
<td>1991</td>
<td>4,366</td>
<td>75</td>
</tr>
<tr>
<td>1992</td>
<td>11,008</td>
<td>252</td>
</tr>
<tr>
<td>1993</td>
<td>27,515</td>
<td>532</td>
</tr>
<tr>
<td>1994</td>
<td>33,767</td>
<td>974</td>
</tr>
<tr>
<td>1995</td>
<td>37,521</td>
<td>2,151</td>
</tr>
<tr>
<td>1996</td>
<td>41,726</td>
<td>2,525</td>
</tr>
<tr>
<td>1997</td>
<td>45,257</td>
<td>3,619</td>
</tr>
<tr>
<td>1998</td>
<td>45,463</td>
<td>2,633</td>
</tr>
<tr>
<td>1999</td>
<td>40,319</td>
<td>2,168</td>
</tr>
<tr>
<td>2000</td>
<td>40,715</td>
<td>2,319</td>
</tr>
<tr>
<td>2001</td>
<td>46,878</td>
<td>3,403</td>
</tr>
<tr>
<td>2002</td>
<td>52,743</td>
<td>3,449</td>
</tr>
<tr>
<td>2003</td>
<td>53,505</td>
<td>4,269</td>
</tr>
<tr>
<td>2004</td>
<td>60,630</td>
<td>5,335</td>
</tr>
</tbody>
</table>

Source: UNCTAD (various issues).

Table 39.3). Indeed India’s FDI regime is reported to be much more liberal than that of China (Nagaraj, 2003). Even so, the volume of FDI in India is lower than that in China by a factor of ten or more, a fact frequently debated in the media.

A number of reasons are offered for the observed differences in the volume of FDI in the two countries including differences in the accounting procedures between the two countries, so-called round-tripping FDI in China, and the horrendous Indian bureaucracy which stiﬂes any sort of enterprise, be it foreign or domestic. There may be other explanations for the observed differences in the volume of FDI the two countries harbour. Yashang Huang (2003) is of the view that a considerable volume of FDI in
China is a substitute for domestic investment. Despite the relatively high savings rate, domestic Chinese investors find it hard to obtain credit lines and borrow from the banks, and they turn to foreign investors for finance. Many of the state-owned enterprises on the verge of bankruptcy also seek FDI, and foreign firms invest in regions of China which local firms hesitate to enter. It may also be the case that the requirements of the Indian manufacturing and services sectors are relatively low compared with those of China, simply because India is better endowed than China with the sort of human skills which FDI provides. In sum the determinants of FDI are rooted in the endowments of the host countries including human skills, the infrastructure facilities they provide and, most importantly, a policy framework which provides distortion-free product and labour markets.

Impact

FDI is defined as ownership of facilities abroad with control over operations. The distinguishing feature of FDI which sets it apart from other forms of capital flows is the control over operations which the parent company exercises over its subsidiaries abroad. As stated earlier, such control over operations is essential to preserve the ownership over advantages the firm enjoys. Such control over operations is exercised by virtue of ownership of equity and possession and control of technology and know-how. The higher the degree of equity participation by the investor firm in an entity, the greater is its ability to exercise control over operations. It is thought possible for a firm to exercise control over operations if it owns say only 4 per cent of the total equity of an entity and the other 60 per cent is disbursed amongst a number of investors. In fact, the International Monetary Fund (IMF) defines any investment by a firm with an equity share of 10 per cent as FDI. The second attribute of FDI – ownership of technology and know-how – is a much more powerful tool for exercising control over operations than majority equity ownership.

The three attributes of FDI, equity ownership, control over operations and transfer of technology, are intertwined. Ownership of equity and technology enables the firm to exercise control over operations and preserve its monopoly over technology and know-how, which in turn enables it to transfer technology and know-how across frontiers. The essential point to note here is that ownership of equity is a means to an end, the end being control over operations and transfer of technology. If the market for technology were perfect and if technology and knowledge were not public goods, the multinational company – the purveyor of FDI – would prefer to enter international markets by contracting technology-licensing agreements. Such agreements, by definition, are bereft of equity participation: the firm transfers technology to the licenses in return for fixed technical fees.
and royalties tied to the profits of the licensee. But because technology can be easily imitated and it is difficult to arrive at a price for most technologies, multinationals engage in FDI, that is, they own equity in entities abroad.

This somewhat long-drawn-out discussion of FDI is to emphasize the fact that it is technology and know-how transfer which is the main attribute of FDI, and it is this attribute which is sought after by countries host to FDI. Admittedly the capital that accompanies technology and know-how is also a benefit to the host countries, but for reasons specified above, transfer of capital is not a major feature of FDI. In any case, the larger the equity share of the foreign firm in an entity, the higher would be the control over operations it exercises, something which most host countries would wish to minimize. Also, relatively poor countries, which have very few of the attributes discussed in the section on determinants, may not be able to attract FDI in large enough volumes to meet their demands for capital.

FDI is also sought for the foreign exchange it provides developing countries, in the form of both the capital that accompanies it and the exports it promotes. The contribution of FDI to China’s exports is well known (Wei, 2004), and also to those of other East Asian countries such as Malaysia and Singapore (Driffield et al., 2004). FDI can also save foreign exchange for the host countries with productive investments in import-substituting industries. All of this, of course, makes for a healthy balance of payments of the country. But here again it should be noted that any investment, be it export-oriented or domestic market-oriented, contributes to the balance of payments as long as it is socially productive. This is simply because the balance of payments, as Kindleberger (1969), commenting on the balance of payments effects of FDI reminded us, is a general equilibrium phenomenon. In other words, the balance of payments is an integral part of the total economy. An FDI project which is socially productive, in the sense that the private rate of return to the investment does not exceed the social rate of return, will contribute to the balance of payments. Host countries which offer various fiscal incentives such as tax holidays and the institution of export-processing zones to lure foreign firms may be giving away income to the foreigners if the private rates of returns exceed the social rates, a conceivable outcome in the presence of market distortions which the incentives are supposed to rectify.

In sum the most significant benefit from FDI to the host countries is the technology and know-how it transfers. It is such transfers which augment the skill endowments of host countries and promote employment for local labour. How does the technology which is transferred get transmitted to the local economy? There are several channels, including imitation, acquisition of skills, competition and various sorts of tie-ups between the
foreign-owned and locally owned firms. Imitation of the products produced by foreign affiliates through reverse engineering, an activity that enables local firms to copy the processes and design of new products, is a recognized channel for spillovers. The acquisition of skills occurs mainly through the movement of skilled labour employed by foreign affiliates to locally owned firms. Such internal migration of labour, trained by foreign affiliates, is a significant channel for spillovers. Labour employed in foreign affiliates may wish to set up their own establishments with the experience and skills gained from their sojourn in the foreign affiliates. Also, foreign affiliates may, either in response to performance requirements imposed by the host country or because of distinct cost advantages, train or establish local suppliers of components and parts. This too would be a channel for spillovers.

Another potent channel for spillovers – or, more to the point, growth of productive efficiency – is competition. The theory here is that the entry of foreign affiliates increases competition in the marketplace and locally owned firms are compelled to increase their productive efficiency. This is the sort of efficiency recognized in the literature as ‘X-efficiency’ rather than allocative efficiency. Finally, locally owned firms may learn marketing techniques and methods of penetrating export markets from export-oriented foreign affiliates. This would count as a specific sort of technology transfer.

These propositions have been extensively tested in the context of FDI in developed and developing countries (Haddad and Harrison, 1993; Blomstrom and Kokko, 1998; Borensztein et al., 1998; Gorg and Greenaway, 2001). These econometric studies have produced a mixed bag of results: some identify positive spillovers from the presence of foreign affiliates in manufacturing industries, and others find them to be either negligible or negative.

These studies identify a number of factors that are likely to promote spillovers of technology and know-how from foreign affiliates to locally owned firms. First, the magnitude of spillovers tends to be high in industry segments in which the gap in technological capabilities between foreign affiliates and locally owned firms tend to be narrow. Second, spillovers are likely to be high when the competition in the marketplace between locally owned firms and foreign affiliates tends to be intense. Third, the extent and magnitude of spillovers differ between industries and host countries. Fourth, several studies show that spillovers are proportional to the magnitude of foreign presence, measured by shares of foreign affiliates in total equity or sales of the relevant industry groups. Fifth, local capabilities (including research and development – R&D – and human skills) sustain high levels of spillovers. Finally, analogous to the last proposition, the liberalization of foreign trade, increased competition and development of local infrastructure all promote spillovers.
The message of all this is clear. Increased volumes of FDI alone are unlikely to generate widespread spillovers. In the absence of competition and cooperator factors such as local R&D and human skills, spillovers from FDI may be limited. Put another way, FDI is a catalyst of technical change and growth; it cannot be expected to be the prime mover. Indeed empirical research suggests that FDI is most effective as an agent of change in economies that possess a threshold level of human capital and skills and in those economies that have attained a threshold level of growth (Blomstrom et al., 1994; Balasubramanyam et al., 1999).

In sum, in the absence of the necessary ingredients and cooperator factors, large volumes of FDI alone may not be effective in promoting growth and may even be counterproductive. For these reasons the exuberance relating to the role of FDI in the growth process and exhortations that developing countries should adopt a wide open door policy towards FDI should be tempered by a recognition of the conditions necessary for the effective utilization of FDI.

Conclusions
Recent literature on FDI reflects the substantial change in attitudes towards FDI by the developing countries. In the past, attitudes towards FDI and its role in the development process ranged to extremes – from hostility to ardent advocacy. In the recent past, there is a growing appreciation of its role in the development process and most developing countries have sought to attract increasing volumes of FDI. Research on FDI reflects this change in attitude. Much of the literature now is centred on econometric testing of the determinants and impact of FDI rather than the polemical debates on the role of FDI. Detailed case studies which provide analytical insights into the nature of FDI and its impact on development, of the sort done by Sanjaya Lall (1983), are unfortunately few and far between.

The message of the recent literature, though, is clear. First, FDI is attracted to countries that can provide the sort of environment which allows foreign firms to establish a foothold and successfully exploit the rents in the advantages they possess. The sort of environment which is conducive to the operations of foreign firms is characterized by a stable macro-economic environment with stable FDI regimes and an assured supply of cooperator factors including human capital. Second, FDI is a superb catalyst of development but not an initiator. The new-found enthusiasm for FDI on the part of developing countries is based on the success stories of the East Asian countries with FDI and the drying-up of alternative sources of finance such as bank credit. It is, though, worth noting that the successful utilization of FDI is contingent upon a number of factors discussed in the foregoing. In this context it is worth recalling Paul Streeten’s observa-
tion written during the 1970s (Streiten, 1971) when controversy on the role of the multinational enterprise as the purveyor of FDI was at its height: ‘it is not sensible to transfer income by attempting to transform the MPE [multinational production enterprise] from what it is – a profit seeking animal – into something it is not – a public service’.

Note
1. The author is grateful to MS Jasleen Sindhu for research assistance with this chapter.

References


Standard economic theory traditionally argued that international private capital flows will make a major contribution to development to the extent that they will flow from capital-abundant industrialized countries to capital-scarce developing countries, and help to smooth spending throughout the business cycle in capital-recipient countries.

In recent years, reality has contradicted both aspects (WESS, 2005). Between 1997 and 2004, developing countries have transferred a large amount of resources to developed countries. In addition to this, private capital flows to developing countries are highly concentrated in a group of large middle-income countries and are particularly insufficient for low-income and small countries. Secondly, private capital flows to developing countries have been highly volatile and reversible; as a consequence, they have been a major factor in causing developmentally costly currency and financial crises. Rather than smooth domestic expenditure, private capital flows seem to have contributed to making it more volatile.

Boom–bust cycles of capital flows have been particularly damaging for developing countries, when they both directly increase macroeconomic instability and reduce the room for manoeuvre to adopt countercyclical macroeconomic policies, and indeed generate strong biases towards adopting procyclical macroeconomic policies (Kaminsky et al., 2004; Stiglitz and Levy, 2005). Furthermore, there is now overwhelming evidence – accepted by institutions like the International Monetary Fund – that procyclical financial markets and procyclical macroeconomic policies have not encouraged growth and, on the contrary, have increased growth volatility in those developing countries that have integrated to a larger extent into international financial markets (Prasad et al., 2003).

The costs of financial volatility for economic growth are high, as it can generate cumulative effects on capital accumulation (Easterly, 2001). Indeed, major reversals of private flows have led to many developmentally and financially costly crises, which have lowered output and consumption well below what they would have been if those crises had not occurred. Eichengreen (2004) estimated that income of developing countries had been 25 per cent lower since 1980 than it would have been had such crises not occurred, with the average annual cost of the crises being just over $100 billion. Griffith-Jones and Gottshalk (2006) have estimated a similar
though somewhat higher annual average cost of crises in the period 1995–2002, of $150 billion in terms of lost gross domestic product (GDP).

These features are by no means inevitable. An appropriate domestic and international environment can improve the capacity of developing countries to benefit from private capital flows. In what follows we examine both characteristics of private capital flows to developing countries and some policy options that would improve their development impact.

Main characteristics of private flows

The volatility and reversibility of capital flows to emerging countries and the marginalization of many of the poorer and smaller developing economies with respect to financial markets are rooted in the combination of financial market failures and basic asymmetries in the world economy (Ocampo, 2001).

Instability is inherent in the functioning of financial markets (Keynes, 1936; Minsky, 1982). Indeed, boom–bust patterns in financial markets have occurred for centuries (Kindleberger, 1978). The basic reason for the existence of these patterns is that finance deals with future information that, by its very nature, is not known in advance; therefore, opinions and expectations about the future rather than factual information dominate financial market decisions. This is compounded by asymmetries of information that characterize financial markets (Stiglitz, 2000). Owing to the non-existence or the large asymmetries of information, financial agents rely to a large extent on the ‘information’ provided by the actions of other market agents, leading to interdependence in their behaviour, that is to say, contagion and herding. At the macroeconomic level, the contagion of opinions and expectations about future macroeconomic conditions tends to generate alternating phases of euphoria and panic. At a microeconomic level, it can result in either permanent or cyclical rationing of lending to market agents that are perceived by the market as risky borrowers. In many cases it is the endogenous behaviour of international financial markets that conditions or strongly influences fundamentals in developing countries. A supply-led, large capital inflow affects the domestic economic situation (for example by generating an asset price bubble or an overvalued exchange rate) in a way that can increase inflows. This can lead to costly macroeconomic crises, which makes regulation and other state intervention in international financial markets essential.

Herding and volatility seem to be accentuated by some features of the functioning of modern markets. An important element in the increased volatility of international bank lending is the use of modern risk management models (such as Value at Risk). As Persaud (2003) points out, the intrinsic problem with market-sensitive risk management systems is that
they incorrectly assume that banks act independently when in fact their decisions are interconnected. When many banks try to sell the same asset at the same time, and there are few or no buyers, prices fall and volatility increases. As prices collapse, for liquidity reasons banks try to sell another asset, which may have been previously uncorrelated with the first. This increases the volatility of the second asset and also correlation. This prompts repeated rounds of selling among agents who use similar models, and generalized herding takes place. The adoption of banks’ own risk management models to determine their required levels of capital in the internal ratings approach, as proposed in the new Basel Capital Accord, could seriously increase banks’ tendency for procyclicality in lending, exacerbating both booms and crashes.

An additional source of concern is the evidence that the Value at Risk (VaR) models first developed by banks are being extensively adopted by fund managers and pension funds, leading to similar herding patterns and to procyclicality in their investment. Therefore, herding is not restricted to one class of actor, but is spreading among many actors. The increasing use of similar market-sensitive risk management techniques (Persaud, 2000) and the dominance of investment managers aiming for very short-term profits, evaluated and paid at very short-term intervals (Griffith-Jones, 1998; Williamson, 2003), seem to have increased the frequency and depth of boom–bust cycles. The downgrade by a rating agency or any other new information available to investors may lead them to sell bonds and stop banks from lending to specific markets; simultaneously, reduced liquidity – owing, for example, to margin calls associated with derivative contracts in these markets – or contagion of opinions about the behaviour of different market segments that are believed to be correlated with a market facing a sell-off, will lead market agents to sell other assets or to stop lending to other markets. Through these and other mechanisms, contagion spreads both across countries and across different flows.

Different types of capital flows are subject, however, to different volatility patterns. In particular, the higher volatility of short-term capital indicates that reliance on such financing is highly risky (Rodrik and Velasco, 1999), whereas the smaller volatility of FDI vis-à-vis all forms of financial flows is considered a source of strength. However, even FDI does have volatile components. A particularly recent concern is that multinational companies, especially those selling in domestic markets, hedge their foreign exchange rate risk. This is particularly problematic when such hedging is done far more when a major devaluation is likely, as this will put additional pressure on the exchange rate and on the reserves. Naturally, such risks tend to become less important as national financial development deepens.
Capital account cycles involve short-term fluctuations, such as the very intense movements of spreads and interruption (rationing) of financing. These phenomena were observed during the Asian and, particularly, during the Russian crisis. However, and perhaps more importantly, they also involve medium-term fluctuations, as the experience since 1980 indicates. During those decades, the developing world experienced two such medium-term cycles that left strong imprints on the growth rates of many countries: a boom of external financing (mostly in the form of syndicated bank loans) in the 1970s, followed by a debt crisis in a large part of the developing world in the 1980s, and a new boom in the 1990s (now mostly portfolio flows), followed by a sharp reduction in net flows since the Asian crisis.

Improved economic conditions in developing countries, as well as the higher global growth, drove a recovery of private capital flows to developing countries in 2003, 2004 and 2005, perhaps signalling the beginning of a new cycle.

More importantly, net transfers of financial resources from developing countries have not experienced a positive turnaround and, on the contrary, continued to deteriorate in 2004 for the seventh year in a row, reaching an estimated $350 billion in 2004 (see Table 40.1). Periods of negative net transfers of financial resources from developing countries (especially from Latin America) have been frequent throughout history; indeed, Kregel (2004) provides evidence that these negative net transfers have been the rule rather than the exception.

Recently, these large and increasing net transfers of financial resources are explained by the combination of relatively low net financial flows and accumulation of very large foreign exchange reserves. Indeed, the most significant aspect of the net outflows from developing countries in recent years has been the growth in official reserves, particularly in Asia (Table 40.1). Accumulation of reserves initially had a large component of ‘self-insurance’ against financial instability, a rational decision of individual countries in the face of the limited ‘collective insurance’, often accompanied by what countries see as undesirable conditionality, provided by the international financial system. However, reserve accumulation in Asia has now clearly exceeded the need in several countries for self-insurance, raising increasing questions about the balance of costs and benefits of additional accumulation, especially if such reserves are invested in low-yielding assets and particularly in a currency, the United States dollar, that may at some point fall quite sharply. At a more fundamental level, the fact that countries like China and India, with very low levels of income per capita and large numbers of poor people – even though they have such dynamic growth – are transferring significant resources to finance developed countries, and especially the US, contradicts theory and is ethically undesirable.
Table 40.1  Net transfer of financial resources to developing countries and economies in transition, 1993–2004 (US$40 billion)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing countries</td>
<td>69.3</td>
<td>35.8</td>
<td>42.9</td>
<td>19.9</td>
<td>-5.2</td>
<td>-37.9</td>
<td>-127.4</td>
<td>-186.5</td>
<td>-153.7</td>
<td>-205.5</td>
<td>-274.8</td>
<td>-353.8</td>
</tr>
<tr>
<td>Africa</td>
<td>1.1</td>
<td>4.0</td>
<td>6.4</td>
<td>-5.8</td>
<td>-4.7</td>
<td>15.6</td>
<td>4.3</td>
<td>-26.2</td>
<td>-14.7</td>
<td>-5.6</td>
<td>-20.2</td>
<td>-32.8</td>
</tr>
<tr>
<td>Sub-Saharan</td>
<td>8.6</td>
<td>6.7</td>
<td>7.4</td>
<td>5.3</td>
<td>7.5</td>
<td>12.1</td>
<td>9.1</td>
<td>3.0</td>
<td>7.9</td>
<td>6.4</td>
<td>6.5</td>
<td>3.9</td>
</tr>
<tr>
<td>(excluding Nigeria and South Africa)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eastern and Southern Asia</td>
<td>18.7</td>
<td>1.0</td>
<td>22.1</td>
<td>18.5</td>
<td>-31.1</td>
<td>-128.2</td>
<td>-142.7</td>
<td>-121.3</td>
<td>-113.1</td>
<td>-142.1</td>
<td>-147.5</td>
<td>-167.8</td>
</tr>
<tr>
<td>Western Asia</td>
<td>33.1</td>
<td>13.2</td>
<td>15.6</td>
<td>5.3</td>
<td>6.2</td>
<td>28.5</td>
<td>-0.9</td>
<td>-39.1</td>
<td>-32.0</td>
<td>-26.7</td>
<td>-47.6</td>
<td>-79.9</td>
</tr>
<tr>
<td>Latin America</td>
<td>16.4</td>
<td>17.7</td>
<td>-1.2</td>
<td>1.8</td>
<td>24.5</td>
<td>46.2</td>
<td>11.8</td>
<td>0.1</td>
<td>6.1</td>
<td>-31.1</td>
<td>-59.5</td>
<td>-73.4</td>
</tr>
<tr>
<td>Economies in transition</td>
<td>1.8</td>
<td>-3.9</td>
<td>-2.3</td>
<td>-6.2</td>
<td>2.7</td>
<td>3.0</td>
<td>-24.0</td>
<td>-48.8</td>
<td>-30.5</td>
<td>-27.0</td>
<td>-34.4</td>
<td>-57.6</td>
</tr>
<tr>
<td>Memorandum item:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavily indebted poor countries (HIPC)</td>
<td>8.5</td>
<td>7.1</td>
<td>6.3</td>
<td>6.8</td>
<td>7.1</td>
<td>8.6</td>
<td>10.1</td>
<td>8.8</td>
<td>8.8</td>
<td>9.9</td>
<td>10.6</td>
<td>11.3</td>
</tr>
</tbody>
</table>

Sources: UN/DESA, WESS (2005).
As private flows recover, an important question for policy-makers in developing countries is whether they will be sufficient and particularly more stable and less reversible than in the past. In this regard, the dominant role of FDI and the fact that it has been relatively stable in times of crises, are positive. However, not all components of FDI are equally stable. Furthermore, multinational companies, especially those producing for the local market, increasingly hedge their short-term foreign exchange risks, particularly when devaluations seem likely. This can lead to major temporary outflows of capital and significant pressure on exchange rates (Ffrench-Davis and Griffith-Jones, 2003; Persaud, 2003). More generally, the increasing use of financial engineering and of derivatives (as well as the growing scale and complexity of derivatives discussed below) seems to make the hypothesis of a hierarchy of volatility, whereby some categories of flows are more stable than others, less clear-cut.

Another potentially positive effect is the greater interest shown by institutional investors (such as life insurers) in investing in emerging countries (European Central Bank, 2005). However, the large rise in ‘carry trade’ – that is to say, investment in high-yielding emerging-market instruments using debt raised at lower cost in mature markets – makes those flows vulnerable to narrowing of interest rate differentials. Furthermore, the large fall in emerging countries’ bond spreads during 2004–05 (while naturally positive in itself for borrowing countries) has raised concerns that this increases the vulnerability of developing countries to international changes.

Finally, there are two structural trends that may add stability. The first is attested by the greater importance of local currency bond markets in developing countries; the second by the fact that international banks have increasingly ‘crossed the border’, lending from their local branches in local currency, and usually fund themselves via domestic deposits. This makes countries less vulnerable to crises, although it also implies that foreign banks are contributing less – or no – foreign savings.

At the time of writing in 2005 there were thus mixed signs in respect of whether the new inflows will be more stable than in the past. Therefore, policy efforts must be made, both in source and in recipient countries, to encourage more stable flows and discourage large flows that are potentially more reversible.

**Measures to counter procyclicality of private flows**

To counter the boom–bust pattern that characterizes private capital flows, several options are available. We here consider two: (1) designing mechanisms to encourage more stable private flows (countercyclical guarantees) or that distribute better the risk faced by developing countries throughout the business cycle (indexed bonds and bonds denominated in the currency
of developing countries); and (2) introducing prudential regulations on the capital account. We also consider the likely effect of the New Basel Capital Accord (Basel II) on patterns of capital flows to developing countries. The procyclical pattern of private capital flows gives a compensatory role also to official financing, in relation to official development financing and to emergency (balance-of-payments) financing, respectively (for a discussion of these, see for example Griffith-Jones and Ocampo, 2003 and WESS, 2005, Chapters III and VI).

**Countercyclical financing instruments**

One way of addressing the problems created by the inherent tendency of private flows to be procyclical is for public institutions to issue guarantees that have countercyclical elements (Griffith-Jones and Fuzzo de Lima, 2004). In this regard, multilateral development banks and export credit agencies could introduce explicit countercyclical elements in the risk evaluations they make for issuing guarantees for lending to developing countries. This would imply that when banks or other private lenders lowered their exposure to a country, multilateral development banks or export credit agencies would increase their level of guarantees, if they considered that the country’s long-term fundamentals were basically sound. When private banks’ willingness to lend increased, multilateral development banks or export credit agencies could reduce their exposure.

There have also been proposals to introduce GDP-indexed bonds. The coupon payments on these bonds would vary in part with the growth rate of the debtor’s economy, being higher in years of rapid growth of GDP (measured in an international currency) and lower in years of below-trend growth. It has been argued that such instruments would improve the cushioning of emerging-market borrowers against adverse shocks by making debt payments more contingent on the borrower’s ability to pay. GDP-indexed bonds would therefore restrict the range of variation of the debt-to-GDP ratio and hence reduce the likelihood of debt crises and defaults. At the same time, they would also reduce the likelihood of procyclical fiscal policy responses to adverse shocks (Griffith-Jones and Sharma, 2006).

Another alternative for better managing the risks faced by developing countries throughout the business cycle consists in the introduction of local currency-denominated bonds. These bonds offer, in particular, a cure against the currency mismatches that characterize the debt structure of developing countries. At the domestic level, the development of domestic capital markets, especially bond markets, also creates a more stable source of local funding for both the public and private sectors, thereby mitigating the funding difficulties created by sudden stops in cross-border capital. In addition to proposals for institutional measures to develop local capital
markets, there have also been innovative proposals to make local currency investments more attractive to international investors. Spiegel and Dodd (2004) have suggested raising capital in international markets by forming diversified portfolios of emerging-market local currency debt issued by sovereign governments. These portfolios of local currency government debt securities would employ risk management techniques of diversification to generate a return-to-risk that competed favourably with other major capital market security indices.

**Prudential capital account regulations**

The accumulation of risks that developing countries face during capital account booms depends not only on the magnitude of private and public sector debts but also on maturity and currency mismatches on the balance sheets of financial and non-financial agents. Thus, capital account regulations potentially have a dual role: as a macroeconomic policy tool with which to provide some room for countercyclical monetary policies that smooth out debt ratios and spending; and as a ‘liability policy’ designed to improve private sector external debt profiles (Ocampo, 2003).

Overall, the experiences with capital account regulations in the 1990s were useful for improving debt profiles, giving governments more latitude in pursuing stabilizing macroeconomic policies, and insulating countries from some of the vagaries of capital markets. There is much evidence that, if well implemented, the benefits far outweigh the costs (Stiglitz and Levy, 2005; Ocampo and Palma, 2005).

One type of capital account regulations are price-based regulations. The basic advantages of price-based instruments are their simplicity and their focus on averting the build-up of macroeconomic disequilibria and, ultimately, preventing crises. A highly significant innovation in this sphere during the 1990s was the establishment in Chile and Colombia of an unre-munerated reserve requirement for capital inflows.

It is noteworthy that institutions such as the International Monetary Fund and the Bank for International Settlements have increasingly concluded that these controls were effective in important aspects. There is broad agreement that they were effective in reducing short-term debt flows and thus in improving or maintaining good external debt profiles. There is greater controversy about their effectiveness as a macroeconomic policy tool. Nonetheless, it can be asserted that reserve requirements helped countries maintain higher domestic interest rates during periods of euphoria in international financial markets.

On the other hand, quantity-based capital account regulations might be preferable when the policy objective is to reduce significantly domestic macroeconomic sensitivity to international capital flows.
The experience of the Asian countries that maintained quantity-based restrictions throughout the 1990s suggests that those restrictions might indeed also be particularly effective in preventing crises. China, India, Taiwan Province of the Republic of China and Vietnam offer successful examples in this regard. Despite the slow and cautious liberalization that has taken place in several of these economies since the early 1990s, the use of such traditional regulations has helped them prevent contagion from the East Asian crisis (see for example, in relation to India, Reddy, 2000).

Malaysia offers an interesting example of the effective use of quantitative regulations during the 1990s. Kaplan and Rodrik (2001) and others provide evidence that Malaysian regulations during the Asian crisis gave the government space within which to enact expansionary monetary and fiscal policies that contributed to the speedy recovery of economic activity.

Although quantity-based restrictions can be effective if authorities wish to limit capital outflows during crises, crisis-driven quantitative controls generate serious credibility issues and may be ineffective in the absence of a strong administrative capacity. A tradition of regulation may be necessary, and the tightening or loosening of permanent regulatory regimes through the cycle may be superior to the alternation of different (even opposite) capital account regimes.

It should be emphasized that capital account regulations should always be seen as an instrument that provides an additional degree of freedom to the authorities with respect to their adopting sensible counter-cyclical macroeconomic policies, but never as a substitute for those policies.

**Basel II and developing countries**

The right regulatory and supervisory regime is essential for maintaining domestic financial stability. In a globalized economy, some common standards of regulation and supervision may be also essential to guarantee global financial stability. This has been the major motivation behind the principles adopted by the Basel Committee on Banking Supervision in recent decades. The second generation of these standards (Basel II), agreed to in June 2004, takes a further step in aligning regulatory capital with the risks in bank lending, and in adapting regulations to the complexities of risk management.

There are fears that Basel II creates the risk of a sharp reduction in bank lending to developing countries, and of an increase in the cost of a significant part of the remaining lending, particularly in the case of low-rated borrowing countries. An equal cause for concern is the danger that Basel II will accentuate the procyclicality of bank lending, which is damaging for all economies, but particularly so for fragile developing ones, which are more vulnerable to strong cyclical fluctuations of financing.
Indeed, the proposed internal ratings-based (IRB) approach of Basel II overestimates the risk of international bank lending to developing countries, primarily because it does not appropriately reflect the clear benefits of international diversification. However, there is a great deal of evidence that by failing to take account of the benefits of international diversification at the portfolio level, capital requirements for loans to developing countries will be significantly higher than is justified on the basis of the actual risks attached to this lending (see, for example, Griffith-Jones et al., 2003).

Therefore, one clear way in which Basle II could be improved so as to reduce the negative and technically incorrect effects on developing countries would be to introduce the benefits of diversification into the internal ratings-based approach. One of the major benefits of investing in developing and emerging economies is their relatively low correlation with mature markets. This hypothesis was tested empirically using a wide variety of financial, market and macro variables (Griffith-Jones et al., 2004a). Every statistical test performed showed that the correlation between developed markets only was higher, in every case, than that between developed and developing markets.

An additional positive effect of taking account of the benefits of diversification is that this makes capital requirements far less procyclical than they otherwise would be. Indeed, if the benefits of diversification are incorporated, simulations show that the variance over time of capital requirements will be significantly smaller than if these benefits are not incorporated. Therefore, introducing the benefits of geographical diversification significantly decreases, though it does not eliminate, the higher procyclicality that the internal ratings-based approach implies. This difference may well be significant enough to prevent a ‘credit crunch’.

However, even if the benefits of diversification are incorporated, the internal ratings-based approach will still be more procyclical than the standardized approach, which is closer to the principles of the first Basel Capital Accord (Basel I). Therefore, as well as introducing the benefits of diversification, it seems desirable to introduce countercyclical measures (for example, countercyclical provisioning against losses) at the same time as Basel II is implemented.

We can conclude that several measures can be taken to reduce boom–bust patterns of private flows, and thus enhance their contribution to development. However, given that the risk of costly crises will remain – even if such measures are introduced – maintaining and improving the supply of countercyclical official liquidity and development finance is essential.

Acknowledgement
The author would like to acknowledge the excellent assistance by Carmen Seekatz.
References


Ocampo, José Antonio (2003), Capital-Account and Counter-Cyclical Prudential Regulations in Developing Countries, United Nations Publications.


Prasad, E., K. Rogoff, S. Wei and M. Kose (2003), Effects of Financial Globalization on Developing Countries, International Monetary Fund, Washington, DC.


UN’s World Economic and Social Survey, 2005.

41 International capital flows to emerging economies: short- and long-run effects

Graciela L. Kaminsky

Introduction
The explosion of capital flows to emerging markets in the early and mid-1990s and their reversal following the crises in Asia, Latin America, and the transition economies have reignited a heated debate on the benefits and drawbacks of financial globalization. Many have argued that globalization has gone too far and that international capital markets have become extremely erratic, with ‘excessive’ booms and busts in capital flows triggering bubbles and financial crises and magnifying the business cycle. In contrast, the traditional view asserts that international capital markets enhance growth and productivity by allowing capital to flow to its most attractive destination.

Even if international capital flows do not trigger excess volatility in domestic financial markets, it is still true that large capital inflows can spark off inflation in the presence of a fixed exchange rate regime. Moreover, transitory capital inflows may distort relative prices, with the domestic economy losing competitiveness as a result of the appreciation of the real exchange rate. Therefore, it is no wonder that policy-makers have used a variety of tools to manage these flows, especially flows of the ‘hot money’ type.

This chapter re-examines the evidence on the characteristics of international capital flows to emerging economies, with particular attention to portfolio equity flows and bank lending around the time of the crises of the 1990s. The results suggest that episodes of surge in capital inflows do, in fact, end abruptly – whether owing to home-grown problems or contagion from abroad. This chapter also reviews the evidence on the short- and long-run effects of financial deregulation on financial and real cycles. Interestingly, the stylized evidence suggests that although financial liberalization may trigger excessive booms and busts in the short run, financial markets tend to stabilize and growth accelerates in the long run, in part because financial globalization seems to trigger institutional reform. The conclusion summarizes what we know about financial globalization and examines policy options.
The behavior of mutual funds

The booms and busts in international capital flows have brought international investors into the limelight. In this section, I examine the behavior of equity mutual funds in emerging markets. I use the information on portfolio allocations provided by Emerging Market Funds Research, Inc., which covers the positions of nearly 1400 international emerging market equity funds, with an average position of about US$120 billion in 1996. It includes United States registered and offshore funds as well as funds registered in Luxembourg, the United Kingdom of Great Britain and Northern Ireland, Ireland, the Cayman Islands, Canada and Switzerland. Both open- and closed-end funds are also included in this data set, which starts at 1995.

Figure 41.1 shows the average quarterly net flows to these regions from 1995 to 1999. Mutual fund flows to emerging markets peaked in the second quarter of 1997, reaching about US$8 billion. Overall, booms in mutual fund flows were followed by reversals. Reversals were not persistent after the ‘Tequila crisis’. Outflows from Latin America reached about US$4 billion in 1995, but mutual funds increased their positions in Latin America by about US$2 billion in the first half of 1996. The Tequila crisis did not have any spillovers in Asia or in transition economies. In fact, flows to Asia ballooned to almost US$11 billion in 1996, while flows to transition economies remained stable throughout 1995–96. The picture changed after the Asian crisis. This time, mutual funds pulled out not only from Asia but also from Latin America, with net outflows in the latter region reaching about US$1 billion in the six months following the collapse of the Thai baht. Mutual fund withdrawals took a turn for the worse in 1998, reaching about US$4 billion in Asia and also in Latin America, with substantial outflows from transition economies after the Russian crisis.

Figure 41.2 assesses the problem of the sudden stops in times of financial turmoil. It reports the average quarterly flows (as a percentage of the mutual funds’ initial positions) to countries in Asia and Latin America, as well as to transition economies in the two quarters following three crises. The top panel looks at the aftermath of the Mexican devaluation in December 1994, the middle panel examines the aftermath of the collapse of the Thai baht in July 1997, and the bottom panel studies the aftermath of the Russian devaluation and moratorium in August 1998. To capture the magnitude of the sudden-stop syndrome, this figure reports total flows relative to average flows (also as percentages of their initial positions) during the whole sample (1995–99). Following the Mexican devaluation, for example, mutual funds sold about 5 percent of their Brazilian positions (relative to their average quarterly buying/selling from 1995 to 1999). Thus, as shown in the first panel in Figure 41.2, Brazil experienced unusual withdrawals of about 5 percent in the aftermath of the Mexican devaluation. As shown in the last
Notes:
Latin America includes Argentina, Brazil, Chile, Colombia, Mexico, Peru and Venezuela.
Asia includes China, Hong Kong, India, Indonesia, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Sri Lanka, Taiwan and Thailand.
Transition economies include Armenia, Azerbaijan, Belarus, the Czech Republic, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Moldova, Poland, Russia, Slovakia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

Source: Kaminsky et al. (2002).

Figure 41.1 Mutual funds: quarterly flows to emerging countries (billions of dollars)
Source: Kaminsky et al. (2002).

Figure 41.2 Mutual fund flows: global spillovers
panel, Malaysia was the country most affected in the aftermath of the Russian crisis, with abnormal outflows of approximately 30 percent.

The extent of the mutual fund sudden stop in the aftermath of the three crises was substantially different. The so-called Tequila crisis was circumscribed to Latin America. Moreover, ‘abnormal’ mutual fund withdrawals in the aftermath of the collapse of the Mexican peso were confined to a handful of Latin American countries, with only Brazil and the Bolivarian Republic of Venezuela – besides the crisis country, Mexico – suffering average withdrawals of 5 and 2 percent, respectively, in the two quarters following the devaluation. In contrast, mutual funds increased their exposure to Asian countries and transition economies, with (above-trend) flows oscillating around 4 percent for Asia and 11 percent for the transition economies.

The aftermath of the collapse of the Thai baht presents a different picture of the international mutual funds industry. It is in this episode that we first observe signs of a more general retrenchment of mutual funds in emerging markets. Mutual fund flows to Asian economies were well below trend in the two quarters following the collapse of the Thai baht. Only flows to China, Pakistan and Sri Lanka were above average. Interestingly, after the collapse of the Thai baht, we observe substantial withdrawals from Hong Kong Special Administrative Region, Singapore and Taiwan Province of the People's Republic of China, with average quarterly withdrawals oscillating at about 12 percent above average in the case of Singapore and Taiwan and about 7 percent for Hong Kong. The retrenchment this time also affected Latin America and the transition economies, with withdrawals reaching about 6 percent for Colombia and 4 percent for the Czech Republic during the two quarters following the outbreak of the Thai crisis. Colombia, the Czech Republic, Chile, Hungary and Peru were the countries most affected in this episode, with sales averaging about 3 percent above average.

The flight away from emerging markets became more pronounced during the Russian crisis, with about half of the countries in the sample experiencing abnormal sales of about 10 percent or even larger. In some cases, withdrawals were massive. For example, average mutual funds sales (relative to trend) in Malaysia reached 30 percent while those in the Czech Republic were in the order of 16 percent. Some Latin American countries were also dramatically affected in the aftermath of the Russian collapse. Colombia and Venezuela, for example, suffered average quarterly outflows of about 8 percent. Mutual funds investments in Mexico and Peru were the only ones that did not suffer following the worldwide turmoil triggered by the Russian default. In fact, inflows to Mexico were 5 percent above the average observed in the 1995–99 period.

Table 41.1 examines in detail why some countries were severely affected by mutual fund withdrawals while others were left unscathed. Three factors
### Table 41.1 The behavior of mutual funds during crises

#### The Mexican crisis

<table>
<thead>
<tr>
<th>Region</th>
<th>Fragility</th>
<th>Liquid financial markets</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia With MF withdrawals</td>
<td>..</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>42</td>
<td>25</td>
</tr>
<tr>
<td>Latin America</td>
<td>67</td>
<td>33</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>67</td>
<td>33</td>
</tr>
<tr>
<td>Transition economies</td>
<td>..</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>75</td>
<td>50</td>
</tr>
</tbody>
</table>

#### The Thai crisis

<table>
<thead>
<tr>
<th>Region</th>
<th>Fragility</th>
<th>Liquid financial markets</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia With MF withdrawals</td>
<td>43</td>
<td>86</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Latin America</td>
<td>75</td>
<td>50</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Transition economies</td>
<td>100</td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

#### The Russian crisis

<table>
<thead>
<tr>
<th>Region</th>
<th>Fragility</th>
<th>Liquid financial markets</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia With MF withdrawals</td>
<td>40</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Latin America</td>
<td>50</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>60</td>
<td>0</td>
</tr>
</tbody>
</table>
are examined: economic fragility, liquidity of financial markets and economic and political risk. Economic fragility is captured using the probabilities of crises in Kaminsky (1998) that measure the likelihood of crises conditional on 18 indicators reflecting macroeconomic vulnerabilities in each country. These indicators provide information about fiscal and monetary imbalances, financial and real vulnerabilities, current account and capital account problems, and world factors. For Table 41.1, I classify an economy as fragile if the probability of a crisis is higher than 50 percent; otherwise it is considered healthy.

Liquidity is captured using four indicators. The first one – the volume traded in the stock market – provides an overall measure of the size and depth of the stock market. The second one – the share of the mutual funds portfolio in each country at the onset of the crisis – is related to mutual funds liquidity in each country, since investors cannot sell in countries in which they have basically no exposure. These first two indicators provide two different pictures of liquidity of financial markets. The third indicator dates the time when firms in emerging markets start to trade in mature and more liquid financial markets. The fourth indicator captures the ability of investors to change their portfolio rapidly in a particular country. In particular, this last indicator evaluates the extent of restrictions to capital mobility in each country. Restrictions could add ‘sand in the wheels’ of capital markets and thus curtail liquidity.

Finally, the risk indicator captures both political and economic uncertainty. The political risk indicator captures uncertainty due to expected changes of authorities or future policy actions, and it also identifies

<table>
<thead>
<tr>
<th>Region</th>
<th>Fragility</th>
<th>Liquid financial markets</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transition economies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With MF withdrawals</td>
<td>50</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Without MF withdrawals</td>
<td>100</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: This table relates the mutual fund (MF) withdrawals (injections) of funds to the emerging markets shown in Figure 41.4 with indicators of fragility, liquidity of financial markets, and economic and political risk in those economies. Data not available.

Source: Kaminsky et al. (2002).
widespread social unrest. In particular, it includes major changes in the political arena or events of political instability that took place six months before and after the crisis. The risk indicator also captures economic risk, such as imposition of restrictions to capital mobility in response to crises. A country is classified as risky when there is at least either political or economic risk.

Table 41.1 shows the characteristics of countries that suffer abnormal withdrawals and injections in the aftermath of the three crises. The table groups the countries into three regions: Asia, Latin America and transition economies. As shown in the first column, countries with fragile economies constitute the bulk of the countries that suffer withdrawals. During the Mexican crisis, for example, Latin America was the only region that suffered withdrawals. Interestingly, 67 percent of the countries that suffered withdrawals in this episode were also countries with deteriorated fundamentals. Again, during the Thai crisis, at least 75 percent of the countries that suffered withdrawals in the transition economies group and Latin America were countries with economic vulnerabilities. Similarly, 43 percent of the Asian countries affected by abnormal withdrawals also had deteriorated economies. The Republic of Korea (South Korea), Colombia, the Czech Republic and Chile, for example, suffered huge withdrawals in the aftermath of the Thai crisis – the Czech Republic and South Korea were the two most vulnerable countries during the Asian crisis (Thailand ranked fourth) in the sample of 25 countries, while Colombia ranked sixth. In contrast, countries that did not experience mutual fund withdrawals were less fragile in general (see Goldstein et al., 2000).

Domestic fragilities were not the only explanation for the sudden-stop syndrome, however. China, for example, did not even suffer a mild hiccup in the midst of the Asian crisis, even when devaluation fears were widespread among investors and the vulnerability of its financial system was widely known. In contrast, Singapore, Taiwan and Hong Kong – countries with the most liquid financial markets in the region – suffered pronounced capital flow reversals even when their economies looked far healthier than that of China. Overall, 86 percent of the countries in the Asia-Pacific region that suffered withdrawals were countries with quite liquid financial markets. In contrast, all the countries in that region unaffected by the Thai crisis had illiquid financial markets.

Finally, risk also had an important role, with 40 percent of the countries most affected by withdrawals also experiencing political and economic risk. In 1994, for example, in the midst of the banking crisis, Venezuela abandoned convertibility. Far from discouraging capital outflows, the implementation of restrictions to capital mobility seems to have also contributed to the fire sales of Venezuelan assets. Similarly, Malaysia suffered
substantial losses in the aftermath of the Russian crisis when it introduced outright controls on capital outflows. Interestingly, the withdrawals may have been triggered by the increased risk – perceived or real – associated with the country.

**The behavior of banks**

Bank-related lending has also been quite volatile since the late 1970s. This section examines the role of European, Japanese and United States banks in spreading the crises of the 1990s. The Bank for International Settlements (BIS) Consolidated Banking Statistics database is used to examine the role of the three international banking clusters. In particular, international claims of reporting BIS banks in emerging economies, including both total cross-border claims and local claims in foreign currency booked by foreign offices, are studied.

As shown in Figure 41.3, bank flows poured into Asia throughout most of the 1990s and accelerated following the Mexican crisis. Bank loans to emerging Asia expanded by 89 percent from June 1994 to June 1997. Part of the rise in lending was due to the European banks’ goal of achieving a higher profile in emerging markets, particularly in South Korea. Much of the lending boom, especially in the case of Thailand, Indonesia and South Korea, was due to a rapid expansion in credit from Japanese banks. Faced with a slumping economy and little domestic loan demand, Japanese banks increasingly looked overseas to the rapidly growing economies of South-East Asia as potential borrowers. United States bank lending to Asia was modest before the crisis. By June 1997, the United States banks’ positions in emerging Asia had only reached US$32 billion and only accounted for 20 percent of all United States bank lending to developing countries. In contrast, by the onset of the Thai crisis, Japanese banks had exposure to Asia four times as much as United States banks (US$124 billion). European bank lending to emerging Asia was also significant and, by the onset of the Thai crisis, the exposure of European banks to Asia surpassed that of Japanese banks, reaching US$161 billion. The exposure of European banks to emerging Asia accounted for about a half of all their lending to emerging markets; South Korea alone accounted for 40 percent of their lending to the developing world.

Japanese banks, heavily exposed to Thailand, were the first to pull out of emerging Asia. Between June and December of 1997, lending by Japanese banks fell by 8 percent. European banks, heavily exposed to South Korea, only began to pull out following the start of the crisis in that country in November 1997. In net terms, European bank lending to Asia continued to increase from June to December 1997. By June 1998, however, lending to emerging Asia was reduced across the board. Bank lending to Asia fell by
Notes:
Asia includes Afghanistan, Armenia, Azerbaijan, Bangladesh, Bhutan, British Overseas Territories, Brunei, Cambodia, China, Fiji, French Polynesia, Georgia, India, Indonesia, Kazakhstan, Kiribati, Kyrgyzstan, Laos, Macau, Malaysia, the Maldives, Mongolia, Myanmar, Nauru, Nepal, New Caledonia, North Korea, Pakistan, Papua New Guinea, the Philippines, the Solomon Islands, South Korea, Sri Lanka, Taiwan, Tajikistan, Thailand, Tonga, Turkmenistan, Tuvalu, US Pacific Islands, Uzbekistan, Vietnam, Wallis Futuna and Western Samoa.

Latin America includes Argentina, Belize, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominica, the Dominican Republic, Ecuador, El Salvador, the Falkland Islands, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Paraguay, Peru, St Lucia, St Vincent, Suriname, Trinidad and Tobago, Turks and Caicos, Uruguay and Venezuela.

Transition economies include Albania, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, the Czech Republic, Czechoslovakia, Estonia, the German Democratic Republic, Hungary, Latvia, Lithuania, Macedonia, Malta, Moldova, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, the Soviet Union, Turkey and Ukraine.

Source: Bank for International Settlements.

Figure 41.3  Bank lending: European banks, Japanese banks, US banks
US$46 billion, with European banks recalling US$12 billion, Japanese banks US$25 billion and United States banks US$9 billion, respectively.

Figure 41.3 also reports bank lending to Latin America and transition economies. Exposure to these regions increased sharply in the mid-1990s (in large part driven by the purchase of domestic banks by European banks), with claims on these regions increasing by about 50 percent from June 1994 to June 1998, immediately before the onset of the Russian crisis. During the 1990s, European banks had the largest exposure to these regions – accounting for 67 percent to Latin America and 84 percent to transition economies. The Russian crisis led to some withdrawals of Japanese and United States lending from both regions, but this was not the case with European banks that had acquired local banks. Total exposure to Latin America by European banks peaked in December 2000.

Figures 41.4 to 41.6 tally country-by-country bank flows originating in European, Japanese and United States banks in the aftermath of the Mexican, Thai and Russian crises. Each figure focuses on the year following the crisis. Figure 41.4 shows that, with the exception of Mexico and Venezuela (which had a banking crisis of its own making), Latin American countries did not suffer major reversals in bank lending following the Mexican crisis. Moreover, within a year of the crisis, lending to Latin America recovered and even surpassed the levels observed before the crisis. Brazil was the prime beneficiary of bank flows during 1995, with lending from European and United States banks reaching US$15 billion. Even in the case of Mexico and Venezuela, withdrawals were not made across the board. Only United States banks recalled loans from these countries. Figure 41.4 also shows that in Asia, the major recipients of capital flows in 1995 were South Korea, Thailand and Indonesia.

Figure 41.5 shows the behavior of bank lending in the aftermath of the Thai crisis. In contrast to the Tequila crisis, the Thai crisis triggered major reversals in bank flows from banks in Europe, Japan and the United States. Thailand, South Korea, Indonesia and Malaysia were the countries that suffered major withdrawals. Contagion was only regional in nature, with almost all of the Latin American countries, and to a lesser degree transition economies, continuing to have uninterrupted access to bank lending.

Figure 41.6 shows contagion from the Russian crisis. As was the case with mutual funds, the reversal in bank lending following the Russian default was not restricted to the Russian Federation or neighbouring countries. This time, the reversal was more widespread, and affected countries as far away as Brazil and South Africa. While Japanese banks continued to recall loans from Thailand, Indonesia and South Korea, reversals were not just restricted to these countries. Japanese banks, as well as United States banks, also recalled loans from Brazil, Mexico, India and South Africa.
Figure 41.4  Bank flows: global spillovers – after the Mexican crisis: December 1994–December 1995

Source: Bank for International Settlements.
Source: Bank for International Settlements.

Figure 41.5 Bank flows: global spillovers – after the Thai crisis: June 1997–June 1998
Source: Bank for International Settlements.

Figure 41.6 Bank flows: global spillovers – after the Russian crisis: June 1998–June 1999
More formal evidence suggests that international banks were at the centre of financial contagion in the late 1990s. For example, Kaminsky and Reinhart (2000) examine contagion during the debt crisis in 1982, the Mexican crisis in 1994 and the Asian crisis in 1997, and find that United States banks were at the core of the contagion during the debt crisis, while Japanese banks spread the Thai crisis to Indonesia, South Korea and Malaysia. Van Rijckeghem and Weder (2003) examine the Tequila, Asian and Russian crises and the flows to 31 emerging countries from 11 creditor countries using BIS banks. Their evidence supports the idea that the degree to which countries compete for funds from common bank lenders is a fairly robust predictor of the incidence of contagion. Finally, Caramazza et al. (2000) extend earlier work on indicators of vulnerability to currency crises by examining the role of financial linkages, while controlling for the roles of internal and external macroeconomic imbalances and trade spillovers. Their results indicate that financial links do matter while exchange rate regimes and controls on capital flows do not seem to.

Globalization and volatility
As discussed in the introduction, the views on the effects of financial globalization have been diverse; there are those who defend capital controls (Rodrik, 1998; Stiglitz, 1999) and those who maintain that capital should be allowed to move freely (Dornbusch, 1998). The rationale for restricting international capital flows is grounded in the belief that market failures and distortions pervade capital markets around the world. One of the most frequently cited distortions is that of asymmetric information, which is rampant in international capital markets due to geographical and cultural differences that complicate the task of obtaining information. In addition, imperfections in international markets are magnified by the difficulties in enforcing contracts across borders. With imperfect information, investors may overreact to shocks, withdrawing massively from countries at the first signs of economic problems, or become euphoric and pour in capital in quantities beyond those justified by ‘good’ fundamentals. On the other hand, those who consider international capital markets to be efficient favor unrestricted capital movements. Financial liberalization is believed to improve the functioning of financial systems, increasing the availability of funds and allowing cross-country risk diversification. Moreover, it is also claimed that financial integration tends to facilitate economic growth.

This section will summarize some of the findings in the literature on the effects of globalization, paying particular attention to the short- and long-run effects of financial integration on real and financial volatility.
Financial markets
The evidence from the crises of the 1990s suggests that crises are preceded by ‘excessive’ capital inflows that, in turn, fuel large expansions in domestic credit and bubbles in financial markets (see, for example, Sachs et al., 1996). There is also evidence that most episodes of banking crises are preceded by financial liberalization (see, for example, Kaminsky and Reinhart, 1999; Demirguc-Kunt and Detragiache, 1999). To reconcile the evidence that globalization is at the heart of financial crises with the hypothesis that international capital markets allow capital to move to its most attractive destination and promote more stable financial markets, I examine the possible time-varying effects of financial liberalization on stock market price cycles.5

Figure 41.7 shows the average amplitude of booms and crashes in stock prices for 14 emerging markets during periods of repression, in the immediate aftermath of liberalization (the four years following liberalization), and in the long run. The evidence in this figure seems to point to excessive cycles, with larger booms followed by larger crashes in the immediate aftermath of financial liberalization. However, liberalization does not permanently bring about more volatile financial markets. If liberalization persists, stock markets in emerging countries become more stable. Kaminsky and Schmukler (2003) argue that these conflicting effects arise because during episodes of financial repression, banks are protected from outside competition and do not have the pressure to run efficiently. Liberalization in this

![Figure 41.7 Average amplitude of booms and crashes in stock prices in 14 emerging markets (in percentage points)](image)

Note: Gray: repression; Black: short-run liberalization; White: long-run liberalization.

Source: Kaminsky and Schmukler (2003).
scenario unveils a new problem, as protected domestic banks suddenly get access to new sources of funding, triggering protracted financial booms. But financial liberalization triggers reforms and better-functioning financial markets as domestic investors, now with access to international capital markets, demand better enforcement rules to continue to invest in domestic financial markets. Moreover, as pointed out by Stulz (1999), the liberalization and gradual integration of emerging markets into international financial markets may help strengthen the domestic financial sector, as foreign investors generally have better skills and more information and can thus monitor management in ways that local investors cannot. Liberalization also allows firms to access mature capital markets. Firms listed on foreign stock markets are in the jurisdiction of a superior legal system with higher disclosure standards that will promote more transparency in the management of the firm and can trigger improvements in corporate governance.

Business cycles and growth
The evidence in the previous section is suggestive of excessive booms and busts in financial markets in developing countries following globalization, but of more stable financial markets in the long run if globalization persists. This section will examine the relationship between globalization and business cycle fluctuations and growth.

Figure 41.8 shows international capital flows to emerging markets in Asia, Latin America and transition economies, as well as annual output growth rates. The panels suggest that capital flows have been procyclical, with large inflows in good times and outflows during recessions. For example, Latin America’s growth rates oscillated at around 4.5 percent in periods of capital inflows, while growth rates were about 1 percent in periods of sudden stops. Similarly, Asia’s economic activity collapsed to about 5.5 percent after the sudden stop in capital flows in the late 1990s, after growing at an average annual growth rate of 8.5 percent during the earlier period of large capital inflows. This evidence contrasts sharply with the prescription that international capital markets should allow countries to smooth out the effect of the business cycle. Countries seem to have lost access to international credit markets during recessions on a systematic basis.

This non-optimal behavior of international capital flows has also been studied by Calvo et al. (2004), who observe that sudden reversals in capital flows to emerging economies lead to large real depreciations and profound downturns. As reported in Kaminsky et al. (2004), macro-policies tend to be procyclical in developing countries while they are countercyclical or acyclical in industrialized countries. That is to say, macro-policies tend to
smooth out the business cycle in industrial countries but magnify it in developing countries, as shown in Table 41.2. The left panel in this table reports the correlation between the cyclical components of fiscal and monetary policy with the business cycle. The right panel shows the correlations of the cyclical components of fiscal and monetary policy with net capital inflows. Interestingly, the evidence suggests that international capital flows to developing countries may trigger procyclical macro-policies. Government expenditure (inflation tax) is positively (negatively) correlated with net capital inflows, indicating that periods of capital inflows are associated with expansionary fiscal policies, and periods of capital outflows with contractionary fiscal policies. While more research is needed, the stylized evidence suggests that international capital flows may trigger more volatile business cycles in emerging economies.

While this evidence points to links between financial integration and output instability over the business cycle, there is also evidence that financial integration promotes growth. A variety of authors have examined the effects of domestic and external deregulation of financial markets in emerging economies and found that they generally trigger sustainable growth in the long run. Bekaert et al. (2005), for example, examine the effects on growth of the opening of the stock market to foreign investors in a sample of about 90 developing countries and find that, overall, liberalization triggers an increase in growth by approximately one percentage point. Similarly, Galindo et al. (2002) study whether financial liberalization promotes economic growth by analyzing its effect on the cost of external financing to firms. They find that the liberalization of domestic and external financial markets reduces the cost of external funds faced by firms. In particular, they find that industries that depend on external finance grow almost 1 percent faster, relative to industries with low external financing dependence, in episodes of globalization compared to episodes of repression. The evidence on the links between financial liberalization and growth is not conclusive, however. Edison et al. (2002), for example, using data from 57 countries from 1980 to 2000, conclude that there is no robustly significant effect of financial integration on economic growth. Similarly, Kraay (1998), using a sample of 117 countries, finds no effect of financial liberalization on growth or, at best, mixed results.

Perhaps the inability of past research to agree on the effects of financial globalization on economic growth lies in the fact that liberalization has time-varying effects on growth. Loayza and Ranciere (2002) present some evidence that suggests this might be the case. These authors estimate transitory and trend effects of financial deepening on growth using a sample of about 80 countries and find that financial deepening, which in general is closely related to financial liberalization, harms growth in the short run but
The countries comprising Asia are Bangladesh, China, Hong Kong, India, Indonesia, Malaysia, Pakistan, the Philippines, Singapore, South Korea, Taiwan, Thailand and Vietnam.

The countries comprising the transition economies are Albania, Armenia, Azerbaijan, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Estonia, Georgia, Hungary, Kazakhstan, Kyrgyzstan, Latvia, Lithuania, Macedonia, Moldova, Mongolia, Poland, Romania, Russia, Serbia and Montenegro, Slovakia, Slovenia, Tajikistan, Turkmenistan, Ukraine and Uzbekistan.

The countries comprising Latin America are Argentina, Brazil, Chile, Colombia, the Dominican Republic, Ecuador, Guatemala, Mexico, Peru, Uruguay and Venezuela.

**Source:** World Economic Outlook.

**Figure 41.8** Private capital flows to emerging markets and GDP annual growth rates
leads to higher growth in the long run. These latest results are closely linked to the evidence from stock market cycles discussed before and suggest that financial liberalization triggers growth in the long run because it fuels institutional reform.

**Conclusions**

The explosion of capital flows to emerging markets in the early and mid-1990s and the recent reversal following the crises around the globe have reignited a heated debate on how to manage international capital flows. Capital outflows worry policy-makers, but so do capital inflows, as they may trigger bubbles in asset markets and lead to an appreciation of the domestic currency and a loss of competitiveness. Policy-makers also worry that capital inflows are mostly of the ‘hot money’ type, which is why capital controls have mostly targeted short-term capital inflows. While capital controls may work, at least in the very short run, the introduction of restrictions to capital mobility may have undesirable long-run effects. In particular, capital controls protect inefficient domestic financial institutions and thus may trigger financial vulnerabilities. Capital controls may also delay improve-

### Table 41.2 Correlations between the cyclical components of macro-policies, real GDP and net capital inflows

<table>
<thead>
<tr>
<th>Countries</th>
<th>Correlations with real GDP</th>
<th>Correlations with net capital inflows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fiscal policy</td>
<td>Monetary policy</td>
</tr>
<tr>
<td></td>
<td>Government expenditure</td>
<td>Inflation tax</td>
</tr>
<tr>
<td>OECD</td>
<td>–0.13</td>
<td>0.16</td>
</tr>
<tr>
<td>Non-OECD</td>
<td>0.33</td>
<td>–0.15</td>
</tr>
</tbody>
</table>

**Notes:**

A positive (negative) correlation between government expenditure (inflation tax) and real GDP indicates pro-cyclical fiscal policy. A negative correlation between lending interest rates and real GDP indicates pro-cyclical monetary policy. A positive (negative) correlation between government expenditure (inflation tax and lending interest rates) and net capital inflows indicates that contractionary macropolicies are linked to episodes of low net capital inflows.

The cyclical component of the various indicators was obtained using the HP filter.

**Source:** Kaminsky et al. (2004).
ments in corporate governance of non-financial firms because, as countries liberalize their capital accounts, domestic corporations start participating in international capital markets, mainly through cross-listing in major world stock exchanges, with higher disclosure standards and under the jurisdiction of a superior legal system. This certainly promotes more transparency in the management of the firm and can trigger improvements in corporate governance (see, for example, Stulz, 1999). Thus, regulation of capital flows may not only provoke financial vulnerabilities but also lower economic growth. Policy-makers have also resorted to sterilization of capital flows to regain control of monetary policy. While sterilization may provide some relief, it may also be quite costly to central banks. Moreover, the ability of governments to control international capital flows or to sterilize them diminishes with globalization.

In conclusion, there is no optimal policy to deal with the risks of volatile international capital flows, as policies that may work in the short run may have adverse effects in the long run. Since there is evidence that currency and banking crises tend to occur in economies with deteriorated fundamentals, conservative macroeconomic policies should be at the heart of dealing with volatile capital flows. Further research should examine whether countries can deregulate financial systems without becoming vulnerable to crises. Since the costs of crises have been quite large, this last question deserves much attention.

Notes
1. This chapter draws on previous research with Richard Lyons, Carmen Reinhart, Sergio Schmukler and Carlos Végh and is a shorter version of a paper entitled ‘International Capital Flows, Financial Stability and Growth’ (Kaminsky, 2006).
2. Liquidity may have an important effect on investors’ portfolio allocations since investors may want to avoid illiquid markets to minimize the price collapses always present when there is no ready market.
3. To identify liquid markets, countries are ranked by region according to their volume traded and according to their share in the mutual funds portfolio at the onset of the crisis. The dummy variable related to volume traded is given a value of one if the country ranks among the top 30 percent of most-liquid countries in the region in that category, and a value of zero otherwise. Similarly, countries are classified as liquid (that is to say, the dummy variable is given a value of one) if they rank among the 30 percent of the countries with the largest share in mutual fund portfolios for the region. A third dummy is created to capture whether emerging-market firms are trading in mature financial markets: the variable is given a value of one if they do, and zero if they do not. Finally, the variable capturing restrictions to entry and exit of foreigners in the stock markets of emerging economies is given a value of one if there are no restrictions, and zero if there are. All of this information is collapsed into a liquidity variable that is the average of the four univariate liquidity dummy variables. Thus, the general index of liquidity, the average of the four components, can have five values: 0, 1/4, 2/4, 3/4 and 1, with a value of one indicating a highly liquid market. I classify a country as having liquid financial markets when this dummy takes a value of 2/4 or higher.
4. See Kaminsky et al. (2002) for a country-by-country detail on fragility, liquidity, risk and mutual fund withdrawals.
5. The results discussed below are from Kaminsky and Schmukler (2003).

6. The 14 emerging economies are Argentina, Brazil, Chile, Colombia, Hong Kong, Indonesia, South Korea, Malaysia, Mexico, Peru, the Philippines, Taiwan, Thailand and Venezuela.

7. Claessens et al. (1998) present evidence that liberalization of the capital account and foreign bank entry lead to improvements in banking system efficiency.

References


Introduction
Controversies about aid effectiveness go back decades. Critics such as Milton Friedman (1958), Peter Bauer (1972) and William Easterly (2001) have leveled stinging critiques, charging that aid has enlarged government bureaucracies, perpetuated bad governments, enriched the elite in poor countries, or just been wasted. They cite widespread poverty in Africa and South Asia despite four decades of aid starting in the 1960s, and point to countries that have received substantial aid yet have had disastrous records such as the Democratic Republic of the Congo, Haiti, Papua New Guinea and Somalia. In their eyes, aid programs should be dramatically reformed, substantially curtailed or eliminated altogether.

Supporters counter that these arguments, while partially correct, are overstated. Nicholas Stern (2002), Joseph Stiglitz (2002), Jeffrey Sachs et al. (2004) and others have argued that although aid has sometimes failed, it has supported poverty reduction and growth in some countries and prevented worse performance in others. They believe that many of the weaknesses of aid have more to do with donors than recipients, and point to a range of successful countries that have received significant aid such as Botswana, Indonesia, Korea and, more recently, Tanzania and Mozambique, along with successful initiatives such as the Green Revolution, the campaign against river blindness, and the introduction of oral rehydration therapy.

This chapter explores trends in aid, the motivations for aid, its impacts, and debates about reforming aid. It begins by examining aid magnitudes and who gives and receives aid. It discusses the multiple motivations and objectives of aid, some of which conflict with each other. It then explores the empirical evidence on the relationship between aid and growth, most (but far from all) of which concludes there is a positive relationship (at least under certain circumstances). It examines some of the key challenges in making aid more effective, including the principal–agent problem and the related issue of conditionality, and concludes by examining some of the main proposals for improving aid effectiveness.
Donors and recipients

What is foreign aid?
The standard definition of foreign aid comes from the Development Assistance Committee (DAC) of the Organisation for Economic Co-operation and Development (OECD), which defines foreign aid (or the equivalent term, foreign assistance) as financial flows, technical assistance and commodities that are: (1) designed to promote economic development and welfare as their main objective (thus excluding aid for military or other non-development purposes); and (2) provided as either grants or subsidized loans.

Grants and subsidized loans are referred to as concessional financing, whereas loans that carry market or near-market terms (and therefore are not foreign aid) are non-concessional financing. According to the DAC, a loan counts as aid if it has a ‘grant element’ of 25 percent or more, meaning that the present value of the loan must be at least 25 percent below the present value of a comparable loan at market interest rates (usually assumed by the DAC – rather arbitrarily – to be 10 percent with no grace period). Thus, the grant element is zero for a loan carrying a 10 percent interest rate, 100 percent for an outright grant, and something in-between for other loans.

The DAC classifies aid flows into three broad categories. Official development assistance (ODA) is the largest, consisting of aid provided by donor governments to low- and middle-income countries. Official assistance (OA) is aid provided by governments to richer countries with per capita incomes higher than approximately $9000 (for example, the Bahamas, Cyprus, Israel and Singapore) and to countries that were formerly part of the Soviet Union or its satellites. Private voluntary assistance includes grants from non-governmental organizations, religious groups, charities, foundations and private companies.

When discussing foreign aid, most people have in mind ODA. Global ODA increased steadily from the 1960s until it reached a peak of $68 billion in 1992, just after the end of the Cold War (Figure 42.1), and then declined sharply to just under $55 billion in 1997. Aid flows began to rebound in the late 1990s following calls for greater debt relief and increased aid to new democracies, and accelerated very sharply after the attacks of 11 September 2001, reaching $120 billion in 2006 (all of these figures would be slightly higher if they included OA). In real terms, total ODA in 2002 was about the same as in 1992, and by 2006 was about 15 percent higher. Measured as a share of donor income ODA fell sharply during the 1990s, and has rebounded only slightly. Donors have pledged to continue to increase aid, most recently in July 2005 when the heads of state of the
Figure 42.1 Global ODA 1975–2006

Group of 8 industrialized countries promised to double aid to sub-Saharan Africa by 2010 and triple it by 2015, but growing budget tensions in donor countries may undermine these pledges.

Who gives aid, and who receives it?
Historically most aid has been given as bilateral assistance directly from one country to another. Donors also provide aid indirectly as multilateral assistance, which pools resources together from many donors. The major multilateral institutions include the World Bank, the International Monetary Fund (IMF), the African, Asian and Inter-American Development Banks, and various United Nations agencies such as the United Nations Development Programme.

In terms of total dollars, the United States has consistently been the world’s largest donor (except in the mid-1990s when Japan briefly topped the list). In 2006 the USA provided $22.9 billion in ODA, with Japan, France the United Kingdom, and Germany the next-largest donors. However, when aid is measured as a share of donor income, the most generous donors are Norway, Denmark, Luxembourg, the Netherlands, and Sweden, each of which provided between 0.81–1.02 percent of GNI in 2006. Saudi Arabia provided aid equivalent to about 0.57 percent of its income. The United States is one of the smallest donors by this measure, at about 0.18 percent of US income in 2006, just over half of the 1970 level of 0.32 percent and less than one-third of the US average during the 1960s. Donors have pledged since the 1960s to devote 0.7 percent of their income as aid, most recently at the Financing for Development Conference in Monterrey in March 2002, but only a handful of small donors have achieved this level of aid.

One hundred and forty-five countries and territories around the world received aid in 2006. Table 42.1 shows the ten largest recipients, each of which received more than $1.8 billion. Nigeria was at the top of the list in 2006, with measured aid of $11.4 billion. But this figure is misleading because it includes $9.5 billion for a one-time debt relief deal. Debt relief is accounted for differently than other components of ODA – the value of debt relief is the charge to the creditor country’s budget for writing off the debt in the year of the debt relief, and does not represent new funding to the recipient (although it does capture a future reduction in debt service obligations). Nigeria’s actual inflow of new finance in 2006 was $1.9 billion. Iraq and Afghanistan together received nearly $12 billion dollars, nearly unprecedented amounts, accounting for about 10 percent of the global total. Amounts to the other countries shown in Table 42.1 are more typical (by historical standards) for large recipients. Total dollar amounts are important, but they do not tell the entire story. On a per capita basis, the aid flows to some of these countries are fairly small. Vietnam received $1.8
Table 42.1 Major aid recipients, 2006

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Total ODA (millions US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Nigeria</td>
<td>11434</td>
</tr>
<tr>
<td>2</td>
<td>Iraq</td>
<td>8661</td>
</tr>
<tr>
<td>3</td>
<td>Afghanistan</td>
<td>3000</td>
</tr>
<tr>
<td>4</td>
<td>Pakistan</td>
<td>2147</td>
</tr>
<tr>
<td>5</td>
<td>Sudan</td>
<td>2058</td>
</tr>
<tr>
<td>6</td>
<td>Congo, Dem. Rep.</td>
<td>2056</td>
</tr>
<tr>
<td>7</td>
<td>Ethiopia</td>
<td>1947</td>
</tr>
<tr>
<td>8</td>
<td>Vietnam</td>
<td>1846</td>
</tr>
<tr>
<td>9</td>
<td>Tanzania</td>
<td>1825</td>
</tr>
<tr>
<td>10</td>
<td>Cameroon</td>
<td>1684</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Aid as % of recipient GNI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Solomon Islands</td>
<td>61</td>
</tr>
<tr>
<td>2</td>
<td>Tuvalu</td>
<td>58</td>
</tr>
<tr>
<td>3</td>
<td>Liberia</td>
<td>54</td>
</tr>
<tr>
<td>4</td>
<td>Burundi</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td>Micronesia, Fed. States</td>
<td>41</td>
</tr>
<tr>
<td>6</td>
<td>Afghanistan</td>
<td>36</td>
</tr>
<tr>
<td>7</td>
<td>Palestinian Adm. Areas</td>
<td>35</td>
</tr>
<tr>
<td>8</td>
<td>Malawi</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>Marshall Islands</td>
<td>29</td>
</tr>
<tr>
<td>10</td>
<td>Guinea-Bissau</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Aid per capita (US$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Palau</td>
<td>1866</td>
</tr>
<tr>
<td>2</td>
<td>Mayotte</td>
<td>1777</td>
</tr>
<tr>
<td>3</td>
<td>Nauru</td>
<td>1740</td>
</tr>
<tr>
<td>4</td>
<td>Cook Islands</td>
<td>1614</td>
</tr>
<tr>
<td>5</td>
<td>Tuvalu</td>
<td>1534</td>
</tr>
<tr>
<td>6</td>
<td>Marshall Islands</td>
<td>786</td>
</tr>
<tr>
<td>7</td>
<td>Palestinian Adm. Areas</td>
<td>426</td>
</tr>
<tr>
<td>8</td>
<td>Solomon Islands</td>
<td>409</td>
</tr>
<tr>
<td>9</td>
<td>Iraq</td>
<td>304</td>
</tr>
<tr>
<td>10</td>
<td>Cape Verde</td>
<td>282</td>
</tr>
</tbody>
</table>

billion in aid in 2004, but this was equivalent to just 3 percent of its gross national income (GNI) or about $22 per person. By contrast, Cameroon received a similar amount, $1.7 billion in 2006, but for its 16.6 million people this was equivalent to about $100 dollars per person. For small countries, a little bit goes a long way. Tiny Solomon Islands received just $205 million, but this translated into 61 percent of GNI and about $409 per person. Aid is typically measured as a share of GNI, but this can be misleading as a high ratio can just as easily be indicative of low GNI as of a large amount of aid.

On a regional basis, sub-Saharan African countries received aid flows of 5.8 percent of GNI in 2006, or $50.2 per person (Table 42.2), although close to one-third of this amount was due to several large one-time debt relief deals that are not new inflows and are not indicative of long-term trends. North Africa and the Middle East received more than $44 per person (largely on account of Iraq), and Europe and Central Asia received about $18 per person. For low-income countries around the world, donors provided aid averaging about $20.2 per recipient in 2006, although once again these figures are inflated by several one-time debt relief deals.

Generally speaking, aid is one of the largest components of foreign capital flows to low-income countries, but not to most middle-income countries, where private capital flows are more important. Aid flows averaged 3.1 percent of GNI in low-income countries in 2004, but just 0.2 percent of GNI in upper-middle-income countries. It is commonly claimed that the decline in aid flows to developing countries in the 1990s was more than offset by a rise in private capital. While this is true for developing countries in aggregate, the rise in private capital flows was heavily

### Table 42.2 Official aid receipts by region, 2006

<table>
<thead>
<tr>
<th>Region</th>
<th>Billion US$</th>
<th>% of GNI</th>
<th>US$ per person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sub-Saharan Africa</td>
<td>38.2</td>
<td>5.8</td>
<td>50.2</td>
</tr>
<tr>
<td>South Asia</td>
<td>9.2</td>
<td>0.8</td>
<td>6.1</td>
</tr>
<tr>
<td>East Asia &amp; Pacific</td>
<td>7.4</td>
<td>0.2</td>
<td>3.9</td>
</tr>
<tr>
<td>Europe &amp; Central Asia</td>
<td>2.6</td>
<td>0.4</td>
<td>17.6</td>
</tr>
<tr>
<td>Middle East &amp; North Africa</td>
<td>14.6</td>
<td>2.0</td>
<td>44.0</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>6.0</td>
<td>0.2</td>
<td>11.0</td>
</tr>
<tr>
<td>Low-income</td>
<td>47.5</td>
<td>3.1</td>
<td>20.2</td>
</tr>
<tr>
<td>Lower-middle income</td>
<td>26.7</td>
<td>0.5</td>
<td>11.0</td>
</tr>
<tr>
<td>Upper-middle income</td>
<td>3.7</td>
<td>0.2</td>
<td>10.1</td>
</tr>
</tbody>
</table>

*Source:* Author’s calculations based on data from OECD 2007 Development Cooperation Report and World Development Indicators 2007.
concentrated in a handful of middle-income countries. In low-income countries, private capital rose much more slowly, and remained significantly smaller than aid.

*Why do donors give aid?*

Donors have a variety of motivations for providing aid, only some of which are directly related to economic development. There is little question that foreign policy and political relationships are the most important determinants of aid flows. During the Cold War, both the United States and the Soviet Union used aid to vie for the support of developing countries with little regard as to whether the aid was actually used to support development. The two largest recipients of US foreign aid (including both OA and ODA) from 1980 until very recently were Israel and Egypt, as the USA provided financial support to back the 1979 Camp David peace agreement. Beginning in 2002 Iraq became the largest aid recipient in the world, and its reconstruction is likely to become the largest single foreign aid program ever recorded. Taiwan and China have used aid (among other policy tools) to try to gain support and recognition for their governments from countries around the world. Many donors provide significant aid to their former colonies as a means of retaining some political influence (Alesina and Dollar, 2000).

Many people see the main rationale for aid as fighting poverty, and although this is less important than political considerations in donor allocation decisions, it still plays an important role. Donors generally provide their most concessional aid to the poorest countries, and some aid programs are designed explicitly with this objective in mind. For example, the World Bank’s concessional financing arm – the International Development Association (IDA) – has an income ceiling ($965 per capita in 2004). Once countries reach that ceiling, in most cases they ‘graduate’ from IDA to non-concessional International Bank for Reconstruction and Development (IBRD) loans. Other programs have less formal graduation rules, but still tend to provide less aid as incomes grow.

Country size matters as well. Large countries, such as Bangladesh, Indonesia, Nigeria and Pakistan receive relatively small amounts of aid on a per capita basis, even though hundreds of millions of people live in poverty in these countries. By contrast, some small countries receive very large amounts. For political reasons, donors generally want to influence as many countries as possible, which tends to lead to a disproportionate amount of aid going to small countries.

Bilateral aid is often designed at least partially to help support the economic interests of certain firms or sectors in the donor country. Multilateral aid is less prone to these pressures, although by no means immune. Many donors ‘tie’ portions of their aid by requiring that certain
goods and services be purchased from firms in the donor’s home country, or that it be used for specific purposes that support groups in the donor countries (such as universities or business consulting firms). Automobiles, airline tickets and consulting services financed by USA foreign aid in most cases must be purchased from USA firms. Tying aid can give it more political support at home, but it can also make it more costly and less effective. If funds must be spent in the donor country, it reduces competition for services so that donors do not always use the least-cost provider. For example, the USA requires that food aid be purchased in the USA and shipped in US carriers to recipient countries, which can be much more expensive and take much longer than if food was purchased in a neighboring country. This means that recipients receive much less value for each dollar of aid allocated than they otherwise could. One study found that tying aid added 15–20 percent to its cost, thus significantly reducing its impact on recipient countries. Donors have begun to reduce the amount of aid that they tie, but the practice is still widespread among some donors. The USA no longer reports the share of its aid that is tied, but historically it has been around 75 percent. Greece ties about 70 percent of its aid, and Canada and Austria more than 40 percent. By contrast, Ireland, Norway, and the UK do not tie any of their aid.

Aid, growth and development
Most foreign aid is designed to meet one or more of four broad economic and development objectives: (1) to stimulate economic growth through building infrastructure, supporting productive sectors such as agriculture, or bringing new ideas and technologies; (2) to strengthen education, health, environmental or political institutions or systems; (3) to support subsistence consumption of food and other commodities, especially during relief operations or humanitarian crises; or (4) to help stabilize an economy following economic shocks.

Despite these broader objectives for aid, growth has always been the main yardstick used to judge aid’s effectiveness. Debate has raged about the relationship between aid and growth for years, but there are some broad parameters of agreement. Even most aid pessimists agree that aid has been successful in some countries (such as in Botswana or Korea, or more recently in Mozambique and Tanzania), that aid has helped improve health by supplying essential medicines, and that aid is an important vehicle in providing emergency relief following natural disasters. Similarly, aid optimists concede that much aid has been wasted or stolen, such as by the Marcos regime in the Philippines and the Duvalier regime in Haiti, and that even under the best circumstances aid can create certain adverse economic incentives. Debate continues on the overall general trends, the conditions
under which aid works or does not work, and on what steps can be taken to make aid more effective. Although the majority of research since the mid-1990s has found a positive relationship between aid and growth, several studies have found no relationship. Three broad views have emerged on the relationship between aid and growth.4

Aid has a positive relationship with growth on average across countries (although not in every country), but with diminishing returns as the volume of aid increases

There are three key channels through which aid might spur growth:

- First, the classic view is that aid augments saving, finances investment and adds to the capital stock. In this view, poor countries are unable to generate sufficient amounts of saving on their own to finance the investment necessary to initiate growth, or at best only enough for very slow growth. In the strongest version of this view, the poorest countries may be stuck in a poverty trap in which their income is too low to generate the saving necessary to initiate the process of sustained growth (Sachs et al., 2004). A related argument is that aid might help relax a foreign exchange constraint in countries that earn relatively little foreign exchange, a view that was popularized through the early ‘two-gap’ models of economic growth.
- Second, aid might increase worker productivity through investments in health or education.
- Third, aid could provide a conduit for the transfer of technology or knowledge from rich countries to poor countries by paying for capital goods imports, through technical assistance, or through direct transfer of technologies such as the introduction of new seeds and fertilizers in the Green Revolution.

Several early studies found a positive relationship between aid and growth (for example, Papenek, 1973; Levy, 1988), but this strand of the literature took a significant turn in the mid-1990s when researchers began to investigate whether aid might support growth with diminishing returns. Oddly – given Solow’s response to the Harrod–Domar model in the 1950s – research until the mid-1990s only tested a linear relationship, a specification which persists in some studies today. A large group of studies that allow for diminishing returns have found a positive relationship although the direction of causality is a subject of ongoing debate.5 These studies do not conclude that aid has always worked in every country, but rather that on average and controlling for other factors, higher aid flows have been associated with more rapid growth. These studies have received much less public attention than
those that have found a zero or conditional relationship. The robustness of the conclusions of several of these studies has been the subject of on-going debate, as has the robustness of the conclusions of several studies that have reached the opposite conclusion, as discussed below. But since the mid-1990s the majority of published research on the topic has found a positive relationship either by allowing for diminishing returns, or by testing for conditional relationships as explored below.

Aid could also have a positive impact on development outcomes other than growth, such as health, education or the environment. Perhaps the best-documented area is health, where aid-supported programs have contributed to the eradication of smallpox, the near eradication of polio, control of river blindness and other diseases, the spread of oral rehydration tablets to combat diarrhea, and the dramatic increase in immunization rates in developing countries since 1970 (Levine et al., 2004). A recent cross-country study found a positive and significant relationship between health aid and infant mortality (Mishra and Newhouse, 2007). Undoubtedly, much aid aimed at health has also been squandered. But beyond the examples listed here, there is little systematic evidence on the relationship between aid and health, education, income distribution or other outcomes.

Aid has no affect on growth, and may actually undermine growth

Peter Bauer was perhaps the most outspoken proponent of this view (for example, Bauer, 1972), but he never provided systematic empirical evidence to support his argument. Many later empirical studies did reach this conclusion, although once again the robustness of these results is the subject of on-going debate. These researchers have suggested a variety of reasons why aid might not support growth:

- First, aid simply could be wasted, such as on limousines or presidential palaces, or it could encourage corruption, not just in aid programs but more broadly.
- Second, it can help keep bad governments in power, thus helping to perpetuate poor economic policies and postpone reform. Some argue that aid provided to countries in the midst of war might inadvertently help finance and perpetuate the conflict, and add to instability.
- Third, countries may have limited absorptive capacity to use aid flows effectively if they have relatively few skilled workers, weak infrastructure or constrained delivery systems. (Aid could help redress these weaknesses, but it may not be aimed to do so.)
- Fourth, aid flows can reduce domestic saving, both private saving (through its impact on interest rates) and government saving (through its impact on government revenue).
Fifth, aid flows could undermine private sector incentives for investment or to improve productivity. Aid can cause the currency to appreciate, undermining the profitability of the production of all tradable goods (known as the Dutch disease). Food aid, if not managed appropriately, can reduce farm prices and hurt farmer income.

The last two points merit further discussion. On aid and saving, while foreign aid adds to total saving (since aid is a form of foreign saving), some studies have shown that a dollar of aid adds less than a dollar to total saving and investment, since domestic savings may fall as aid increases. Some of these studies conclude that aid is ineffective because it ‘leaks’ to consumption. This approach is not particularly helpful in the aggregate since large portions of aid are in fact designed specifically to directly increase consumption and not investment, including food aid, immunization programs, purchases of textbooks, technical assistance, and the like. Nevertheless, even where aid is aimed at investment, the impact could be partially offset by a reduction in either private saving (through a decline in the rate of return on private investment) or government saving (through a fall in tax revenues). There is a wide range of estimates of the offset effect, but most find that $1 in aid translates to an increase in investment in the range of 33 to 67 cents. Much depends on the particular country, the type of aid, and other factors.

Aid also could undermine incentives for private sector activity. Aid can spur inflation and cause a real appreciation of the exchange rate, which reduces the profitability of production of all tradable goods, creating ‘Dutch disease’ effects. Aid flows can enlarge the size of the government and related services supporting aid projects, drawing workers and investment away from other productive activities such as agro-processing, garments or footwear exports. To the extent that these tradable activities are a key source of productivity gains, long-term growth may suffer. Similarly, food aid can sometimes undermine local food production if an influx of food drives down prices (it has less adverse impact on production when it displaces food imports).

The empirical studies that have found no relationship between aid and growth have been influential. However, very few published studies have reached that conclusion since the mid-1990s (a recent exception is Rajan and Subramanian, 2005a). Most of those that do use restrictive models that impose constraints such as a linear relationship between aid and growth, ruling out by assumption the possibility of diminishing returns. Most also only examine aggregate aid, imposing the restriction that all aid has a similar impact on growth, which is not particularly realistic, since famine
relief, immunization programs and road projects are all likely to have very different impacts on growth.

Aid has a conditional relationship with growth, helping to accelerate growth under certain circumstances
This view holds that aid supports growth in some circumstances but not others, and searches for key characteristics associated with the difference. This ‘conditional’ strand of the literature has three subcategories, with the effectiveness of aid depending on the characteristics of the recipient country, the practices and procedures of the donors, or the type of activity that the aid supports.

Recipient-country characteristics Isham et al. (1995) found that World Bank projects had higher rates of returns in countries with stronger civil liberties. Burnside and Dollar (2000), in a very influential study, concluded that aid stimulated growth in countries with good policies, but not otherwise. Other researchers have proposed different country characteristics that might affect the aid–growth relationship, including export price shocks, climatic shocks, the terms of trade, macroeconomic and trade policies, institutional quality, warfare, type of government and location in the tropics.8 All of these studies rely on an interaction term between aid and the variable in question, and (not surprisingly) many of the interaction terms are fragile. Easterly et al. (2004) find that the original Burnside and Dollar results do not hold up to modest robustness checks. Roodman (2007) tests several other ‘conditional’ studies and find most of them to be relatively fragile, although the conclusions of Dalgaard and Tarp (2004) are more robust.

Nevertheless, the view that aid works better (or in a stronger version, aid works only) in countries with good policies and institutions has become the conventional wisdom among donors, partly based on this research and partly due to development practitioners that believe this to be the case based on their own experience. The appeal of this approach is that it can explain why aid seems to have supported growth in some ‘well-behaving’ countries but not others. These findings have had an enormous impact on donors (World Bank, 1998). The concept feeds directly into the World Bank’s Performance-Based Allocation (PBA) system for distributing concessional IDA funds, and was the foundation for the United States’ new Millennium Challenge Account (Radelet, 2003).

Donor practices Many analysts have argued that donor practices strongly influence aid effectiveness. For example, multilateral aid might be more effective than bilateral aid, and ‘untied’ aid is thought to have higher
returns than ‘tied’ aid, as discussed previously. Many observers argue that donors that have large bureaucracies, do not coordinate with other donors, or have poor monitoring and evaluation systems undermine the effectiveness of their own programs. Two influential and overlapping views argue that aid would be more effective if there were greater ‘country ownership’ or broader ‘participation’ among government and community groups in recipient countries in setting priorities and designing programs. There has been substantial debate about these issues, and in some cases these ideas have begun to change donor practices. But to date there has been very little systematic research connecting specific donor practices to aid effectiveness.

**Type of aid** Different kinds of aid might affect growth in different ways. Clemens et al. (2004) disaggregated aid into types most likely and least likely to affect growth within a few years, if at all. They separated aid into three categories: (1) emergency and humanitarian aid (likely to be negatively associated with growth, since aids tends to increase sharply at the same time growth falls following an economic shock); (2) aid that might only affect growth after a long period of time, if at all, and so the relationship may be difficult to detect (such as aid for health, education, the environment, and to support democracy); and (3) aid that is directly aimed at affecting growth (building roads, ports and electricity generators, or supporting agriculture). It found a strong positive relationship between the third type of aid (about half of all aid) and growth, a result which stood up to a wide variety of robustness checks. As expected, the relationship with the other types was less detectable.

To summarize the aid and growth research, it appears that aid has been successful in some countries but not others. The overall trend is a subject of debate, but most research has found a positive relationship although the direction of causality is not always clear. This research is only beginning to scratch beneath the surface and investigate what types of aid are most effective and the conditions under which aid has the largest impact on growth. Since disputes continue about the determinants of economic growth more broadly, perhaps it is not surprising that the aid–growth relationship continues to be a matter of sharp debate.

**Donor relationships with recipient countries**

The criticisms about aid have led to debates about how aid programs can be improved to support growth and development more effectively. But the challenge is not easy. Aid programs face some inherent difficulties in trying to achieve a wide range of objectives, provide financial oversight and ensure results.
The principal–agent problem

A key issue facing aid agencies is that there is only an indirect and distant relationship between the people actually providing the financing – taxpayers in donor countries – and the intended ultimate beneficiaries of aid projects – poor people living in low-income countries. In most aid programs, there is a long and complex chain of principal–agent relationships, starting with the taxpayers that delegate authority to elected officials, who in turn become principals that delegate authority to a new set of agents, the heads of aid agencies, which delegate to agency employees, contractors and consultants. In the recipient country, there are similar relationships between citizens, their government and those that actually implement programs. The objectives, incentives and information available to these agents are not always well aligned with the objectives of either the taxpayers or the beneficiaries.

All public sector agencies and many private companies are faced with these principal–agent problems, but the international dimension and physical separation between the original taxpayers and ultimate beneficiaries makes it an even greater challenge for aid. In domestic public programs (such as rubbish collection or local schools) the taxpayers and ultimate beneficiaries are the same people, so they have clearer information about success or failure and can reward or penalize their agents accordingly by re-electing them or voting them out of office. But this feedback loop is broken for aid agencies. Taxpayers cannot tell if their money is well spent, beneficiaries sometimes do not even know about local programs, and each have limited mechanisms for penalties and rewards.

The principal–agent problem affects nearly all aspects of aid delivery including program design, implementation, compensation, incentives, evaluation and allocation of funding. The problem can never be fully solved – private companies face similar issues between owners, managers and employees, as do private aid foundations and charities. The challenge is to design institutions and incentives that mitigate these problems as much as possible to clarify goals, objectives, incentives and rewards. In this regard, one of the key challenges for donors is if, when, and how to apply conditions to their aid, a subject to which we now turn.

Conditionality

Partly as a result of the principal–agent problem, donors often apply conditions on aid programs to encourage recipients to act more in accord with the donors’ (and possibly the ultimate beneficiaries’) interests. Donor conditions on recipient actions or policies are among the most controversial aspects of aid. Policy conditionality is most often associated with the IMF and World Bank, but all donors use conditions to some extent.
The rationale for economic policy conditions is straightforward: donors believe that certain policies and actions in different countries are important for growth and development, and that without them providing aid is futile. If government policies have led to high rates of inflation, massive inefficiencies and waste of public spending, and extensive corruption, then providing aid – whatever the specific purpose – without requiring fundamental change would provide no benefits and perhaps could perpetuate damage. Some even argue that the primary purpose of aid is not the money, but for aid to act as a lever for the policy reforms.

There are three key problems with conditionality. First, it is not always clear what policy conditions are the most appropriate to ensure sustained growth and development. Development doctrine has swung from a state-led approach in the 1950s and 1960s, to basic human needs in the 1970s, to a macroeconomic approach focused on open markets in the 1980s and 1990s, to a greater focus on institutions beginning in the mid-1990s. As a result, the list of conditions is constantly evolving. Debate has raged for decades about whether specific IMF and World Bank conditions are justifiable and whether they support or hurt stabilization, growth and development. And who should bear the costs if donor-imposed conditions make things worse?

Second, while donors are often criticized for imposing too many conditions, they are almost as often criticized for not imposing enough conditions. Some advocates that criticize the IMF for imposing too much fiscal austerity also insist that it should require governments to spend a minimum amount on health and education. The World Bank is often asked to add conditions to force governments to take specific actions, for example on projects that have potential adverse environmental consequences.

Third, conditionality does not seem to work. Most analysts agree that governments implement reforms only when it is in their interests to do so, and donor conditions have little if any impact on that decision. Many donors continue to disburse aid even when recipients fail to meet conditions, sometimes repeatedly so. Donors are faced with their own internal incentives to continue to disburse aid to support the contractors and recipients that depend on it. They also face a ‘Samaritan’s dilemma’ that withdrawing aid would create short-term pain for the very people it is aimed to help.¹⁰

The nature of conditionality has changed over time as the most pressing issues have changed and as donors continue to wrestle with the best ways to apply conditions. During the 1980s, most conditions focused on macroeconomic issues, trade reforms and privatization, as reflected in IMF and World Bank-sponsored structural adjustment programs. During the 1990s as macroeconomic imbalances improved and following the end of the Cold
War, attention shifted to governance, corruption and institution-building. Debate has re-emerged as to whether aid should be conditioned on democratic reforms in recipient countries. Whether governance-focused conditionality is a good idea, or whether it will be more successful than structural and policy conditionalities, remains to be seen.

There are no clear-cut rules for conditionality. Striking the right balance between responsible oversight and accountability on the one hand, and ensuring against high bureaucratic obstacles and the imposition of unnecessary controls or unwarranted policy changes on the other, requires flexibility, judgment and the ability to balance multiple objectives – none of which are easy for aid agencies to achieve.

Improving aid effectiveness
The debates about the strengths and weaknesses of aid have led to specific ideas for change, some of which donors have begun to put into practice. Four stand out.

Country selectivity One influential idea is that donors should be more selective about the countries to which they provide aid, based on the view that aid works best in countries with good policies and institutions. In the strongest version, aid should be provided only to countries that meet these criteria. A more moderate view is that more aid should be allocated to countries with stronger policies and institutions, but some aid should be targeted to countries with weaker policies, especially post-conflict countries. This proposal turns the conditionality debate: instead of providing aid to encourage reforms, give it to countries that have already demonstrated a desire to implement key reforms. In the language of the principal–agent problem, donors should spend less time trying to write contracts that force an alignment of incentives and instead give more aid to countries that on their own demonstrate similar motivations and objectives. Some donors have begun to be more ‘selective’, including the World Bank in the allocation of its concessional IDA funds, some European donors in terms of providing budget support, and the USA with its new Millennium Challenge Account. But since so much aid is allocated for political, security and other foreign policy reasons, there are limits to how far donors are likely to go in this direction.

Recipient participation Many analysts argue that aid has been weakened by donor domination in setting priorities, designing programs and implementing projects, and push for either a more ‘country led’ approach in which recipient governments take a stronger role, or a ‘participatory’ approach in which various groups in recipient countries (government,
NGOs, charities, the private sector) play a more active role. The idea is to eliminate some of the problems in the long chain of principal–agent relationships, and more tightly integrate the ultimate beneficiaries in key aspects of the aid delivery process. The World Bank and IMF (by requiring Poverty Reduction Strategy Papers), the Global Fund to Fight AIDS, Tuberculosis and Malaria, and the Millennium Challenge Corporation have all moved towards greater local participation in designing and implementing the programs they finance. This approach is new, so there is no evidence yet on the extent to which (or the circumstances under which) it improves aid effectiveness. There is a clear and inescapable tension between country ownership on the one hand, and donor priorities and conditionality on the other. Donors are more likely to facilitate a participatory approach in countries in which governments show a strong commitment to sound development policies, and less so in countries with corrupt and dictatorial governments.

Harmonization and coordination Managing aid flows from many different donors is a huge challenge for recipient countries, since different donors usually insist on using their own unique processes for initiating, implementing and monitoring projects. Recipients can be overwhelmed by requirements for multiple project audits, environmental assessments, procurement reports, financial statements and project updates. According to the World Bank, developing countries typically work with 30 or more aid agencies across a wide variety of sectors, with each sending an average of five missions a year to oversee their projects. The donors all want to meet with the same top government officials, leaving them with much less time to deal with pressing matters. These concerns have led to numerous suggestions for donors to coordinate their activities more closely, harmonize their systems or ‘pool’ their funds (Kanbur and Sandler, 1999). But while there has been some progress, the pace of change amongst the donors seems glacial.

Results-based management The emphasis on demonstrating the effectiveness of aid has led to calls for improved monitoring and evaluation and results-based management. In this view, aid programs should aim to achieve very specific quantitative targets, and decisions about renewing or reallocating aid going forward should be based on those results. There are three basic objectives: (1) helping donors allocate funds towards programs that are working; (2) detecting problems at an early stage to help modify and strengthen existing programs; and (3) improving the design of future programs. Stronger monitoring and evaluation would help improve principal–agent relationships so that aid agencies have clearer incentives and
taxpayers have better information about the impact of aid on its intended beneficiaries.

**Summary and conclusions**

Aid flows fell in the 1990s after the end of the Cold War and aid was widely attacked for being ineffective in spurring growth and development. However, aid began to grow again in the late 1990s and indications are that it will continue to grow throughout the first decade of the twenty-first century, although probably less rapidly than donors have pledged.

Most empirical research on aid and growth conducted since the mid-1990s has found a positive relationship, in contrast to popular perceptions, particularly studies that have allowed for diminishing returns and have controlled for other factors that affect growth. Some studies have found that the aid–growth relationship is conditional on the policy or institutional environment, but many of those results have been fragile. Some studies have concluded that there is no relationship or even a negative one, but while influential, these studies are few in number and tend to use restrictive assumptions. Recent research that has explored how different types of aid might have different impacts on growth has suggested one key reason why earlier research has reached mixed conclusions.

Nevertheless, there is little doubt that aid has been less effective in spurring development than is often expected. Aid can keep bad governments in power for too long, and can undermine incentives for saving, tax collection and private sector production. Aid relationships are made much more difficult by a complex chain of principal–agent problems that weaken information flows, introduce myriad motivations for different actors, and make monitoring and accountability more difficult. Attempts to solve the principal–agent problem through conditionality have not been very successful. The newest wave of reform efforts aims to solve some of the weaknesses of aid and the principal–agent problem through greater donor selectivity in choosing aid recipients, increased recipient participation in setting priorities and designing programs, streamlining aid bureaucracies, increasing donor coordination, and establishing clearer goals for aid and stronger monitoring and evaluation of aid-financed activities. These ideas have been very influential in designing aid programs in recent years, but there is no systematic evidence at this point as to whether these changes will lead to greater aid effectiveness.

**Notes**

1. This chapter draws heavily from Chapter 14 of Perkins et al. *Economics of Development*, 6th edn, 2006 (New York: W.W. Norton & Co.), (used by permission), and from Radelet et al. (2006). I thank Bilal Siddiqi and Sami Bazzi for their research assistance, and Amitava Dutt and Jaime Ros for comments on an earlier draft. I also thank the William and Flora Hewlett Foundation for financial support.
2. Non-concessional loans from donor agencies are counted as part of official development finance, but not as official development assistance.

3. More precisely, assistance to countries with per capita incomes (for three consecutive years) above the World Bank’s ‘high income’ threshold, but the DAC makes some exceptions.

4. This summary draws heavily from the review in Clemens et al. (2004). For another recent review of the literature see Hansen and Tarp (2001).


References


for Research in Economic Development and International Trade, University of Nottingham.


43 International migration and the brain drain

Francisco L. Rivera-Batiz

Introduction

International migration flows have expanded by historical proportions in recent years. In 1960, there was a stock of slightly over 75 million people residing in countries other than their country of birth. By 2005, this number had grown to 190 million. For many developing countries, international labor flows have become a natural outcome of the globalization process, as much as trade and investment flows. The money the migrants send back home has become a major source of income for families in developing countries. In 2005, migrant remittances amounted to $600 billion. For some countries, the income received from the services of workers abroad is now a major item of the balance of payments.

But international migration flows are not without their costs. For many years, the issue of the brain drain has been studied by international economists. From software engineers in India to doctors in the Philippines and nurses in sub-Saharan Africa, the exodus of skilled migrants has been a policy concern for many developing countries. The impact on the skilled labor forces of some countries has been substantial, especially in sub-Saharan Africa. In Ghana, for example, over 40 percent of persons with a college degree or more have migrated to other countries. In Gambia, the corresponding proportion is close to 65 percent, and in Somalia it is 59 percent. What are the consequences of these labor outflows? What benefits or costs do they impose in the sending nations?

This chapter provides an analysis and survey of the key recent trends in international migration, their determinants and major consequences. The next section presents data on the growth of global migration flows since the mid-1950s. The section after that proceeds to examine the causes of these migration flows, focusing on the main economic factors involved. The subsequent section analyzes the consequences of international migration for source countries. The concluding section discusses implications for future research.

Trends in international migration flows

International migration refers to the movement of people across national borders. Data on migration are available directly from national immigration
authorities in recipient countries. However, the data available for many countries are sketchy and difficult to compare with data for other countries due to differences in migration policies, definitions of what constitutes immigrants, the presence of undocumented migration, and so on. Some international organizations gather cross-country data on migration and seek to provide more uniform, comparable statistics. The Organisation for Economic Co-operation and Development (OECD) has for many years collected information on the migration of OECD countries and has an extensive database for these countries (see, for example, OECD, 2006). The United Nations has the most comprehensive worldwide database on the number of migrants residing in different countries (United Nations, 2007). There are also data collection efforts supported by the World Bank (Docquier and Marfouk, 2006), the International Monetary Fund (Carrington and Detragiache, 1998), and other institutions such as the Development Research Centre on Migration, Globalisation and Poverty at the University of Sussex (see Parsons et al., 2007).

Table 43.1 shows the massive growth in the estimated number of international migrants between 1960 and 2005. In 2005, there were 190 million people residing in countries other than their country of birth, up from 75 million in 1960. Most of these migrants originated in developing countries: it is estimated that 123 million migrants, or 65 percent of the total, were born in developing countries. Mexico had the highest number of persons residing abroad, equal to over 10 million in 2000, largely in the United States. This was followed by India, which had close to 9 million persons residing abroad in 2000, with the country’s diaspora spread all over the world. Other countries with mass emigration include Bangladesh (6.6 million emigrants), the Philippines (3.4 million), Pakistan (3.4 million), Turkey (3.0 million), Afghanistan (2.7 million), Morocco (2.6 million), Egypt (2.5 million) and Algeria (2.1 million), among others.

The destination of emigrants from developing countries is evenly split between high-income and developing countries. In 2005, close to 62 million people born in developing countries were residing in high-income economies, but 61 million resided in other developing nations. Among high-income countries, those with the largest number of immigrants included the United States (38.4 million), Germany (10.1 million), France (6.4 million) and Canada (6.1 million). Developing countries with the highest immigrant populations included Russia (12.1 million, who moved mostly from elsewhere in the former Soviet Union), Ukraine (6.8 million), Saudi Arabia (6.4 million), India (3.3 million) and the United Arab Emirates (3.2 million).

The emigration of skilled workers is one of the major concerns relating to migration flows in developing countries. To measure the magnitude of
skilled emigration or brain drain, a country’s labor force is divided into those who are skilled – generally considered to be those who have some tertiary education, that is, 13 years of schooling or more – and those who are less skilled, who have achieved less than tertiary education. The emigration of the skilled has been rising, just as global migration flows have increased in general. In OECD countries, for example, there were 12.5 million skilled immigrants of working age in 1990, but by 2000 the number had risen to 20.4 million, equal to 34.6 percent of the total number of immigrants in these countries.

The skilled emigration rate is defined as the stock of skilled migrants from a country (all persons with tertiary education living abroad) calculated as a percentage of the total skilled labor force in the source country augmented by the skilled migrants themselves. This shows the percentage of workers with tertiary education who were born in a country but are residing outside its borders. Hence, it is a measure of the relative impact of the emigration on the sending country’s skilled labor market. The highest skilled emigration rates in the world prevail in the Caribbean, where in 2001

### Table 43.1 Estimates of the Stock of International Migrants in the World, 1960–2005

<table>
<thead>
<tr>
<th>Year</th>
<th>Stock of migrants</th>
<th>Change between years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960</td>
<td>75 463 352</td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td>78 443 933</td>
<td>2 980 581</td>
</tr>
<tr>
<td>1970</td>
<td>81 335 779</td>
<td>2 891 846</td>
</tr>
<tr>
<td>1975</td>
<td>86 789 304</td>
<td>5 453 525</td>
</tr>
<tr>
<td>1980</td>
<td>99 275 898</td>
<td>12 486 594</td>
</tr>
<tr>
<td>1985</td>
<td>111 013 230</td>
<td>11 737 332</td>
</tr>
<tr>
<td>1990</td>
<td>154 945 333</td>
<td>43 932 103</td>
</tr>
<tr>
<td>1995</td>
<td>165 080 235</td>
<td>10 134 902</td>
</tr>
<tr>
<td>2000</td>
<td>176 735 772</td>
<td>11 655 537</td>
</tr>
<tr>
<td>2005</td>
<td>190 633 564</td>
<td>13 897 792</td>
</tr>
</tbody>
</table>

as much as 42.8 percent of the region’s tertiary labor force resided outside its borders (Docquier and Marfouk, 2006). Skilled emigration has also been enormous for a number of countries in Africa. In East Africa, the skilled emigration rate was close to 20 percent in 2001.

The determinants of international migration flows
What has caused the mass migration documented in the previous section? There is a massive literature examining the determinants of migration flows. This section presents the main forces and conceptual approaches that seek to explain migration flows.

The economic approach to migration: costs and benefits
At the level of theory, the classical economic model of the decision to migrate was formalized by Sjaastad (1962) and has been extended in a number of directions (see, for example, Lucas, 1985; Borjas, 1999). In this approach, the decision to migrate is seen as an investment decision that depends on individual assessments of the net balance of the present and future costs and benefits of migration. For a worker $i$, the net gain from migration, $G_i$, is equal to the present discounted value of the benefits minus the costs of migrating:

$$G_i = \sum_{t=1}^{T} \left[ \frac{B_{it}}{(1 + r)^t} - \frac{C_{it}}{(1 + r)^t} \right]$$  (43.1)

where $T$ is the lifespan of the worker, $r$ is a discount rate, $B_{it}$ are the benefits at any given time $t$, generally in the form of higher wages or improved employment opportunities in the destination country, and the costs, $C_{it}$, include the direct costs of the move (transportation costs), the foregone earnings when the individual migrates, and any utility losses associated with leaving the homeland.

Note that different individuals will face varying costs and benefits, and the probability of a person migrating from an origin to a destination area will vary. Older workers, for example, may suffer the greatest net losses in foregone earnings and may also face higher psychic costs of leaving the source country. As a result, the likelihood of migration declines with age. Location also matters. Indeed, most migration flows occur among geographically close, often neighboring, countries, where the costs of migration are lower. Empirically, the role of geographical and cultural proximity in determining migration movements has been emphasized by the so-called ‘gravity model of migration flows’, for which there is ample supporting evidence (see Karemera et al., 2000; Hatton and Williamson, 2005).
Labor markets and the returns to international migration

Due to the difficulties of estimating future costs and benefits, most studies examining the decision to migrate focus on the current labor market gains from migrating, $I_i$, given by:

$$I_i = \log W_{iD} - \log W_{iO} \quad (43.2)$$

where $W_{iD}$ is the wage rate that individual $i$ can obtain in the destination and $W_{iO}$ is the equivalent wage at home. These wage rates are influenced by the various characteristics of workers, including their schooling, $Ed_i$, their on-the-job experience, $Ex_i$, their motivation, $M_i$, health, marital status, number of children, and so on. Suppose, for simplicity, that the following characterizes the wages of individual $i$ in the destination and origin regions:

$$\log W_{iD} = \alpha_D + \alpha_1D Ed_i + \alpha_2D Ex_i + \alpha_3D M_i \quad (43.3)$$
$$\log W_{iO} = \alpha_O + \alpha_1O Ed_i + \alpha_2O Ex_i + \alpha_3O M_i \quad (43.4)$$

where the $\alpha$’s represent how the various individual characteristics (education, experience, and so on) are rewarded in each country. The net gain from migration is thus:

$$I_i = \log W_{iD} - \log W_{iO} = (\alpha_D - \alpha_O) + (\alpha_1D - \alpha_1O) Ed_i + (\alpha_2D - \alpha_2O) Ex_i + (\alpha_3D - \alpha_3O) M_i \quad (43.5)$$

For any individual, the incentive to migrate will vary on the basis of the skills (education, experience, and so on) that the worker is endowed with, his or her motivation, and how skills as well as motivation are rewarded in relative terms in the source and destination countries.

A large body of research has now accumulated studying the labor market outcomes of developing-country emigrants in their destinations (see, for example, the collection of research in Zimmermann and Constant, 2004, as well as the surveys by Borjas, 1999 and Hanson, 2006). This literature suggests that the economic returns to migrating are substantial for most workers. But the economic progress of migrants varies according to the characteristics of the migrants themselves (age, schooling, immigration status, and so on), the timing of the migration, and the country of destination. In some European countries, for instance, immigrants have substantially lower rates of labor force participation as well as higher unemployment rates than the native-born population. The relatively poor labor market outcomes of some immigrants are partly related to lack of skills, but they also depend on labor market conditions in – and
time spent in – recipient countries (Fertig and Schmidt, 2002; Rivera-Batiz, 2007).

Migrant selectivity

One of the most discussed issues in the international migration literature is whether emigrants are the most qualified, skilled workers in the origin economy or not. If migrants are positively selected, then they will be more likely to succeed abroad, but their exit from the source country will drain the most capable, most skilled population from the nation, with a potentially negative impact on its economy and society.

What determines whether migrants are positively or negatively selected? There are forces that favor a positive selectivity and others that encourage a negative selectivity. The best-known hypothesis is that emigrants tend to be positively selected because in order to compensate for the substantial costs of migration, only those who have the strongest drive and motivation – and the expectations of great rewards – will actually undertake the migration process (see Chiswick, 1978, 1999). Of course, if costs of migration decline, then this aspect of selectivity will tend to become less significant. In addition, as equation (43.5) suggests, the impact of motivation on migration depends on whether motivational skills are more strongly rewarded in the destination region. Indeed, one suspects that holding constant the distribution of motivational skills in a population, if these skills are poorly rewarded at home but richly rewarded abroad, the incentives to migrate from source to destination regions will increase.

A second hypothesis (referred to in the literature as the ‘Roy model’) is that, if those with greater skills or abilities are rewarded more highly compared to the less-skilled in the origin area when compared to the destination region, this will generate less incentives for those at the top of the skills or ability distribution to emigrate compared to those at the bottom of the distribution, causing a negative selectivity of migrants. Therefore, the emigrant contingent will be positively or negatively selected depending on the relative inequality of the distribution of income at home and abroad. For instance, the more unequal the income distribution in the origin area as compared to the destination, the lower the incentives for the highly skilled to emigrate relative to the less skilled. This result is easily obtained from equation (43.5). Since \( \frac{\partial I_i}{\partial \text{Ed}_i} = (\alpha_{1D} - \alpha_{1O}) \), if greater inequality in the source country is associated with a rate of return to education in the origin area that exceeds the rate of return in the destination country, then \( (\alpha_{1D} - \alpha_{1O}) < 0 \). This implies that, as the education of the prospective migrant rises, the incentives to migrate tend to decline. Note also that under asymmetric information, employers in the destination region may not be able accurately to assess the skills of the migrants and they may offer lower
wages to the highly skilled migrants, when compared to employers in the source country, that are able to assess more accurately the potential migrants’ skills and pay them wages more consonant with their skills. This will again reduce the rate of return to education received by the emigrants in the destination region relative to the source country \((\alpha_{1D} - \alpha_{1O})\) will decline), shrinking the incentive to migrate of the highly skilled relative to the unskilled (see Stark and Taylor, 1991).

Some evidence appears to support the implications of the Roy model (see Borjas, 1987, 2006). For instance, recent research finds a negative selectivity in the migrant contingent from Mexico to the United States (that is, the emigrant group tends to have lower average skills than the population left behind), as would be expected from the relatively more unequal distribution of income in Mexico relative to the United States (see Fernandez-Huertas Moraga, 2007, for this analysis and Chiquiar and Hanson, 2005, for alternative results).

**Income differences and international migration**

Overall, the evidence on the importance of economic factors in motivating migration flows, as presented in equations (43.1)–(43.4) is extensive. Both documented and undocumented migration flows have been found to be strongly correlated to the relative economic conditions in recipient and source countries (see, for example, Adams, 1993; Hanson and Spillimbergo, 1999; Drinkwater, 2003; Castaldo et al., 2005).

But despite the widespread support for the hypothesis that increased income differentials between recipient and source countries stimulates migration, there is also ample support for the view that this connection does not always work and may actually hold in reverse (see Hatton and Williamson, 1998, and the review by Waddington and Sabates-Wheeler, 2003). For instance, in recent research seeking to determine the impact of differences in income per capita on migration flows in the world, Hatton and Williamson (2005, p. 240) find that in sub-Saharan Africa, increases in income at home increase migration. Indeed, the evidence suggests that the relationship between income per capita in source countries and emigration has an inverted-U shape. For poor countries, as income rises, migration actually increases. But as the wealth of a country grows, at some point, further increases in income per capita actually reduce emigration.

One explanation for this behavior is the fact that, at low levels of income per capita, a large part of the population just simply cannot afford the monetary costs of migrating. However, as per capita income in a country rises, this allows some people to save enough to pay for the transport and other costs of migration, thus increasing migration flows (see Hatton and
Williamson, 2005). An additional explanation is that the massive structural changes occurring in the early industrialization of an economy (the shift from agriculture to industry and services, from urban to rural areas, and so on) leads to a dislocation of the population that fosters international migration. As the economic development process matures, however, these changes diminish and migration declines. A third hypothesis for the paradoxical rise of migration flows as income increases in many economies is based on ‘relative deprivation theory’. It suggests that, as inequality rises in the early years of economic development (a trend first noted by economist Simon Kuznets), those who become relatively poor will become increasingly dissatisfied with their relative standing in the community. This will stimulate them to emigrate in order to improve their standard of living (see Stark and Taylor, 1991).

Families and the decision to migrate
Despite its powerful role in explaining migration flows, another problem of using a simple economic approach based on income differences across countries is that it cannot explain temporary migration. If there is a significant and persistent wage and employment differential between origin and destination regions, why do so many migrants wish to stay in the destination only for short periods of time?

One explanation is offered by the so-called new economics of labor migration (see Stark and Bloom, 1985). In this approach, it is understood that migration decisions are often made by families and households, not individuals. In contrast to the analysis in equations (43.1)–(43.4), the decision-making is not in the hands of an individual who maximizes his or her utility. Rather, migration decisions are made by families or households that maximize household utility over time. This utility, $U$, is derived from a stream of consumption by family members located in the home and/or destination regions, $C_D$ and $C_O$, added all the way up to the time horizon of the family, $T$, and discounted to the present time, $t$:

\[
U = \int_t^T \exp[-\rho(\tau-t)] \log U[C_D(\tau), C_O(\tau)] \, d\tau \quad (43.6)
\]

where $\rho$ is a rate of discount. Equation (43.6) is maximized subject to a budget constraint establishing that the present discounted value of income earned in the source and destination areas is equal to or greater than the present discounted value of family consumption spending.

Visualizing migration as an intertemporal family or household decision can help understand that the migration of some household members may be part of the savings and investment decisions of a family. The idea is that, if low-income households encounter capital market imperfections at home,
which exclude them from access to the financing of investments in housing, durable goods or in self-employed businesses, migration abroad may lead to the accumulation of remittances that can then be used to finance those purchases and investments. Migration becomes a short-term activity needed by households to raise funds in the absence of local financing. Another application of this approach is that the migration of one household member may reduce the costs of migration for other household members. This is what ‘network or chain migration theory’ suggests (see Piore, 1979; Bauer et al., 2000).

The political economy of immigration policy

The discussion so far has described the wide array of forces that may generate a supply of potential migrants. However, in an international context, destination-country governments exert an enormous force in restricting immigration flows. As a result, the volume of migration between developing countries and high-income economies is often determined by the immigration policies imposed by the latter.

Research on the determinants of immigration policies follows a growing literature in political economy that seeks to answer the question of how the policies of a country are generated through the interaction of economic, political and social forces (Mayda and Patel, 2006). A wide array of economic forces may lie behind the setting of immigration policies. For instance, if immigration reduces wages and raises the profits of employers and owners of capital, then persons who own firms or farms or own relatively large amounts of capital will be in favor of immigration, but those who do not have wealth and only have their labor will be against immigration (see Benhabib, 1996). But if the distribution of capital in a country is highly concentrated, with a great part of the workforce laboring at low wages, then there may be very few persons supporting immigration and many opposing it. If immigration policy is determined by influence of voters, immigration policy restrictions may be high. This force is magnified if the immigrants are unskilled since in this case the immigrants may be perceived as competing with the poor, generating stronger cries for immigration restrictions (Hatton and Williamson, 2005).

The available evidence confirms the role that economic forces play in affecting the immigration attitudes of natives in recipient countries (see Gang et al., 1999, 2002; Scheve and Slaughter, 2001; Mayda, 2006). But economic factors are not the only ones affecting attitudes towards immigrants. Social phenomena, such as xenophobia and bias against foreigners, can have a major influence on immigration policies (see Gang and Rivera-Batiz, 1994; and Gang et al., 2002, for analyses of the determinants of attitudes towards immigrants).
International migration and economic development

What are the economic consequences of mass migration for the developing world? What changes in the source countries are generated by migrants? This section examines the existing theory and evidence on how emigration affects developing nations.

The distributional and welfare impact of emigration

The simplest theoretical framework that can be used to examine the effects of international migration is one that focuses on the aggregate economy, within a simplified setting where there are only two inputs: capital and labor (see Bhagwati and Rodriguez, 1975; Borjas, 1999; Mishra, 2006). Figure 43.1 shows the labor market in the source country before and after emigration. The demand for labor (the economy’s marginal value product curve) is $L^D$ and the supply of labor before emigration is depicted by $L^s$, where it is assumed that there is an inelastic supply of labor given by the country’s labor force. Under the assumption that the labor market is perfectly competitive and generates full-employment, the equilibrium before emigration is at point $A$, with a wage rate equal to $W^*$ and employment $E^*$. If the number of emigrant workers is given by $M$, the labor force declines, shifting the aggregate supply of labor to the left, to $L^s - M$. The equilibrium after migration occurs at point $B$. The shortage of labor induced by emigration reduces employment from $E^*$ to $E^{**}$, and raises...
wages rates from $W^*$ to $W^{**}$ (this assumes that the domestic capital stock is unaffected by emigration).

The impact of emigration can be depicted in Figure 43.1 by noting first that national income is given by $Y = WL + rK$, with $K$ equal to the economy’s capital stock and $r$ the rate of return to capital. Diagrammatically, after emigration, the income received by the workers that remain in the economy is equal to $W^{**}E^{**}$, as represented by the rectangle $W^{**}BE^{**}O$. This rises compared to the situation before emigration, when the income of these workers was $W^*E^*$. The gain in income by labor is $(W^{**} - W)E^{**}$. But although labor gains by emigration, capital is hurt by it. The income received by capital is given by $Y - WL$. Diagrammatically, the value of national income, $Y$, is equal to the area below the demand for labor curve up to the level of employment. After emigration, the value of national income is $DBCE^{**}O$. As a result, after emigration, the income received by capital is $Y - WL = DBCE^{**}O - W^{**}E^{**} = DBW^{**}$. The income derived by capital before the emigration was $DBACW^*$. Hence, the income of capital declines by $DBACW^* - DBW^{**} = W^{**}BAW^*$.

In this context, emigration results in a redistribution of income from capital to labor. But the emigration has also a net, overall negative impact on the income of those left behind. Adding the loss of capital and the gain to labor leaves a net loss represented diagrammatically by the area $BAC$. Algebraically, this loss can be approximated by:

$$\frac{(\Delta Y)}{Y} = - \frac{1}{2}(\Delta W)M = - \frac{1}{2}S_L \eta_{LL} \left(\frac{M}{L}\right)^2 \quad (43.7)$$

where $\Delta W$ is the change in wages($\Delta W = W^{**} - W^*$), $M$ is the loss of labor($M = - \Delta E = E^* - E^{**}$), and where $\eta_{LL} = - (\partial W/\partial E)(E/W)$ is the negative of the elasticity of the labor demand curve with respect to wages. The effects on the income of labor and capital are then: $(\Delta WE)/Y = S_L \eta_{LL} (M/L)(1 - (M/L)) > 0$ and $(\Delta rK)/Y = - S_L \eta_{LL} (M/L)(1 - (M/2L)) < 0$.

Although in this simplified setting emigration results in a gain to labor (the mobile factor), a loss to capital (the fixed factor) and a net loss to the overall economy, these conclusions can be reversed in more complex frameworks. First of all, the model is a partial-equilibrium, closed-economy model but most modern economies are both highly diversified and engage in substantial international trade. In a general equilibrium framework with trade, such as the Hecksher–Ohlin–Samuelson (HOS) model, emigration has no lasting impact on the source country, whether on income distribution or in terms of the net impact on economic welfare. The explanation is that the emigrants themselves induce a reduction in the demand for labor in the source country. As the migrants leave, they do place upward pressure on wages, as noted in the earlier model. But this makes employment in
capital-intensive sectors, such as heavy manufacturing, more profitable. As production shifts away from labor-intensive products and into capital-intensive sectors, the overall demand for labor in the economy shrinks. This puts downward pressure on wages, reversing the initial impact of the emigration. This result, where an outflow of labor leads to no change in wages, is based on the Rybczynski effect in the trade literature (see Bhagwati and Rodriguez, 1975; Rivera-Batiz, 1983).

But the assumptions of the HOS model under which this result is derived are stringent. Besides the assumptions of perfect competition and no distortions, the framework assumes the absence of externalities and increasing returns. In addition, the model does not distinguish between skilled and unskilled labor emigration. Once these assumptions are relaxed, the theoretical analysis may yield complex and ambiguous effects of emigration (see, for example, Dutt, 2005). Furthermore, when considering costs and benefits from migration one should consider as well the overall, global impact of migration flows. If there are net world gains, there are then strong reasons for governments in recipient and destination areas to coordinate their migration policies, so that both developing countries and high-income economies can profit from the migration flows (see Pritchett, 2006; Kapur and McHale, 2005).

The empirical evidence on the welfare and distributional effects of emigration on source countries is scant (the analysis of the impact of immigration has been more extensive). Recently, however, some studies have utilized Census data over time to examine the issue. Mishra (2006) examines emigration from Mexico to the USA, concluding that the outflow of Mexican workers to the United States between 1970 and 2000 has increased worker earnings in Mexico by 5.9 percent of GDP and has reduced the income of the owners of fixed factors by 6.4 percent of GDP, with a small, negative net impact on overall economic welfare of those left behind (see also Borjas, 2006). On the other hand, there are other potential impacts of emigration to consider that are not examined in this literature, as discussed next.

Remittances and the impact of emigration

One of the most visible impacts of the migrants on source countries is connected to the remittances that they send back home. These flows of resources have grown exponentially in recent decades. Measured in 2000 international purchasing power parity (PPP) dollars, developing countries received $50 billion in migrant remittances in 1980, but by 2005 this had multiplied to $605 billion (World Bank, 2007). The regions receiving the most remittances were East Asia and the Pacific, and South Asia (each receiving about $171 billion in 2005), followed by Latin America and the
Caribbean (with $86 billion), Europe and Central Asia ($78 billion), the Middle East and North Africa ($75 billion) and sub-Saharan Africa ($24 billion).

The significance of remittances in many developing countries can be seen by comparing the value of remittances with the value of the merchandise goods exported by the source countries. Table 43.2 shows how significant migrant remittances can be, rising in some countries to over 100 or 300 percent of exports.

Remittances clearly constitute an improvement in the standard of living for family members who are recipients of such income. And recent evidence suggests that remittances are connected to lower poverty levels (see Acosta et al., 2006; Adams, 2007). Some questions have been raised as to the extent to which the remittances simply raise current consumption instead of stimulating investment and future economic growth. Recently, however, a number of studies have documented that, first, a significant portion of so-called consumption spending consists of household investments in

### Table 43.2 Migrant remittances in developing countries, largest recipients, 2005

<table>
<thead>
<tr>
<th>Country/region</th>
<th>Remittances in current $</th>
<th>Remittances in PPP-adjusted $</th>
<th>Remittances as % of merchandise exports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Developing countries</td>
<td>179,425</td>
<td>605,678</td>
<td>6.0</td>
</tr>
<tr>
<td>Mexico</td>
<td>21,772</td>
<td>32,222</td>
<td>10.3</td>
</tr>
<tr>
<td>India</td>
<td>19,843</td>
<td>105,564</td>
<td>26.2</td>
</tr>
<tr>
<td>Philippines</td>
<td>11,634</td>
<td>45,605</td>
<td>29.2</td>
</tr>
<tr>
<td>Lebanon</td>
<td>5,722</td>
<td>5,493</td>
<td>335.3</td>
</tr>
<tr>
<td>Morocco</td>
<td>4,221</td>
<td>12,325</td>
<td>42.5</td>
</tr>
<tr>
<td>Serbia/Montenegro</td>
<td>4,129</td>
<td>9,868</td>
<td>103.8</td>
</tr>
<tr>
<td>Pakistan</td>
<td>3,955</td>
<td>14,277</td>
<td>29.5</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>3,583</td>
<td>15,407</td>
<td>43.1</td>
</tr>
<tr>
<td>Brazil</td>
<td>3,540</td>
<td>7,682</td>
<td>3.0</td>
</tr>
<tr>
<td>Colombia</td>
<td>3,345</td>
<td>10,704</td>
<td>15.8</td>
</tr>
<tr>
<td>Egypt</td>
<td>3,341</td>
<td>8,018</td>
<td>43.5</td>
</tr>
<tr>
<td>Guatemala</td>
<td>2,592</td>
<td>5,962</td>
<td>88.2</td>
</tr>
<tr>
<td>El Salvador</td>
<td>2,564</td>
<td>5,718</td>
<td>77.8</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>2,471</td>
<td>6,671</td>
<td>43.0</td>
</tr>
<tr>
<td>Algeria</td>
<td>2,460</td>
<td>7,577</td>
<td>7.9</td>
</tr>
<tr>
<td>Jordan</td>
<td>2,288</td>
<td>5,010</td>
<td>58.9</td>
</tr>
</tbody>
</table>

*Source:* Data for remittances in current dollars are taken from World Bank (2007); other indicators are author’s calculations using PPP adjustments and exports from World Bank (2007).
housing, automobiles and durable goods, whose long-term wealth-raising capacities are substantial; second, the use of remittances for community investment projects is not insignificant and also acts to stimulate local development; and, third, the multiplier effects of the increased consumption spending in generating local economic activity may be substantial (see, for example, de la Garza and Lowell, 2002; Adams, 2007).

On the other hand, remittances do tend to be spent largely on internationally non-tradable goods, which can result in rising prices of these goods compared to internationally traded goods, generating an appreciation in the value of the domestic currency, which adversely affects domestic export sectors (Rivera-Batiz, 1986). Evidence of this Dutch disease-type effect has emerged recently (see Amuedo-Orantes and Pozo, 2004; Acosta et al., 2007). A vicious cycle may develop, where emigration leads to remittances that then deteriorate domestic competitiveness and growth, leading to further emigration, and so on.

The impact of the brain drain
The contribution of migrant remittances for economic development must be weighted against any externalities generated by the migration flows. In countries where emigration leads to the loss of the most talented and skilled, the so-called brain drain, migration could result in significant negative externalities (see Bhagwati, 1979; Bhagwati and Rodriguez, 1975).

If the emigration of skilled labor is substantial and these workers are employed in local service sectors, the result can be acute shortages in the supply of essential services, from school teachers to professors and nurses. Note that the emigration of workers employed in sectors that produce exports and imports is not subject to these effects because local consumers can import these products from abroad when the laborers leave the country. But when the workers are employed in service sectors that produce internationally non-traded goods, the impact of emigration is more significant and potentially disastrous, because domestic consumers can only obtain those services locally (see Rivera-Batiz, 1982). If doctors and nurses emigrate, the supply of health services can collapse, resulting in higher prices and acute shortages. A brain drain can therefore reduce sharply the economic welfare of those left behind.

On the other hand, there may be positive externalities of skilled emigration on source countries. First of all, some researchers have recently suggested that a brain drain may actually raise the level of schooling of the population in the source country, at least in the long run. There are several reasons for this. One hypothesis is that the brain drain will raise the rate of return to education and, as a result, more young people in the country will decide to pursue higher education, thus raising educational attainment.
Indeed, the point that the emigration of skilled workers may lie behind the rising relative wages of skilled labor in many developing countries has frequently been made. However, evidence that these changes have stimulated local investments in human capital has not been produced yet (see Schiff, 2005).

Another possible positive externality associated with the brain drain relates to the fact that emigrants may generate international networks that could enhance the scientific and technological capacities at home. One example is the role that has been played by the Indian diaspora in Silicon Valley and elsewhere in the growth of India’s information technology sector. By generating greater flow of skills and information among countries, and by raising the reputation of a domestic sector internationally, this type of emigration can have significantly positive effects at home (see Kapur and McHale, 2005).

Implications for future research
Both the theory and empirical evidence on the determinants and consequences of international migration have grown enormously in recent years. This research matches the growing importance of migratory flows. But despite the progress, the literature so far has generally failed to consider the dynamic effects of emigration, focusing instead on analyzing – and estimating – static models. This is an especially relevant issue to discuss in relation to the brain drain since there are a number of possible mechanisms through which the emigration of the skilled can affect a country’s economic growth.

First of all, the mainstream analysis of economic growth, as developed by Solow and Swan, suggests that a drop in population growth should lead to an economic expansion for those left behind since the steady-state amount of capital per worker in the economy would rise, increasing per capita income. But there are a number of caveats to this story. First, the emigration of working-age people means that the dependency rate in the economy rises, which tends to absorb resources that would otherwise be dedicated to the accumulation of capital and economic growth. The working-age skilled emigrants are also more likely to have higher savings rates than the rest of the population. Indeed, the empirical evidence available is consistent with a negative impact of a brain drain due to these demographic effects (Bloom and Williamson, 1998).

Another issue is that the emigration of skilled labor may reduce the human capital available to those left behind. This can potentially have a devastating effect on economic growth. One of the main sources of economic growth is technological change, which depends to a large extent on human capital (Romer, 1990). But if human capital flees a country, then the
ability of those left behind to sustain innovation and technical change may be compromised, thus reducing an economy’s economic growth (Rivera-Batiz, 1996).

On the other hand, a growing literature suggests that in countries where the quality of the public sector governance is low, many educated workers tend to be employed in activities that are not necessarily high-productivity activities. In other words, in economies where the government is highly bureaucratic, where corruption is rampant, and the rule of law does not apply, many highly educated workers will not be able to be gainfully employed (see Rivera-Batiz, 2002). In this case, the potentially negative impact of the brain drain on technological change and, therefore, on economic growth diminishes. It is a matter for future research to examine theoretically and empirically in greater detail how emigration is connected to technical change and economic growth.

References


Introduction

International technology transfer (ITT) refers to any process by which a party in one country gains access to technical information of a foreign party and successfully absorbs it into its production process. The importance of ITT for economic development is widely recognized and it has been argued that barriers to technology adoption help explain the income gap between developed and developing countries (Parente and Prescott, 1994). Such barriers include regulatory and institutional constraints that entrepreneurs must overcome as well as low levels of human capital. Furthermore, the technology frontier is a moving target – new technologies are continually being introduced. To close the technology gap, developing countries must adopt new technologies, at a faster rate than they are being created. Both market forces and government policies have an important role to play in accomplishing this formidable task.

At the heart of ITT is the exchange of information and knowledge. Technology may be codified (for example, in blueprints) or uncodified (for example, know-how of engineers). It may be embodied in products or people, or disembodied in ideas or services. ITT often occurs between unrelated partners in market-based transactions. However, information also flows internationally between related parties on a non-market basis, within the boundaries of firms and joint ventures. Given the multifaceted nature of technology transfer, there exist numerous channels through which technology flows across international boundaries. One major channel is trade in goods and services. All exports bear some potential for transmitting technological information. Trade in capital goods and technological inputs can directly improve productivity by being integrated into production processes. Another major channel of ITT is direct trade in knowledge via technology licensing, which may occur within firms, among joint venture partners or between unrelated firms. The focus of this chapter is the channel of foreign direct investment (FDI).

All these channels may facilitate imitation and reverse engineering. Because imitation does not require compensating technology owners, it can be an attractive option for developing economies. As Hoekman et al. (2005)
note, the temporary migration of students, scientists and managerial and technical personnel to universities, laboratories and conferences also plays an important role in encouraging ITT. Furthermore, ITT can also result from the temporary movement of professionals and other service suppliers who enter a developed country to perform specific services and in the process acquire additional knowledge and skills that are transferred back to the home country upon completion of the contract. While such channels of ITT are no doubt of crucial importance we do not discuss them here, in order to limit the scope of this chapter.

Although use of the word ‘transfer’ in the phrase ‘international technology transfer’ seems to suggest that the process of ITT is somehow smooth and automatic, nothing could be further from the truth. The fact that developing countries lag behind the technology frontier merely creates the potential for ITT. For ITT actually to occur, providers and acquirers of new technologies have to undertake deliberate and often significantly large investments.

Investment costs are not the only hurdle facing ITT. The market for technology is hampered by at least two significant market failures: the presence of asymmetric information and/or market power. In fact, the superior information possessed by sellers when protected by intellectual property rights is often what creates the market power. By keeping transactions within one firm, technology transfer via FDI can lessen some of the difficulties that confront arm’s-length exchange of technology, yet many problems remain. Even within the same firm, Teece (1976) finds the costs of transferring technology to a foreign plant average 20 percent of the total investment required for the plant.1

Fuller benefits for local technological capacity are realized if the technologies introduced from abroad diffuse locally. The first step – getting the technology into the country – is ITT, and the second step – getting the technology into the possession of local firms – is what we call technology diffusion. While the first step is typically a deliberate act, the second step can often be an unintended consequence. What makes the role of FDI especially important is that multinational companies dominate global research and development (R&D) and therefore act as important conduits of ITT. Yet multinationals are in the business of maximizing profit, not the economic development of host countries. However, available models and empirical studies argue that achieving a balance between the objectives of multinationals and host countries is not as difficult as it may appear on casual observation.

**Foreign direct investment as a channel of international technology transfer**

Today, intra-firm trade (that is, trade between subsidiaries and headquarters of multinational firms) accounts for roughly one-third of total world
trade, and sales of subsidiaries of multinational firms exceed worldwide exports of goods and services. Thus, FDI is the dominant channel through which firms serve customers in foreign markets. While much of FDI occurs between industrial countries, developing countries are becoming increasingly important host countries for FDI. Approximately 33 percent of the global stock of FDI today is in developing countries (UNCTAD, 2003).

FDI is growing in importance as a channel of ITT. Multinational activity occurs primarily in industries that are characterized by a high ratio of R&D to sales and by large shares of professional, scientific and technical workers (Markusen, 1995). A basic tenet of the theory of the multinational firm is that such firms rely heavily on intangible assets, such as superior technology and well-established brand names, to offset the logistical and other disadvantages of operating in multiple countries as well as to compete successfully with local firms that are better acquainted with the host-country environment. In 1995, of all transactions in royalty and license fees, transactions within the same firm made up in excess of 80 percent, so most explicit trade in technology takes place within multinational firms (UNCTAD, 1997).

Virtually all empirical studies of FDI find that foreign-owned plants in developing countries are typically more productive than purely domestic ones. For example, a recent paper by Arnold and Javorcik (2005) provides direct evidence on the impact of FDI. Using data from Indonesia’s manufacturing sector during the period 1983–86, the authors focus on the effects of acquisitions of local firms by foreign ones. They find that foreign ownership leads to significant improvements in acquired plants: after three years, the acquired plants outperform the control group in terms of total factor productivity by 34 percent.4

ITT through FDI can be either horizontal or vertical in form. When horizontal, FDI transfers the full technology needed to produce the good. When vertical, different stages of the production process are split across countries so only the technology for the stage (or stages) being produced in the host country is transferred. For developing countries, often the more labor-intensive stages are shifted abroad.

Of course, instead of opting for FDI, a firm may sometimes decide to contract with a local firm for production of components in a developing country rather than forming its own production subsidiary there. Due to the participation of local firms, licensing and joint ventures can yield both ITT and technology diffusion. As might be expected, it has been shown that costs of transferring technologies across countries work against FDI (and licensing) as mode choices. Norback (2001) confirms, using Swedish data, that high costs of technology transfer discourage production abroad in favor of exports.
Studies such as Mansfield and Romeo (1980) and Smarzynska (1999) have found that newer technologies are transferred through FDI, whereas older technologies are transferred through joint ventures and technology licensing. Mode choice may be one way that firms attempt to maintain their technological advantage by avoiding modes with high technology diffusion until technologies become somewhat dated. Or perhaps costs of transferring technologies via arm’s-length channels are larger for newer technologies due to greater information asymmetries. Moral hazard considerations can also be important in this context. For example, Ramachandran (1993) has shown that subsidiaries receive greater resources than partially licensor-owned or independent firms once the incentives for both sides to invest in transferring technology are considered. Strategic incentives can also reinforce moral hazard and asymmetric information considerations: Fosfuri (2000) constructs a model with the feature that firms strategically use the vintage of technology to deter imitation by licensees, so that more recent technology is transferred to affiliates than to outsiders (see also Saggi, 1999).

**Foreign direct investment and local technology diffusion: friends or foes?**

An important consequence of FDI is that shifting production to a developing country can reduce technology adoption costs for indigenous local firms. The degree to which imitation costs are lowered by FDI might be higher for process than for product technologies. For product technologies, reverse engineering may be the main way that imitation costs are reduced. Since better process technologies tend to be difficult to deduce from inspection of the final good, first-hand experience with the technology may be required. Multinational firms bring production to the host country, providing workers with experience using the new technology. Workers then often leave to work for rival local firms or to start their own firms. Either way, such worker turnover generates knowledge flows that may lead to local firms adopting some aspects of the ways in which the new technology is better than the old. Also, any degree to which multinational firms adapt technologies to the local economic environment reduces costs of technology adoption for local firms.

Since technology advantages are often needed to survive as a multinational firm, why do multinationals not do anything and everything possible to curtail diffusion of their technologies to rival firms? As argued in Glass and Saggi (2002a), when the gains to local firms are great, the costs of preventing leakage of technologies to rivals (the wage premiums required to keep workers from leaving) are apt to be great as well. Additionally, the presence of multiple multinational firms in an industry likely leads to positive externalities among them: costly efforts undertaken
by any one multinational to curtail spillovers to local competitors would benefit all multinationals. A multinational firm that seeks to protect its technology through litigation, for example, bears the full cost but not the full benefits of its action. As a result, multinationals might very well under-invest in activities that curtail the local diffusion of their technologies.

Focusing on vertical technology transfer from a multinational to its suppliers, Pack and Saggi (2001) have shown that technology diffusion among suppliers can benefit foreign firms sourcing components. Thus, fully integrated multinational firms would be expected to be more averse to technology diffusion than firms involved in arm’s-length production deals with local firms. Mexico’s maquiladoras appear to have benefited from the transfer of sophisticated production techniques and backward linkages, especially in the automobile industry. Goh (2005) finds, however, that diffusion of knowledge to other potential suppliers can either encourage or discourage technology transfer depending on the incumbent supplier’s cost of technological effort. Using firm-level data from Lithuania, Javorcik (2004a) finds evidence of spillovers from foreign affiliates to their local suppliers in upstream sectors, but only for projects with shared domestic and foreign ownership (not for fully foreign-owned investments).

Evidence regarding whether technologies transferred by multinational firms diffuse to competing local firms is mixed. Finding that sectors with more foreign involvement have higher productivity or faster productivity growth could stem from FDI being attracted to those sectors rather than FDI improving productivity or accelerating productivity growth. Plant-level studies are required to help alleviate any selection bias in industry-level studies. Haddad and Harrison (1993) find that foreign firms have higher levels of total factor productivity (TFP) but lower TFP growth than domestic firms in Morocco. A stronger positive effect of FDI in low-tech sectors than in high-tech sectors may indicate that local firms in high-tech sectors lack absorptive capacity. Or perhaps multinationals in high-tech sectors take more actions to preserve their technological advantages.

Aitken et al. (1996) provide an empirical assessment of the hypothesis that technology spillovers ought to increase the marginal product of labor, and this increased productivity should show up as higher wages. Their study employs data from manufacturing firms in Venezuela, Mexico and the United States. For both Mexico and Venezuela, a higher share of foreign employment is associated with higher overall wages for both skilled and unskilled workers. Furthermore, royalty payments to foreign firms from local firms are highly correlated with wages. Most importantly, the study finds no positive impact of FDI on the wages of workers employed by domestic firms. In fact, the authors report a small negative effect for domestic firms, whereas the overall effect for the entire industry is positive.
These findings differ from those for the United States, where a larger share of foreign firms in employment is associated with both a higher average wage as well as higher wages in domestic establishments. Putting Aitken et al.’s (1996) findings into the context of previous work, it is clear that wage spillovers (from foreign to domestic firms) are associated with higher productivity in domestic plants. Conversely, the absence of wage spillovers appears to accompany the existence of productivity differentials between domestic and foreign firms.

Using annual census data on more than 4000 Venezuelan firms, Aitken and Harrison (1999) provide a plant-level test of the spillover hypothesis. They find a positive relationship between foreign equity participation and plant performance, implying that foreign participation indeed benefits local plants that receive such participation. However, this own-plant effect is robust for only small plants, that is, those plants that employ fewer than 50 employees. For larger plants, foreign participation results in no significant improvement in productivity relative to domestic plants. More interestingly, they find that productivity in domestic plants declines with an increase in foreign investment – that is, they find evidence of negative spillovers from FDI. The authors suggest that these could result from a market-stealing effect: foreign competition may have forced domestic firms to lower output and thereby forgo economies of scale.6

However, the results of Haskel et al. (2007) contrast with those of Aitken and Harrison (1999). Haskel et al. (2007) use plant-level panel data for all UK manufacturing from 1973 to 1992 to re-examine the issue of spillovers from FDI. As the authors note, there can be little doubt that local firms in the UK possess sufficient absorptive capacity to benefit from the introduction of newer technologies by multinationals. So if spillovers do not materialize, they cannot be attributed to the limitations of domestic firms. Across a wide range of specifications, the authors find that there are positive spillovers from FDI at the industry level. More precisely, they find that a 10 percent increase in foreign presence in a UK industry raises the total factor productivity of that industry’s domestic plants by about 0.5 percent. However, the authors also note that the large tax breaks and incentive packages given to multinationals seem out of proportion relative to the magnitude of spillovers they generate.

While some studies have cast doubt on the optimistic view that FDI generates positive spillovers for local firms, others have reached different conclusions. Regardless of one’s view of these findings, it is worth stressing that domestic firms should be expected to suffer from an increase in competition that often results from FDI; in fact, part of the benefit of FDI is that it can help weed out relatively inefficient domestic firms. Resources released in this process will be put to better use by foreign firms with superior
technologies, efficient new entrants (both domestic and foreign) or some other sectors of the economy. However, such resource reallocation does not occur instantaneously. Existing studies of spillovers do not cover a long enough period to be able to determine accurately how FDI affects turnover rates (entry and exit). Furthermore, horizontal studies miss spillovers that may result from FDI in industries other than the one in which FDI occurs.

In a critical discussion of the plant-level studies of horizontal spillovers from FDI, Moran (2004) argues that there is a substantial difference in operating characteristics between subsidiaries that are integrated into the international sourcing networks of the parent multinationals, and those that serve protected domestic markets and are prevented by policy restrictions (such as mandatory joint venture and domestic content requirements) from being so integrated. These different operating characteristics include size of plant, proximity of technology and quality control procedures to industry best practices, speed with which production processes are brought to the frontier, efficiency of operations and cost of output. He argues that while the former have a positive impact on the host country, often accompanied by vertical backward linkages and externalities, the latter may actually have a negative impact. Drawing upon a wealth of case studies and econometric evidence, Moran (2004) argues that this contrast in performance holds across different industries, countries and time periods. He astutely notes that the failure to differentiate between export-oriented FDI and import-substitution FDI, or between foreign investors free to source from wherever they wish and foreign investors operating with domestic content requirements, or between foreign investors obliged to operate as minority shareholders and those with whole- or majority-ownership, accounts for the inability of earlier studies to isolate the influence(s) of FDI on host-country welfare.

We noted earlier that arm’s-length technology transfer is usually of lower quality than its intra-firm counterpart. But is greater involvement of local firms, such as in the form of joint ventures, more likely to lead to diffusion? While this appears plausible, there is little empirical evidence in support of this idea. For example, Blomstöm and Sjoholm (1999) find that the degree of foreign ownership did not affect the productivity of local partners or spillovers to domestic firms in Indonesia for 1991. Yet having any foreign participation at all did matter: plants with no foreign participation were less productive. These findings could represent selection at the plant level – FDI is attracted to more productive plants; or a threshold effect – that foreign participation, not the degree of participation, is what matters most.

Although the extent of technology diffusion resulting from FDI is unresolved, that FDI stimulates economic growth in the host country enjoys strong empirical support. Balasubramanyam et al. (1996) find the
growth-stimulating effects of FDI are stronger for countries that pursue export promotion rather than import-substitution policies. So trade policy seems to affect the benefits of FDI, although trade orientation could proxy for other unmeasured differences across countries. For export-promoting countries, FDI stimulated growth more than domestic investment. Borensztein et al. (1998) find that FDI contributes more to economic growth than domestic investment for countries that have a sufficient stock of human capital. Countries with insufficient human capital presumably lack the ability to absorb technologies. Xu (2000) finds that countries need to achieve a minimum level of human capital in order for the technology transferred by US multinational firms to contribute to productivity growth, but most less-developed countries do not satisfy the required threshold.

**Policy options for acquiring and absorbing new technologies**

Separating the concept of ITT from that of local technology diffusion is especially important for analyzing policy choices. When assessing the impact of a policy option, the effects on both ITT and technology diffusion should be considered. Some policies might promote ITT but not technology diffusion. Others might promote technology diffusion but then adversely affect ITT through discouraging FDI. Multiple policy instruments are likely to be needed to achieve the ideal combination of ITT and technology diffusion. Too much technology diffusion, and there may not be much technology to diffuse since the potential for local spillovers may deter FDI. Too much ITT, and few of the advanced technologies may ever be used by indigenous firms.

Many countries such as Japan, South Korea and China have historically restricted FDI, often in favor of technology licensing or joint ventures. Foreign firms were often required to share technologies with local firms in order to conduct business in Japan. It is difficult to judge whether countries restricting FDI would have performed better or worse than if they had taken a more liberal approach, since the counterfactual is not observed. The findings that newer technologies are transferred through FDI rather than through joint ventures and technology licensing call into question the wisdom of policies that favor technology licensing or joint venture over FDI. Even if more technology diffusion results, the technology obtained may be far below the state of the art. It is possible that policy interventions could act to improve the terms of licensing contracts for local firms by removing alternatives (or making the alternatives far less desirable) for the foreign firm.

More recently, developing countries have become quite eager to attract FDI. Part of this eagerness may stem from enhanced awareness that FDI can serve as an important channel of ITT, although employment issues
surely also play a role. Many countries, both developed and developing, offer fiscal and financial incentives to attract FDI. Eliminating restrictions on FDI is likely to be beneficial, at least at the world level, since foreign firms would be freer to choose between modes without interference. However, when it comes to promoting FDI, a few words of caution may be in order. You can have too much of a good thing. Similar to free trade being best and export subsidies being distortionary, care is needed to avoid overstimulating FDI. Incentives could lead to FDI being attracted to the wrong countries – countries where costs will be higher than alternative locations. Excessive competition for FDI between locations could bid away all potential benefits.

As noted above, empirical evidence on technology spillovers from FDI is mixed, so benefits to local firms might not be realized. If governments use incentives to try to obtain the right kind of FDI, one has to question whether the government can indeed pick industries with the best potential for spillovers. Adequate human capital and infrastructure are needed for absorption, and any bureaucratic impediments to technology adoption should be minimized. So much focus on FDI risks overlooking opportunities to improve the diffusion of technologies.

Stronger protection of intellectual property (IP) has often been suggested as a means of attracting FDI. The thought is that firms will avoid FDI in favor of exports to countries with weak protection of IP (although they may also shift from licensing to FDI). Theoretical studies such as Glass and Saggi (2002b) and Glass and Wu (2007) based on the quality-ladder model of growth cast doubt on the idea that FDI rises with stronger IP protection, once the repeated nature of innovation is captured. Taylor (1993) has suggested that poor protection of IP may lead firms to mask their technologies in order to make them harder to imitate. However, in a recent paper Branstetter et al. (2006b) have argued that in a variety expansion North–South product cycle model with endogenous Northern innovation, Southern imitation and FDI, intellectual property rights (IPR) reform in the South leads to increased FDI from the North, as Northern firms shift production to Southern affiliates. Furthermore, this increased FDI drives an acceleration of Southern industrial development, as the South’s share of global manufacturing and the pace at which production of more recently invented goods shifts to the South both increase. In addition, their empirical results show that US-based multinational corporations (MNCs) expand the scale of their activities in reforming countries after IPR reform, and this effect is disproportionately strong for affiliates whose parents rely strongly on patented intellectual property as part of their global business strategy. Furthermore, they also provide evidence from highly disaggregated trade data that suggests that the expansion of multinational activity leads to a higher net level of production shifting...
to developing countries, more than offsetting any possible decline in the imitative activity of indigenous firms.8


Stronger IP protection may be more important for technology licensing than for FDI due to the risk of opportunistic behavior and difficulty enforcing contracts. Yang and Maskus (2001) consider a model in which stronger IP protection increases the licensor’s share of rents and reduces the costs of licensing contracts. Thus, better IP protection may indeed stimulate licensing and technology transfer.

Concluding remarks
International technology transfer is a complex, multifaceted phenomenon. In this short chapter, we have chosen to highlight the role foreign direct investment and multinational corporations play in the process of international technology transfer, paying little attention to international trade of goods and services, movement of workers and professionals, and other related phenomenon. While we do feel that FDI is the most important channel, we do not mean to imply that the other channels are not of considerable importance as well.

That multinational companies are pivotal in introducing new technologies to host countries is fairly well established. However, lagging countries have not only to obtain foreign technology but also to learn how to use it to its fullest potential. In this context, we have found it useful to make a distinction between initial international technology transfer and subsequent technology diffusion within host countries. This distinction can be important since some policies could promote technology transfer but deter technology diffusion, or promote technology diffusion but deter technology transfer. With respect to the contribution of FDI, there is some good news and bad news. First the bad news: multinationals will usually lose from further horizontal diffusion of their technologies and should be expected to take actions that thwart that process. The good news is that technology transfer to local suppliers is compatible with the motives of multinationals and a plethora of empirical evidence indicates that vertical linkages between multinationals and their local suppliers play a crucial role in the industrial development of host countries.
An important policy conclusion of this analysis is that host countries are better off facilitating processes that are compatible with the motives of multinationals. In other words, a developing country should perhaps be less concerned about being able to produce an automobile of its own and more concerned about developing a competitive network of suppliers that can serve (and gain from) well-established foreign firms. It is in this mutually beneficial exchange that the most productive policy intervention might lie. Of course, if both sides are indeed willing participants, policy intervention required would be ‘light’ as opposed to ‘heavy’. Furthermore, it would not be targeted in nature. Instead it would ensure that local businesses have access to adequate infrastructure and skilled workers, and that their expansion or downsizing decisions are not hampered by burdensome regulations. In our view, this is another plus of pursuing policies that take proper account of the incentives multinational firms have (and do not have) to encourage industrial development in host countries.

Notes
1. See also Mansfield and Romeo (1980) and Ramachandran (1993).
2. That FDI leads to ITT may seem obvious, yet Glass and Saggi (1999) have argued that whether FDI creates ITT in aggregate depends crucially on whether substitute channels of ITT, such as imitation, exist. FDI could merely displace imitation that otherwise would have occurred, leaving ITT essentially unchanged. See also Glass and Saggi (1998) for a model in which narrowing the technology gap induces multinationals to transfer state-of-the-art technologies.
3. A nice feature of their approach is that they control for the self-selection problem – that is, while it is true that multinationals typically acquire firms that are relatively more productive, they further contribute to the future productivity of acquired firms.
4. See Saggi (2002) for an extensive discussion of this literature.
5. Cheng et al. (2005) examine the impact of the ability of workers to absorb foreign technologies on the extent of production by multinational firms.
6. Nevertheless, on balance, Aitken and Harrison (1999) find that the effect of FDI on the productivity of the entire industry is weakly positive. They also note that similar results are obtained for Indonesia, except that the positive effect on own plants is stronger, whereas the negative effect on domestic plants is weaker, suggesting a stronger overall positive effect.
7. See also Taylor (1994) for the effects of IP protection on technology transfer.
8. In a related paper, Branstetter et al. (2006a) have shown that there is a significant increase in technology transfer following reforms among affiliates of firms that make extensive use of the US patent system.

References


Introduction

Those international financial institutions (IFIs) most important for development nowadays were not established for this purpose. Keynes drafted the Bretton Woods system as the allied counter-proposal to Nazi Germany’s ‘New Order’ of a European Economic Community, basically structured like the present EU, even including the prospect of a European currency union, though no European ‘Parliament’ without real parliamentary powers (Raffer and Singer, 2001 [2002, 2004], pp. 1–2). The OECD (1985, p. 140) describes the initial tasks: ‘The IBRD was there to guarantee European borrowing in international (North American) markets; the IMF was there to smooth the flow of repayments.’ Because Southern delegations demanded resources for development, ‘and Development’ was added to ‘International Bank for Reconstruction’ (IBRD).

When the communist threat made European economic recovery a very important issue, other programs, mainly the Marshall Plan, took over. Apparently, IBRD loans were considered inappropriate for successfully reconstructing Europe, which raises questions whether they can develop much poorer countries. The addition ‘and Development’ allowed the IBRD to turn South (Caufiled 1998, p. 56). Somewhat later the IMF also shifted totally to the South in spite of its more general mandate.

The International Trade Organization, the last pillar of the Bretton Woods system, did not come into being. Only the ‘provisional’ General Agreement on Tariffs and Trade (GATT) regulating trade in manufactures was established. Keynes’s ideas of developmental interest, such as stabilizing commodity prices, were not implemented. The GATT’s successor, the World Trade Organization (WTO) is characterized by a sharply different view on development, based on the Washington Consensus.

The Bretton Woods institutions (BWIs) and the WTO reflect global power asymmetries. Unlike the Inter-American Development Bank, where regional developing members must have 50.005 percent of total votes, industrialized countries (ICs) control the BWIs by clear voting majorities. The WTO’s one-country-one-vote principle is in practice overturned by the consensus approach, bilateral pressure on developing countries (DCs), and
the so-called ‘Green Room’, the practice of backroom negotiations to which only a few countries are invited, whose results are then presented to the rest for ‘consensus’, usually under time and other pressure. At Seattle this triggered strong protests. Southern discontent has also surfaced during the Doha Round.

**Shifting focus**

The IBRD (2005) and its low-income-country window IDA (International Development Association) declare ‘global poverty reduction and the improvement of living standards’ as their ‘mission’. The bank is not ‘a bank in the common sense’ (ibid.). These statements differ fundamentally from its initial businesslike approach precluding financing social activities. It not only refused ‘messing around with education and health’ or a water treatment plant, but even forced Columbia not to accept a French loan for waterworks (Caufield, 1998, p. 64). McNamara was the first president focusing on poverty. McNamara had the merit of giving credibility to the idea that helping the poor is not wasting resources, but makes economic sense. Brought about by Euromarket lending, the debt problem shifted the orientation of the BWIs fundamentally. The IBRD and IDA moved from project financing towards program lending, which should be exceptional pursuant to their Articles of Agreement.

The IMF was not designed as a development organization, but to enable members of the Bretton Woods system with short-term balance-of-payments problems to stay within the agreed parity bands. The demise of Bretton Woods left the IMF with very few remaining tasks, such as the Compensatory Financing Facility (CFF). These could have been transferred to another institution. Arguably, the IMF should have been dissolved. The debt crisis 1982 provided a new role.

Until the Cologne Summit entrusted both BWIs with the Highly Indebted Poor Country Initiative II (HIPC II), which explicitly includes anti-poverty measures, the IMF had usually and rightly stated that this was not its mandate. Turning debt managers, both BWIs gained strong control over development policies, characterized by the term ‘conditionality’. Conditionality became part of the IMF’s statutes as late as 1969. Until its introduction the IMF fulfilled a highly useful role of emergency lending. Forced to get the BWI ‘seal of approval’ in order to get urgently needed new loans, DCs in distress have to fulfill a wide range of conditions, not all visibly connected to economic necessities. BWI loans might carry over 100 conditions, which raises questions whether all can be complied with. All debt relief measures have increased their leverage.

Introduced to compensate export earnings shortfalls beyond the member’s control, the CFF illustrates the strengthening of conditionality
over time. Initially, a statement sufficed to cooperate with the IMF where required. Eventually, the ‘Fund has increasingly come to the realization that even though a country’s export shortfall was both “temporary” and largely beyond its control the country might still have balance-of-payments difficulties attributable to inappropriate policies and that large amounts of unconditional credit might cause the country to delay adopting needed policy adjustments’ (Polak, 1991, p. 9). Polak (1991, p. 12), an influential IMF theoretician, is outspoken: ‘The purpose of the Fund’s conditionality is to make as sure as possible that a country drawing on the Fund’s resources pursues a set of policies that are, in the Fund’s view, appropriate to its economic situation in general and its payments situation in particular’ – even if the country’s economic policy is not at all the reason for temporary problems.

The IBRD has never made unconditional loans. Conditions requiring policy changes have even been attached to projects (cf. Mosley et al., 1991, p. 27). ‘Programme lending’ increased conditionality: ‘[T]he Bank felt that it needed a place at the top policy-making table’ (Mosley et al., 1991, p. 34) beyond what it could expect from project monitoring. Stern (1983, p. 91), the IBRD’s Senior Vice-President, praised structural adjustment lending as enabling ‘the Bank to address basic issues of economic management and of development strategy more directly and urgently’, as a ‘unique opportunity to achieve a comprehensive and timely approach to policy reform’ (Stern, 1983, p. 104), the response to a ‘feasible . . . call for increased sacrifices’ (Stern, 1983, p. 91).

The IMF started adjustment measures in sub-Saharan Africa after 1973. After decades of adjusting debtor countries and ‘appropriate’ development strategies, no country regained economic viability. Attempts to prove success econometrically were given up long ago. Often no statistically significant difference between program and non-program countries was found. Khan (1990), an IMF econometrician, found significantly reduced growth in program countries; as Polak (1991, p. 42) points out, a predicted reduction in the growth rate of at least 0.7 percent of GDP each year a country had an IMF program. After years of ‘Structural adjustment’ the IBRD (1989, p. 6) found a lack of ‘an integrated analytical framework to understand better the links between a program and its expected macroeconomic outcomes’. In spite of little success, official creditors have steadily increased the role of the BWIs.

Rodrik’s (1996) analysis of neoliberal reforms might explain this behaviour better. He sees the debt crisis as an opportunity seized for a ‘wholesale reform of prevailing policies’, offering the chance ‘to wipe the slate clean and mount a frontal attack on the entire range of policies in use’ (ibid., p. 17). A crisis brought about by overspending, overlending and the sudden

Amitava Krishna Dutt and Jaime Ros - 9781848442818
Downloaded from Elgar Online at 07/19/2019 02:48:07AM via free access
change of economic policy in ICs, which sent interest rates skyrocketing, was declared to stem from disliked policies in DCs.

The BWIs have strongly pushed for policy change while refusing to subject themselves to market mechanisms and basic legal principles. This produced an economically inverted incentive system absolutely at odds with market incentives. Although IFIs (co-)determine their clients’ policies, they refuse to share the risks involved appropriately, insisting on full repayment, even if damages caused tortuously by their staff should have occurred (Raffer, 2004). Borrowers have to pay for such damages. IFIs may gain financially from tortuous behavior and errors by extending new loans necessary to repair damages done by prior loans. New and larger crises increase their importance. Grave negligence creating damages leads to new loans correcting such damages, increasing IFI incomes. The IMF’s proposal of a Sovereign Debt Restructuring Mechanism would have perpetuated this situation, granting IFIs de jure preferred creditor status and increasing the Fund’s role in debt management (cf. Raffer, 2005).

The WTO

The WTO (2005a) presents itself as a ‘negotiating forum’, ‘a set of rules’ and a help ‘to settle disputes’. It sees its mandate in preventing ‘self-defeating, destructive . . . protectionism’ (WTO, 2005b). Theoretically it rests its case solely on comparative advantages (ibid.), although this theorem only works in a two-countries-two-goods world and constant returns to scale are necessary to guarantee welfare gains (cf. Viner, 1937, pp. 470–79).

Mattoo and Subramanian (2005, p. 19) argue that the WTO seems to be the ‘best vehicle’ for advancing Northern corporate interests, seeking ‘the opening of markets in developing countries for manufactured goods’. While able to retain high protection where wanted, ICs have managed to restrict or outlaw protection where it could be in the interest of DCs. The principles of the WTO are in many important respects the very opposite of the ideas behind the creation of UNCTAD.

In spite of rhetoric, tariff escalation continues to exist: ‘OECD tariffs on finished industrial products are about eight times higher than on raw materials . . . These barriers delay entry into the export-oriented industries, which are most accessible to developing countries’ (OECD, 2000, pp. 31–2). Average tariffs on Southern manufactured exports continue to be a multiple of those on imports from other ICs. The Doha Development Round again pressures for more market access of ICs. Market access for non-agricultural products is, for example, hardly of overwhelming developmental interest. DC interests can hardly be identified except, arguably, in trace elements. Special and differentiated treatment practically abolished by the WTO is again discussed. The Doha Declaration calls for a review to
strengthen it, without practical noteworthy effects so far. The WTO’s record has triggered proposals for making it more development-friendly, even from within the BWIs (Hoekman, 2005).

Heavy subsidizing of domestic Northern agriculture and agro-exports conform to WTO obligations. US exports are priced 65 percent below production costs. The EU exports sugar and beef at less than half their production costs. WTO cotton and sugar panels legally established that ICs had failed to abide by the loose rules on subsidies they had crafted during the Uruguay Round. ‘Comparative access to subsidies, not comparative advantage’ (Oxfam, 2005, p. 9) shapes ‘world markets’. Institutions interlink. In the name of economic efficiency the IBRD pressured Mali to pay local cotton producers this subsidy-determined ‘world market price’ in 2004. The government ultimately refused to bankrupt domestic peasants.

‘Voluntary Export Restrictions’ have become legalized. The Trade-Related Investment Measures (TRIMS) treaty restricts developmental options to industrialize. Politics successfully pursued by Asia’s dragons are now outlawed. The Trade-Related Intellectual Property Rights (TRIPS) agreement does not strictly speaking protect intellectual property, because the host of tribal knowledge in many DCs remains unprotected. It ‘increased the monopoly power of patent holders and limits the ability of generic producers to compete’ (Mattoo and Subramanian, 2005, p. 20), enabling pharmaceutical companies to raise prices far above what many poor people can afford. DCs are pressured not to use those WTO safeguards to protect public health, which the USA or Canada have used. Over the years the USA has threatened trade sanctions against countries revising their legislation to incorporate TRIPS safeguards, such as Thailand and South Africa. Complaints were launched against Brazil and South Africa using their WTO rights to fight AIDS via affordable drugs. DC protests brought about change for the better. At Doha the right of WTO members to use, to the full, the provisions in the TRIPS agreement was ‘reaffirmed’. This would have been unnecessary if DC rights under TRIPS had been respected.

Dispute settlement was one of the main chocolates on the tray to convince smaller and weaker countries to sign the WTO treaties, promoted as a rule-based system protecting the rights of the weak. It had been argued that the WTO would substitute bilateral (and GATT-violating) measures such as the US Super 301, a unilateral measure implemented in breach of international treaties. Meanwhile, the WTO accepted Super 301 (WTO, 2000, pp. 67–8). A review process had been agreed at Marrakesh, but ICs blocked any change. Although Doha is called a ‘single undertaking’ encompassing all issues, dispute settlement, a main concern of DCs, is not part of it.
Dispute settlement has no mandate to protect members’ rights. Article 3.7 of the Understanding on Rules and Procedures Covering the Settlement of Disputes states: ‘Before bringing a case, a Member shall exercise its judgment as to whether action under these procedures would be fruitful. The aim of the dispute settlement mechanism is to secure a positive solution to the dispute.’ The probability of success, not the rule of law, is explicitly established as the guiding principle. This is a unique and unfortunate approach. Suing big guys is often fruitless, not least because of the way ‘relief’ is organized. There is no right to compensation for damages suffered by violations of contractual obligations. The winning party may be authorized to suspend WTO concessions subject to strict and constraining rules. After winning against the USA, Antigua, whose exports of Internet games were blocked in breach of contract, was authorized to sanction the USA. The positive side is that DCs can and do win cases and can often obtain relief.

Less agricultural export subsidies and reviewing dispute settlement are not the only unfulfilled promises made while the signature of DCs was coveted. Before Marrakesh, the countries of the Organisation for Economic Co-operation and Development (OECD) apparently perceived a need to assure DCs of relief measures against expected higher food prices. After ratification net importers were referred to existing BWI facilities. A list of net food-importing developing countries exists meanwhile, but being listed does not confer automatic benefits. Donors and international institutions want a role to play. Estimates of the WTO’s benefits to DCs were ‘exaggerated and its costs were underplayed’, ‘liberalization assumptions were disconnected from what the [Uruguay] Round actually achieved’ (Mattoo and Subramanian, 2005, p. 21). Calling Doha a ‘Development Round’ continues this tradition.

The WTO is used to cleanse trade relations from disliked historical obligations. WTO compatibility is presently the EU’s argument to remove those remnants of the Lomé framework that had been adopted in favor of and due to pressure by DC signatories. The Final Act of the Uruguay Round also increased IFI influence. IFIs are to cooperate more closely with the WTO to achieve greater coherence in economic policy. Mali’s example above illustrates how this is done. Structural adjustment lending forced DCs to open and liberalize their economies to the extent of making the ‘WTO process a “victim” of the success of the World Bank and the IMF’ (Mattoo and Subramanian, 2005, p. 20). The WTO treaties are now useful to lock in liberalization, to restrict the options of DCs. Simultaneously, bilateral treaties try to gain further concessions, which are likely to become enshrined into WTO treaties at a later stage.

Unlike in other cases where signing means accepting all obligations of treaties, accession to the WTO means prior bilateral agreements with all
members. These are used to extract further, individual concessions. Small countries have little choice, unlike big ones. The EU demanded that Russia increase its domestic oil price as a precondition for being allowed to join the WTO. Russia declined. Apart from China, few DCs could defend issues important for them as easily.

Massive protests, the events of Seattle, and the forming of the G20 have improved the position of DCs somewhat. Especially Brazil and India have acquired new importance. But it remains to be seen how much influence any DCs will finally have.

**Statutory rights and developing members**

Critical minds point at the considerable difference between statutory rights and the policy space DCs enjoy de facto, as well as at the costs resulting from this discrepancy. These rights, the fear that bilateral arrangements might be worse, promises (such as large cuts in agro-export subsidies), and pressure seem to have enticed DCs into signing the WTO treaties. Once members, they are safe from bilateral accession agreements.

The WTO does not fully protect the rights of weaker members, as pharmaceuticals illustrate. Big players have a choice and cannot be forced to honor contractual obligations. When the EU complained against the Helms–Burton Act, the US observed that this would not lead to resolving the dispute but pose serious risks for the WTO. After agreeing bilaterally not to apply Helms–Burton to EU corporations the complaint was dropped. In a dispute with Brazil, Canada simply refused to provide information it was obliged to disclose promptly and fully pursuant to Article 13.1 of the Dispute Settlement Understanding. Expressly mentioning this and its potential to undermine the dispute settlement system, the WTO (2000, p. 59) found against Brazil, as Canada’s WTO-violating behavior could not be proved because of Canada’s additional violation of WTO rules on providing information.

DCs exercising contractual rights may raise the WTO’s concern. Though ‘not extensively used’ in Asia after the 1997 crisis, selective tariff increases by some countries remaining ‘within the flexibility allowed by bindings under the WTO agreements’ gave ‘cause for concern to the extent they may distort the pattern of production and trade’ (WTO, 1998, p. 28). The WTO has never voiced similar concern on potential distortions regarding ICs, including agrarian subsidies or the long phasing-out of GATT-inconsistent restrictions.

Capital controls are a membership right pursuant to the IMF’s constitution explicitly restricting the use of Fund resources to finance ‘large and sustained’ outflows. Even current transfers can be restricted with the Fund’s approval. The IMF may, but is not obliged to, request controls. Its
statutes clearly show that it is not supposed to press for liberalization of capital movements. Asian countries had the right to control capital outflows – as the IMF admitted when Malaysia exercised it (cf. Raffer and Singer, 2001 [2002, 2004], p. 157). IMF programs financing large and sustained outflows by speculators violated the IMF’s constitution, causing damages to DCs while increasing IMF drawings and thus earnings. Pressure to liberalize capital accounts has made increased stocks of international reserves necessary. These have become ‘one of the widely-used targets of poverty reduction strategies in Africa’ (UNCTAD, 2002, p. 31). This money is lost for poverty eradication, debt service or financing the Millennium Development Goals (MDGs). As DCs pay higher interest than they receive on reserves (often US Treasury Bonds), they face substantial annual costs which they could avoid by exercising statutory rights.

The IBRD’s statutes demand debt relief in the case of default, without any conditions. Article IV.4.c confers a right onto members suffering from an acute exchange stringency to ‘apply to the Bank for a relaxation of the conditions of payment’. Article IV.7 contains the obligation to reduce claims in the case of default. The statutes of the Asian and Inter-American Development Banks are similar. The African Development Bank’s new statutes do not contain this clause. The European Bank for Reconstruction and Development writes off losses, proving that IFIs can survive while doing so. By simply refusing to acknowledge default, even if countries have not paid anything for six or seven years, the IBRD does not act according to its statutes. This creates damages by delaying solutions.

**Attempts at reforms by the South**

The preponderance of Northern interest in most international institutions triggered Southern attempts to create new global institutions more in line with developmental needs and less under Northern control. Success has remained limited.

Early attempts focused on mechanisms within the UN, the most prominent being SUNFED (Special UN Fund for Economic Development). Inspired by the generosity of the Marshall Plan, the ‘wild men’ at the UN advocated setting up a UN Fund to administer large-scale soft aid. Theoretically, this idea was based on the then generally accepted Keynesian consensus that capital availability determined growth, which in turn was needed to improve the lot of the poor. It was easy to argue that European countries had a moral obligation to help as they had been helped by the US. ICs opposed the idea of the UN administering large funds. Demanding harder terms, nearer to the market, the IBRD was particularly strongly opposed to the principle of soft financing on economic
grounds. Once it became clear that soft multilateral financing would be done by an institution administered by the IBRD, the Bank dropped all reservations. Helped by the Cuban revolution, which sparked a wave of US ‘generosity’, and US interests in disposing of embarrassingly high, practically useless holdings of inconvertible currencies, IDA was established in 1960, attached to the IBRD, which ICs control. ICs preferred IDA to any UN Fund, ‘because the structure of the World Bank ensured weighted voting in their favour’ (OECD, 1985, p. 141). The Marshall Plan’s participatory and recipient-friendly approach was not to be repeated. The OECD (1985, p. 146) observed a certain differentiation regarding multilateral aid: ‘by and large the largest donors have favoured the World Bank, while the smaller donors have favoured the United Nations’. The UN obtained two valuable consolation prizes: the UNDP and the World Food Programme.

The UN Economic Commission for Latin America (ECLA; Spanish: CEPAL) took the lead in defending Southern interests by publishing Raúl Prebisch’s (1949) findings on the effects of trade on DCs, one pillar of the Prebisch–Singer thesis. This publication angered the USA to the point of attempting to close ECLA (Toye and Toye, 2003, p. 463). ECLA’s economists, the ‘Cepalistas’, continued over decades to produce divergent views, a source of heterodox theories emanating from the South.

Prebisch was also the engine behind the establishment of the United Nations Conference on Trade and Development (UNCTAD) created in order to reduce or eliminate damaging effects of world trade on DCs. Its ideas on appropriate trade-related development policies differ pronouncedly from the WTO approach. Commodity agreements were established, all of which collapsed later on. In the 1970s the South demanded a Common Fund to stabilize commodity prices. Its ‘Second Window’ was to finance projects such as research and development (R&D) and economic diversification. After protracted bargaining the Common Fund was agreed on before the UNCTAD meeting of 1979. Voting shares were allocated to the North (42 percent), the Group of 77 (that is, the South, 47 percent), communist countries (8 percent) and China (3 percent). ICs (East and West) stipulated financial contributions higher than their shares in voting rights. It took roughly a decade until enough countries had ratified to bring the Fund into existence, although the Organization of Petroleum Exporting Countries (OPEC) bankrolled the contributions of poor DCs. Inadequate resources prevented the Common Fund from playing any role. Several of this Fund’s ideas were taken up by an eminent person’s report on commodities in 2003 (Khor, 2005).

A highly useful recent activity of UNCTAD is its Debt Management Financial Advisory Services (DMFAS) program helping DCs to improve
debt management. As effective debt management is part of good governance, and accounting errors by creditors to the detriment of DCs are documented, one might have expected the BWIs to help debtors in establishing appropriate controlling. They did not. Advice by UNCTAD is preferable because UNCTAD is, unlike the BWIs, not a creditor.

There exist other multilateral institutions, notably those established by OPEC or Arab countries, such as the OPEC Special Fund. Established in 1977 with strong political and financial support of OPEC to finance projects in favor of the rural poor, IFAD (International Fund for Agricultural Development) was a successful attempt to make multilateral structures more democratic (Maurizio, 1983). Active participation by the projects’ beneficiaries was sought. Votes were equally split among the North, OPEC countries and other DCs, separating votes and contributions. This did not make IFAD particularly popular with ICs. The first replenishment led to quarrels. Soon ICs demanded a restructuring of IFAD so as to align votes and contributions. Meanwhile over 56 percent of all votes are distributed in accordance with cumulative convertible currency contributions. IFAD, a child of the South’s drive for a New International Economic Order in the 1970s, was finally brought in line with present realities. At a time when the power of ICs has reached its apex since decolonialization, a multilateral institution where they have only one-third of the votes is simply unacceptable.

Before 1997, Asian countries wanted to establish an Asian Monetary Fund as a regional counterweight to the IMF. The Asian crisis put an end to this, at least for some time. It is interesting to note that the IBRD (1999, p. 2) knew years before that the liberalization policies it encouraged in Asia would lead to disaster (cf. Raffer and Singer, 2001 [2002, 2004], pp. 150–51). This helps us to understand why conspiracy theories abounded. Even as distinguished an economist as Bhagwati (1997) spoke of a ‘Wall Street–Treasury complex’ dictating the agenda.

Conclusion

Evaluating the record and policies of international institutions does not suggest that they fully support development. Attempts by DCs to establish more development-friendly international institutions corroborate this conclusion.

At present one sees an interlinking of important institutions. The BWIs force debtor DCs to liberalize quickly and strongly. The WTO then protects and perpetuates the results. DCs suffering from negative WTO effects may draw on IMF resources, thus increasing their dependence on the Fund. Bilateral treaties go beyond the WTO, using the greater leverage ICs have vis-à-vis most DCs, justifying DC fears that bilateralism might protect them even less. But bilateral treaties also prepare the ground for further
concessions by DCs that may eventually become WTO obligations. Typically, more recent treaties extend the definition of ‘investment’ to loans, thus increasing creditor leverage. The WTO process moves so quickly that small DCs simply have no chance to implement agreements before new changes are negotiated or even implemented. Unequal treatment of members depending on their political clout is another fact.

Chang (2002, p. 139) argues that ICs are ‘kicking away the ladder’ of development by ‘insisting that developing countries adopt policies and institutions that were not the ones that they had used in order to develop’. Chang feels this may be done ‘out of genuine (if misinformed) good will’, nevertheless with catastrophic results. He might as well have quoted List’s recommendation of North–South relations. Better known for his opposition to the ‘English philosophy’ of free trade as harmful to Germany in its early development stages, List (1920 [1841], p. 211) advocated joint exploitation of DCs as ‘promising much richer and more certain fruits than the mutual enmity of war and trade regulations’. International institutions seem to show such ‘neo-Listian’ (Raffer, 1987) tendencies.

Important international institutions are constructed in such a way that they cannot but implement the wishes of ICs. Within IFIs, ICs usually enjoy comfortable voting majorities. Within the WTO, unequal economic and political influence asserts itself in spite of ‘one-country-one-vote’. Even if they wanted to, international institutions could not act against their powerful members. ICs seem more interested in gaining economic and political advantage than in supporting development. As selective WTO liberalization proves, ICs do not wish to approximate free market conditions globally, but seek export and investment possibilities where this is in their interest. Thus, Northern FDI in the South is said to be good for anyone needing WTO protection. Chinese FDI in the USA, however, is not – and is treated differently.

Finally, international institutions have self-interests. They want to gain influence and enlarge their mandates. This seems to explain the BWIs’ role in debt management particularly well. Regarding the three most important institutions, these factors combine to produce effects that hinder rather than foster development.

References


46 North–South issues

Amitava Krishna Dutt

Introduction

It has become established practice in development economics to view the world as being divided into two parts, one comprised of rich, developed countries and the other comprising poor, less-developed countries. Although other terms have also been used to describe this division – including Center (or Core or Metropole) and Periphery, the West and the Rest, the First World and the Third World – the North–South terminology, despite its less than complete geographical accuracy, has become the most popular one, as reflected by its usage in North–South (or Brandt) Commission, ‘North–South’ negotiations in international organizations and other arenas, and the literature on ‘North–South’ models.

The relevance of the division between the North and South has sometimes been questioned because of the lack of homogeneity within the North and especially the South (as reflected, for instance, in the emergence of the newly industrialized countries or NICs from the ranks of the South, with characteristics very different from the least-developed and often stagnating economies of Africa and elsewhere) and because of divergent interests within them which has reduced their ability to bargain as unified blocs of countries. Despite this, however, the North–South dichotomy has remained an enduring one – in both descriptive and analytic senses – for a number of reasons. First, there are great differences in levels of economic development – however measured – between rich and poor countries, and it is of enormous ethical importance to examine whether such differences are persistent or not, and if they are, what can be done to remove them and bring about greater international equality. Second, one of the central issues in all of economics is why poor countries remain poor and what can be done to raise their levels of income and production, and it is possible that much can be learnt by comparing the experiences of rich and poor countries. Third, it is widely perceived that there are major asymmetries in the economic characteristics of rich and poor countries, and that this may well imply that trade, factor movements and other interactions between them will have asymmetric impacts on them. As the interaction between countries seems to be growing due to the widely noted process of globalization, it is important to explore the effects of such interaction on rich and poor countries. Fourth, there are many issues, including those of trade
policy, the activities of transnational corporations (TNCs), international labor migration, intellectual property rights and the global environment, on which the interests of rich and poor countries appear to be opposed, and on which cooperation between them is arguably required for their mutual benefit. All these issues can be examined and analyzed most simply by dividing the world into a rich and a poor part. It is little wonder that large and growing theoretical, empirical and policy-oriented literatures have developed, analyzing the world economy in terms of the North–South dichotomy.

This chapter provides a flavor of the main issues raised in these discussions. First, it discusses the main findings on whether the gap in levels of development between the North and the South have been growing or falling. Second, it briefly reviews some contributions to the theoretical literature on North–South models. Following that, it examines in turn the implications of, and issues concerning, trade, capital flows, labor migration, technology transfers and environmental interactions between the North and the South. The final section mentions other North–South issues and some issues which are downplayed by dividing the world into a North and a South.

North–South inequality

A large body of empirical research has examined changes in the level of international inequality and the gap between rich and poor countries. This issue has been investigated in various ways, mostly using purchasing power parity-adjusted per capita real GDP figures. Many observers have employed standard indicators of inequality – such as the Lorenz curve, the Gini coefficient and the Theil index – to show that inequality across countries has increased; Sala-i-Martin (1996) shows that the standard deviation of the log of per capita real GDP for 110 countries increased more or less steadily between 1960 and 1990, implying what is called $H^2$-divergence. An alternative technique regresses growth rates of per capita GDP for the 1960–90 period on the logarithm of initial level of per capita GDP to find a positive coefficient, implying that richer countries on average grow faster, so that there is $\beta$-divergence (Sala-i-Martin, 1996). Quadratic regression equations involving the same variables find an inverse U-shaped relationship, implying a positive relationship between starting income level and per capita growth for most of the sample, and a negative one for a small group of high-income countries. Quah (1993), examining the distribution of per capita GDP levels (relative to the world average), finds that the distribution tends over time to one with a thinning middle and accumulation at the two tails (the so-called twin-peaks phenomenon), and that countries very seldom move from low to high ends of the distribution. In sum, these
figures suggest growing inequality among countries, with convergence among a group of rich countries. Overall growth figures for groups of countries tell a similar story: according to World Bank data, the richest one-third of countries on average grew by an annual rate of 1.9 percent between 1970 and 1995, whereas the middle third grew by only 0.7 percent and the bottom third showed hardly any growth at all (Scott, 2001, pp. 162–3). Pritchett (1997), using different plausible estimates for initial levels of income for poor countries (for which hard data are not available), finds that there has been ‘divergence, big time’ between rich and poor countries over the last 150 years.

These findings, however, have been disputed by some analysts, who point out that by treating each country as one observation, most studies do not give adequate weight to the large low-income countries, China (which is sometimes left out of the sample due to lack of data) and India, which have experienced relatively high rates of growth in recent years (Dollar and Kraay, 2002; Bhalla, 2003). They argue that correcting the problem yields the result that poor countries – as a whole – are growing faster than rich countries. However, there may be a case for giving each country an equal weight, since each country follows a particular set of policies and is thus a single observation. Thus, if we are interested in the relative performance of countries (rather than individuals) it is more appropriate to give equal weights to countries. If we are interested in inequality among people in the world, population-weighted measures are appropriate, but because such measures implicitly assume that there is no inequality within countries, they have to be supplemented by data on within-country inequality to measure inequality among people (see Milanovic, 2005).

Even if the North and South are experiencing divergent patterns of development, it does not imply that this divergence is due to the nature of their interaction. Divergence could be occurring due to factors internal to them, as explored in models of poverty traps which formalize the earlier literature on vicious and virtuous circles in closed economies (see Azariadis and Stachurski, 2005). Since the empirical exploration of the relationship between North–South interactions and divergence is still in its infancy, we may turn to theoretical North–South models to examine the implication of such interaction.

North–South models
North–South models are not new. The classical economists, including Adam Smith (who discussed the importance of increasing productivity due to the division of labor as economies grow) and David Ricardo (who examined the role of trade in postponing the arrival of the stationary state in rich countries by enabling cheap food imports and in having the opposite effect
for poor countries), discussed North–South issues. Subsequently, Marxist writers on imperialism and dependency theorists, and development economists more generally, stressed the role of the South in providing markets and investment outlets for the North, and examined the problems of surplus transfers from the South, the deterioration of the Southern terms of trade, and of uneven development. For instance, Lewis’s (1954), pioneering contribution on economic development with unlimited supplies of labor examined the determination of the terms of trade between rich and poor countries in a world with three goods: a Northern good, a Southern good and a non-traded good which both regions produce. Lewis showed that technological change in the Southern goods sector, low and stagnant productivity of the non-traded goods sector (food) in the South, and high and rising productivity in that sector for the North, have the effect of turning the adverse Southern terms of trade further against the South.

More mathematically explicit general equilibrium North–South models which endogenize capital accumulation in the two regions emerged from the early 1980s. Findlay (1980) examines capital accumulation in a global economy with the North growing with full employment as in Solow’s model and the South with unlimited supplies of labor at a fixed real wage as in the Lewis model. Taylor (1983) allows for unemployment in the North as well, assuming that effective demand determines Northern growth as in the Kaleck–Keynes approach. These and other models can be thought of as special cases of a general framework (see Dutt, 1990) in which the North and South are completely specialized in the production of their good, where the Northern good is a consumption-cum-investment good, and the Southern a pure consumption good. The models embody specific behavioral and institutional assumptions for the North and the South, thereby highlighting their structural differences, and assume that fixed fractions of consumption expenditure in each region are spent on the two products (that is, preferences are homothetic), and trade between the two regions is balanced. The framework examines short-run equilibria with given stocks of capital with markets for both goods clearing, and long-run equilibria in which the stocks of capital in the two regions grow at the same rate due to saving and investment. The models are used to examine the effects of changes in such things as technology, consumption expenditure patterns and savings rates. Of particular interest are results which demonstrate that Southern growth depends on Northern growth, which is determined independently of the South (which imply that attempts by the South to grow faster by saving more merely imply a deterioration in its terms of trade, revealing its dependent status), the relation between the Southern terms of trade and Southern growth, and the possibility of uneven development (reflected by a rise in the relative capital stock of the North to the South).
due, for instance, to technological change, changes in consumer preferences and industrial concentration in the North (see Dutt, 1990). These results confirm some of the informal ideas of earlier writers on uneven development, but they depend on some of the specific assumptions made about the structures of the Northern and Southern economies. The models have also been extended to incorporate issues such as international capital flows and technology transfers.

The models stressing structural asymmetries between the North and the South do not explain why such asymmetries arise. Implicitly, they assume that events in the past, such as the Atlantic slave trade (Darity, 1992) or colonial domination and consequent policy regimes (Dutt, 1992a), related to the balance of military power (Findlay, 1992) create and lock in these structural differences. Models which assume identical structures for the two regions have also been developed to show how ‘small’ historical events can make one region (the North) end up exporting goods exhibiting increasing returns to scale and productivity-enhancing learning effects, while the other region (the South) becomes more oriented towards goods exhibiting constant returns and relatively technologically stagnant sectors, so that there is uneven development (Krugman, 1981, 1990).

The models discussed so far can be seen as reactions to the dominant neoclassical Heckscher–Ohlin–Samuelson (HOS) trade models which contain optimizing agents, are usually static in nature, and assume that markets clear so that labor and other resources are fully employed everywhere. However, the neoclassical approach has also contributed to the development of North–South models. Most of the North–South models from the neoclassical perspective, however, have followed the contributions of new growth theory, which emphasize the role of increasing returns and externalities in the growth process (see Darity and Davis, 2005). Many new growth theory models imply economic divergence between rich and poor countries due to economies of scale along Smithian lines even without any interaction between the two. These results often carry over to models with North–South trade, as we shall see below.

North–South trade
According to the canonical HOS theory of trade, with its assumptions of constant returns to scale and perfect competition, countries export goods which use their abundant factor intensively. It is generally assumed that the North is capital-abundant and the South labor-abundant, and trade with the North exporting capital-intensive goods and the South labor-intensive ones results in gains from trade through the reallocation of resources according to comparative advantage. With the addition of the assumption of identical technology across trading partners and some other conditions,
the theory also implies – according to Samuelson’s factor price equalization theorem – that trade equalizes factor returns across countries even without any factor mobility between them, since in effect abundant factors move abroad embodied in traded goods. These implications are frequently invoked as proof of mutually beneficial North–South trade, and of convergence. Modifications of the approach, which focus on skilled and unskilled labor as the two factors and assume that the North is skilled-labor abundant, have also been used to examine how the North can suffer increasing inequality and (with rigid wages) the unemployment of unskilled labor (Wood, 1994).

The HOS approach, however, does not imply that countries necessarily gain from trade or from moving to freer trade. If economies are ‘distorted’, for example because of the existence of production externalities or factor market rigidities, the economy may well lose from trade (see Bhagwati and Srinivasan, 1983). If such ‘imperfections’ are more prevalent in the South, as commonly supposed, North–South trade may not benefit the South. Dynamic extensions to the approach which incorporate some of the insights of new growth theory even imply uneven development. For instance, when trade liberalization leads to a rise in the wages of skilled workers or the return to capital in the North and reductions in the South, it can speed up human and physical capital accumulation in the North and slow it down in the South, resulting in a growing gap between the two regions (Baldwin, 1992; Stokey, 1991).

More drastic departures from the HOS approach may also imply divergent growth patterns. North–South models in which the North is specialized in goods with higher income elasticities than Southern goods imply that balanced growth for the two regions creates an excess supply of the Southern goods and a deterioration of the Southern terms of trade (Thirlwall, 1979). Long-run equilibrium in models with such a property – which seems empirically plausible – leads to uneven development (see Dutt, 2003). Models such as Krugman’s (1981 [1990]), in which the North specializes in the production of relatively sophisticated goods which exhibit increasing returns and dynamic learning and spin-off effects, while the South specializes in goods which do not have these properties, also imply uneven development and possible losses from trade for the South, a result which is found in a range of similar models (see Ethier, 1982; Boldrin and Scheinkman, 1988).

If insights such as these have any validity, the policy implication for the South is to attempt to change its pattern of specialization with industrial and trade policies towards goods with favorable demand and technological characteristics. Whether individual Southern countries can do so effectively is, of course, another matter. But the experiences of late industrializers in
the past, such as the USA and Germany (where such policies were espoused by Alexander Hamilton and Friedrich List) and Japan, and more recent success stories of South Korea and Taiwan, certainly point to their importance (see Amsden, 2001).

**North–South capital movements**

According to standard neoclassical theory the mobility of factors of production leads to production convergence. The argument can be made in its simplest form for a one-good, two-factor – capital and labor – world with diminishing returns to factors of production, perfect competition, flexible prices and profit-maximizing behavior for the case of capital movements (see Bhagwati, 1979). With the North being capital-abundant, the marginal productivity of capital is low compared to that of the South, where capital is scarce (assuming superior Northern technology does not outweigh the strength of diminishing returns). With the rental rate equal to the marginal product of capital, capital will move from the North to the South, increasing per capita production in the South and reducing it in the North, increasing production in the world as a whole, and increasing per capita income (taking into account payments to capital) in both regions.

In fact, capital does not move as much from the North to the South as is suggested by this theory, as shown by the fact that for substantial periods of time there has been a reverse transfer of financial resources from the South to the North. To the extent that capital does move from the North to the South, it moves to a small group of countries. Modifications of the neoclassical model provide reasons why this may be so. If we replace the assumption of diminishing returns by increasing returns, introducing external economies, or internal scale economies with imperfection competition, it is no longer the case that the capital-abundant North has a lower return to capital than the South, and capital will move in the reverse direction (Lucas, 1990). If lenders do not know exactly what borrowers do with borrowed funds and can only observe outcomes of their activity, while borrowers know what they are doing, we have the problem of asymmetric information, and lenders will require collateral to ensure that borrowers do not willfully default. The implication of this is that borrowers in rich countries who have higher initial endowments of capital will be able to borrow more than those in poor countries, because they can put up collateral to overcome moral hazard problems, while borrowers in poor countries are less able to do so (Gertler and Rogoff, 1990; Matsuyama, 2004). This may imply that capital will flow from poor to rich countries, making rich countries even richer, resulting in a process of uneven development.
one-good assumption, and allow the North and South to trade in different products, capital flows from the North to the South can turn the terms of trade against the South (Singer, 1950), and lead to Southern immiserization of the type discussed by Bhagwati (1956). These insights have been incorporated into some North–South models. Burgstaller and Saavedra-Rivano (1984) introduce capital mobility into Findlay’s model and find that with capital mobility (compared to the case without it) Northern per capita income will be higher, but Southern income per worker will be lower (because of the payments that have to be made for foreign capital), and the terms of trade of the South deteriorates. Relative Southern employment will also fall if the Northern propensity to spend on the Southern good is less than that out of Southern profits, since there is a reduction in the demand of the Southern good due to the redistribution of income caused by capital mobility. In Blecker’s (1996) model in which Northern growth is driven by aggregate demand, greater capital mobility leads to uneven development in the sense of an increase in the stock of Northern capital to Southern capital, due to Southern terms-of-trade deterioration. Such outcomes, however, depend on the assumption that capital flows increase the production of the Southern good, thereby resulting in a deterioration of its relative price. If, as pointed out in some empirical studies, capital flows in the form of foreign direct investment feature deep integration, involving the production of typically Northern goods in the South, capital flows could lead to more even development patterns as the North loses markets to the South and experiences greater excess capacity and unemployment and lower growth, and lower profits encourage more capital flows to the South (see Dutt, 1996). TNCs bring in not only capital, but also technology and exporting capabilities, which make these effects more likely. Such outcomes are more supportive of those who fear Northern stagnation due to capital outflows.

However, North–South capital flows may not bring about Southern development because they may not be productively invested. Borrowing by the South in the past, often due to loan-pushing by Northern banks, led to the accumulation of Southern debt but often went right back to the North through capital flight by corrupt elites (Darity and Horn, 1988). Debt-ridden Southern countries are then forced to make interest payments and repay loans by running current account surpluses and reducing growth. Default can threaten financial stability in the North as well. Portfolio flows and bank loans often finance short-term investment in the South in stock markets and real estate, leading to speculative bubbles which, when they burst, bring about sharp capital outflows, currency crises and macroeconomic contraction (see Stiglitz, 2002).
**Labor migration**

The standard one-good neoclassical model with diminishing returns and perfect competition discussed earlier also implies convergence of per capita production due to the movement of people from the labor-abundant South to the labor-scarce North. Such an approach, which implies that migration leads to a reduction in the Northern wage, can be used to explain some of the opposition to immigration in the North. It can also be used to explain income losses to those who are left behind in the South, because of the disappearance of the surplus produced by the workers who migrate (see Bhagwati, 1979).

Such a simple framework does not capture the complexities of real-world migration, however. Countries of the North heavily restrict the migration of unskilled workers from the South, allowing mainly the legal immigration of skilled workers. If one distinguishes between skilled and unskilled workers, and introduces scale economies and imperfect competition, uneven development due to brain drain from the South may occur. For instance, if unskilled workers and intermediate services are used in the production of the final good under conditions of constant returns to scale and perfect competition, while non-traded intermediate services are differentiated products, each produced by a monopolistic competitor under conditions of increasing returns with skilled labor as the only factor of production, the migration of skilled labor from the South to the North reduces the number of intermediate goods, quantity of the final good produced, and per capita income in the South, and has an opposite effect in the North, implying uneven development (Dutt, 2005). The endogenization of skilled labor supply, due to tax-financed government educational spending or the decision by workers to accumulate human capital, into new growth theory models can also imply divergence growth patterns (Haque and Kim, 1995; Wong and Yip, 1999).

**Technology transfers**

If knowledge is something that all countries can share, it may be supposed that the South will eventually catch up with the North. Lucas (2000) develops a simulation model in which countries take off in sequence (depending on their internal conditions) and in which latecomers grow faster than leaders because they have access to technology and policy experiences of the latter. His model implies that although international inequality increases initially, ‘sooner or later everyone will join the industrial revolution . . . economies will grow at the rate common to the wealthiest economies, and . . . percent differences in income levels will disappear’ (Lucas, 2000, p. 166).

Such an outcome, however, does not occur even in standard neoclassical models of technology transfer. In Krugman’s (1979 [1990]) model of
product innovation which formalizes the product life cycle approach, and which assumes that Northern ‘new’ goods become ‘old’ goods produced by the South at a constant rate of ‘radioactive’ decay, the ratio of Northern to Southern income stabilizes in equilibrium, but does not become unity. In models of technology transfer involving process innovation, where the change in Southern productivity depends positively on the North–South productivity gap (reflecting more opportunities for transfer), a narrowing of the technology gap occurs if there is initially a large gap, but complete technological catch-up does not occur unless transfers occur at an infinite rate.

Models in which rates of technology transfer are either constant or monotonically related to the relative technology gap (a higher technology gap leading to a faster rate of technology transfer) may not capture the realities of the process of technology transfer. Since much technological knowledge is tacit, and requires constant modification and adaptation, the process of transferring technology is not so very different from that of innovation, and successful technology transfer requires the development of some amount of social and technological capability of the South (Abramovitz, 1986; Bell and Pavitt, 1993). This can have a number of consequences. First, if we measure (relative) technological capability by the South–North productivity gap, a large productivity gap may increase the potential for technology transfer, but reduce the ability of the South to do so successfully. In this case, even standard models will imply convergence of the technology gap if the gap is not too large, but divergence if the North and South are too far apart (Verspagen, 1991), which may explain the formation of convergence clubs of rich countries and the exclusion of many poor countries. Second, technology transfer may be linked to other aspects of North–South relations. To the extent that technological capability is enhanced by the presence of TNCs, FDI can speed up technology transfer. However, an adverse impact of indigenous technological development cannot be ruled out if competition between TNCs and domestic firms hurt the latter. Trade liberalization and its consequence on the composition of production may, as mentioned earlier, have adverse effects on learning by doing and slow down the accumulation of technological capability (see Van der Klundert and Smulders, 1996).

In addition to these issues, the protection of intellectual property rights (IPRs) can serve as a barrier to technology transfers. Indeed, the protection of IPRs has been a major source of conflict between the North and the South, as reflected in the Uruguay Round discussions of the WTO. Since most technological innovation occurs in the North, it is not surprising that the North, especially its innovating firms, has an interest in protecting IPRs internationally, while the South, which is more interested in the speedy
diffusion of technology from the North to the South, is against such protection. These insights are confirmed by simple partial equilibrium models of product and process innovations in terms of consumer and producer surpluses in the North and South (see Chin and Grossman, 1990; Deardorff, 1992).

Although the interests of the North and South may be in conflict, it is often argued that the international IPR protection may be good for the world as a whole, because in its absence there would be insufficient innovation, since innovating firms would receive smaller rewards for their research and development activity and therefore conduct less of it. However, the partial equilibrium models imply that the effect of international IPR protection on world welfare (measured by the total surplus accruing to the North and the South) is ambiguous: if the North is a large part of the economy and protects IPRs internally, there will be sufficient innovation in the North to make world welfare higher without international IPR protection than with it. This result can carry over to a dynamic general equilibrium setting: in Grossman and Helpman’s (1991) model weaker protection of IPRs will not only speed up technology transfers, but also accelerate innovation in the North, because Northern resources are devoted more to innovation, rather than to production, which increasingly moves to the South.

While it is often the case that tighter international IPR protection slows down technology transfer, this may not be true under all circumstances. For instance, with IPR protection TNCs may more willing transfer better technology abroad, while without it they are likely to hold it back, fearing imitation (see Lai, 1998). However, Glass and Saggi (2002) find that although stronger IPR protection makes TNCs and Northern firms safer from imitation, the greater difficulty in imitation results in more resources being absorbed in imitative activity which reduces FDI, as well as Northern innovation.

**Environmental issues**

The literature on the environmental Kuznets curve suggests that the rich North is able to deal better with environmental problems than the South, where poverty, population growth and lax pollution control have an adverse environmental effect which can in turn have an adverse effect on growth and other development indicators, such as health. North–South interactions through trade and capital movements may exacerbate such problems, if they lead to the movement of the production of relatively dirty and resource-intensive goods from the North to the South, for instance, because of less-stringent environmental regulation; although, if capital-intensive Northern goods are more pollution-intensive, this may not necessarily occur (Copeland and Taylor, 2003).
Further North–South interactions regarding the environment occur because of the global public goods nature of several environmental issues, such as global warming, depletion of the ozone layer and loss of biodiversity (Sandler, 1997). The North blames the South for its increasing contribution to damaging the global environment, because of high population growth and pollution-intensive growth, while the South blames the North because of its high per capita contribution to this damage. The solution of global environmental problems requires such conflicts to be resolved by mutually beneficial international agreements between the North and the South which take into account the poverty of the South.

Conclusion
The North–South approach can be criticized for downplaying the differences within rich and poor countries, and the possibility that poor countries may grow and join the ranks of the rich. However, two-region models have been extended to include a third, consisting of newly industrialized countries, to explore the causes of its growth and to analyze whether its growth results in the end of uneven development or the exacerbation of the gap between the rest of the South and the North (see Dutt, 1992b). Moreover, multi-region models have been developed to show how the countries can endogenously sort themselves into Northern and Southern groups (see Matsuyama, 1996, 2004).

This brief review has confined attention to only a few of the major North–South issues, not examining other areas of interaction (for instance, due to cultural and political influences), conflict (such as those over labor standards) and cooperation (related to the existence of global public goods such as global health and security). Nevertheless, it suggests that the North–South framework remains a useful way of looking at many important and interesting development issues.

Notes
1. I am grateful to Jaime Ros for his helpful comments on an earlier draft of this chapter.
2. More details can be found in the other chapters in Part VI of this Handbook.

References


PART VII

DISTRIBUTION AND POVERTY
There is a rising level of concern about income inequality and poverty in developing countries. The development process has brought rising inequality within many countries and has also increased the gap between the poorest and richest countries. All this has led to an increasingly vocal demand for a more equitable style of development. Certainly no one would wish for more inequitable development, but it is still important to stop and ask oneself what distribution measures really measure, and what comparisons of these measures over time or across countries tell us. In particular, what do trends in inequality mean in developing countries? How are they affected by the development process? Are they something that governments in those countries can affect and should worry about?

I am going to address these questions by first considering at the individual or family level what should be included in a distribution measure. Once one has a satisfactory measure at the individual level, one needs a number or index which aggregates or summarizes all the individual distribution data across some reference group. The best-known is the Gini coefficient, but there are many others whose properties will be discussed further on. But in addition to this one has to decide the group over which to aggregate the individual distribution data. What group should that be? Should it be a country, a region within a country, the whole world, or should it be confined to a group such as an age cohort within a country? Responses to this question will have a big impact. My focus in this chapter is going to be on the distribution within countries rather than for the world as a whole, or between countries, because I believe that the national state is the basic economic and administrative unit which is able to influence the distributions that matter to its citizens.

Next I want to consider briefly some problems in the interpretation of the aggregate measures. What do changes over time or comparisons across countries tell us? Finally I will discuss several of the determinants of the distribution and the ways that they are likely to affect the distribution during the development process.
Measuring inequality at the individual or family level

Even though in some ultimate sense ‘fairness’ may imply an equal distribution of welfare across a population, I am going to limit my discussion to the distribution of goods or income produced by an economy. I am going to assume that a more equal distribution of goods or income is in the economic sense a fairer distribution, even though I have no way of knowing or of proving that a fairer economic distribution is also a more equitable distribution of welfare. I am bringing welfare into the discussion here because it will guide me in my choice of what to measure and, more importantly, what the reference group should be for my aggregate distribution statistics.

All of my statistics on distribution come from household surveys which are now available for a large number of countries. These measures can be based on income or on consumption. The World Bank favors consumption because it believes that it is both more accurately measured and a better proxy for lifetime income than observed income. If one thinks that people are concerned with their lifetime income rather than their income in any particular year, then the distribution measure should also be based on lifetime income. While I cannot measure lifetime income, it is likely that consumption is a better proxy than observed income in any single year. Families smooth out temporary fluctuations in income by either running down saving during recessions or adding to their assets when they have a temporary windfall. In some situations consumption is not a perfect measure, however. Suppose one is interested in the relationship between stabilization or recession and the distribution of welfare. Poor families may maintain their consumption by distress selling of assets or by borrowing at high interest rates. In either case the distribution of welfare is made less equal, even though the distribution of consumption may not show that, or may understate the rise in welfare inequality. If one is interested in short-run fluctuations and their distributional impact, it is probably better to use an income-based distribution measure. For longer-run questions, the consumption-based measure is better.

Every family receives income in kind. While the statisticians have learned how to include the imputed value of home-grown food or the rent on owned houses in the total value of family consumption, they do not generally include as consumption the value of government services such as public health care or education. These can have a fairly sizable effect on measures of the distribution.

For labor market questions, the distribution of earnings is relevant. However, since welfare is based on consumption, and since the family is the social unit which converts the earnings of some of its members into consumption for all its members, one should use the distribution of income among families. I can convert that into an individual distribution for
poverty analysis by dividing total family income or consumption by the number of family members.\textsuperscript{2}

**Aggregation problems**

I now want to address several aggregation questions.\textsuperscript{3} The first is what group the distribution should be based on. Using welfare as our criterion for answering this question, the group should be the one to which the individual belongs and across which he or she makes comparisons. That turns out to have important implications. It is customary to calculate distributions at the national level at a point in time. That is reasonable if national relative position is an important consideration to individuals in a society. But in large countries it may well be that a regional distribution is more consistent with how people see their own welfare. If inequality indexes are thought to be good proxies for social tension or the likelihood of populist demands for redistribution, it is important that the aggregation reflects the interpersonal comparisons made by people in a society.\textsuperscript{4}

Another possible sub-national group classification is by age or cohort. Suppose that people's primary concern is their relative position in their own age cohort. Young people presumably know that there is a positive age–earnings profile. Incomes rise with age and experience. It that is so, it would seem reasonable that purely age-based inequality would be less significant as a source of social tension than inequality within an age cohort. To take this to an extreme, suppose the individuals are concerned only with intra-cohort inequality. Now imagine an economy in which each age cohort has exactly the same expected lifetime income and age–income profile. In such an economy, there would be no income inequality provided that we were able to base the distribution measure on permanent income. That would not be true if the measure was based on consumption, if consumption, like earnings, has an inverted U-shape, rising through most of an individual's working life, and then falling in retirement. In that case any aggregate distribution measure based on observed income or consumption at a point in time will show a good deal of inequality even when there is no intra- or inter-cohort inequality. In any year the younger members of the society will be earning or consuming less than their elders. Observed inequality will be higher, the steeper the age–consumption profile, the greater the return to experience or the more credit-constrained the younger members of the society are. Apparent inequality will also be higher, the faster the rate of population growth, not because younger cohorts have lower expected lifetime incomes, but simply because there will be more of them at the bottom of the income pyramid at any point in time.

Statistically in this hypothetical economy there is inequality between the old and the young. But the significance or meaning attached to that
inequality depends on the attitudes or preferences of the members of that society. If the comparator group for each member is his or her own age cohort (or their own region of the country) then a national distribution of income or consumption will be a misleading indicator of what one might call ‘socially relevant’ inequality.

Aggregate measures of inequality

Once a choice has been made of what to measure, one is left with the problem of how to aggregate the information into a small number of understandable statistics. There are many ways of doing this, each of which implicitly assigns certain weights to each individual observation. The most straightforward statistic is the shares of total income going to different shares of the population such as the poor and the rich, or the bottom and top quintiles.

Perhaps the best-known measure is the Gini coefficient, now used around the world to compare distributions between countries or within countries over time. The Gini coefficient is the ratio of the area between a 45º line and a curve called the Lorenz curve showing the shares of total income accruing to different shares of the population. The Gini varies between zero and one. If there is perfect equality, each individual or family earns the same income, and the actual Lorenz curve overlays the 45º line. Here the Gini coefficient is zero since the gap between the lines is zero. Maximum inequality is when one individual or family owns all the income. In this case the Lorenz curve has a backward L-shape, and the Gini is equal to one since the gap between the Lorenz curve and the 45º line is equal to the entire area under the 45º line.

There are a number of alternative aggregate measures of the distribution found in the literature. One is the coefficient of variation (CV). It is defined as the square root of the variance of income divided by the mean to make it, like the Gini, independent of the level of income. The CV has the somewhat counterintuitive property that it gives equal weight to transfers between individuals regardless of their incomes. That is, inequality changes by the same amount if changes in income occur close to or further away from the mean, provided only that the changes are of the same size:

\[ CV = \frac{\text{Var}^{1/2}}{\mu} \]

If one wishes to put more weight on what happens at the bottom of the distribution, one can transform income into the log of income and then take the standard deviation of this transformed income measure. Since the log of income falls rapidly for small values of income and since the standard deviation uses the square of the deviation from the mean, the
transformation will increase this measure of inequality when there are many people with levels of income far below the mean. The measure is defined as:

\[ SDY = \left\{ \Sigma (\log \mu - \log X_i)^2/n \right\}^{1/2} \]

**Distribution comparisons among countries and over time**

Two of the main uses of distribution statistics are to compare distributions across countries or to measure changes in a single country over time. Deininger and Squire (1996) at the World Bank recently collected a set of comparable national Gini coefficients for 108 countries around the world. Many of these countries have estimates going back over 30 years. Table 47.1 gives their estimates of the median Gini by region and decade.

Other than the very large differences between Latin America and most other regions of the world, what is striking is that there is no convergence over time especially in Latin America. Latin inequality fell slightly in the 1960s, but then it reversed after 1970. Meanwhile there was some improvement in the Middle East, Africa and South Asia, and worsening in Eastern Europe in the 1990s after the fall of communism. In the developed countries inequality has remained roughly constant at a low level. In the other developing regions where inequality was once high, it has fallen quite sharply. Only in Latin America has inequality remained near its high initial level.

The differences between regional Gini coefficients translate into large differences in the amount or share of income going to the rich and poor. In the 1990s on average in Latin America the top 5 percent of the population received 25 percent of total income while the bottom 30 percent got only 7.5 percent. In South-East Asia the top 5 percent received only 16 percent of income while the bottom 30 percent got 12.2 percent. The comparable

**Table 47.1 Decadal median Gini coefficients by region**

<table>
<thead>
<tr>
<th>Region</th>
<th>1960s</th>
<th>1970s</th>
<th>1980s</th>
<th>1990s</th>
</tr>
</thead>
<tbody>
<tr>
<td>E Europe</td>
<td>25.1</td>
<td>24.6</td>
<td>25.0</td>
<td>28.9</td>
</tr>
<tr>
<td>S. Asia</td>
<td>36.2</td>
<td>33.9</td>
<td>35.0</td>
<td>31.9</td>
</tr>
<tr>
<td>OECD &amp; high-income</td>
<td>35.0</td>
<td>34.8</td>
<td>33.2</td>
<td>33.7</td>
</tr>
<tr>
<td>E. Asia &amp; Pacific</td>
<td>37.4</td>
<td>39.9</td>
<td>38.7</td>
<td>38.1</td>
</tr>
<tr>
<td>M. East &amp; N. Africa</td>
<td>41.4</td>
<td>41.9</td>
<td>40.5</td>
<td>38.0</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>49.9</td>
<td>48.2</td>
<td>43.5</td>
<td>46.9</td>
</tr>
<tr>
<td>Latin America</td>
<td>53.2</td>
<td>49.1</td>
<td>49.7</td>
<td>49.3</td>
</tr>
</tbody>
</table>

figures for Africa show the top 5 percent got 24 percent, the bottom 30 percent got 10.1 percent. In the developed countries the top 5 percent got only 13 percent of total income while the poorest 30 percent got 12.7 percent. These shares imply that in Latin America the average income of the richest 5 percent is 20 times that of the poorest 30 percent. In South-East Asia the richest 5 percent have average incomes only 7.9 times that of the poorest 30 percent.

**Problems in comparisons across countries**

One of the problems with interpreting differences in any of the aggregate measures between countries is that they force one to make value judgments about the weights that one gives to welfare or income at different income levels. One can see this most easily for the Gini coefficient. Lorenz curves for different countries may cross as may the curves for the same country over time. The Gini will say unambiguously which country has a more equal distribution. But even assuming that one thinks that more equal distributions are preferable, can one really be sure that one prefers the more equal distribution in all cases? Take the case of two countries with the same mean income. Suppose that in country A there are relatively few poor people because the middle class has been taxed to support a generous safety net. But there are a lot of rich people. In country B the poor have a much smaller share of income than in country A, but there is a big middle class and a relatively small group of the rich. Here the Lorenz curves of the two countries cross. The curve for country A lies above that of country B at the bottom of the distribution, but below it at the top of the distribution. If one is really concerned about the welfare of the poor, one may prefer the distribution of country A, even though its Gini is larger (more unequal) than the Gini of country B.

Large countries are likely to be more heterogeneous and have significant backward regions and significant differences in regional incomes. That will make their measured income distribution less equal than in smaller countries. Similarly, national distributions are generally less equal than purely urban distributions because on average, rural incomes are lower than urban. Countries with big indigenous populations are also likely to have more unequal distributions as will be seen below. In all these cases the reason for higher inequality is the existence of a large group within a national border which does not fully participate in the process which generates economic growth. Note that none of this would be particularly relevant to cross-country comparisons if comparator groups were truly national. As mentioned above, welfare statements about distributions are based on the position of individuals relative to others in their comparator group. If that group is local, or limited to members of an indigenous group
or the rural population, then comparisons of national distributions between countries may have misleading welfare implications.

Comparisons for the same country over time
Comparisons over time in single countries may have the same ambiguities and problems of interpretation that I noted above in the discussion of crossing Lorenz curves. They have additional interpretation problems in economies in which the population is growing. In that situation, by definition, over time the populations whose distributions are being measured are not the same and it is important to make a distinction between what I will call the base period population and the later group to which it will be compared. Apart from any questions about comparator groups, changes in the observed distribution will be affected by the earnings of the new entrants. Apparent earnings of groups such as the rich or the poor will appear to rise or fall even if the earnings of the base period population do not change. Distribution statistics are based on shares of total income received by different shares of the population. When there is an increase in the income-earning population, both income shares and population shares are affected. Total and average incomes are affected by where these new entrants come into the income pyramid.

When studies of the distribution report statistics on the incomes or income shares of the rich or the poor, they mean the group at the top or the bottom of the income pyramid at different moments. They do not mean the base period rich or poor. Where population growth is rapid or where comparisons are made over long periods of time, the distinction between the distribution or growth rate of income of the base period population and the observed population is significant (see Morley, 1981).

Determinants of the distribution
I turn now to a short discussion of the determinants of the distribution of income. One should distinguish between what I will call the primary or earnings distribution and the family distribution. This first is the distribution of income to the owners of the factors of production that produce it. The family distribution is derived from the primary distribution according to what factors of production each family owns and what each of those factors earns in economic activity. Economic factors mainly affect the primary distribution while demographic factors play a key role in the corresponding family distribution.

A useful abstraction here is the factor market where the demand and supply of each factor determines its earnings. For simplicity consider four factors, skilled and unskilled labor, capital and land. Like other economic markets, prices are determined in factor markets by the interaction of the...
supply of each factor and the demand for factor services as well as institutional factors such as the minimum wage. Markets clear at a price at which someone is willing to use the available supply of each of the factors. That set of market-clearing factor prices determines the relative price of skilled and unskilled labor and the rate of return on capital and land. In the labor market, it may well be the case that demand is so low that many are forced to work part-time in the informal sector or are unemployed altogether because the legally prescribed minimum wage in the formal sector exceeds the marginal product of the fully employed labor force.

On the demand side, economic growth shifts out the demand for each of the factors, which tends to raise each of their prices. What happens to relative factor prices depends on the nature of the growth process. If it is skill-intensive, skill differentials widen. If it is led by sectors such as agriculture which use mainly unskilled labor, the reverse should occur. Structural reforms such as trade liberalization also affect factor demand by changing relative goods prices and the composition of output.

The supply side is critical to understanding the dynamics of the process. In the short run the supply of factors is fixed because each of the factors is a stock which produces a flow of services. That stock can change, but that takes time. Thus in the short run there is a fixed stock of factors determined by past investment decisions, and a set of factor demand curves. The two together produce the short-run set of factor prices. While this process sets the valuation of the factors of production, one needs to know the distribution of ownership of the factors of production in order to determine the primary earnings distribution, since it is the owner that receives the payment for the factor services. The primary distribution in the short run is determined jointly by the relative factor prices that come from the factor market and by the pattern of ownership of the factors of production.

The dynamics of the process come from investment. Investment increases the stock of physical capital, while the education system produces graduates who enter the labor force and change the supply of both skilled and unskilled labor. Training and experience also increase the stock of skilled labor or what I will call from here on, ‘human capital’. Migration also changes the position of factor supply curves, and that could be for either unskilled or skilled labor.

It is important to remember that investment takes time and that the amount that can be added to the stock of any of the factors is relatively small over any short-run time period. Thus, in any period as short as say a year, it is impossible to shift the factor supply curves out by more than 3–5 percent. That means that in the short run, it is changes that come from the demand side that are most likely to explain observed changes in the primary
distribution. Over time these stocks change because of additions to the physical capital stock, education, retirements and so on.

An important feature of the distribution process is the dynamic feedback between factor prices and changes in factor supplies through investment. Relative prices or returns on physical capital determined in the factor market affect investment and education decisions. If there is a rise in the rate of return on capital, investment in physical capital increases. Similarly if there is an increase in the wage differential between skilled and unskilled labor, or between university and high school graduates, that will tend to increase the demand for university education. Fewer people will enter the labor force after high school. More will continue on to earn a university degree. Over time, those decisions will increase the supply of educated labor in the labor force. If there were no changes on the demand side, this increase in the supply of skilled or educated labor should reduce the skill differential. One could thus think of the initial rise in the skill differential or the return to capital as a dynamic signal which sets in motion an investment process which eventually adjusts factor supplies and relative factor prices to a long-run equilibrium. By definition that equilibrium is one in which additions to supply are just sufficient to offset changes in the demand for factor services coming from economic growth and technological change.

From the point of view of the primary earnings distribution, a rising skill differential or profit rate which tends to increase inequality in the short run is also a signal which expands the supply of educated labor or physical capital in the long run. These long-run changes on the supply side may well reverse the short-run rise in inequality that induced them since they tend to drive down the rate of return to capital, both human and non-human.

There is thus an important distinction or ambiguity between the short-run and the long-run meaning of a rise in the skill differential or the rate of return to capital. In the short run an increase in either of these two is almost surely regressive. But in the long run, so long as the supply side reacts positively to these changes in the rate of return, the change could be progressive either because of upward mobility, or because the increase in physical capital drives down the rate of return and raises the productivity and the average wage of workers.

The ambiguity I am discussing here is a specific example of the dual function of income in a market system. On the one hand relative income determines the distribution at each point in time. Any relative increase in the income of the rich is regressive. But on the other hand income is also the signal or incentive by which economic agents are encouraged to change their behavior. A rise in the skill differential induces socially desirable education investment. Similarly a rise in profits induces investment and a shift of productive resources from less-desirable to more-desirable uses. One
makes a serious error of interpretation if one concentrates only on the short-run regressive effect of changing factor returns without taking into account progressive long-run supply responses.

**Inequality and economic growth**

In a classic article, Simon Kuznets empirically analysed the relationship between growth and equity, using the historical experience of England and the United States (Kuznets, 1955). He found that over the course of the nineteenth century, inequality rose as both countries grew, but somewhere around the beginning of the twentieth century the trend reversed and inequality began to decline. This non-linear, inverted U-shaped relationship between income and inequality has been labelled the Kuznets curve and has been the subject of a vast empirical literature looking for similar relationships in other countries and in international cross-sections between countries.

Kuznets’s rationale for the relationship is based on the important idea that growth starts somewhere specific, in either a region, a sector or a city. From that starting point its effects spread through a variety of linkages to the rest of the economy. In the first stage of the process, inequality rises. Later, as the effect of the growth stimulus spreads out in the population, inequality begins to decline.

In the Kuznets study, the growth process was the Industrial Revolution. Industrialization started in the urban centers of agricultural economies. Incomes in the small industrial sector were much higher than those in the agricultural sector. As industry expanded, inequality at first increased because the higher urban wages benefitted only a small fraction of the population. This is the stage when growth and inequality are positively related or when the Kuznets curve has a positive slope. At some point, when the urban sector got big enough, inequality trends were determined by those left behind in low-wage agriculture. At that point, further industrial urban growth began to reduce inequality because it reduced the population share in low-wage agriculture.

The point here is that the growth process starts somewhere specific, after which its effects spread by a variety of linkages to the rest of the economy. In the first stages of this process inequality is almost certain to rise. Later, as the effect of the growth stimulus spreads in the population, inequality will begin to decline. How long this takes to happen, or indeed whether the spread effect is big enough to offset the initial increase in inequality, depends on the strength of linkages. Linkages are the connections between other sectors or economic actors and the sector in which growth is occurring.

Linkages are a key determinant of the relationship between growth and inequality. They determine the ‘spread effect’ of growth in the economy.
The stronger they are, the faster and further the benefits of growth will spread out, and the more equitable growth is likely to be. This notion can be applied in a variety of important ways. In some economies there are big backward regions or indigenous populations which are only weakly or marginally connected to the modern, dynamic sector where growth is occurring. The areas themselves have a significant proportion of total population, which means that their relative income levels will have a noticeable effect on inequality. But for reasons that are not entirely understood, growth in the dynamic sectors does not induce much forward or backward linkage activity. As a result, when these countries grow, there is not much of a spread effect to their backward or poor regions. Growth under these conditions tends to be inequitable. One could say that in these countries inequality is high because of growth, in the sense that if the entire country had remained in the same conditions as its backward regions, inequality would be lower. Here, growth leaves behind significant proportions of the population. Growth always leaves some people behind, if one accepts the idea that growth starts in a particular sector or area of the country. The key thing here is that if the linkages are weak and the areas left behind are large, then the interval in which growth is inequitable is likely to be a long one.

For the same reason, inequality is likely to rise with growth in countries with large indigenous populations. Typically the links between indigenous people and the rest of the economy are weak. When growth occurs, it provides little stimulus to incomes of the indigenous. Conversely one could say that inequality is likely to be lower the smaller and more homogenous the economy. Small countries with homogenous populations are unlikely to have backward regions or groups which are disconnected from the modern economy. Examples are countries like Uruguay or Hong Kong where the bulk of the population lives in a small number of interconnected urban areas. When this sort of country grows, a greater share of the population benefits because most people are linked either directly or indirectly to the sector where the growth stimulus began.

Two examples illustrate the point I am making here. In Brazil growth was very rapid between 1960 and 1980, but it was centered in the south-east, both because of rapid industrialization in the area around Sao Paulo and also because of the expansion of the agricultural frontier. The north-east, which in 1960 contained 33 percent of the population, also grew, but far more slowly than the south and south-east. Partly as a result, the Gini for Brazil rose from 0.53 in 1960 to around 0.58 in 1980, one of the fastest increases in inequality observed anywhere. Similarly China has had a period of explosive growth since the mid-1980s. But that growth was concentrated on the coast, leaving the interior of the country relatively untouched. Between 1990 and 1999 average income in the coastal provinces
and in urban centers grew far faster than in the inland provinces and the rural sector. This was a major contributor to rising national inequality indexes. For the country as a whole the Gini increased from around 0.26 in the mid-1980s to 0.37 in 2000 (Kanbur and Zhang, 2005). Over one-half of that increase was caused by the increase in the gap between rural and urban incomes, and about one-third by the rising inland–coastal gap (Kanbur and Zhang, 2005).

Despite the predictions of Kuznets that at some point in all countries inequality should decline, there seem to be at least two features of growth under current conditions that may reverse this trend. First, modern growth has become increasingly skill- and capital-intensive. That has raised the return to both capital and skilled labor. Wage differentials for the educated have widened considerably and profits have gained relative to labor. Second, and this is more controversial, removing the barriers to capital mobility, an important part of the current generation of economic reforms, has increased the bargaining power of capital and its ability to extract wage concessions under the threat of moving to lower-cost developing countries. At the same time trade liberalization has lowered the prices of simple manufactures produced in developing countries. That has also put pressure on the wage differential in developed countries. All of this may make the world distribution more equal by narrowing the wage differential between developing and developed countries, but at the same time the within-country distributions that I have been discussing may well get less equal in both the advanced and the developing countries.

Governments can do a number of things to make the distribution more equal. The burden of taxes can be shifted toward the more affluent. Government can spend more or subsidize goods and services such as education and health care that benefit the poor. It can create a safety net financed out of general revenue to reduce or even eliminate extreme poverty. It can sponsor public works programs to absorb the unemployed and build useful infrastructure. But probably the two most important things that government can do are to maintain a sustainable and stable growth rate for the economy and to reduce the ranks of unskilled labor by effective programs of education and training.

Conclusions
In this chapter I have discussed how to measure the distribution, how distributions are likely to change as countries develop, and what government can do to make the distribution more equitable. I argued that the effect of growth on the distribution is determined jointly by the skill-intensity of growth, the structure of the economy and its factor markets and by how region- and factor-specific it is. The stronger the linkages between the
leading sector, factor or region and the rest of the economy, the more equitable growth will be.

I paid particular attention to several problems of interpretation of changes in the distribution as indicators of social welfare either within countries over time or between countries. To be a socially relevant measure, it is important that the distribution be defined over the appropriate comparator groups – those with whom people actually make welfare comparisons. The typical aggregation is by country. But distributions by regions or by age cohorts may be better indicators of welfare, even at the national level.

Comparisons of distributions over time are particularly troublesome. First, the populations are different and the lifetime distribution of any cohort is likely to be very different from the national distribution. Second, the distributions that are observed are short-run and any growth process that drives up skill differentials and profit rates will be judged regressive in the short run. But it is unclear how to evaluate this short-run rise in inequality if it generates a progressive factor supply response in the long run by increasing either the supply of skilled labor or employment-creating investment.

Notes
1. The author would like to thank the editors for comments on a previous version of this chapter.
2. This measure is called family income or consumption per capita
3. For a fuller discussion of aggregation problems and measures see Sen (1973).
4. The distinction I am making here is closely related to the notion of horizontal equity discussed by Ravallion (2004). National distribution statistics implicitly assume that welfare is invariant with respect to where income is earned, or in other words that welfare is unaffected by equivalent transfers among groups with the same income, such as the rural and urban poor. A trade reform which lowers the price of food may leave aggregate national distribution statistics unchanged, by helping the urban poor and harming the rural poor. That could exacerbate social tensions even if the national distribution of consumption became more equal.
5. One could cite a third example. Bourguignon and Morrisson (2002) showed that there was a very large increase in world inequality between 1820 and 1950, almost all of which was caused by an increase in the gap between the fast-growing industrial economies in the USA and Europe and those in the rest of the world. Weak linkages between these economies and those on what many have called the periphery were presumably responsible for this result.

Bibliography


Introduction
Until the end of the Cold War, most development economists were not particularly concerned with the distribution of income, but instead with understanding growth and reducing absolute poverty in the developing world. For one thing, Kuznets (1955) had suggested that a deterioration in the distribution of income might be the natural outcome of the early stages of development, as people begin the shift from low-productivity subsistence agriculture to high-productivity sectors. And mainstream economists’ starting assumption, rooted in the Smithian trade-off between efficiency and equity, was that in the other direction of causation, inequality resulting for example from increased security of property rights would enhance growth by encouraging investment and savings and creating a necessary incentive for individuals to work hard.²

But beginning in the 1990s, as panel data on changes in the distribution of income in developing countries became available, as mainstream development economists became more concerned with political economy analysis, and – perhaps – once the fall of the Berlin Wall liberated the mainstream from the taboo of Marxian analysis, economists became more interested in assessing the effects of income distribution on growth. In the last 15 years a major focus of new theoretical and empirical work has been the effects of income inequality on growth and other indicators of inequality in the developing world. Much of that work has been ably reviewed in major reports of the United Nations Development Programme (UNDP), the Inter-American Development Bank and the World Bank.³

Still there is no consensus among economists that income inequality matters, and little attention among development practitioners to policies to address inequality as opposed to growth and poverty reduction.⁴

Obviously if people care about their relative income status then *ipso facto* inequality matters. That they do, to some extent, has long been remarked; consider Adam Smith, who noted that a man to retain his dignity may in one society need enough income to buy a linen shirt, and Veblen (1970) who noted that the absolutely well-off worry about their status relative to the more absolutely well-off.⁵ Hirschman (1973) observed that people stuck in
a tunnel in a stopped lane of traffic eventually become deeply frustrated if the other lane, but not theirs, has been inexplicably (and presumably unfairly) moving – quite independent, to extend the metaphor, of the kind of car they drive. Easterlin (1995) noted that happiness (or subjective well-being, or utility, to use the economists’ term) varies directly with one’s own income and inversely with the income of others, that is, that relative as well as absolute income matters. He comes to that conclusion in part based on surveys of happiness within countries over time; the average level of happiness has not increased even where average incomes have increased substantially. It is possible in fact that inequality of income reduces the utility or happiness not only of the relatively poor but of the better-off, who may enjoy their own affluence less if others are visibly worse off.

In this review I focus, however, on the instrumental reasons why a highly unequal distribution of income matters in developing countries. I review a large body of work, primarily of economists, indicating that beyond some level inequality in developing countries matters because: (1) where markets are underdeveloped, high income inequality is likely to inhibit growth; (2) high income inequality can discourage the evolution of the economic and political institutions associated with accountable government (which in turn enable a market environment conducive to investment and growth); and (3) high income inequality can undermine the civic and social life that sustains effective collective decision-making, especially in multi-ethnic settings.

Theory and some empirical work suggest that inequality does not undermine growth directly. Instead it is the interaction of inequality with imperfect markets or with unaccountable or incompetent governments (increasingly labeled weak ‘institutions’ in the latest literature on growth – see for example Acemoglu et al., 2000) that harms growth. In addition, and conceptually different, inequality (that is high enough) may directly create conditions that lead to or exacerbate poor governance and thus poor economic policy, and/or weak social and economic institutions and thus ineffective implementation of stable and sound policies – reducing growth through the effect on economic, political and social institutions. Weak institutions broadly defined are increasingly viewed as the key cause of low growth in developing countries. Since weak markets, poor governance and underdeveloped institutions might be said to be the very characteristics that define a country as ‘developing’, it follows that inequality is a key factor in understanding the dynamics of growth and institutional development in the developing world.

The reader will note that I do not discuss the effect of growth on inequality, the subject of the Kuznets hypothesis, nor the evidence that inequality and growth may each be simultaneously affected (Lundberg and Squire, 2003), either similarly or differently, by still other economic and non-economic
variables such as inflation and increased access to education. Once panels of household data enabled analysis of changes in the distribution of income over time within countries, the existence of a stylized Kuznets effect was not supported by the evidence (for example, Deininger and Squire, 1996), almost certainly because so many other country-specific factors compound any fundamental relationship there might be.

**Effect of inequality on economic growth and poverty: theory and evidence**

Two stylized facts emerge from the growing literature on the effects of inequality on growth. First, the evidence suggests that inequality above some level is more likely to reduce growth. Second, theory and empirical work suggest that high levels of inequality are more likely to harm growth in developing than in developed countries.

Barro (2000), in a study of the determinants of growth, was among the first to report a structurally different relationship of inequality to growth in developing compared to developed countries. Across developed and developing countries combined, he found no clear effect of inequality on growth. However, dividing the sample into the two groups he found the relationship is structurally different. In higher-income developed countries inequality may indeed be associated with higher growth (as often referred to in contrasting the USA and countries of Western Europe). Below a certain income level (about $2000 US 1985 dollars – equivalent to about US $3200 in 2000 dollars), higher income inequality is associated with lower growth. (The simple relationship is illustrated for developed and developing countries in Figure 48.1.) Cornia et al. (2004), using data from a more comprehensive set of household surveys, tested the relationship between changes in inequality and growth over almost four decades for 25 countries. They report a positive effect on growth as the Gini coefficient increases from very low levels (from the .15 typical say of subsistence economies and of the former socialist economies to .30) and a negative effect as the Gini coefficient rises from .45 (typical in Latin America and sub-Saharan Africa) to higher levels.8

The specific thresholds should not be taken too seriously, given poor measurement particularly of the distribution of income. However, they allow for a rough assessment of how widespread across people and countries in the developing world the resulting vulnerability might be. The critical thresholds of a Gini at or above .45 and income per capita at or below $3200 affect a significant number of countries and people in the developing world. Virtually no developing or transitional economies have income Gini coefficients below .30, though India and China did at about that level for much of the post-World War II period until the 1990s. About 15 percent of the population of the developing world currently lives in countries (33
Figure 48.1  Inequality and per capita income growth in developing and rich countries, 1970–2000
countries) with reported Gini coefficients of .45 or higher and per capita income below $3200 (in 2000 dollars), mostly in Latin America and sub-Saharan Africa. But that percentage mounts to 40 percent if China, whose reported 2003 Gini coefficient was 44.9, is included, and rises further to 44 percent if Brazil, whose per capita income now just exceeds $3200, is included. Other countries with per capita income below the Barro threshold where the income Gini has risen in the last 15 years and is now above .40 are Bangladesh and Pakistan. In India and Vietnam, inequality has also risen rapidly since the 1990s but reported Gini coefficients are still below .40.9

These findings are broadly consistent with theory. Why might some level of inequality enhance growth? First, inequality can be too low, as when it was imposed in state-managed economies where planning and controls replaced price and other market signals, encouraging ‘shirking’ and free-riding. A certain degree of inequality may be necessary to permit the incentives that induce individuals to work hard, innovate and undertake risky but productive investment projects, resulting in higher output and productivity, and therefore higher average incomes and growth rates. (For economists, these incentive effects are the backbone of the moral hazard argument against tax-financed distribution; Okun, 1975). Second, some concentration of income could encourage growth if high rates of saving enable more investment, and if savings rates are greater where income is concentrated in the hands of the rich whose marginal propensity to save is higher than that of the poor (Galenson and Leibenstein, 1955; Kaldor, 1961). A related idea is that investments in infrastructure and industry critical to development are large and indivisible; in the absence of well-functioning capital markets, wealth and income need to be highly concentrated to generate the minimum required resources to undertake new investment projects.10 (Recent ‘endogenous’ models of growth, however, rely much more heavily on the incentive effects of institutions and policy than on high savings and investment as the keys to sustained growth.)

The incentive effects of inequality can be thought of as the outcome of ‘constructive’ inequality, that is, income inequality that reflects solely differences in individuals’ responses to equal incentives or opportunities, and is thus consistent with efficient resource allocation.11 In contrast would be ‘destructive’ inequality, reflecting inefficient privileges for the rich, social and economic discrimination which reduces incentives for effort, investment and innovation by some groups, and in general reduced potential for productive contributions of the already poor. In a kind of tautology, destructive inequality can be defined as that inequality which results in lower, rather than higher economic growth (Birdsall, 2001).

The idea of destructive effects of inequality is consistent with the empirical evidence noted above of lower growth at very high measured levels of
inequality. Theory suggests that inequality is also more likely to be destructive in developing countries (as Figure 48.1 suggests). The remainder of this section sets out why that is likely to be the case. In summary it is because inequality tends to undermine growth when it combines with or interacts with weak markets and poor government policy. In general in developing compared to developed countries, financial and other markets are less complete and public policy is less effective in addressing market failures and imperfections.

**Imperfect credit and other markets**

Benabou (1996) and Aghion et al. (1999) develop models in which inequality exacerbates the effect of capital and other market failures on growth. When creditworthy borrowers cannot borrow because they lack collateral to comfort lenders (given imperfect information, a market failure in itself), then their lack of income or wealth limits their ability to invest. In addition, given limited liability (the borrower cannot repay more than his or her net worth), borrowers with less wealth have less incentive to exert effort to ensure success of an investment since they must pay lenders a higher portion of their returns (a moral hazard effect). In this case redistributing wealth has no adverse incentive effects – on the contrary it creates a positive incentive – and will be growth-enhancing. Weak or non-existent insurance markets will also force those without assets to bypass high-return projects. Galor and Zeira (1993) and earlier Loury (1981) suggest that the distribution of wealth affects output due to the indivisibility of investments in human capital. When it is difficult to borrow, lack of liquidity limits investments in human capital despite prospective high returns; this obviously affects the poor but may also affect the large majority of middle-income people in developing countries with a high concentration of income at the top of the income distribution. Birdsall et al. (1998) note that even where the poor are credit-constrained, they can exploit an increase in the return to potential new investment (in education or their own farm or business) by increasing their work effort. They will do so as long as the returns to their labor are adequate – as was the case in Korea and Taiwan in much of the post-war twentieth century. If labor markets are functioning well, and returns to education or other investments are rising, the credit market may not matter as much. Or in those countries, lower overall inequality of wealth, income and land (well below the Gini of 0.45 on the distribution of income), may have minimized the negative effect on growth of an interaction between inequality and weak markets.¹²

Obviously weaknesses in capital markets are greater in developing countries, as are compensatory policies such as enforcement of creditor rights. They are also more likely the lower average income is and the higher the
proportion of poor people, making it difficult to distinguish empirically between the negative effect of inequality per se (whether of income, wealth, education or land) interacting with weak markets from the negative effects of high rates of poverty. In any event, whether or not because capital markets are weaker and more people are poorer, it is not surprising that inequality undermines growth in developing countries though not necessarily in developed countries.

In these models, it is not actually income inequality but inequality of financial wealth or other assets that interacts with weak capital markets to reduce growth. (Only recently has household level data on financial wealth of reasonable comparability across countries become available.) But increasing evidence suggests that other assets – land and education – tell the same story. Latin America still appears to bear the costs of its historic land inequality. Carter and Coles (1998) show that concentration of land ownership is associated over long subsequent periods with concentration of income, even in countries where the economic relevance of agriculture has declined. Birdsall and Londono (1997) show that across countries inequality in the distribution of education reduces growth, and that once inequality of land and education are accounted for, inequality of income washes out as a factor affecting growth. In that respect, market economies in Latin America compared to East Asia, discussed below, do not operate differently – it is just that they operate in a context of high concentration of land and education.

Ineffective or corrupt institutions of the state and resultant poor public policy

As with weak markets, weak governments and poor public policy are likely to exacerbate the effects of inequality (of income, assets, education and so on) on growth. Behrman et al. (2000) show that differences across countries in social mobility, measured by differences in the effect of parents’ income and education on children’s education, are robustly and systematically affected by differences in two factors: public spending on primary education and the depth of financial markets. Repressed interest rates and directed credit programs that end up limiting access to credit except for privileged insiders worsen the effect of inherently imperfect capital markets on growth. Lack of adequate public spending on basic health and education means that public policy is not correcting for the inherent inability of markets alone to compensate for differences across households in endowments of all kinds. Growth is then lower than it could be since aggregate accumulation of human capital is reduced.

If income inequality interacting with poor policy reduces growth, then it is implicated in reduced poverty reduction – given that empirically, growth
has seemed necessary if not sufficient for reducing poverty, and since whatever growth occurs will help the poor less in an accounting sense the less equal the distribution of income (Ravallion, 1997, 2001). There may also be a more substantive link of inequality to the persistence of poverty where state institutions and government policy fail to ensure equal opportunities for the poor, even when there is income growth on average. Birdsall and Londono (1997) report that across countries in the period 1960 to 1990 greater land and education inequality reduced the income growth of the poorest quintile about twice as much as they reduced average income growth for all quintiles. In the extreme, unequal distribution of land may cut off altogether the usual effect of growth in agriculture on reduction of rural poverty. Some evidence suggests that agricultural growth in Latin America in the 1970s and 1980s failed to reduce poverty at all (De Janvry and Sadoulet, 2000), as large landowners captured most of the benefits. In contrast, in Indonesia, where small farmers provide the bulk of agricultural production, growth was good for the rural poor even in the days of Sukarno, and still better in the days of Suharto (Timmer, 2006a, 2006b).

**Political instability and social conflict**

Initial theorizing put any negative effect of inequality on growth not on its interaction with weak markets or poor public policy, but through a direct effect in the political sphere, as Benabou put it: ‘where asset markets are complete and distributional effects arise solely from the balance of power in the political system’. Economists suggested that higher inequality causes lower growth because the median voter, who is relatively poorer where inequality is high, votes for inefficient redistribution financed by growth-reducing higher taxes (Persson and Tabellini, 1994; Alesina and Rodrik, 1994). Their cross-country tests were not, however, convincing. Moreover, the median voter theory did not square with anecdotal evidence that policies in unequal countries are often shaped not by the relatively poor median voter (even where there is democracy) but by a more politically influential elite, and with lack of any evidence that redistributive policies, measured in terms of the marginal tax rate, are associated with lower growth (Easterly and Rebelo, 1993).

An alternative political explanation blames political instability on ‘social discontent’ (associated with inequality among other things) (Alesina and Perotti, 1996). Socio-political differences that reduce the security of property rights and the expected return on investment thus reduce investment and subsequently growth. In a test of the determinants of growth collapses after 1975, Rodrik (1999) found that high inequality and the quality of institutions that manage conflict were key underlying factors – not the size nor the intensity of external shocks. He argued that with high inequality,
distributional conflicts would be more difficult to resolve, delaying fiscal and monetary adjustment and diverting productive resources to bargaining over distributional changes. Benabou (1996) notes that if the rich understand the implication for growth of rent-seeking in unequal societies and of populist revolts, it may be in their collective interests collectively to transfer wealth to the poor through land reform, education subsidies or trade protection. The problem may be (as experience in Latin America and Africa suggests) that for such transfers to be efficient and growth-enhancing requires effective institutions of the state.

Effect of inequality on political and economic institutions

A large literature is concerned with the importance of effective institutions for growth (for example North, 1990; Acemoglu et al., 2004). Does inequality (in some ‘initial’ state) contribute to the failure of effective institutions to emerge in some societies? Engerman and Sokoloff (1997, 2002) suggest that differences in the factor endowments of colonial North and South America contributed to differences in the concentration of income which in turn affected the evolution of different economic and political institutions. Abundant slave or indigenous labor, and soil and climate conducive to large plantation agriculture in the south, and the opportunities for extraction of mineral wealth, were conducive to the high concentration of income, human capital and political power. The elite in the south then tended to create and sustain institutional arrangements that protected their interests but did not encourage broad-based investment, for example in education or productive economic activity. In contrast were the smallholder farms of the north, where the soils and climate were conducive to wheat, for example, and cheap labor was not available. In these settings, more democratic institutions evolved, property rights were broad-based, and a thriving smallholder class supported public financing of education and in general created local governments that were accountable to most citizens.

Public-choice models similarly attribute poor public policy to government regimes in which bureaucrats and insiders face no real checks on the pursuit of their own interests (Buchanan and Tollison, 1984). If the rich favor public policy that preserves privileges independent of their economic efficiency, inequality may not only inhibit growth by interacting with government failure and poor public policy, as set out above, but may contribute to poor institutions and government failures in the first place. The problem seems especially great when concentration of income at the top is combined with substantial poverty at the bottom, and there is not a large middle class to demand accountability from government. Easterly (2001) and Easterly et al. (2006), use country-level data on size of the middle class (instrumented
by differences in commodities produced, recalling Engerman and Sokoloff),
to study the determinants of good ‘institutions’ (measured in terms of
survey results on accountability, corruption, property rights, and so on).
They conclude that a small middle class is implicated in weak institutions,
and through weak institutions in low growth.

An example is the apparent relationship between a high concentration of
income in a society and differences across countries in the policy and institu-
tional capacity that ensure access to education – as in the difference
between East Asia and Latin America in educational opportunities for the
poor (Birdsall et al., 1997). Supply of publicly subsidized education is likely
to be limited where the rich resist a large tax burden to finance services
which they can purchase privately. Targeting social services to the poor can
help reduce the fiscal burden of greater public spending, but easily leads to
loss of political support from the working and middle class. Without
middle-class interest and pressure, the quality of public services deterio-
rates (and the middle class resorts to private services). Thus it is possible
for income inequality to contribute to poor public policy and institutions
even where there is little or no absolute poverty – for example in US cities.

It is also likely that high income inequality encourages rent-seeking by
the rich through bribes and extortion in the political sphere, and populist
and protectionist policies when those who feel disadvantaged acquire a
that income inequality is associated with weakening of the protection of
property rights.

In short, not only does theory and some evidence suggest inequality
harms growth in interaction with poor public policy, but it is plausible that
high inequality more directly undermines good public policy by delaying or
stalling the emergence of the political and economic institutions (property
rights, an independent judiciary, accountability to voters and checks on
abuse of privileges and power) – institutions that are increasingly viewed as
fundamental to sustaining growth.

**Effect of inequality on social institutions, social capital and collective
decision-making**

Amartya Sen places considerable emphasis on individuals’ ‘capability’ to
participate in the life of the community as an aspect of development inde-
pendent of any implications for economic growth (Sen, 1992, 1999). Partici-
ipation in the life of the community suggests there are assets that are
held not individually but only in relation to others; Putnam (1993) defines
the asset of social capital in terms of trusts, norms and networks that can
improve the efficiency of society, ‘facilitating coordinating actions’. Social
capital may also have economic value to the extent that it reduces the cost
of transactions and of contract enforcement, and as Rodrik (1993, 1999) argues, reduces resistance of losing groups to political compromises.

There is good evidence from microeconomic analyses that income inequality adversely affects some of the inputs or correlates of social capital. In Tanzania, informal insurance is higher in communities where income inequality is lower (La Ferrara, 2000). Among sugar cooperatives in India, where land ownership is more unequal, cooperatives are less productive (Banerjee et al., 2001). The literature on local public finance addresses the same issue indirectly, in assessments of the link between income levels and the formation of communities with different amounts of heterogeneity. A typical finding is that the quality of publicly provided education is inversely related to income inequality, controlling for average income (Fernandez and Rogerson, 2003).

Finally there is the evidence from studies of crime and violence. Fajnzylber et al. (2002) assessed the impact of inequality on homicide rates in a cross section of 39 countries over the period 1965–95. Income inequality measured by the Gini coefficient had a significant and positive effect on homicide rates, robust to a variety of specifications. Ratios of income of contiguous quintiles starting with the second quintile (that is, third to second, fourth to third, and fifth to fourth) exacerbate crime, and at an increasing rate. In other words, it was not poverty nor inequality at the bottom that explained crime, but the disparity between the middle strata and their richer counterparts. It was not absolute but relative income that mattered.

It is difficult to distinguish conceptually between the effects of inequality on political and economic institutions and on such ‘social’ institutions as social capital and shared civic customs and habits. To some extent that may be because across societies such ‘institutions’ as broad-based property rights, democracy with checks on abuse of power, and ‘trust’ among citizens, tend to be correlated with each other. In any event, evidence suggests that in each category, such institutions have evolved less successfully where income inequality has been high.

**Inequality and growth in East Asia versus Latin America**

In 1960, average real per capita income in Latin America was higher than in East Asia. Since then, average per capita income has risen almost tenfold in East Asia whereas in Latin America it has less than doubled (Table 48.1). In 1960, income and land inequality were significantly higher and income concentration much more extreme in Latin America compared to East Asia. (Table 48.2); Taiwan and Korea both benefited from externally imposed land reform after World War II. The contrast over four decades between fast-growing East Asia, with its relatively low inequality in 1960
### Table 48.1 Inequality, Income and Growth in Latin America and East Asia, 1960 and 2000

<table>
<thead>
<tr>
<th>Region</th>
<th>Income share of poorest 10% of population</th>
<th>Income share of richest 10% of population</th>
<th>GDP per capita (constant 2000 US$)</th>
<th>Average real GDP per capita growth rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin America¹</td>
<td>0.51</td>
<td>0.53</td>
<td>1.7</td>
<td>1.1</td>
</tr>
<tr>
<td>East Asia²</td>
<td>0.42</td>
<td>0.43</td>
<td>2.6</td>
<td>2.1</td>
</tr>
<tr>
<td>China</td>
<td>0.32</td>
<td>0.39</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.53</td>
<td>0.54</td>
<td>1.3</td>
<td>1.1</td>
</tr>
</tbody>
</table>

**Notes:**
- All group averages are unweighted.
- n/a indicates data not available.
- Latin America includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. 1960 and 2000 income Gini data not available for Ecuador, Guatemala, Nicaragua and Paraguay.
- East Asia includes Hong Kong, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand, but excludes China.
- East Asia excludes Singapore, Latin America excludes Bolivia, Chile, Honduras and Peru.

**Sources:** WDI (2005) and WIDER (2005).
(compared to Latin America, and in particular in fast-growing Korea and Taiwan compared to Thailand and Indonesia) and slow-growing Latin America, with its very high inequality (Figure 48.2), is consistent with the theory and evidence reported above: that high inequality in developing countries, where it is likely to be combined with imperfect and weak markets and poor government policy, reduces an economy’s growth prospects; and that high ‘initial’ inequality puts at risk the development of the economic, political and social institutions that support deeper markets, better government and sustained growth.23

Rapid growth in East Asia is associated with the region’s early export push, supported by high savings and investment and healthy rates of total

---

Table 48.2 Income, Education and Land Inequality in Latin America and East Asia, 1960 and 2000

<table>
<thead>
<tr>
<th></th>
<th>Income Gini³</th>
<th>Education Gini⁴</th>
<th>Land Gini⁵</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin America¹</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.51</td>
<td>0.53</td>
<td>0.53</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.06</td>
<td>0.06</td>
<td>0.13</td>
</tr>
<tr>
<td>East Asia²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean</td>
<td>0.42</td>
<td>0.43</td>
<td>0.58</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.04</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>China</td>
<td>0.32</td>
<td>0.39</td>
<td>n/a</td>
</tr>
<tr>
<td>Mexico</td>
<td>0.53</td>
<td>0.54</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Notes:
All group averages are unweighted.
n/a indicates data not available.
1. Latin America includes Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, Uruguay and Venezuela. 1960 and 2000 income Gini data not available for Ecuador, Guatemala, Nicaragua and Paraguay.
2. East Asia includes Hong Kong, Indonesia, Malaysia, Singapore, South Korea, Taiwan and Thailand.
4. Education Ginis for population aged 15 years and over.
5. East Asia average excludes Hong Kong, Singapore, Taiwan and Malaysia for which data are not available. Latin America average excludes Bolivia, Chile, Ecuador, El Salvador and Guatemala.

factor productivity growth in manufacturing (World Bank, 1993). Behind export success were other factors rooted in rapid changes in household decisions and behavior. Those other factors included unprecedented gains in small farmers’ agricultural productivity, high demand for schooling including of girls, and declines in fertility far more rapid than and at lower income levels than had occurred in the industrialized economies (Birdsall and Sabot, 2002). Governments generally ensured that exchange rates were competitive and that fiscal discipline kept inflation low. Governments also favored public investment in basic (primary and secondary) education.

In Latin America, inflation and overvalued exchange rates penalized agriculture, and were combined with tariff and other protection of industry and subsidies to capital that may have reduced the demand for labor. Spending on education was comparable to that in East Asia but was much more concentrated on highly subsidized university education for a select few, responding to the demands of richer households. In the 1960s, educational attainment of the adult population was at roughly the same levels in East Asia and Latin America, and inequality of education (measured in terms of number of years of schooling achieved) was actually higher in East Asia. Since then, educational attainment has risen more quickly in

Sources: WDI (2005), WIDER (2005), and authors’ calculations.

Figure 48.2 Income Inequality and GDP per capita growth in East Asia and Latin America, 1960–2000
East Asia and education inequality has fallen faster (Birdsall and Londono, 1998). Broad-based investment in basic education in East Asian countries led to substantial growth of labor productivity and enabled firms to acquire and adapt new technologies and move up the value chain as increasingly skilled cohorts of workers became available (Schultz, 1961; Romer, 1994). The export-push, labor-demanding strategy chosen by East Asian countries generated the conditions for a savings and investment boom in middle-income and poor households and farms (Birdsall et al., 1998).

It seems plausible that one region’s lower inequality compared to the other, among other things, affected the difference in the two regions’ subsequent trajectories of growth, inequality and investment in human capital. The story is not straightforward. Latin America has a longer history of democracy, for example. But the differences do suggest that the potential negative effect of inequality – of income, land and other assets – on growth and on the evolution of institutions that support the development process, deserves continuing attention.

Notes

1. I am grateful to my former and current research assistants: Gunilla Pettersson, Christine Park, and Karelle Samuda; and to Lyn Squire, John Williamson and Amitava Krishna Dutt for their thoughtful comments on an earlier draft. This chapter was meant to be co-authored with Richard Sabot, with whom I enjoyed spirited discussions about its content and emphasis before his untimely death in 2005; this chapter reflects inadequately those discussions and the contribution he would have made.

2. For example, Finis Welch entitled his 1999 address to the American Economics Association ‘In Defense of Inequality’. The reference to the Smithian trade-off is to Smith’s Theory of Moral Sentiments published in 1759 (Smith, 1982 [1759]). Kaldor (1961) noted that a higher profit share would encourage savings on the assumption that capitalists have a higher propensity to save, from which it follows that when income is more concentrated, savings and investment and thus the equilibrium rate of growth will be higher.


4. Lyn Squire (personal correspondence; and see Lundberg and Squire, 2003) makes the point that policy recommendations for addressing inequality may not be much different from those meant to address poverty in a country with an egalitarian distribution of income. (An exception might be tax policy, which ideally might be more progressive in the former setting, if only to sustain politically open markets. In addition greater priority in the face of political and administrative constraints might go to anti-trust and anti-monopoly programs in high-inequality settings.) I do not try to address this point in this chapter since it is not focused on policy per se but on a review of the implications of inequality for the dynamics of growth in the developing world.

5. Graham and Felton (2006) provide a survey of recent studies linking measures of ‘well-being’ (or reported ‘happiness’ in surveys of individuals) to prevailing levels of inequality.
Results depend on setting, definition of reference group, and the particular measure of well-being. In Europe and the USA inequality has generally negative effects on reported measures of well-being.

6. Graham and Pettinato (2002) make the point that what is important is people’s perceptions about their current and future income relative to others. Graham and Felton (2006) report based on happiness surveys that people in Nigeria are as happy as people in France despite the huge discrepancy in per capita incomes.

7. Much of what I say about income inequality applies to consumption inequality, and much theory reviewed below applies better to wealth than to income and consumption inequality. In principle ‘income inequality’ as I use it refers to ‘permanent income’, though in fact empirical work on income inequality is almost always based on current income, and sometimes on wages and other pre-tax income. Elsewhere I have used the term ‘money inequality’ to distinguish income and consumption inequality from inequality of ‘opportunity’ (which is difficult if not impossible to measure) and of land, education and other non-monetary assets. See Birdsall (2001).

8. See Chapter 47 in this volume for a discussion of the Gini coefficient and other measures of distribution. The studies referred to all use panels of country observations and employ country fixed-effects estimations, so that they are assessing changes over time within countries, not differences across countries.

9. Reported Gini coefficients are from the WIDER (WIID2a) database; see http://www.wider.unu.edu/wiid/wiid.htm and WIDER (2005). Income per capita is from the World Bank World Development Indicators (http://www.worldbank.org/data). For the statements in this paragraph, I used Gini coefficients from as many countries as possible. For some countries only Gini coefficients of the distribution of consumption are available. The distribution of consumption will be more equal than the distribution of income so that the number of countries and people in the categories I defined may be higher than stated here.

10. With this in mind, many developing countries embraced the need for the state to assume the commanding heights of the economy and used tax and donor resources to finance state-led industrial investments throughout much of the post-World War II twentieth century. This approach almost certainly, and ironically, led to increased concentration of income. Worse, in some countries the later privatization of those investments further increased income concentration, though there is also good evidence that privatization of water, electricity and other utilities has improved access to these services by the poor (Nellis and Birdsall, 2005).

11. Rawls (1971) argued that unequal systems of incentives and rewards may be justified if they improve the position of the least advantaged. His fundamental point was that an increase in inequality can only be justified if the outcome is an improvement in the welfare of the worst-off.

12. A closely related more Keynesian point is that greater inequality may depress aggregate demand, and thus investment incentives and growth – even where markets are otherwise functioning well. See Chapter 14 in this Handbook.


14. Ravallion (2001) reported an average elasticity of poverty reduction with respect to growth of $-2.5$, implying that for every 1 percent increase in the growth rate in average income, the proportion of the population living below $1/day falls by an average of 2.5 percent.

15. See also Deininger and Squire (1996). These findings contrast with the conclusion of ‘Growth is Good for the Poor’ in which Dollar and Kraay (2002) find that average incomes of the poorest quintile rise proportionately with average incomes in a sample of 92 countries spanning the last four decades.


17. De Mello and Tiongson (2006) find no evidence that governments of highly unequal countries are more likely to attempt to redistribute income.

18. Similarly it is often in the collective interest of an ethnic or racial majority to support anti-discrimination and other policies and programs to reduce horizontal inequalities, that is, inequalities among groups in political, economic and social dimensions, as these
otherwise can provide the basis for inter-group animosity and fuel civil conflict (Stewart, 2001; Ostby, 2003). See Chapter 63 in this Handbook.

19. On the demand side, low public spending combined with pressure to maintain or expand enrollments has led to low-quality schools, reducing the economic returns to poor families of sending children to school who can otherwise help at home or work. In effect schooling could be analyzed in terms of a two-sector model, with poor families confined to one sector with low returns, and the rich going to the other sector where returns are high. The difference in returns between poor and rich would explain the high dropout rates throughout much of Latin America, even in the face of high returns on average to those who manage to complete secondary school (Behrman and Birdsall, 1983).

20. The importance of institutions in development is discussed further in Chapter 61 in this Handbook.

21. In the USA the percentage of households that participate in various membership organizations is higher in metropolitan areas with lower income inequality – controlling for racial and ethnic heterogeneity, income, education and other household characteristics. The effect is substantial. An increase in the Gini coefficient of inequality by one standard deviation leads to a reduction in the probability of participation of 24 percentage points – more than twice the effect on participation of an individual going from the status of high school dropout to high school graduate or higher (Alesina and La Ferrara, 1999).

22. Land inequality is still extremely high in Latin America.

23. Rapid growth in East Asia without accompanying increases in inequality also contravenes the pattern suggested by Kuznets. More recently in China, rapid growth has been accompanied by rising inequality.

References


Understanding poverty in all its dimensions
The eradication of poverty is a fundamental goal of development. Ameliorating the lot of the poor is central not only in its own right but because improvements in their health, education and access to credit, infrastructure and markets unleash the untapped potential of large sections of the population, thereby contributing to overall growth and development. Progress in reducing poverty is thus a key indicator of development. Webster’s Dictionary defines poverty as ‘the state of having little or no money and few or no material possessions’. Taking this definition as a starting point, one person out of every six worldwide struggles to survive each day on less than the price of a cup of coffee in the United States. This statement contains two immensely important messages: a large part of the world’s population is obliged to live on the merest pittance; and the disparities between the rich and the poor are truly staggering.

Prestigious though it undoubtedly is, Webster’s Dictionary may not, however, capture current thinking well about the term ‘poverty’ nor its current usage in much of the professional literature. As a result of intensive research on the subject, economists and other social scientists have developed new and richer perspectives on the manifestations of poverty that better capture the dire straits that constitute daily life for so many people. For example, Amartya Sen has emphasized the importance of an individual’s capabilities to determine their own lives (Sen, 1999), of which control over resources (income) is only one aspect. Others have explored the factors that empower the poor to lead full and productive lives (Stern et al., 2005). In addition to low incomes and inadequate consumption, poverty in these interpretations is understood to encompass lack of education, poor health, insecurity, violence, social and political exclusion, as well as absence of basic rights and human dignity. This broader conceptualization of poverty has pushed development experts to expand the conventional indicators of poverty to reflect today’s more comprehensive characterization of the phenomenon. With this in mind, this chapter undertakes three tasks.

It first describes the attempts to measure poverty as the definition of the term has expanded from an initial focus on subsistence to a broader appreciation of the many elements constituting a person’s well-being. As we shall see in the next section, even unidimensional measures of poverty based
exclusively on income or expenditure pose many conceptual problems as well as the ever-present difficulty of securing appropriate data. Moving to richer, multidimensional measures poses additional problems, among which the allocation of appropriate weights to the constituent elements is perhaps the most daunting.

The second task is the presentation of the best available evidence. Conceptual difficulties notwithstanding, statistics can be marshaled to demonstrate two important points. First and foremost, progress in reducing poverty on almost all fronts has been greater since 1980 to date than in any other similar period in the recorded history of mankind. And second, progress has been uneven in several respects: across countries; within countries; and across the various dimensions of poverty. Evidence on each of these points is presented in the third section. If progress has been uneven around the world – and it has – then it is of interest to explore why some countries or regions have done better than others and to share the lessons of their success with those that have fared less well. This, the third task, is undertaken in the fourth section. It concludes with some observations on what the available evidence suggests for future efforts to attack poverty.

From unidimensional to multidimensional poverty

A unidimensional focus on income

Consider first how to measure poverty when this is confined to meaning lack of income. For purposes of exposition assume that we have a universally accepted measure of income (as we shall see, matters are not so simple). Call this measure \( y \). Let us also assume that we know the income of each individual in a population so that they can be ranked according to income beginning from the individual with the least income and proceeding in order to the individual with the greatest income. Finally, let \( y^* \) be the poverty line, that is, the income threshold below which people will be said to be poor.

Even with all the elements apparently in place to measure poverty, the analyst still faces the task of deciding how to present the evidence on poverty in a single indicator. Here I present three well-known indices to illustrate the range of choice and the issues to be considered. Perhaps the simplest is to count the number of individuals for whom \( y < y^* \). Known as the ‘head count’, this indicator represents the number of individuals in poverty and is the most commonly used measure.

The head count, however, says nothing about the depth of poverty or the extent to which an individual’s income falls below the poverty line. To address this issue, the ‘poverty gap’ measures the average income needed to bring the incomes of all poor people up to the poverty line or \( y^* - \overline{y} \) where \( \overline{y} \)
is the mean income of all individuals below the poverty line. The poverty gap however makes no allowance for the severity of poverty because it treats an extra dollar for the poorest person among the poor as equal in value to an extra dollar for the person just below the poverty line. The Foster–Greer–Thorbecke index overcomes this deficiency by appropriate weighting. Thus, the measure of severity for the $i$th individual is given by $\left( \frac{y_i^* - y_i}{H_1} \right)^\alpha$ where $y_i$ is the $i$th individual's income and $\alpha$ is the weight. As $\alpha$ increases beyond unity, the weight of those whose incomes are well below the poverty line is magnified in the summary index for all individuals (Foster et al., 1984). This feature, however, introduces a degree of arbitrariness because the appropriate value of the weight is inevitably a matter of judgment.

While the preceding points should be born in mind, much of the effort to track progress in reducing poverty worldwide has focused on the head count, a practice that we continue in the third section. Before looking at the numbers, however, I will first examine some of the difficulties in arriving at a measure of income and deciding on a poverty line, difficulties I had assumed away in the forgoing discussion.

**Measuring income and expenditure**

Researchers seeking to measure income in developing countries face many daunting challenges. Some of the main difficulties include: how to capture non-marketed and non-priced subsistence output such as crops grown by households for their own consumption; how to allow for free social services and pure public goods; whether to use data on income or expenditure; how to accommodate differences in the demographic composition of households; how to distinguish between temporary and permanent poverty; how to manage seasonal variations in income; and so on (Kanbur and Squire, 2001; Lok-Dessalien, 1999).

While there is no universally accepted procedure for dealing with each of these issues, in practice most analysts follow some conventions. Thus, expenditure is usually preferred to income since the former is thought to be a better indicator of permanent income. Since data are usually collected at the level of the household, most analysts present information on expenditure per household member by dividing household expenditure by the number of household members. Thus no allowance is usually made for the different consumption needs of children and adults although ‘equivalence scales’ are available and are used on occasion. Nor is any allowance made for distributional rules within the household based on gender, age or working status.

The researcher’s ability to deal with these issues depends crucially on the quality of the data, which are usually obtained from household surveys. In
addition to the difficulties noted above arising from the number of activities, products and services that are unrecorded, estimates of household expenditure are also affected by the limitations of sample surveys (Deaton, 2003). Recall errors, short reference periods and the exclusion from the sampling frame of people in remote areas and other marginal groups who are most likely to be poor are common problems. The quality of enumerators and the consistency of surveys over time are other factors complicating the task of the researcher. That said, surveys of reasonable quality are now available for 97 countries covering 93 percent of the population in the developing and transition worlds (Chen and Ravallion, 2004).

Setting the poverty line

Poverty lines are commonly distinguished according to whether they are absolute or relative. While arguments can be advanced that absolute measures best capture the notion of individual capabilities incorporated in the broader definitions of poverty, the issue is more open in the case of poverty confined to income or expenditure. In this case, ‘absolute poverty’ refers to subsistence below some minimum, socially acceptable norm, usually established on the basis of nutritional requirements plus other essential goods. ‘Relative poverty’ compares the proportion of the population below a poverty line defined relative to mean income or some measure of the overall well-being of the population. Relative poverty lines thus adjust to changing circumstances. Absolute poverty lines on the other hand remain unchanged and are therefore better able to track changes in poverty over time and, where they are based on equivalent real baskets of goods and services, can also be used to aggregate across countries to track global poverty. For this reason, the remainder of this chapter focuses on measures of absolute poverty.

The absolute poverty line currently used to track global poverty is the well-known figure of a $1 a day. Introduced first in the 1990 World Development Report (World Bank, 1990), this figure is based on actual poverty lines then prevailing in some of the poorest countries in the world (World Bank, 1990). These measures typically are based on the cost at local prices of a diet providing the minimum nutritional requirements for subsistence, plus some allowance for other basic needs, usually confined to clothing and shelter. To translate these figures into dollars, purchasing power parity (PPP) conversion factors are used. Although not designed to capture the purchasing power of the poor, they remain the best instrument for translating local currencies into dollar equivalents for purposes of international comparisons. This weakness notwithstanding, the comparison revealed a marked bunching of these national lines around the dollar-a-day mark and resulted in its selection as a reasonable indicator of absolute
poverty, at least as perceived by those in the countries suffering most acutely from low incomes. For the remainder of this chapter, we will use this poverty line to indicate extreme poverty but rely as well on a higher figure – roughly $2 a day – to provide a measure of poverty that is more than mere subsistence.

**Multidimensional measures of poverty**

Estimates of poverty based exclusively on income or expenditure fail to capture significant aspects of deprivation experienced by the poor. Hence there is a need to go beyond traditional methods of measuring poverty and address the broader multidimensional nature of poverty by including measures of health status, educational attainment, political voice and social inclusion as well as measures of control over material resources. While conceptually straightforward, this extension encounters several practical problems of implementation (Falkingham and Namazie, 2001; Boltvinik, 1999).

**Data issues**

A major difficulty is that many of the required measures are not readily available or, where they are available, suffer from problems of interpretation. Consider political voice or social inclusion. Neither has been measured on a routine basis or according to a widely endorsed procedure. While data collection efforts will undoubtedly continue to improve, it is currently difficult to track progress in all dimensions, especially for historical periods. Even where measures are available on a regular basis and are collected according to generally accepted conventions, quality remains an issue. For example, educational attainment is often represented by the net primary school enrollment rate but such measures are silent on the quality of education actually provided.

Moreover, many of the measures of interest are reported only as national aggregates with no distributional breakdown. Consequently, it is not possible to develop measures of who is poor or deprived and who is not for many of the non-income indicators. At best, one can report national averages. Increasingly, however, household surveys are incorporating questions about health status, educational attainment, and so on. A common result emerging from these surveys is that the poor in one dimension are, as one might expect, often the poor in other dimensions as well. For example, school attendance rates for 6- to 17-year-olds for the poorest decile of the population ranked by income were 31 percent in Pakistan, 32 percent in Nepal and 51 percent in Vietnam, compared with rates of 71 percent, 88 percent and 84 percent for the richest decile in the same countries (Appleton and Song, 1999). Evidence on health status tells a similar story. Stunting, as measured by height-for-age, affected 52 percent of preschoolers in the poorest decile in Pakistan, 55 percent in Nepal and 58
percent in Vietnam. In contrast, the corresponding rates for the richest segment were 32 percent, 27 percent and 23 percent (Appleton and Song, 1999).

While quantitative information on health status and educational attainment by households ranked according to income or expenditure is beginning to appear, more qualitative approaches are required to capture political voice, social exclusion, measures of self-respect, and so on. Interviews and focus-group discussions with villagers, urban slum dwellers and marginalized communities remain the best means of securing the fullest and most complete understanding of the plight of the poor (Narayan et al., 2000).

To aggregate or not  Putting the data issues to one side, the focus on multiple dimensions of poverty raises the question of the most useful form of presentation. One possibility is to combine the various individual indicators into a single index. Thus, if there are now $n$ indicators with corresponding weights represented by $\alpha_n$, then $\sum \alpha_n \gamma_n$ provides an overall measure of well-being. In principle, households could then be ranked according to their value of this measure, and then those falling below some minimum could be identified as the poor, in the same manner as with the single indicator. In fact, this is never done because as noted above distributional data do not exist for many of the dimensions of interest. Nevertheless, attempts have been made to combine income and social indicators at the national level to provide a more complete measure of well-being.

The best-known such index is the Human Development Index introduced by the United Nations in 1990 (UNDP, 1990) and its subsequent extension, the Human Poverty Index. The latter aggregates three separate measures: longevity (the percentage of people expected to die before age 40); literacy (the percentage of adults who are illiterate); and deprivation in overall economic provisioning both public and private (the percentage of people without access to water and health services and the percentage of underweight children below the age of five). The basic problem with such aggregates is that there is no satisfactory way of arriving at the weights (Srinivasan, 1994). The United Nations Development Programme (UNDP) has assumed weights of one-third for each of the three measures without any real justification other than convenience. That said, composite indices are considered valuable at the global level for advocacy purposes, but less so for country-specific policy-making purposes since they effectively conceal knowledge on the individual measures through the process of aggregation. For this reason, I report an array of indicators in the remainder of this chapter to capture the main dimensions of poverty but stop short of using an aggregate index.
Reducing poverty: the record since 1980 to date

Progress in the aggregate: a success story

The previous section has introduced the basic proposition that poverty should be understood as a multidimensional phenomenon: it is the failure to meet minimally acceptable standards, not only with respect to income or expenditure, but also with respect to health, education, and social and political values. This richness carries a cost: it is difficult to arrive at simple statistics that fully capture all aspects of such a broad concept. We have also seen that attempts to aggregate individual indicators into composite measures pose insuperable problems and in fact conceal information. Accordingly, I present in Table 49.1 several indicators, each of which reflects an important dimension of poverty.

The table reveals two decades of genuine progress. Perhaps the most dramatic improvement is the virtual halving of the percentage of the population living in poverty in low-income and middle-income countries (following World Bank classifications). In Table 49.1, the poor, defined as those subsisting on less than $1 a day, accounted for 40 percent of the population in 1981 but only 21 percent in 2001.2 In terms of absolute numbers, however, the outcome is not so impressive. Because of population growth (roughly an increase of 1.5 billion people), the number of poor only fell by 390 million. On the other hand, had the percentage remained at its 1981 level, the number of poor would have increased to over 2 billion, implying that as many as 1 billion people, or one-fifth of the population, avoided poverty relative to what might have happened based on an extrapolation of 1981 circumstances.

Table 49.1 also presents the progress of low- and middle-income countries in reducing poverty in its non-income dimensions and in improving the

Table 49.1  Measures of aggregate well-being in low-income and middle-income countries

<table>
<thead>
<tr>
<th>1981</th>
<th>1990</th>
<th>2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of extreme poor</td>
<td>40.4</td>
<td>27.9</td>
</tr>
<tr>
<td>Number of extreme poor (billion)</td>
<td>1.48</td>
<td>1.22</td>
</tr>
<tr>
<td>Life expectancy (years)</td>
<td>60</td>
<td>63</td>
</tr>
<tr>
<td>Under-5 mortality rate per '000</td>
<td>131</td>
<td>103</td>
</tr>
<tr>
<td>Literacy (%)</td>
<td>61</td>
<td>68</td>
</tr>
<tr>
<td>Net primary school enrollment (%)</td>
<td>78</td>
<td>95</td>
</tr>
</tbody>
</table>

Source:  Chen and Ravallion (2004) and World Bank (various issues).
welfare of their people. It shows an increase of four years in life expectancy at birth for this group of countries during the two decades to 2001. To see an increase of one year in life expectancy every five years is remarkable progress. While child mortality rates also witnessed significant progress, they remain unacceptably high. The literacy rate also shows steady improvement in this group of countries and net primary school enrollment is now almost 100 percent.

Failures and setbacks

The progress in aggregate does not mean that everyone advanced and in all dimensions. Indeed, more disaggregated evidence reveals that some groups not only failed to participate in the overall progress, but they saw their situation worsen in one respect or another. At least four groups can be identified where progress in one or more dimension of poverty departs significantly from the overall progress revealed by Table 49.1: those in Africa; those in marginalized regions or social groups despite country-wide progress; those just above the poverty line; and those where indicators for health or education have improved while poverty as measured by expenditure has worsened, or vice versa. I briefly consider each in turn.

The overall progress revealed in Table 49.1 notwithstanding, one region of the world – Africa – has seen poverty worsen. Recall that the number of extremely poor in China fell by 420 million during the two decades of the 1980s and 1990s. This is more than the worldwide decline in the extremely poor – 380 million. Thus, in the rest of the developing and transition world the number of poor increased slightly during this 20-year period, and if some regions saw declines in poverty, as South Asia did, then other regions must have experienced substantial increases in extreme poverty. This happened in sub-Saharan Africa, and dramatically so: the number of extremely poor almost doubled from about 160 million in 1981 to 315 million in 2001 (Table 49.2). Sub-Saharan Africa is the only region in the developing world to see the head count index increase – from 41.6 percent in 1981 to 46.9 percent in 2001. As a result, Africa’s share of the world’s extremely poor increased from 11 to 29 percent during this period. Worse yet, not only has the number of Africans in extreme poverty increased, but the severity of their poverty has also increased. While the rest of the world’s poor saw their daily expenditure increase from $0.70 in 1981 to $0.77 in 2001, those in Africa saw theirs fall from $0.64 to $0.61. Poverty is worsening in Africa and becoming more concentrated in Africa (Chen and Ravallion, 2004).

Uneven progress is apparent among social indicators as well. Take for example life expectancy, perhaps the most fundamental measure of overall well-being. Table 49.1 reveals steady progress. This has been true of almost all countries, but 26 countries saw life expectancy fall during the 1990s and
in some cases precipitously so. For example, life expectancy fell by 17.8 years in Botswana, 16.3 years in Zimbabwe and 14.1 years in South Africa (World Bank, 2004). These figures depict a rapid undoing of steady progress over many years. Nineteen of the 26 countries are in Africa, and four are in the republics of the Former Soviet Union. While causes vary across countries, the advent of the AIDS crisis has played a major role, especially in Africa. In several countries the prevalence of HIV among the 15–49 age group in 2001 exceeds one-fifth: Botswana (38 percent), Lesotho (29.6), Namibia (21.3), South Africa (20.9), Swaziland (38.2) and Zimbabwe (24.9) (UNAIDS, 2004).

This phenomenon of uneven progress across regions of the world is also evident within countries. A second group therefore comprises those living in certain areas or belonging to certain ethnic groups who have not participated fully in their country’s overall progress. Consider the case of India. Using the official poverty line, the number of Indian poor fell from 320 million in 1993–94 to 260 million in 1999–2000. But not all states prospered equally. Indeed, two states – Madhya Pradesh and Orissa – saw the numbers in poverty increase, albeit only slightly (Government of India, 2002). Groups within states can also experience different outcomes. For example, the incidence of poverty in the coastal districts of Orissa declined from 46 percent to 32 percent during this period. The experience of the southern districts, however, was the reverse: the incidence of poverty increased from 43 percent to 50 percent. And the districts with the highest initial incidence of poverty – the northern districts – experienced a significant increase from 66 percent to 81 percent (de Haan, 2004). Isolation in economic terms, political terms, social terms and racial terms underlie many of these examples of lagging regions or lagging social groups.

A third, major group to witness a deterioration in their well-being is the less poor, or those just above the extreme poverty line. The dollar-a-day poverty line is usually characterized as depicting extreme poverty; it allows mere subsistence at best. Consequently, a higher poverty line – $2 a day – is also frequently employed. This provides important new information. Using this measure of poverty, the number of poor actually increased from

Table 49.2   Extreme poverty in sub-Saharan Africa

<table>
<thead>
<tr>
<th></th>
<th>1981</th>
<th>1990</th>
<th>2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of extreme poor</td>
<td>41.6</td>
<td>44.6</td>
<td>46.9</td>
</tr>
<tr>
<td>Number of extreme poor (million)</td>
<td>163</td>
<td>227</td>
<td>316</td>
</tr>
</tbody>
</table>

Source:  Chen and Ravallion (2004).
2.4 billion in 1981 to 2.7 billion in 2001 (see Table 49.3). Given that the number below the dollar-a-day line fell, the net result is a significant increase in the number of people surviving on expenditure per day of between $1 and $2. In fact, the number increased from about 1 billion in 1981 to 1.7 billion in 2001. This is significant for at least two reasons. It suggests that the factors – be they growth, redistributive policies, or whatever – that produced the decline in extreme poverty have been less successful in helping those just above the extreme poverty line. And it suggests that a large number of people remain highly vulnerable to droughts, economic crises, conflicts or other calamities that could easily push them into the ranks of the extremely poor.

The final group comprises those who have experienced uneven progress among indicators. While Table 49.1 shows that in aggregate the various indicators all move in a positive direction, in some cases progress in reducing poverty as measured by increases in expenditure has co-existed with deteriorating outcomes for the social indicators. Thus, eight of the 19 African countries that saw life expectancy fall in the 1990s enjoyed increases in real gross domestic product (GDP) per capita. Since HIV was prevalent in all eight countries, it appears that the disease strikes life expectancy more or sooner than GDP. In still other cases, real GDP per capita has declined while life expectancy has continued to increase. Indeed, life expectancy increased in just over half of the 48 countries that saw real GDP per capita fall in the 1990s (World Bank, 2004). While many factors come into play, it is noteworthy that 22 out of the 26 countries that saw continued progress in life expectancy spent 2 or more percent of GDN on health. The often referred-to cases of Sri Lanka and the state of Kerala, India, are powerful illustrations of how sustained investment in nutrition and health can result in unusually long life expectancies at relatively low levels of income. Thus, life expectancy in Sri Lanka was 73 years in 2000 despite its relatively modest level of GDP per capita.

Table 49.3  ‘Vulnerable’ population

<table>
<thead>
<tr>
<th></th>
<th>1981</th>
<th>1990</th>
<th>2001</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of poor</td>
<td>66.7</td>
<td>60.8</td>
<td>52.9</td>
</tr>
<tr>
<td>Number of poor (billion)</td>
<td>2.45</td>
<td>2.65</td>
<td>2.74</td>
</tr>
<tr>
<td>% of ‘vulnerable’</td>
<td>26.3</td>
<td>32.9</td>
<td>31.8</td>
</tr>
<tr>
<td>Number of ‘vulnerable’ (billion)</td>
<td>0.97</td>
<td>1.43</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Source: Chen and Ravallion (2004).
Learning from the past
The review of progress in reducing poverty in all its dimensions presented in the previous section revealed substantial but uneven progress. Overall the well-being of the poor has improved, but many have been bypassed or marginalized and others have seen their situation worsen. This suggests two broad questions: What is driving the general progress? And what can be done to ensure that progress is more widespread and inclusive?

In answering these questions, the starting point for most analysts is the role of GDP growth. Table 49.4 illustrates the association between growth in GDP per capita and selected indicators of poverty. For the 13 countries that enjoyed growth rates of at least 2.3 percent a year in GDP per capita in the 1980s and 1990s, the head count index fell by seven percentage points, illiteracy by six percentage points, and life expectancy increased by more than three years. Thus, the people in these 13 countries saw average conditions improve in three important dimensions. At the other extreme, those living in the 39 countries that experienced low growth in GDP per capita saw the head count index increase, illiteracy fall by over seven percentage points, and life expectancy increase by 1.4 years.

The following interpretation of these aggregate trends draws on a wide range of other material including country studies and case studies as well as cross-country analyses that cannot be reported here because of space limitations. Table 49.4 plus the evidence of many other studies indicates a strong but imperfect relationship between growth in average incomes and the incomes of the poor. This reflects the tendency for inequality in national

<table>
<thead>
<tr>
<th>Change in indicator: comparing 1980s and 1990s</th>
<th>Unit</th>
<th>Period</th>
<th>High growth</th>
<th>Moderate or improved growth</th>
<th>Low growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Head Count Index % with less than US$1 a day</td>
<td>%</td>
<td>1990s</td>
<td>24.1</td>
<td>31.4</td>
<td>36.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1980s</td>
<td>31.0</td>
<td>32.1</td>
<td>30.2</td>
</tr>
<tr>
<td>Illiteracy %</td>
<td>%</td>
<td>1990s</td>
<td>17.2</td>
<td>31.2</td>
<td>31.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1980s</td>
<td>22.9</td>
<td>37.6</td>
<td>38.8</td>
</tr>
<tr>
<td>Life expectancy Years</td>
<td>Years</td>
<td>1990s</td>
<td>70.0</td>
<td>62.9</td>
<td>59.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1980s</td>
<td>66.8</td>
<td>60.6</td>
<td>58.4</td>
</tr>
<tr>
<td>Number of countries</td>
<td></td>
<td></td>
<td>13</td>
<td>53</td>
<td>39</td>
</tr>
</tbody>
</table>

income to change only slowly (see Li et al., 1998) implying that incomes of the poor increase more or less by the same percentage as the incomes of everyone else (Dollar and Kraay, 2002). The strong influence of growth is seen clearly in Table 49.4. Results of this sort, however, say nothing about policy or causality. Thus, the observed transformation of growth into a reduction in the head count index may or may not require a wide range of redistributive policies (Kanbur, 2003). The evidence presented thus far is silent on this issue. And with respect to causality, both growth and inequality are outcomes of the same economic system and all the factors that influence that system (Lundberg and Squire, 2003). Thus, growth and poverty as measured by the head count index are jointly determined.

Reductions in illiteracy, on the other hand, are not obviously correlated with growth and are therefore presumably driven by other factors. For countries with universal enrollment, reductions in illiteracy are largely determined by each country’s population dynamics and the share of school-age children in the population. And for those without universal enrollment, the capacity of the school system is the decisive factor. The key policy instrument in these circumstances is the provision of adequate support from the public budget to maintain schools, improve quality and expand capacity where necessary. While funding alone is not sufficient – many other factors influence the delivery of public services (World Bank, 2004) – it is nevertheless an essential ingredient. As long as countries can maintain budget outlays for schooling, we should not therefore expect to see a strong relationship between movements in the growth rate and reductions in illiteracy. In the extreme, however, a collapse in GDP as has occurred in failed states or countries in major transitions will inevitably undermine the educational system. For example in countries like Albania, Bosnia, Congo, the former Yugoslavia, Rwanda, and so on, various shocks and ethnic conflicts have resulted in a colossal destruction of human and social capital (World Bank, 2000).

Improvements in longevity, however, do appear to be linked with growth in GDP per capita (see Table 49.4). Nutritional intake, a key factor influencing morbidity and longevity, depends to a considerable degree on the incomes available to households to buy food of increasing variety and quality. At the same time, life expectancy, like literacy, reflects the quantity and quality of public spending on, in this case, health services. Extraordinary examples like Sri Lanka and the Indian state of Kerala reveal the power of intensive and sustained public support for health services. By the same token and as we have seen above, major new diseases like HIV/AIDS can quickly undermine years of progress in extending people’s lives.

The key question facing the policy-maker is, of course, the policies and programs required to achieve growth in GDP and to ensure that all, and
especially the poor, benefit. While it is not possible to go into detail, the broad elements of a strategy consistent with the existing evidence can be sketched. In short, the evidence from the success stories suggests a ‘two-part strategy’ that on the one hand harnesses market incentives, social and political institutions, infrastructure and technology to promote growth, and on the other hand, supports the poor through the provision of health, education and other basic services (World Bank, 1990). The first part of the strategy provides income-earning opportunities for the poor, while the second part strengthens the capabilities of the poor to take full advantage of those opportunities. In this sense, the two parts of the strategy are mutually reinforcing.

A recent assessment of development experience that takes a very broad definition of poverty as its starting point has elaborated and refined this ‘two-track’ approach (Stern et al., 2005). The first pillar of this strategy calls for creating a supportive investment climate to encourage firms and farms, small and large, to invest, create jobs and increase productivity. The investment climate is determined by several factors that can be classified under three broad groups: macroeconomic and trade policies, infrastructure, and governance and institutions. The second pillar involves empowering and investing in poor people by enhancing their health, education and security and by fostering mechanisms for them to participate in the growth process of the economy. This approach is broadly consistent with the evidence reported here.

In conclusion it is worth making two cautionary remarks. First, useful though it is to have some general view on strategy, country circumstances vary so much that careful and possibly major tailoring may be required to arrive at an appropriate national strategy. For example, many developing countries have suffered from ethnic strife. In Sierra Leone, the prolonged strife has exacted a heavy human and economic toll. Other countries have been ravaged by the AIDS epidemic. Any national poverty-reduction strategy would have to recognize and respond to such local circumstances. Second, special actions may be required for groups which may otherwise be excluded from the national strategy’s reach such as ethnic minorities, HIV/AIDS-positive individuals and those damaged by war, or which have suffered temporary setbacks arising from price fluctuations, unemployment or natural disasters.

Notes
1. The valuable research assistance of Partha P. Sahu, Intern, global development network (GDN), is acknowledged with much appreciation.
2. The exact figure is $1.08. The original figure of $1 was recalculated using the new PPP of 1993 (Chen and Ravallion, 2004).
3. Eastern Europe and Central Asia also saw the head count index rise, but it remains negligible; the paucity of survey data for this region in the 1980s should not be forgotten.
Thus the estimates are heavily based on interpolations, which do not allow for any changes in distribution (Chen and Ravallion, 2004).

4. The exact figure is $2.15. The original figure of $2 was recalculated using the new PPP of 1993 (Chen and Ravallion, 2004).

References
World Bank (various issues), World Development Report.
Introduction
Two questions are at the heart of economic analysis of gender equality issues in development:

1. Are the fruits of economic growth and development in developing countries fairly shared between women and men, girls and boys?
2. Does gender inequality promote or hinder economic growth and development in developing countries?

There is a vast literature addressing the first question, which will be considered in the next section; and a much smaller but growing literature addressing the second question, which will be considered subsequently. The challenges of designing macroeconomic policies to promote gender-equitable growth and development will be briefly considered in the final section.

Gender and the distribution of the fruits of growth and development
A gender analysis of the distribution of the ‘goods’ produced by development (not only income and wealth but also capabilities) goes beyond a focus on women as an isolated group and beyond mere disaggregation by sex. It situates distribution of these ‘goods’ in the context of the social construction of gender (that is, the social construction of norms of masculinity and femininity), which shape the choices made by people; and the consequences of those choices. It recognizes not only differences between the sexes, but also structural inequality between the sexes, embedded in institutions, including not only families but also markets, businesses and states (see for example, Sen, 1983; Folbre, 1986; Sen, 1996; Agarwal, 1997; Harriss-White, 1998; Elson, 1999).

Prior to the 1970s, there was little discussion of gender issues in economic development. If the topic was mentioned, there was a tendency to assume that women were an underutilized factor of production which could be mobilized for structural transformation and economic growth, and that economic growth and structural transformation would in turn be good for women. An example is the work of Arthur Lewis. In his famous model of ‘economic development with unlimited supplies of labour’, he identified ‘the wives and daughters of the household’ (in the subsistence
sector), as a source of labour for the modern ‘capitalist’ sector, arguing that this would lead to gains for women: ‘because most of the things which women otherwise do in the household can in fact be done much better or more cheaply outside, thanks to large scale economies of specialization, and also to the use of capital (grinding grain, fetching water from the river, making cloth, making clothes, cooking the midday meal, teaching children, nursing the sick etc.)’ (Lewis 1955, p. 404). In his book on the theory of economic growth, Lewis was in no doubt about the benefits to women: ‘Women benefit from growth even more than men . . . Woman gains freedom from drudgery, is emancipated from the seclusion of the household, and gains at last the chance to be a full human being, exercising her mind and her talents in the same way as men’ (Lewis, 1955, p. 422).

This optimistic view was challenged by Ester Boserup (1970), who argued that modernization of economies in Africa and Asia had marginalized women. In the agricultural sector, she argued, women had been deprived of access to training, land rights, education and technology, by both colonial and post-colonial administrators, who could not conceive of women being farmers in their own right, even though in much of sub-Saharan Africa and South-East Asia women enjoyed a significant autonomous role in traditional agricultural production. This lack of access to resources meant that while men’s productivity in farming increased, women’s productivity did not.

In the industrial sector, she argued, women accounted for a much lower percentage of the industrial labour force in large-scale modern factories than they did in home-based handicraft manufacturing. She pointed to obstacles on the demand side, including labour market regulations, and employers’ prejudiced perceptions of women’s capacities and work commitment; and on the supply side, she suggested that women had difficulties combining work in the modern sector with their family responsibilities, and were hindered by the view that work outside the home was not proper for women. Above all, women were hampered by their lack of appropriate skills, stemming from their lack of formal education. As a result of all these factors, women had been left marginalized and excluded from development. Boserup’s remedy for this was investment in more and better education and training for women – planners must change their view that women were primarily housewives, and train women to compete equally with men in the marketplace, so that women could be included in economic modernization.

As Naila Kabeer (1994) points out, Boserup’s book laid the foundations for a large body of ‘Women-in-Development’ literature, and a large number of policy initiatives aimed at ‘integrating women into development’. Irene Tinker (1990), in describing the making of the field of ‘Women-in-
Development’, calls Boserup’s book ‘the fundamental text for the UN
Decade for Women’ (1975–85) (Boserup, 1970). Boserup’s ‘marginaliza-
tion’ thesis found support from other authors, such as Saffiotti (1978), who
examined the implications for women’s employment of import-substitution
industrialization in Brazil and found that during the 1950s and 1960s, while
women’s industrial employment increased overall, their share relative to
men in the industrial labour force declined.

Addressing women’s marginalization through education has been a con-
stant theme in the ‘Women-in Development’ literature. There have been
significant increases in girls’ enrolment in primary school since the 1970s,
and by 2000 almost all girls (and boys) were enrolled in primary school in
developing countries in East Asia and the Pacific, Europe and Central Asia,
and Latin America and the Caribbean (UNESCO, 2004). However, in the
Middle East and North Africa, South Asia and sub-Saharan Africa,
significant numbers of children remained out of school, of which 54 per
cent were girls. In South Asia, the gender gap was particularly wide, with
Even in regions with parity in enrolment, girls drop out of school more
than boys, so that, with the exception of Latin America and the Caribbean,
boys’ completion rates remained higher than those of girls.

At the secondary level, by 2000, no region had succeeded in enrolling all
children, and girls’ enrolment lagged that of boys in all regions, with the
exception of Latin America and the Caribbean, where the reverse was true.
Moreover, girls’ secondary enrolment rates remained substantially lower
than their primary enrolment rates. In sub-Saharan Africa and India, girls’
secondary enrolment was particularly low, at 29.7 per cent and 47.1 per cent
respectively. The comparable figures for boys were 35.6 per cent and 53.7
per cent (UNESCO, 2004).

Boserup’s concern about the lack of visibility of women farmers and
their lack of resources, including training, has been another continuing
theme in the ‘Women-in-Development’ literature. The importance of agri-
culture as a source of livelihood has declined in most developing countries,
though it remains high in some countries and regions. The proportion of
the economically active population in agriculture in 2000 was, on average,
21 per cent in Latin America and the Caribbean, 59 per cent in Africa and
36 per cent in Asia. In the latter region there was wide variation, with only
10 per cent of the economically active population in agriculture in South
Korea, while the figure for China was 67 per cent and for India, 60 per cent
(UNRISD, 2005, p. 91).

There is still a lack of reliable data on women’s share of the economically
active population in agriculture, since women’s work is still undercounted
in censuses and labour force surveys, despite some improvements (Beneria,
1992, 2003). In particular, women's 'own-account' or 'self-employed' farming is undercounted, and women are more likely to be enumerated as 'unpaid family workers' contributing to farms managed by their husbands, and 'agricultural wage workers', working for larger-scale commercial farms, than as farmers in their own right.

In much of sub-Saharan Africa, women and men in the same household farm and manage separate plots, while at the same time supplying labour inputs to each other's plots, so that many women are both 'own-account farmers' and 'unpaid family workers'. Moreover, migration, war and HIV/AIDS have reduced rural male populations in sub-Saharan Africa, and about one-third of all rural households are headed by women, leading to the 'feminization of agriculture', according to the UN Food and Agriculture Organization (FAO, 2005). In many Latin American countries, there has been a feminization of agriculture, as men have migrated from rural areas in search of better incomes, and women have taken over the management of family farms, and do the bulk of farm labour (UNRISD, 2005, p. 96). In Asia, similar factors have led to more women in China, India and South-East Asia taking on the management of farming activities, though there is not full agreement on whether farm management could be said to be 'feminized' (UNRISD, 2005, p. 97). In India, for instance, 48 per cent of self-employed farmers are women; and in dairying and animal husbandry, women farmers far outnumber men (Ministry of Agriculture, Government of India, 2005).

Farming women throughout the developing world have less rights over land than do men, as a result of a variety of factors, such as unequal inheritance practices, registration of land titles only in the name of male household heads, and land reforms that are biased against women (Grown et al., 2005, p. 75). Comprehensive data on the size of the gender gap in land ownership are not available. A study covering five Latin American countries found that women account for between 11 and 27 per cent of landowners (Deere and Leon, 2003). Women farmers in sub-Saharan Africa lack secure rights to the land they farm, and frequently lose their land when they are widowed or divorced (World Bank, 2001, pp. 121–2). In South Asia, few women own agricultural land; and of those that do, few exercise full control over it (Agarwal, 1994).

Since the early 1990s there have been some improvements in women's formal land rights in Latin America and sub-Saharan Africa. Joint titling of land to couples has been introduced in Brazil, Colombia, Costa Rica, Honduras and Nicaragua (Deere and Leon, 2003). In Uganda, Kenya, Tanzania, South Africa and Rwanda women's formal statutory land rights have been strengthened – but customary law is still a powerful obstacle, generally preventing women from owning or inheriting land in their own name.
(Grown et al., 2005, p. 81). Moreover, in other parts of sub-Saharan Africa, women’s customary use rights over land have been weakened through the introduction of individual ownership of land (Lastarria-Cornhiel, 1997). Women’s land rights were strengthened by new legislation in India in 2005 but many challenges of implementation remain (Agarwal, 2005).

Formal ownership rights do not necessarily ensure gender equality in livelihoods. For instance, land markets themselves are not gender-neutral: women buyers have lower bargaining power than men, and face discrimination (Deere and Leon, 2003). Moreover, women farmers still receive less technical support from agricultural extension workers, and only a tiny percentage of extension workers are women (World Bank, 2001, p. 52). In the case of women smallholders, many aspects of the disadvantages depicted by Boserup still persist. But efforts to promote gender equality are unlikely to yield substantial gains in terms of improved standards of living for women smallholders without more effective strategies for the agricultural sector as a whole (UNRISD, 2005, p. 104).

Some women have been moving into new areas of ‘modern’ commercial agricultural production. In the production of flowers, fruit and vegetables for export from Latin America and sub-Saharan Africa, women comprise between 45 and 90 per cent of contract workers (UNRISD, 2005, p. 98). However, women are mainly employed in temporary or seasonal low-paid jobs, with long hours of work, poor health and safety conditions and no social protection (UNRISD, 2005, p. 98). This illustrates a key shortcoming of the Women-in-Development (WID) ‘marginalization’ thesis: it identifies gender issues in development in terms of women’s exclusion from development, whereas the problem is increasingly the terms and conditions of women’s integration in development.

This type of criticism of the WID approach was initiated in the early 1980s, especially by women scholars from developing countries who argued that women were subordinated to development rather than marginalized from it (see, for example, Beneria and Sen, 1981). The disadvantage experienced by women in the process of development was not, they argued, primarily the result of the persistence of ‘traditional’ cultural practices and prejudices, but of the way in which development has created new job structures, in which gender inequality is embedded. For instance, women are concentrated at the lowest rungs, with lower pay, and less job security and social protection than men; and are expected to combine paid work with ongoing responsibility for the unpaid work of caring for families and communities, reproducing both the labour force and the social fabric. This approach argued for changing development strategies to make structural transformation more egalitarian (Sen and Grown, 1987).
This critique was reinforced by the emerging evidence that, unlike import-substitution industrialization, export-oriented industrialization did not marginalize women, but actually increased their share of manufacturing employment. Women’s share of employment in the growing service sector also increased. By 2003, women’s share of wage employment in non-agricultural sectors in Latin America and the Caribbean had reached 43.5 per cent, approaching that of the developed regions (46.4 per cent). In Eastern Asia, women’s share was 40 per cent, in South Eastern Asia 38.6 per cent and in sub-Saharan Africa 35.8 per cent. Lower shares prevailed in Southern Asia (18 per cent), Northern Africa (21.5 per cent) and Western Asia (aka the Middle East) (20.2 per cent) (data from UN Statistics Division, Millennium Indicators Database).

The rising trend in women’s share of employment in industrial and service sectors has been accompanied by a lively debate on the extent to which such employment has improved women’s lives and reduced gender gaps in well-being (see, for instance, for an early contribution, Elson and Pearson, 1981; and for a recent contribution, Kabeer, 2000; for a survey of the arguments, see Razavi, 1999). Gender inequalities persist in pay and conditions of employment, with most women segregated in a few occupations in which the vast majority of those employed are women (Anker, 1998). At the end of the twentieth century, on average, the hourly wages of women in developing countries were 73 per cent of those of men (compared to 77 per cent in developed countries). In both developed and developing regions, more than 80 per cent of the gender wage gap could not be explained by measurable differences in workers’ characteristics, and probably indicates discrimination in the labour market (World Bank, 2001, pp. 55–6). Women workers in developing countries are more concentrated than men in informal employment that lacks social protection; and within informal employment, in the more precarious types, with lower incomes (Chen et al., 2005). Informal employment has been growing as a share of total employment, as labour markets have become both more flexibilized and more feminized (Standing, 1999).

The conventional wisdom is that, despite problems of the quality of women’s paid employment, their earnings give them greater bargaining power within their households, and more influence over how household resources are allocated (see for example, Kakwani and Son, 2006). However, this is not universally the case (Elson, 1999). Women in Uganda report that when a woman starts to earn an income of her own, her husband is liable to withdraw his financial support, and shift to her the responsibility for paying school fees, medical bills, and buying food and clothing and other necessities for the whole household (Ellis et al., 2006, p. 24). Moreover, in most parts of the world, the division of unpaid domestic work between men
and women in the household is rarely renegotiated when women start earning (Kabeer, 2005). Case studies reveal that it matters where women obtain their earnings: for instance, in Ahmedabad, India, home-based paid work does not give women as much say in household decisions as employment outside the home (Kantor, 2003).

Women’s entry into the labour market was accelerated in the 1980s and early 1990s by stabilization and structural adjustment policies (Çağatay and Ozler, 1995). It has been suggested that this acceleration was widely linked to ‘distress sales’, as women are forced to try to make good shortfalls in household income following the loss of employment by male household members (for example, Moser, 1989; Gonzalez de la Rocha, 2000). Elson (1991) argued that stabilization and structural adjustment polices implicitly assumed unlimited supplies of female labour, available to make good through unpaid work in families and communities any shortfalls in provision of public sector non-tradeable services (such as health, education, water and sanitation), and to increase production of exports, while at the same time maintaining household food security and the social fabric of family and community networks. Moreover the theory ignored the gender norms that structure the division of labour, and mean that men’s labour tends not to be reallocated to ‘women’s work’, where there is a decrease in what is considered to be ‘men’s work’ (for example construction) and an increase in what is considered to be ‘women’s work (for example garment-making, unpaid care work). Instead, a more likely outcome is unemployment and underemployment for men (who do less paid work but little or no more unpaid work), and overwork for women (who do additional paid work as well as unpaid work). Failure to take this into account in designing adjustment policies, argued Elson, results in extra burdens for women, and risks deterioration in health, nutrition and education.

A number of case studies provide empirical backing for Elson’s argument (for example Moser, 1989; Lim, 2000; Tanski, 1994), but lack of data and methodological problems hinder definitive conclusions about whether women and girls have, on average, borne a greater share of the costs of adjustment (Haddad et al., 1995). Nevertheless, there is widespread empirical support for the conclusion that women are disadvantaged in the adjustment process, unless specific measures are taken to address pre-existing gender inequalities (Haddad et al., 1995).

One of the important gaps in data is comprehensive information on the extent of men’s and women’s unpaid work. Time-use surveys have been used for more than two decades in developed countries to remedy this. In the late 1990s such surveys began to be introduced in a growing number of developing countries. They confirmed what small-scale case studies had suggested: women and girls spend more time on unpaid work than men and
boys; and when both paid and unpaid work is taken into account, women and girls have a longer total working day than men and boys. For example, in South Korea, in 1999, the total working time for males was almost six hours a day on average, while for females it was almost seven hours a day. Males spent on average only 50 minutes per day on unpaid work, while females spent almost four hours (Tae-Hong, 2001, p. 8). A similar picture was revealed by a time-use survey for six states of India in 1998–99: on average female total working time was just over 7.5 hours per day, while for males it was 6.5 hours per day. Females spent almost five hours a day in unpaid work, and males spent only about 30 minutes (calculated from Chakraborty, 2005, Table 3). Charmes (2006, Table 3.2) shows a similar picture for sub-Saharan African countries. For example, in Benin in 1998 the total working time for females was on average almost 7.5 hours a day, while for males it was about five hours a day; females spent almost 3.5 hours per day on unpaid work, while males spent just over one hour. In Madagascar in 2001, the total working time for females was on average almost 6.5 hours per day, while for males it was almost 5.5 hours per day; females spent just over 3.5 hours on unpaid work, while males spent almost 50 minutes. In Mexico, in 1995, the total working time for females was on average just over 8.5 hours a day, while for males it was almost 7.5 hours a day. Females spent four hours in unpaid work, while males spent just over 1.5 hours (calculated from Elson, 2000, p. 102). Very few developing countries have conducted time-use surveys at regular intervals, so it is not possible to examine trends. Cross-country analysis that includes both developed and developing countries shows that higher gross domestic product (GDP) per capita is associated with a decline in time spent on unpaid work, and smaller gender gaps in total hours worked and in time spent in unpaid work (World Bank, 2001, p. 185). A key factor in this is investment in infrastructure and public services. Nevertheless, development does not eliminate unpaid work, as people value time to care for their family and friends (Folbre, 2001).

By the beginning of the twenty-first century, the context for discussion of gender issues in development had become the acceleration of globalization, the growth of income inequality between countries and between households within countries, and the growing differences in the trajectories of developing countries, with some, such as China and India, experiencing rapid structural transformation and growth, others, such as many countries in Latin America and the Caribbean, growing only slowly, and yet others, especially in sub-Saharan Africa, suffering deteriorating conditions.

Millions of women in China and India were not, however, benefiting from the rapid increase in national income because they were ‘missing’, in the sense of either having died prematurely, or not having been born at all.
This phenomenon was first brought to public attention by Amartya Sen (Sen, 1990a). More recent data (China 2000 and India 2001, see Klasen and Wink, 2003) confirms its persistence, and also its prevalence in middle-income developing countries such as Taiwan and Korea (Klasen and Wink, 2003). Preference for sons has not been reduced by rapid development. Indeed, in India the sex ratio imbalance is higher in the higher-income states than in the lower-income states.

Despite rapid growth in some developing countries, millions of people throughout the developing world continue to live in poverty. There has been considerable debate on whether poverty is ‘feminized’ in the sense of females being disproportionately represented among the poor (Çağatay, 1998; World Bank, 2001, p. 63). Measurement of poverty is dominated by a focus on consumption poverty, calculated on a household basis. This has led to a preoccupation with comparing the poverty rates of male- and female-headed households. The evidence is mixed: in some countries female-headed households are disproportionately represented in households below the poverty line; in others they are not (World Bank, 2001, p. 64). An alternative comparison is between proportions of the male and female populations that are in households below the poverty line. Again the evidence is mixed. In some countries a higher proportion of the female population than of the male population lives in households below the poverty line, but not in others (see for instance, ECLAC, 2002, Tables 6a and 6b). However, there is general support for the view that women are more vulnerable to poverty in old age than are men, as many more women are widows than men are widowers (World Bank, 2001, p. 67).

Moreover, women are certainly over-represented among the adult population who have no income of their own, because their participation in paid work remains lower than that of men. This limits their bargaining power within households and their capability to live a life of dignity, even if the household in which they live has an average income above the poverty line (Sen, 1990b). There is also plenty of evidence that poverty is differently experienced by males and females, and that gender gaps in education, health and work burdens tend to be larger in households in poverty (World Bank, 2001, pp. 61, 66).

Some studies have found that economic growth narrows gender gaps. For instance, Dollar and Gatti (1999) found that a higher level of per capita GDP was associated with greater gender equality in secondary schooling, in life expectancy and in representation in parliaments. However, when the labour market is brought into the picture the results are different for different groups of countries. Seguino used a composite gender equality indicator that includes relative labour force participation rates and female share of technical, professional and administrative positions, as well as the
indicators used by Dollar and Gatti, and clustered countries into four groups, ranging from poorest to richest in per capita terms. She found that in the highest and third-highest groups, there was a positive relation between growth and gender equality, but there was a negative relation in the lowest and second-highest income groups. Her conclusion was that economic growth is not sufficient by itself to achieve gender equality (UNRISD, 2005, p. 58).

These conclusions are reinforced by two regional studies of the impact of growth on gender equality. In Asia, in the period 1970–90, gender equality (assessed using a composite quality-of-life indicator that includes allowance for ‘missing women’) was highest in those countries that grew slowest (Seguino, 2002). In Latin America and the Caribbean, for the period 1970–2000, economic growth was found to have little beneficial effect on closing gender gaps in well-being, though the share of the manufacturing sector in GDP and the share of government expenditure in GDP were positively related to the reduction of gender gaps (Seguino, 2007).

Attention is now beginning to be focused on inequalities between women in developing countries. If inter-household inequality is rising, then it seems likely that inequality between women is also rising. There is as yet no comprehensive study on this topic. A recent study of maternal mortality and poverty in ten developing countries, based on demographic and health surveys, found that the proportion of women dying from maternal causes increases with the poverty of the households in which they live. For instance, in Indonesia, the probability of maternal death was three to four times greater in the poorest than in the richest quintile. In the Philippines and Tanzania the probability was two to three times greater (Graham, 2004). More research needs to be done on inequalities between women.

Gender and the determinants of economic growth and development

In this section, we examine whether gender inequality hampers or promotes economic growth and development. The 1990s saw the emergence of a growing literature on this topic, though it is still very small compared to the literature on the distribution of the fruits of growth and development. A number of cross-country econometric studies relating gender inequality to economic growth in developing countries is now available. They have produced a variety of results, depending on the dimensions of inequality considered, the selection of countries and the specification of the equations (World Bank, 2001).

Barro and Lee (1994) found a negative relationship between female secondary schooling and economic growth, though the relation between male secondary schooling and growth was positive. However, subsequent studies have found a positive relation between economic growth and gender
equality in education (for example Hill and King, 1995; Dollar and Gatti, 1999; Esteve-Volart, 2000; Klasen, 1999, 2002). Dollar and Gatti (1999) used data from over 100 countries covering three decades, and found that an increase in one percentage point in the proportion of adult women who have secondary education is linked to an increase in growth rates of per capita income of 0.3 percentage points per year. Klasen (2002) finds that the higher gender gaps in education in sub-Saharan Africa, compared to East Asia, and their slower reduction, accounted for 0.6 percentage points in the 3.5 percentage points difference in the growth rates in the two regions in the period 1960–92. Closing the gender gap in education enrolment by 2005 is a Millennium Development target, adopted by the UN General Assembly at the Millennium Summit in 2000. An estimate of the impact on the economic growth of countries that were not on track to meet this target found that they would have grown faster by about 0.1 to 0.3 percentage points if they had been on track to close the gap (Abu-Ghaida and Klasen, 2004).

Increasing the level of education of girls may increase growth in per capita incomes directly, by increasing the participation of women in the labour market and the productivity of women’s labour, and indirectly, by facilitating a transition from a high to a low rate of fertility (since educated women have fewer children). During the fertility transition, the working-age population grows at a higher rate than the dependent-age population. Bloom and Williamson (1998) refer to this as a ‘demographic gift’ and argue that it raises the rate of growth of per capita income during the transition (provided policies are in place to employ the working-age population productively). They estimate that it accounts for between 1.4 and 1.8 percentage points of growth in per capita income in East Asia, in the period 1965–90. Sub-Saharan Africa has yet to experience the transition. Its fertility remains high and its working-age population has not grown faster than its total population. A recent study of Uganda estimates that a period of sustained fertility decline could boost medium-term per capita growth rates by 0.5 to 0.6 percentage points a year (Klasen, 2005).

Nevertheless, questions remain about the direction of causation between the education of girls and economic growth. For instance, Robbins (1999) argued, in a study of six Latin American countries, that causation goes from increases in growth to increases in education of girls, rather than vice versa. He found that economic growth leads to rising educational attainment by drawing more women into the labour force, increasing the opportunity cost of women’s time, and thus reducing fertility and leading families to invest more in the education of their (fewer) children, girls as well as boys.

There is a general agreement that increasing the level of girls’ education improves outcomes for their children. Studies based on household data
show that the more educated are mothers, the lower their children’s mortality, controlling for household income and other aspects of socio-economic status; and child immunization rates rise with mother’s education (World Bank, 2001, pp. 79, 80). Cross-country regression analysis indicates that higher levels of female enrolment in school, and lower gender education gaps, are associated with lower rates of infant mortality (Hill and King, 1995). Increases in women’s education accounted for 43 per cent of declines in child malnutrition in 1970–95 (Smith and Haddad, 2000). A recent estimate of the costs of failing to achieve gender parity in educational enrolment by 2005 found that by 2015, such countries would have on average 15 per 1000 higher rates of under-five mortality and 2.5 percentage points higher prevalence of underweight children under five (Abu-Ghaida and Klasen, 2004).

Focusing on labour markets, rather than education, produces mixed results: greater gender equality in participation in the labour market seems to promote faster growth, but greater equality in wages does not. Recent studies on the Middle East and North Africa (Klasen and Lamanna, 2003) and India (Esteve-Volart, 2004) suggest that growth would be higher if the gender gap in labour market participation were reduced (through more women entering the market). However, cross-country regression analysis of growth and the gender wage gap in 20 semi-industrialized economies (such as those in East Asia) in the period 1975–95 found that, controlling for gender differences in educational attainment, gender wage inequality was a stimulus to growth. A 0.1 percentage point increase in the gap between female and male returns per year of secondary education is associated with a 0.1 percentage point increase in the growth of per capita GDP (Seguino, 2000). There is a strong inverse relation between the gender wage gap and the educational attainment gap in semi-industrialized countries (Seguino, 2005, Table 2). This implies that although narrowing gender gaps in education (through increasing the education of girls) tends to raise productivity, women’s lack of bargaining power in the labour market holds their wages down. Seguino argues that this stimulates growth through higher profits in female labour-intensive manufactured exports, higher investment and higher foreign exchange earnings. Of course, if high gender wage gaps held back household investment in girls’ education, the dynamic would be different, but this does not seem to have happened in semi-industrializing countries, especially in East Asia (Seguino, 2005, p. 22). Instead the combination of educated but cheap female labour has been critical to the expansion of exports of manufactures. Seguino has extended her analysis to compare the impact of different kinds of inequality on growth in 37 semi-industrialized countries in the period 1975–99. She finds that whereas income inequality between households is negatively associated with
growth, the gender wage gap is positively associated with growth (Seguino, 2005, p. 23).

However, in agrarian economies, in which self-employment and family labour are more important than wage labour, it seems likely that gender inequality in production can hamper economic growth. Lack of aggregate data means that cross-country regression analysis has not been used to investigate this relationship, but micro-level studies suggest that gender inequality in access to and control of assets and income is likely to hamper growth. For instance, several studies of smallholder agriculture found that agricultural productivity could be increased if female farmers had the same levels of inputs (such as fertilizer, land and labour) and education and training as male farmers (World Bank, 2001, pp. 85–6). The following examples are highlighted by Blackden and Bhanu (1999). In Burkino Faso, output could be increased by 10–20 per cent by shifting resources from men’s plots to women’s plots within the same household. In Kenya, increasing the education and input levels of female farmers to those of male farmers could increase yields obtained by women farmers by as much as 22 per cent. In Tanzania, reducing the amount of unpaid work that women have to do could increase household cash incomes of smallholder coffee and banana growers by 10 per cent, labour productivity by 15 per cent and capital productivity by 44 per cent. In Zambia, if women farmers enjoyed the same level of investment in agricultural inputs as men farmers, agricultural output could increase by up to 15 per cent.

Gender inequality in control of resources and division of responsibilities within households can hamper the expansion of agricultural exports, including non-traditional agricultural exports, by smallholders. In sub-Saharan Africa, the production of export crops often requires inputs of women’s labour on plots controlled by their husbands, while the cash income from this production is controlled by their husbands. The Poverty and Social Impact Assessment of Uganda’s Strategic Exports Initiative (Booth et al., 2003) found that such intra-household inequalities limit the export supply response in Uganda because women prefer to put more of their labour into producing crops on their own plots for household use and sale in local markets, rather than into producing export crops controlled by their husbands. This is corroborated by other studies on Uganda (for example, Muhereza, 2001), and there is evidence of similar effects in other countries, such as Zambia (Wold, 1997) and Burkino Faso (Smith and Chavas, 1999).

It is important not to overstate the role of household inequalities in constraining export expansion and productivity in smallholder economies (Whitehead, 2005). There are many extra-household gender inequalities that are also important, including unequal access to extension services,
markets, transport and credit, and unequal laws and social norms. Such inequalities also constrain the contributions to growth made by self-employed women in non-agricultural activities. This is important for future development, since the share of agriculture in employment and income generation has been declining and is expected to decline further. Self-employment is particularly important in sub-Saharan Africa. In this region (excluding South Africa), informal employment comprises 78 per cent of non-agricultural employment and self-employment represents 70 per cent of informal employment (ILO, 2002).

Many hopes have been invested in new forms of microfinance as a means of enabling self-employed women to both increase their own incomes and contribute to economic growth. During the 1990s there was a large expansion in poor women’s access to small loans from microfinance institutions, many of them based on the social collateral of group liability, rather than the economic collateral of individual assets, such as land. Optimism about the impact of microfinance reached a high in the 1997 Microcredit Summit, which issued a Declaration stating that:

> empirical evidence has shown that women, as a group, are consistently better in promptness and reliability of repayment. Targeting women as clients of microcredit programs has also been a very effective method of ensuring that the benefits of increased income accrue to the general welfare of the family, and particularly the children. At the same time, women themselves benefit from the higher status they achieve when they are able to provide new income (quoted in Mayoux, 2000, p. 3)

However, detailed evaluations of a range of programmes in Africa and Asia have shown that there are a number of limitations. Mayoux (2000, pp. 12–13) summarizes them as follows: women may not have control over the loans they get, but act as de facto intermediaries between male family members and microfinance institutions; even if women control the loans, they may not generate significant increases in income, because women are crowded into a narrow range of traditionally female low-return activities; even if there is an increase in women’s income, men may control its use, or men may reduce their own contributions to meeting household expenses. Kabeer (2005) finds that microfinance institutions that combine financial services with other forms of support, and which build the organizational capacity of poor women, are more likely to be empowering. Poor women need more than just loans and savings accounts.

To make a significant impact on their own incomes and on the national economy, women need to be able to move beyond informal microenterprises and develop small businesses that are registered and eligible for loans from the formal financial sector, and for support from government
agencies. There remain many legal and social barriers which prevent women, more than men, from formalizing and growing their businesses, as shown in detail in the case of Uganda by Ellis et al. (2006).

Facilitating women’s increased participation in the market economy, either in self-employment or in wage employment, will only promote economic growth if there is sufficient demand for their products and their labour. This raises the question of appropriate macroeconomic polices, to which we turn briefly in the next section.

**Macroeconomic policy challenges for gender-equitable development**

At the sectoral level, there is a great deal of knowledge about economic policies that work to improve the position of poor women, to reduce gender inequality and to promote growth and development (Grown et al., 2005). Less attention has been paid until recently to appropriate macroeconomic policies. The importance of examining macroeconomic policies from a gender perspective was emphasized in contributions to two special issues of *World Development* ( Çağatay et al., 1995; Grown et al., 2000). However, the World Bank paid little attention to macroeconomics in its 2001 policy research report, *Engendering Development*.

In this final section we briefly consider some aspects of monetary and fiscal policy and trade and investment policy that need to be addressed if the relations between growth, development and gender equality are to become mutually reinforcing.

Removing deflationary bias in monetary and fiscal policy is one important challenge. Employment has ceased to be a goal of macroeconomic policy. Instead, the focus is solely on financial variables, such as inflation, the fiscal deficit and debt-to-GDP ratios. Rates of inflation have been brought down to much lower levels than in the 1980s, but in many regions, this has been at a huge sacrifice in public investment, economic growth and decent jobs (for evidence, see, for example, UNRISD, 2005, p. 30). Deflationary bias in macroeconomic policy was identified as an important issue for women in the UN *World Survey on the Role of Women in Development* (UN, 1999), which argued that women disproportionately bear the costs of this bias. Much of the research and policy development on gender equality in employment has been focused on measures to enable women to compete with men on an equal basis. These measures are important, but they are not sufficient. To the extent that they are successful, they will simply redistribute some jobs from men to women. This will reduce gender gaps, but not in a way that provides ‘full and productive employment and decent work for all’, as called for by the UN Secretary-General (UN, 2006, p. 6). In order for gender equality to be realized in ways that ‘equalize up’, rather than ‘equalize down’, there needs to be an expansion
of the total number of decent jobs, as well as an improvement of women’s access to them.

Women are particularly likely to be disadvantaged by deflationary bias because it interacts with, and reinforces, other policy biases, such as male breadwinner bias, the assumption that men are more deserving of decent jobs because they are assumed to be the principal economic support of families, while women’s incomes are wrongly perceived to be merely supplementary, and not essential to family well-being (Elson and Çağatay, 2000, pp. 1354–56). Seguino (2003) finds this to be important in explaining why women are much more likely to be unemployed than men in Barbados, Jamaica and Trinidad and Tobago, using data from the period 1980–99. This inequality cannot be explained in terms of women being less educated than men, since women have a higher unemployment rate than men with the same education; nor in terms of the different sectoral distributions of men’s and women’s employment. While both male and female unemployment rates fell in economic upturns, male rates fell more than female rates. Male workers were the first to be hired in economic upturns, even in the female-intensive service sector.

Global unemployment rose from 5.6 per cent in 1993 to 6.2 per cent in 2003, and the female unemployment rate was slightly higher than the male rate; among young people, the gender gap was bigger (ILO, 2004). There were some regional differences: in the Middle East and North Africa, and Latin America and the Caribbean, the female unemployment rate was higher than the male, while in sub-Saharan Africa and East Asia the reverse was true (ILO, 2004, p. 2). However, in many countries, female unemployment rates are likely to underestimate the true extent of women’s unemployment because women are more likely to be ‘discouraged workers’ who respond to their failure to find jobs by ceasing to search for one actively, although they would like to have a job if one were available. The discouragement and underemployment of women appear to have been significant in the aftermath of the Asian financial crisis in 1997–98. For instance, in South Korea, the rate of job loss for women was higher than for men, but subsequently, male unemployment rates appeared to be higher than female rates, while a higher proportion of women than before were employed in various types of informal employment (UNRISD, 2005, p. 42).

In a pioneering study, Braunstein and Heintz (2006) investigated the link between monetary policies and gender equality in employment in 17 low- and middle-income countries in the period 1970–2003. They identified episodes when monetary policy led to contractionary inflation reduction, when the growth of employment fell below its long-run trend; and episodes when it led to expansionary inflation reduction episodes, when the growth of employment was faster than its long-run trend. (The contractionary
episodes were associated with real interest rates being maintained above their long-run trend – an indicator of deflationary bias.) The study found that in 67 per cent of the contractionary inflation reduction episodes, the female-to-male employment ratio fell below its long-run trend, indicating that women were disproportionately affected by the slowdown in employment. However, in expansionary inflation reduction episodes, there was no clear disproportionate effect on either women or men. The female-to-male employment ratio increased faster than the trend in 53 per cent of cases, and at or below the trend in 47 per cent of cases.

Braunstein and Heintz concluded that a policy of responding to inflationary pressures by raising positive real interest rates above their long-run trend, and reducing real money supply below its long-run trend, tended to be associated with a greater loss in female than in male employment (relative to long-run trends in both). They noted that in 33 per cent of contractionary inflation reduction episodes, women’s employment was not disproportionately affected by deflationary policies. They found that in these episodes, the real exchange rate either depreciated or remained at its long-run trend. They concluded that ‘maintaining a competitive exchange rate may offset some of the gender bias observed during contractionary inflation-reduction’ (Braunstein and Heintz, 2006, p. 12).

Inadequate levels of employment and decent work are also related to an emphasis on managing government budgets by cutting expenditure rather than raising tax revenues. Globalization has made it harder for governments to raise tax revenue. Çağatay (2003) summarizes the key aspects of this fiscal squeeze. Trade liberalization cuts import duties and export taxes, key sources of revenue in many poor countries. Competition to attract multinational corporations and their highly paid executives leads to cuts in corporation and capital gains taxes, and tax holidays and other exemptions and to cuts in top rates of income tax. Development cooperation grants have fallen as trade is supposed to replace aid. Governments have been encouraged or pressured into turning to sales taxes like value-added tax (VAT) to raise revenue, but such taxes fall most heavily on poor people and worsen the distribution of income. Revenue has also been raised through the sale of public enterprises and other public assets, but this only gives a one-time boost to revenue, and may result in costs for services, like water, that poor people cannot afford.

With revenue limited and debt burdens rising, the pressure has been on governments to make their budgets sustainable by cutting back on expenditure. This pressure has come from the public international financial institutions like the International Monetary Fund (IMF) and the World Bank, and also from private investors, who have seen budget deficits as harbingers of inflation, signals that the value of their assets would be eroded. In order
to build a reputation for ‘sound finance’ in financial markets, many governments have enacted legislation (such as balanced budget laws) that severely limits the fiscal space (Bakker, 2002).

More room for a variety of fiscal policies would not by itself ensure that fiscal policy is used to promote gender equality. Recognizing this, since the late 1990s, a series of gender budget initiatives (GBIs), in all parts of the world, have sought to improve the distribution, adequacy and impact of government budgets at national, regional and local levels; and to secure greater transparency in the use of public money; and to secure greater accountability to women as citizens. A number of tools have been developed for analyzing the gender dimensions of government budgets (Budlender and Sharp, 1998; Elson, 1998). By 2002, up to 50 countries in all parts of the world had hosted some kind of gender budget initiative (Budlender and Hewitt, 2002, p. 8). There is no one template: GBIs have taken place at all levels of government, involving regional and local government budgets as well as national budgets. Moreover, a multiplicity of actors have been involved: government ministers and officials (especially women’s ministries, sometimes Ministries of Finance), parliamentarians, women’s organizations and academic researchers (Budlender et al., 2002; Budlender and Hewitt, 2002). An important area of concern has been whether budgetary policies are reducing or increasing the amount of unpaid domestic work that poor women have to do; and whether they make it easier or more difficult for people (both men and women) to combine paid work and caring for their families without enduring excessive hours of work.

The most effective GBIs have produced some institutionalization of gender equality concerns in one or more stages of the budget cycle in one or more ministries, or have resulted in an ongoing public scrutiny of the budget from a gender equality perspective. Their aims have included: raising awareness and understanding of gender impacts of budgets and the policies they fund; making governments accountable for their budgetary and policy commitments; and changing and refining government budgets and policies to promote gender equality. Many examples of success in achieving the first two goals have been identified, though fewer in achieving the third (Sharp, 2002)

As well as the challenge of securing more gender-equitable public finance, there is also the challenge of securing policies on industry, trade and finance that enable increases in women’s wages without jeopardizing economic growth. This challenge is analysed by Seguino and Grown (2006), who identify the need for an approach they label ‘industrial policy under conditions of strategic openness’. This entails an industrial policy that promotes a shift of female employment to high-quality exports with a low price elasticity (so that higher wages for women will be less likely to have a
negative effect on exports). This should be complemented by policies that slow down the mobility of foreign direct investment, providing incentives for firms to respond to higher wages by investment in technological improvements, rather than by relocation. In addition, the maintenance of exchange rates should be maintained at competitive levels. Such a strategy would require a greater latitude for special and differential treatment of developing countries in international trade agreements.

It is also important to consider the impact of trade policies on unpaid domestic work. Pioneering research has been done on this topic by Fontana (2003), who used a computable general equilibrium model to simulate the impacts of trade liberalization on the paid and unpaid work of women and men in Bangladesh and Zambia. This technique is able to show the quantitative implications of the expansion of paid employment of women for the time they spend in unpaid domestic work and leisure, and the conditions under which there may be some redistribution of unpaid work between women and men.

The policy challenges identified above were considered by the United Nations Research Institute for Social Development in its comprehensive report on gender equality (UNRISD, 2005) The report concluded that to achieve greater gender equality, a new package of macroeconomic policies is required that puts more emphasis on redistributive taxation, gender-responsive public spending, the creation of decent work, universal social protection schemes, and policies to enable people more easily to reconcile their paid and unpaid work responsibilities, all in the context of a more just and equitable system of international economic relations (UNRISD, 2005). The challenge for the future is to ensure that not only micro and sectoral polices take account of gender issues, but also macroeconomic policies. This will require a more systematic incorporation of the unpaid work of the household sector into macroeconomic analysis, alongside the paid work of the public and private sectors.

References


UN (1999), World Survey on the Role of Women in Development, Division for the Advancement of Women, Department of Economic and Social Affairs, New York: United Nations.
Children and development

Paul Glewwe and Amy Damon

Introduction
In developing countries, 30 percent of the population is less than 15 years old. Thus any discussion of the well-being of the population in any developing country must examine the welfare of children. In general, the relationship between children and economic and social development is primarily one of the impact of development on children, rather than the impact of children on development. Yet today’s children will become adults in one or two decades, and events during their childhood will have a strong impact on their lives as adults, including their contribution to economic and social development.

This chapter summarizes recent research by economists on the status of children in developing countries. It begins by examining the status of children in those countries, and proceeds by assessing the impact of economic growth on children. It then presents policies that are likely to improve children’s health, education and employment status. The final section presents some concluding remarks.

Child welfare in developing countries
By virtually any measure, the welfare of children in developing countries has improved markedly over the past half-century, not only for developing countries as a whole but for all regions as well. This section reviews this progress and presents the current status of children with respect to their health, educational status and employment status.

Health
In almost every dimension the health status of children is improving in nearly all developing countries. Progress in reducing infant and child mortality rates is shown in Tables 51.1 and 51.2. These rates show how many children died, out of every 1000 born, before their first and fifth birthdays, respectively. For developing countries as a whole, the infant mortality rate has dropped dramatically, from 126 in 1960 to 53 in 2000, and the child mortality rate has declined sharply, from 197 to 78, over the same time period. This decline has occurred in all regions of the developing world.

Despite this progress, infant and child deaths are still common in developing countries. The situation in sub-Saharan African countries is especially
worrisome. In particular, progress in reducing infant and child mortality was much slower between 1980 and 2000 than it was from 1960 to 1980. While rates were similar in sub-Saharan Africa and South Asia in 1960 and 1980, and both made good progress in reducing rates in those two decades, South Asia was able to reduce infant and child deaths much more quickly between 1980 and 2000. The most obvious explanation is the advent of AIDS in sub-Saharan Africa, but slow economic growth is also likely to have played a role.

**Education**

Schooling is another area in which child welfare has increased in the last half-century, as seen in Tables 51.3 and 51.4. These tables present gross enrollment rates, which are defined as the number of children enrolled as primary (or secondary) students divided by the number of children in the age range associated with that level of schooling. Note that it is possible for these enrollment rates to exceed 100 because it is common for ‘over-age’ children to be in a particular level because of delayed initial enrollment or grade repetition.

In developing countries as a whole, the primary school gross enrollment rate increased from 68 percent in 1960 to 99 percent in 2000. These

### Table 51.1 Infant mortality rate (per 1000 live births)

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>116</td>
<td>65</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>14</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>65</td>
<td>46</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>19</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>95</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>25</td>
<td>26</td>
<td>32</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>151</td>
<td>83</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>13</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>166</td>
<td>119</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>162</td>
<td>118</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>38</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>126</td>
<td>82</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>117</td>
<td>135</td>
<td>147</td>
</tr>
</tbody>
</table>

*Source: World Bank (2002b).*
Table 51.2  Child mortality rate (per 1000 live births)

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>175</td>
<td>94</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>14</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>88</td>
<td>57</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>19</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>139</td>
<td>77</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>25</td>
<td>26</td>
<td>32</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>237</td>
<td>117</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>13</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>265</td>
<td>182</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>8</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>268</td>
<td>191</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>38</td>
<td>45</td>
<td>47</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>197</td>
<td>124</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>117</td>
<td>135</td>
<td>147</td>
</tr>
</tbody>
</table>


Table 51.3  Primary school gross enrollment rate

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>77</td>
<td>102</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>8</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>101</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>21</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>89</td>
<td>101</td>
<td>112</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>23</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>60</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>8</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>33</td>
<td>79</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>6</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>42</td>
<td>75</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>37</td>
<td>38</td>
<td>36</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>68</td>
<td>89</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>103</td>
<td>110</td>
<td>107</td>
</tr>
</tbody>
</table>
increases occurred in all regions of the developing world (except in Europe and Central Asia, where the enrollment rate was already 101 percent in 1960). Increases in the secondary school gross enrollment rate are even more striking, increasing from 19 percent in 1960 to 61 percent in 2000. These rates also increased in all regions.

Despite these gains, there are some troubling patterns. First, primary school enrollment rates are still below 100 percent in sub-Saharan Africa, South Asia and the Middle East and North Africa. Indeed, heavy grade repetition exaggerates the percentage of children who are actually in primary school. Sub-Saharan Africa and South Asia also lag behind in their secondary school enrollment rates, although they are increasing rapidly over time. Second, there is evidence that children in developing countries learn much less than children in developed countries for a similar amount of time in school. These issues are discussed in detail in Glewwe and Kremer (2006).

Employment
School-age children work in many poor countries, sometimes so much so that they leave school. Most children who work in developing countries work alongside of their parents in agricultural activities, but some work in factories or in other activities in urban areas. Concerns regarding child

Table 51.4 Secondary school gross enrollment rate

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>15</td>
<td>41</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>52</td>
<td>91</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>21</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>18</td>
<td>47</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>23</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>12</td>
<td>37</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>10</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>4</td>
<td>17</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>36</td>
<td>36</td>
<td>27</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>19</td>
<td>42</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>102</td>
<td>108</td>
<td>91</td>
</tr>
</tbody>
</table>
labor have received increased attention in policy circles in the last one to
two decades.

While child labor has received increased attention, child labor itself has
decreased in almost all developing countries. The school enrollment trends
discussed in the previous subsection suggest that this is the case, and the data
in Table 51.5 confirm this. In developing countries as a whole, the labor force
participation of children aged 10–14 has been reduced by almost half from
1960 to 2000, from 24.4 percent to 13.5 percent. Yet it is still high in two
regions, South Asia and sub-Saharan Africa, which are the two regions with
the lowest school enrollment rates. This may reflect lower economic growth
in those two regions, an issue that is explored further in the next section.

The impact of economic growth on child welfare

Economic development has often been equated with income growth, but
development includes not only income growth but also better health out-
comes, higher levels of education, better housing (including potable water
and hygienic sanitation conditions), and perhaps even democracy and
respect for human rights. While it is very rare for income growth not to be
accompanied by these other aspects of the quality of life, the nature of
economic growth can determine how quickly economic growth leads to

<table>
<thead>
<tr>
<th>Region</th>
<th>Statistic</th>
<th>1960</th>
<th>1980</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>East Asia</td>
<td>mean</td>
<td>28.4</td>
<td>21.3</td>
<td>12.8</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Europe/Central Asia</td>
<td>mean</td>
<td>3.8</td>
<td>1.1</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>27</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>Latin America/Caribbean</td>
<td>mean</td>
<td>14.5</td>
<td>9.8</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>Middle East and North Africa</td>
<td>mean</td>
<td>17.2</td>
<td>10.9</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>South Asia</td>
<td>mean</td>
<td>40.0</td>
<td>31.5</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Sub-Saharan Africa</td>
<td>mean</td>
<td>40.6</td>
<td>35.2</td>
<td>27.7</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>All developing countries</td>
<td>mean</td>
<td>24.4</td>
<td>19.3</td>
<td>13.5</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>134</td>
<td>134</td>
<td>134</td>
</tr>
</tbody>
</table>
improvements in other areas. This section presents evidence on the potential role that economic growth can play in leading to improvements in child welfare, first by showing that per capita income is positively correlated with indicators of child welfare and that income growth is correlated with improvements in child welfare, and then citing several studies that examine more carefully the causal impact of income on child welfare.

**Correlation between per capita income and child welfare**

Nations with higher per capita incomes have, on average, higher levels of child welfare, as seen in Table 51.6. More specifically, in the year 2000 in low-income countries (those with per capita income levels of $755 or less), 79 children out of every 1000 born died before their first birthday, and another 42 died before their fifth birthday. In contrast, in middle-income countries (those with per capita income levels above $755 but below $9266) only 27 out of 1000 children died before their first birthday, and only another seven died before their fifth birthday.

Middle-income countries also have much lower rates of child labor force participation than do low-income countries: 4 percent of children aged 10–14 work in middle-income countries, but 22 percent work in low-income countries. Primary and secondary school (gross) enrollment rates are also higher in middle-income countries; the rates for those countries are 110 percent and 77 percent, respectively; while the rates in low-income countries are much lower, namely 88 percent and 40 percent.

Not only is child welfare higher in countries with higher income, but improvements in child welfare are also positively correlated with the rate of

### Table 51.6 Income levels and child welfare

<table>
<thead>
<tr>
<th>Income level</th>
<th>Statistic</th>
<th>Infant mortality rate (per 1000 live births)</th>
<th>Child mortality rate (per 1000 live births, children &lt; 5 yrs)</th>
<th>Child labor rate (children 10–14)</th>
<th>Primary school gross enrollment rate</th>
<th>Secondary school gross enrollment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income</td>
<td>mean</td>
<td>79</td>
<td>121</td>
<td>22</td>
<td>88</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>72</td>
<td>72</td>
<td>69</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>Middle income</td>
<td>mean</td>
<td>27</td>
<td>34</td>
<td>4</td>
<td>110</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>$n$</td>
<td>72</td>
<td>72</td>
<td>59</td>
<td>56</td>
<td>51</td>
</tr>
<tr>
<td>Low and middle</td>
<td>mean</td>
<td>53</td>
<td>77</td>
<td>13</td>
<td>100</td>
<td>61</td>
</tr>
<tr>
<td>income</td>
<td>$n$</td>
<td>144</td>
<td>144</td>
<td>128</td>
<td>104</td>
<td>89</td>
</tr>
</tbody>
</table>
income growth. This is seen in Table 51.7. Data from about 90 countries are used to divide countries into the 25 percent that had the slowest rates of growth from 1980 to 2000, the 25 percent that had the fastest economic growth, and the 50 percent with ‘intermediate’ rates of economic growth. Child labor and infant mortality drop faster in countries with higher economic growth. For example, from 1980 to 2000 labor force participation of children aged 10–14 dropped by 55 percent in countries with high economic growth but only by 43 percent in countries with low economic growth. Similarly, the infant and child mortality rates decreased by almost half in countries that had high economic growth while decreasing only by about one-quarter in countries with low economic growth.

The causal impact of income on child welfare

The correlations shown in the previous subsection are consistent with the hypothesis that income growth causes improvements in child welfare, but they do not constitute proof of a causal relationship. This subsection briefly reviews several recent studies that use microeconomic (household survey) data to provide more convincing evidence of a causal relationship.

Income growth can lead to improvements in child welfare, and in social welfare more generally, by two distinct pathways. The first, and most obvious, pathway is that households with higher incomes can purchase goods and services that improve children’s health and education outcomes, and improve child welfare in other ways. The theoretical literature has developed formal models that show how these income effects can occur (see, inter alia, Basu, 1999; Glewwe, 2002). Second, households with higher incomes, and more generally higher-income economies, generate more tax revenue (via either direct or indirect taxes) that governments can use to provide health, education and other services. Household survey data can be used to search for causal relationships that operate through one or both of these pathways.

Table 51.7 Changes in child welfare by rate of per capita income growth

<table>
<thead>
<tr>
<th>Growth rate</th>
<th>Statistic</th>
<th>% Δ Child labor</th>
<th>% Δ IMR</th>
<th>% Δ CMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow growth (bottom 25%)</td>
<td>mean</td>
<td>−42.7%</td>
<td>−23.0%</td>
<td>−23.6%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Medium growth (25th–75th percentile)</td>
<td>mean</td>
<td>−49.3%</td>
<td>−40.5%</td>
<td>−43.2%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>46</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>Fast growth (top quartile)</td>
<td>mean</td>
<td>−55.0%</td>
<td>−47.9%</td>
<td>−51.9%</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>
Four recent studies using data from Vietnam show causal impacts of household income and/or provision of education and health services on child health, education and child labor outcomes. Glewwe and Jacoby (2004) present a dynamic model of school attainment that focuses on the role played by household wealth. They use panel data from the 1990s to show that increases in household wealth over time lead to a substantial and statistically significant increase in years of schooling, even after controlling for changes in the quality of schooling, the rate of return to education and the opportunity cost of (child) labor.

A study by Glewwe et al. (2004) using the same data examines the impact of economic growth on children’s nutritional status, as measured by their height for age. Unlike the case with education, household income by itself has little impact on children’s nutritional status. This suggests that improvements in the nutritional status of Vietnamese children in the 1990s were primarily due to improvements in health care services. Unfortunately, there are no data available on changes in the quality of health care services over time, so the paper cannot present strong evidence in favor of this conjecture. However it does present some evidence that higher-quality health care facilities lead to improvements in children’s nutritional status.

Research by Wagstaff and Nguyen (2004) examines the factors that affect child mortality rates in Vietnam. The authors find that access to safe drinking water, vaccination campaigns and access to trained medical personnel during childbirth reduce child mortality in that country. In contrast, but consistent with Glewwe’s result for child nutritional status, they find no impact of household income on child mortality.

Finally, Edmonds (2005) shows that children in better-off households in Vietnam are less likely to work, and that households whose incomes increase over time are more likely to keep their children in school and less likely to put them to work. This pattern is seen in other countries as well, as discussed in Edmonds and Pavcnik (2005).

**Policies to promote child welfare**

The evidence from the previous section demonstrates that both increases in households’ disposable incomes and increased government spending on social services lead to improvements in child welfare. This section reviews what developing countries can do to improve child welfare using policies that work through both types of causal pathways.

**Promote economic growth**

Macroeconomists and other economists have had, and continue to have, long debates on the best way to promote economic growth. For recent summaries of the evidence, see Easterly (2005), Rodrik (2005) and the World...
Bank (2002a). Unfortunately, there is disagreement on what has been learned, but there are some areas of agreement. This chapter is too brief to delve into the details of the best policies to promote economic growth, but there is broad agreement among economists that economic growth is necessary for large, sustained improvements in child (and adult) living standards in developing countries.

**Health policies**

A recent book by Wagstaff and Claeson (2004) provides a detailed assessment of the effectiveness of health policies in developing countries. Policies that are effective in raising children’s nutritional status include improved hygiene and sanitation, dietary supplements that provide iron and vitamin A (for both mothers and children), and provision of deworming drugs to school-age children. To reduce infant and child mortality, the authors recommend improved hygiene sanitation, provision of mosquito nets that are treated with insecticide, child immunizations, dietary supplements of zinc and vitamin A, and improved facilities for childbirth.

Finally, it is important to realize that child health has large implications for children's educational outcomes. Several recent studies have found sizeable and statistically significant positive impacts of child health on education outcomes. Thus there is growing evidence of a causal impact of child health on education. Note as well that there is no clear evidence of large gender differences in the impact of child health on education outcomes. For details on the relationship between health and education, see Glewwe and Miguel (2008).

**Education policies**

This subsection summarizes recent research on policies that lead to increases in the years children spend in school and in the skills learned while in school. For details, see Glewwe and Kremer (2006).

Several education policies have been found to be effective in increasing the number of years that children spend in school, as well as their daily attendance during those years. More specifically, programs that reduce the costs of schooling faced by parents or provide incentives for daily attendance (either explicitly or implicitly through school meals) have sizable impacts on school enrollment and attendance. Randomized evaluations of school-based health programs, for example a deworming program in Kenya, suggest that, in some situations, these programs can be an extraordinarily cost-effective means of increasing the amount of time that children in developing countries are in school.

Evidence concerning the impact of education policies, such as provision of textbooks and additional training for teachers, on the skills that children
acquire in school is more mixed. In general, studies based on cross-sectional data suggest that most education policies have had limited impacts on the academic skills of schoolchildren in developing countries. Evidence from recent ‘natural experiments’ in middle-income countries suggests that reducing class size can raise academic achievement but that providing computers has little effect. Recent randomized trials conducted in low-income countries provide a more mixed picture.

The evidence suggests that the most effective forms of spending on education are likely to be those that respond to inefficiencies in schooling systems. Providing textbooks written with atypical students in mind will benefit only atypical students, whereas remedial education may be extremely effective in an environment in which many students fall behind and are no longer able to follow their teachers’ lessons. Providing radio mathematics education or computer-based education may be effective when teachers attend irregularly.

Schools in developing countries face significant institutional problems: distortions in education budgets often result in inefficient allocation and spending of funds; weak teacher incentives lead to problems such as high rates of teacher absenteeism; and, given the difficulties faced by these school systems, curriculums are often inappropriately matched with the level of the typical student. Yet reform initiatives can easily have unintended consequences. The details of these programs are critical for determining their effects on the incentives faced by teachers and others (principals, parents, and so on). Governance reforms and allowing school choice appear to hold more promise than simply providing monetary incentives to teachers based on test scores, but much more empirical evidence is needed on the impact of these reforms as well.

Employment policies

Of particular interest in recent years has been the issue of child labor. Clearly, children who work long hours cannot attend school, and there are many children who do work that could be directly harmful to their welfare. On the other hand, most children who work are working for their parents either on the family farm or on a rented plot of land, and their contribution towards their families’ income may have important welfare benefits for them and for other family members. This raises the issue of when, if ever, it is appropriate for governments to oppose choices that parents make for their children.

Two related policies that should reduce child labor without attempting to interfere with the choices that parents make for their children are to improve school quality and reduce the cost of attending school. Better and less-expensive schools will make schooling more attractive and thus should
persuade parents to allocate more of their children’s time to schooling and less to child labor. In contrast, Edmunds and Pavcnik (2005) argue that attempts to ban child labor are likely to be ineffective and, if effective, may harm poor families. For further discussion, see Edmunds and Pavcnik (2005) and the references cited in that paper.

Concluding comments
Child welfare has improved in all regions of the developing world since the 1960s, and continued improvements in the future are likely. Even so, there are still opportunities to increase children’s schooling, raise their health and nutritional status, and reduce the amount of time they spend working to support their families. Policies to promote economic growth are one general avenue to accelerate progress, but there are other policies that are also effective, as briefly summarized in the preceding section. Efforts should focus on South Asia and sub-Saharan Africa, where the welfare of children is lowest.

A final important point is that improved child welfare today translates into better economic and social development in the future. Children who have higher levels of education and better health are more productive workers and thus contribute more not only to their own welfare as adults, but also to overall economic growth. Lastly, more-educated and healthier children will also be better able when they are adults to provide a better life for their own children.

References


The economic geography of ethnic and racial inequality

Virtually all countries have ethnic/racial divisions. Where those divisions are present they are universally characterized by dense gaps in economic status between the respective groups. Intergroup division and inequality occur in countries at all levels of per capita income or development more broadly construed. To the extent that intergroup inequality is significantly driven by discriminatory practices, those practices have a persistence that defy the standard prediction of conventional economics that discriminators will be driven from the marketplace because their actions are unprofitable. Furthermore, despite the strong and widely popularized claims of some scholars, there is no clear general relationship between ethnic/racial division within a country and its overall economic performance. Nor is there any sound reason to believe that causation is typically unidirectional from ethnic division to economic performance. But high rates of economic growth do not necessarily erode ethnic division.

Research that one of the authors of this chapter has undertaken with other collaborators (Darity and Deshpande, 2000; Darity and Nembhard, 2000) demonstrates the international prevalence of ethnic/racial differentiation and the strong correspondence between such differentiation and economic disparity. Darity and Nembhard (2000), in particular, demonstrate that intergroup differentiation and inequality is present in countries with large and small populations (for example the USA and Belize), countries with high and low average income levels (for example Japan and India), countries with a recent experience of rapid economic growth and slow economic growth (for example Malaysia and New Zealand), and countries with comparatively high degrees and low degrees of general inequality (for example Brazil and Australia). Indeed, even countries that are more inclusive toward women in their national political processes or that display greater gender equality do not consistently display reduced levels of ethnic/racial inequality.

Moreover, inequalities between ascriptively differentiated groups can be detected using a variety of measures of disparity. Intergroup gaps exist in income, occupational status, access to quality education, access to quality health care, wealth, and related measures of well-being. With respect to wealth, for example, in the USA, one of the world’s most affluent countries,
blacks and Latinos have approximately a mere one-tenth of the net worth of whites at the median (Kochhar, 2004).

Ethnic and racial inequality and the Human Development Index
In the 1994 edition of the United Nations Development Programme’s (UNDP) *Human Development Report* ethnic/racial inequality was described across a number of countries by calculating disaggregated values of the UNDP’s Human Development Index (HDI). The HDI is a measure of well-being that includes not only per capita income as a component but also indicators of educational attainment and health status in a population. Again, cross-national comparisons suggest that intergroup differentiation and inequality is evident in countries with scores at all levels of the HDI (Darity, 2002).

Within countries intergroup disparity measured by the HDI is widely evident as well. In 1992, toward the end of the apartheid era, South Africa’s overall HDI score was 0.65. However, the score estimated separately for white South Africans was 0.88, a score that would have placed white South Africa 24th out of approximately 180 countries worldwide (at the same level as Spain). In contrast, blacks in South Africa would have had an HDI score of only 0.46, placing them 123rd out of 180 countries, ranking slightly above the Congo.

Disaggregated estimates for Brazil were provided by region rather than by race or ethnicity. But the specific regions utilized in the UNDP, the south and the north-east of Brazil, provide information about racial/ethnic inequality there. The population of the south is disproportionately white, consisting largely of Euro-Brazilians, while the north-east is disproportionately black and mulatto or Afro-Brazilian. The overall HDI for Brazil as a whole was 0.76 in 1992, placing the country 63rd internationally. However, southern Brazil had a score of 0.84 which would have placed that region 43rd in the HDI rankings. North-eastern Brazil’s score was 0.55; the region would have ranked in the lower half of the world on the basis of HDI.

Similarly for Nigeria the *World Development Report* did not provide direct information on HDI scores by ethnic group. Scores were provided for each of the 19 Nigerian states. This afforded indirect information about intergroup inequality in Nigeria because each of the states are relatively ethnically homogeneous internally. It is striking that while Nigeria’s overall HDI was a very low 0.348 in 1992, ranking the country 139th in the world, the state of Bendel had a score of 0.67, higher than Sri Lanka or Cuba, while the state of Borno had an HDI of only 0.16, beneath any country in the world.

Canada’s HDI in 1992 of 0.93 was the highest in the world. Nevertheless, even there the evidence of intergroup disparity was dramatic. While it was
not possible to construct separate HDIs for Canada’s ethnic/racial groups with the available data, the data ‘do show that the “aboriginals” (the Indians, the Inuit, and the Metis, constituting 2.3% of the population) have a life expectancy 5.6 years lower than the rest of the population, and their real income is one-third less’ (UNDP, 1994, p. 100). Indigenous people were far more likely to be subjected to violence, experience depression and undergo unemployment. Their unemployment rate of 20 percent was twice the national average in Canada at the time (UNDP, 1994, pp. 25–6, 32).

The record also indicates that discrimination generally plays an important role in maintaining conditions of economic disparity, particularly discrimination in employment and housing. Plus, despite the theoretical presumption of conventional economics, the historical record gives little reason to believe that such discriminatory practices will disappear with the passage of time, even in market-based economies. In some countries, for example Brazil, there has been evidence of an increase in discriminatory differentials against blacks and mulattos over time (Darity, 1998). In others, like the USA and South Africa, where legal regimes of segregation have been overturned, there is evidence of a decline in discrimination in the period immediately following the regime change, but still high levels of discrimination persisting thereafter (Darity and Deshpande, 2000; Darity, 2002). Statistical estimates actually indicate that measured discrimination increased against Puerto Rican, Mexican and native American ancestry men in the USA between 1980 and 1990 (Darity, 2002).

Ethnic conflict, neoinstitutionalism and economic growth

A provocative and influential paper by William Easterly and Ross Levine (1997) oriented the discussion of the role of ethnicity toward an examination of its impact on economic development, rather than its impact on intergroup disparity. Easterly and Levine argued that the low rates of economic growth characteristic of African nations’ economies since the 1970s are attributable to the high levels of ethnic diversity there. Ethnic variation in a country, in and of itself, would lead to rent-seeking practices that would prove to be predatory on effective governance and contribute directly to slower growth. In the most recent versions of the argument (Easterly, 2001; Easterly et al., 2006), taking a tone very similar to Robert Putnam’s (2000) approach to conditions for community health, unity and participatory democracy, ethnic diversity becomes a critical factor undermining ‘social cohesion’ and thereby undermining the quality of ‘institutions’ that might otherwise promote economic growth.

One implication of the Easterly and Levine position is African governments would be larger – inefficiently larger – relative to the scale of their respective economies than governments in parts of the world with less
ethnic diversity. But Dani Rodrik (2000) finds that the public sectors in African countries are not generally comparatively larger by international standards, and he shows that there is at most a weak correlation between the size of the public sector and the magnitude of rent-seeking activity and corruption in a country.

Easterly and Levine focus on ethnic diversity – simply the magnitude of ethnic variation that exists in a country. They give considerably less attention to the forms of ethnic antagonism (for a catalogue of these forms see Bardhan, 1997) that might play out in different settings and their consequences. One can easily conclude that racial/ethnic differentiation that results in genocidal violence and/or conditions of civil war will not be good for economic growth. But this requires more than mere diversity; it requires high levels of group consciousness and high levels of between-group conflict. Indeed, in the context of a model utilizing evolutionary game theory it can be demonstrated that the continuation of group consciousness is dependent upon unequal resources being associated with identification with and membership in each group (Darby et al., 2006).

During the midst of the Burundian genocide of the early 1990s – in this case directed by the Tutsis against the Hutus – Leonce Ndikumana (1993, p. 30) described the country as possessing ‘a rigid ethnic stratification and unequal distribution of power along ethnic lines. While the Hutu make up the majority of the population (about 85 percent), the minority Tutsi (about 14 percent) control the government, the military, and the economy. The third ethnic group, the Twa, has assumed a role of second-class citizens with little integration in the economic and political system’. The Burundian genocide at that time was associated with an ‘annual decline in agriculture (value added) [that] reached −10.5% in 1994, exceeding the decline experienced after the 1972 massacre (−10%), [and] total production (GDP) . . . declined at an annual rate of over 6% in 1993 and 1994’ (Ndikumana, 1993, p. 30). Between 1960 and 1998 Nkurunziza and Ngaruko (2002) estimate that Burundian per capita income fell from $620 to $370. Without the civil wars they estimate that per capita income would have been $667 in 1997 instead of $397.

In neighboring Rwanda, five years of civil war culminated in the 1994 genocide directed by the majority Hutus against the minority Tutsis. Gross domestic product (GDP) fell in three of the five years; in 1994 there was an astronomical fall in GDP of 40 percent. Only the infusion of massive amounts of foreign aid led to a 9 percent growth rate in 1995 (US Department of State, 2007).

Burundi and Rwanda are extreme cases. Such extreme forms of inter-ethnic conflict are not unique to the African continent. Inter-ethnic strife leading to genocidal violence has occurred in the former Yugoslavia, in
Indonesia in 1997 and 1998, and in Cambodia under the Khmer Rouge, all contributing to economic decline. The UNDP (1994, p. 47) identifies Turkey, the United Kingdom (especially Northern Ireland), Iraq, Iran, Israel, Lebanon, Colombia, Guatemala, Bangladesh, India, Laos, Myanmar, Pakistan, the Philippines, Sri Lanka and Tajikistan as non-African nations where ethnic conflict has turned violent with adverse consequences for economic performance. France’s ethnic violence in late 2005 also fits the model.

Generally, Robert Bates (2000) argues that the level of political violence represented by riots, demonstrations, revolts and assassinations is lower in African countries than might be expected, given the level of ethnic division there. While the short-term effects on growth of political violence, particularly genocidal violence, are negative, the long-term effects on economic growth are ambiguous:

Even a genocidal process of ethnic homogenization of population or ethnic homogenization of control over a nation’s resources is not inimical to prosperity, at least for the ‘winners’ and their descendants. Indeed, wealth seizures in the form of conquest of native peoples and appropriation of their lands, coupled with the use of captive and slave labor, laid the basis for the affluence of today’s richest nations, for example, the United States, Australia, Britain, France, Belgium and the whites of southern Africa.

Theft via conquest has long constituted an effective mechanism for achieving redistribution of wealth among groups. Industrialization, by destruction of lives of indigenous peoples, has been a commonplace event during the past half millennium. Violence is the historic adjunct to compulsory wealth redistribution across racial or ethnic lines. (Darity, 2002, p. 133)

Even in the near term, if the population being exterminated is located on an ‘undeveloped’ frontier, it may even be possible for economic activity to proceed in the ‘developed’ region of a nation undisturbed by the genocide. Indeed, do the genocidists even count the persons who are being eliminated as part of their relevant national population when they are computing the economic consequences of their actions? It is also possible that if the genocidal violence reduces total population sufficiently, in principle, it could offset a decline in national output sufficiently to produce an increase in per capita income. This perverse possibility reinforces the importance of Amartya Sen’s (1997) warning that one should not confuse increases in income with improvements in human well-being.

The significance of forced intergroup wealth redistribution for the long history of economic growth and uneven development is the core theme of Eric Williams’s (1994 [1944]) now classic study Capitalism and Slavery. Acemoglu et al. (2001) and Acemoglu (2003) edge onto this story but back away with a neoinstitutionalist explanation of the divide in the world.
between rich and poor nations. In his paper ‘Root Causes’, Acemoglu (2003, p. 27) proposes that there are two principal explanations of ‘the fundamental causes of prosperity between countries . . . geography and institutions’. For Acemoglu (2003, p. 27):

Good institutions [that promote economic development] have three key characteristics: enforcement of property rights for a broad cross section of the society, so that a variety of individuals have incentives to invest and take part in economic life; constraints on the actions of elites, politicians, and other powerful groups, so that these people cannot expropriate the incomes and investments of others or create a highly uneven playing field; and some degree of equal opportunity for broad segments of society, so that individuals can make investments, especially in human capital, and participate in productive activities.

Presumably, one must assume that industrialization in the United States in the midst of the consolidation of a regime of legal segregation was not a sufficient violation of the institutional conditions that Acemoglu lauds to invalidate his hypothesis.

Precisely why he settles on these two as the central explanations is not made clear. In the process Acemoglu eliminates from consideration the Williams perspective that emphasizes the hothouse effects on European economic development produced by the Atlantic slave trade and the slave plantation system in the Americas. Williams’s perspective places the stress on the role of colonialism in explaining variations in prosperity in the world economy today. There is at least one additional explanation which, thankfully, Acemoglu does not invoke as an option – variations in national cultures.

When all is said and done, Acemoglu (2003, pp. 27, 29) contends that institutional strength trumps geography in the following key passages:

if you look around the world today you’ll see almost no wealthy country achieves this position without institutions protecting the property rights of investors and imposing some control over government and elites.

geography neither condemns a nation to poverty nor guarantees its economic success. If you want to understand why a country is poor today, you have to look at its institutions rather than its geography.

Note that the ‘equal opportunity’ feature of ‘good institutions’ in Acemoglu’s comment about ‘look[ing] around the world today’ is notably absent.

But what determined whether a country developed ‘good institutions’ on two out of three dimensions? For those regions of the world that underwent the process of colonialism, Acemoglu argues that it is the type of colonialism they experienced that set the path for the quality of their institutions. Colonies where Europeans established ‘extractive societies’
tend to have a poor institutional framework today, while colonies where Europeans established ‘settler societies’ tend to have a positive institutional framework. Hence the places where resource extraction took place, but Europeans did not migrate in large numbers, tend to be poor while the places where Europeans relocated tend to be more affluent. Certainly this would be fuel for the Eurocentric cultural determinist, but Acemoglu avoids that trap. Instead he argues that the form that colonialism took shaped the incentives faced by local elites for institutional development that pushed the two types of colonial systems on distinct long-run paths.

From Eric Williams’s perspective the distinction between colonies of extraction and colonies of settlement is artificial. In the latter there consistently were forms of extraction that took place – expropriation of the land by the settlers, the exploitation of native and slave labor, as well as other forms of coercion. The division of the world into rich and poor countries went hand in hand with the racialization of the colonial process. The colonizers were enriched whether they remained in their home country or moved to a site of settlement; the colonized were impoverished. These are the parallel elements of uneven development – development for some and underdevelopment for others. Simultaneously, ethnic divisions were crystallized and sustained to make it possible for the ‘winners’ of the colonial game to continue winning. The fascinating sixth chapter of Williams’s (1942) *The Negro in the Caribbean*, entitled ‘The Middle Class and the Racial Problem’, provides a rich explication of the emergence of the interaction of color and class stratification as a consequence of racialized colonialism.

In another paper Acemoglu et al. (2005, pp. 546–7) highlight the centrality of Atlantic trade in the period between 1500 and 1800 as the foundation for European economic development. They contend that countries best able to take advantage of the growth opportunities afforded by the Atlantic economy were those ‘with relatively nonabsolutist initial institutions, most notably in Britain and the Netherlands [i]n contrast [with] countries where the monarchy was highly absolutist, such as Spain and Portugal’. According to Acemoglu et al., non-absolutist states that did not experience rapid growth, like Venice and Genoa, did not have adequate physical access to the Atlantic to gain from the cross-oceanic trade.

This argument should imply that Britain, in particular, should have demonstrated rapid economic growth from the point at which significant constraints were imposed on absolutism, the signing of the Magna Carta in 1215, a full three centuries before the interval that Acemoglu et al. (2005) identify as the period of the ‘rise of Europe’. An alternative to the neoinstitutionalist account is the argument that the British and the Dutch were simply the winners of the game of mercantilist rivalry (Darity, 1990).
Indeed, many of the pecuniary benefits of the Portuguese and Spanish colonial systems were transferred to Britain via intra-European trade (Darity, 1990). For example, it has been estimated that during the eighteenth century trade surpluses with Portugal brought 50,000 pounds of bullion into London weekly (Birnie, 1935, pp. 175, 180).

Acemoglu et al. (2005) are aware that their analysis touting the importance of the Atlantic trade for European economic development could connect directly with arguments, like Williams’s, about the importance of ‘the associated profits from colonialism and slavery’. But they immediately seek to sever the connection with the following observation:

> It is undoubtedly true that colonial relations with the New World and Asia contributed to European growth. Nevertheless, quantitative analyses, for example, Engerman (1972), Engerman and O’Brien (1991) [sic: O’Brien and Engerman (1991)], O’Brien (1982), and Bairoch (1993, ch. 5), suggest that the volume of trade and the profits generated by the Atlantic trade appear to be too small to account for much of European growth directly. (Acemoglu et al., 2005, p. 562)

Instead, they characterize the institutionally promoted gains from trade for European development as indirect effects: ‘the rise in Atlantic trade enriched and strengthened commercial interests outside the royal circle and enabled them to demand and obtain the institutional changes necessary for economic growth (Acemoglu et al., 2005, p. 550). They dismiss the direct effects by invoking what one of the co-authors of this chapter has dubbed the ‘small ratios’ argument.

They do not appear to be aware that the ‘small ratios’ argument has been critiqued in a counter set of quantitative analyses to such an extent that the opposite position can be sustained. The volume of trade and profits from the Atlantic economy, particularly the slave trade itself, were enormous by comparative historical standards (Solow, 1985; Bailey, 1986; Darity, 1990; Cuenca Esteban, 1997; Inikori, 2002). Even O’Brien and Engerman (1991) concede that the available trade statistics indicate that colonial trade was of paramount importance for England, at least in the eighteenth century, if not the seventeenth century as well.

Easterly et al. (2006) also subscribe to a variant of neoinstitutionalism, but offer a characterization of ‘good institutions’ closer to Putnam’s notions of social cohesion, and a quite different set of factors as contributors to ‘good institutions’. One of the factors that they hypothesize will make for ‘bad institutions’ is, again, ethnic diversity. However, in the earlier paper on Africa, Easterly and Levine (1997) attribute the extreme ethnic diversity that they claim is present there to colonialism – perhaps providing a basis for rapprochement with Williams. However, they never explore or consider the direct enrichment effects on Europeans of the colonial
process, the task which Williams pursued in depth, from the crucible of the Atlantic slave trade to the mid-twentieth century.

**Measuring ethnic diversity or ethnic polarization?**

Easterly and Levine (1997) presume that ethnic diversity necessarily has a negative impact on economic growth. In contrast, Alesina and La Ferrara (2005) propose that there are ways in which ethnic diversity might prove beneficial for economic growth. They acknowledge that there are potential costs to diversity including ‘[c]onflict of preferences, racism, and prejudices . . . lead[ing] to policies that are at the same time odious and counterproductive for society as a whole [and] [t]he oppression of minorities may[be] lead[ing] to political unrest or even civil wars’ (Alesina and La Ferrara, 2005, p. 762). But they also say that ‘a diverse ethnic mix also brings about varieties in abilities, experiences and cultures that may be productive and may lead to innovation and creativity’ (Alesina and La Ferrara, 2005, p. 762). Robert Bates (2000) has also argued that ethnic group social capital can promote human capital formation among their own to levels that would not have taken place in the absence of group identification and membership – with potential positive effects on the society as a whole. Paul Collier (2000) has suggested that the effects of ethnic diversity on growth are negative in non-democratic societies and positive in democracies. Ultimately, Alesina and La Ferrara (2005, p. 763) conclude that the impact of diversity on economic performance is an empirical question since, at the theoretical level, it is unclear whether ‘the benefits of “diversity”’ will outweigh ‘the costs of heterogeneity of preferences’.

To address this question, an appropriate measure of ethnic diversity is needed. This is the point where a ferocious debate has raged among social scientists examining the relationship between ethnicity and economic development. The debate is simultaneously about which measure is most appropriate from a conceptual standpoint and which measure has the best technical attributes.

Easterly and Levine (1997) initiated the empirical work in this area by deploying the ethno-linguistic fractionalization (ELF) index. Using the Herfindahl concentration formula, a country’s ELF score represents the likelihood that any two people chosen at random from the population will speak different languages. The likelihood has a minimum value of zero when there is no ethno-linguistic variation in a country and a maximum value of 100 when the odds are perfect that any two persons drawn at random will speak a different language. They find that the ELF index (expressed as a percentage) is positively and statistically significantly related to their measures of unproductive macroeconomic policies.
Still, some peculiarities are immediately evident with this measure; the extent of ethnic tension and violence definitely is not reflected in the index score. Haiti’s ELF score, for example, is 1 in Easterly and Levine’s (1997) study. This is due to the high degree of linguistic homogeneity, providing no indication of the high level of conflict conducted by the social elite and the military toward the majority of civil society (UNDP, 1994, pp. 41–2). Burundi’s ELF score of 4 is attributable to the near universal use of French. The index could not have been used effectively to forecast the long cycle of genocidal violence there. The same is true of Rwanda.

Daniel Posner (2004) criticizes both the design of the ELF index as a measure of ethnic diversity and its use in growth regressions. The ELF index was built on coding decisions made by a Soviet ethnographer in the 1960s. Insofar as ethnic identities undergo some fluctuation in importance and intensity, definitions now a half-century old may be misleading. More significant, the identification of ethnic groups solely on the basis of linguistic differences ignores other critical markers of distinction, such as phenotype, religion, attire and cultural practices. Furthermore, the ELF index does not take into account the spatial distribution of ethnic groups within a country, nor does it incorporate any information that would capture the depth of ethnic division.

Nonetheless, Posner’s primary objection to the ELF index is the absence of consideration of political organization and contestation by the relevant racial/ethnic groups in a society. The original Easterly and Levine (1997) hypothesis that greater ethnic diversity leads to slower economic growth rests on an intermediate causal step involving the formation of macroeconomic policies. Greater ethnic diversity is supposed to increase interest group polarization, motivating rent-seeking, overspending and financial repression and subverting the provision of public goods. Posner argues that the ELF index is an inappropriate measure for testing this hypothesis because there is no necessary relationship between the existence of ethnic divisions and the forms of political competition that lead to an institutional environment that inhibits growth. He contends that many ethnic groups do not operate as independent political actors because they lack the political strength either to influence policy directly or to mobilize as part of larger coalitions and parties.

Posner proposes a revised measure of ethnic diversity called the Politically Relevant Ethnic Groups (PREG) index. Using secondary sources, Posner identifies those ethnic groups that have been politically active and have engaged in competition over social policy in each of four decades. He applies the Herfindahl formula to his new measure of ethnic groups and develops index values for 42 African countries in each decade. Although correlated with other comparable indices, the PREG index...
generates important differences in the rank ordering of the most- and least-fractionalized countries.

The most common technique for analyzing the economic effects of ethnic diversity, in the aftermath of the Easterly and Levine (1997) study, is to include a measure of ethnic fractionalization as an explanatory variable in a cross-country growth regression. Posner compares the performance of the ELF index with the PREG index in explaining variations in economic growth across the African countries. First, a series of policy measures are regressed separately on the ELF and PREG indices. Surprisingly, although the ELF index has a statistically significant effect on many of the policy variables in the world sample, the effect disappears when the analysis is restricted to the African sub-sample. In contrast, the PREG index has a statistically significant relationship with the black market premium and with fiscal surpluses in the anticipated direction.

Posner then regresses growth rates separately on the ELF and PREG indices. The early Easterly and Levine (1997) results can be replicated with the global sample, but it appears to have no effect on the growth performance of the countries in the all-African subsample. The PREG index does have a statistically significant effect on African growth rates in the expected direction, that is, the higher the PREG value, the lower the country’s growth rate.

Finally, Posner regresses rates of growth separately on the ELF and PREG indices with other policy variables included as controls. Peculiarly, the PREG index loses significance in the presence of these additional controls, while the ELF index now has a significant effect, even in the African subsample. Posner’s interpretation is that the PREG index successfully captures the indirect effect of ethnic diversity on economic growth via macro-economic policies but does not have a direct effect on growth. The ELF index, in contrast, has a direct effect on growth independent of other policy practices. Posner speculates that the ELF index may be picking up some other mechanism through which ethnic diversity affects economic development. In the end, Posner concludes that ethnic fractionalization is negatively and strongly related to economic growth in Africa, and offers his results as evidence that macroeconomic policies are an important channel through which ethnicity influences economic development.

James Fearon’s (2003, p. 198) construction of an alternative measure to the ELF index attempts to locate ethnic groups based upon ‘what people in the country identify as the most socially relevant ethnic groupings’. As Alesina and La Ferrara (2005, p. 792) point out, Fearon’s basis for development of his index – Fearon’s determination of what people in a given country decide are ‘the most socially relevant ethnic groupings’ – depends significantly upon Ted Gurr’s (1996) Minorities at Risk Project at the University of Maryland.
And, indeed, the criteria that Fearon utilizes predicated upon Gurr’s database comes far closer to capturing inter-ethnic tension-cum-violence than does the ELF index. Gurr (1996) defines a ‘minority’ (not necessarily a numerical minority but any group that is less than 100 percent of the population) at risk as a communal group that: (1) faces political and/or economic discrimination; and (2) acts on its own behalf collectively in the political process. The entire population consists of minorities at risk in Burundi, Chad and South Africa on the Gurr criteria. In a global sample Gurr finds that the African continent has the largest share of population comprised of minorities at risk, a result potentially inconsistent with Bates’s observations about the levels of political violence in Africa relative to the levels of ethnic differentiation.

Alesina et al. (2003) show that the Fearon index, based largely upon the Gurr scale, is closely correlated with their version of the ELF index, an extended version that includes ethnic groups defined by other characteristics such as skin color. The ‘more comprehensive’ version of the ELF index captures salient ethnic differences in Latin America that the language-only index would not capture:

In [Latin America], the language index shows more homogeneity because the language of the former colonizers (Spanish, Portuguese, English) is often spoken by most, but the index based on skin color or ethnic origin (say black, mulattos, white, mestizos, Indian, etc.) shows more heterogeneity. (Alesina and La Ferrara, 2005, p. 792)

Alesina and La Ferrara seem to find the extended ELF and the Fearon indices both to be quite satisfactory, although quite different in design.

Jose Montalvo and Martha Reynal-Querol (2005) have proposed the polarization index as an alternative to the ELF measure that is conceptually quite distinct from the others discussed here. The ELF index presupposes that a country is more fractionalized, the greater its number of separate groups. A country with two similarly sized groups facing each other in a cauldron of hostility would not be depicted as highly fractionalized by the ELF index. The Montalvo and Reynal-Querol index reaches its peak value when a country consists of two equally sized groups, and then declines in value as the number of groups increases, departing from the half-and-half split. Alesina and La Ferrara (2005, p. 793) describe the comparative assessment of the polarization index with the ELF index as follows:

[Montalvo and Reynal-Querol] show that this index is highly correlated with ethno-linguistic fractionalization (ELF) at low levels of ELF, uncorrelated at intermediate levels, and negatively correlated at high levels. In a cross-country
regression analysis, they find that ethnic polarization has a positive impact on the likelihood that a civil war occurs and a negative effect on a country’s growth rate. They do not find an independent effect of ethnic fractionalization. Using a different data set, Alesina et al. (2003) compare the results of the polarization index RQ and the fractionalization index ELF, and find that fractionalization works slightly better as a determinant of policies and economic outcomes. While the apparent inconsistency between the two sets of results may be due partly to different parameterization and partly to different data sources, it is between the two measures at low levels of fragmentation.

Tade Okediji (2005) focuses his criticism on the dimensional limitations of the ELF index. He argues that linguistic differences are only one of several possible cleavages that are associated with ethnic division. Racial and religious identities also form the basis of ethnic differentiation, and the ELF index neglects such groupings if they do not coincide with linguistic differences. Moreover, many countries are characterized by complex interactions of racial, religious and linguistic fractionalization, leading to overlapping identities and variation in the salient factors that distinguish one group from another. To counter these limitations of the language-only ELF index, he advances an alternative index that is quite similar conceptually to the more comprehensive ELF index developed by Alesina et al. (2003). But he then performs a comparison with the language-only ELF index that is quite original and has interesting implications.

Okediji proposes the Social Diversity Index (SDI), a measure of fractionalization intended to capture the multidimensional nature of ethnicity – a sort of Human Development Index for ethnicity. Also using secondary sources, he partitions each country into primary racial, religious and linguistic groups. His measure enables him to classify each individual in a country according to all three characteristics, although the person can only belong to one category within each characteristic. When compared with the ELF index, the SDI has a higher mean and a lower standard deviation across the same sample of countries. Okediji concludes that ethnically diverse societies are far more common than the ELF index would suggest, and that the variation in the degree of ethnic fractionalization across countries is much smaller than suggested by Easterly and Levine in 1997. Thus, Okediji returns the discussion to the point of origin of this chapter – the near universal presence of salient ethnic/racial differentiation across the nations of the world and the near universal presence of ethnic/racial inequality.

The research on the relationship between ethnic fractionalization and economic performance using cross-country regressions invariably treats ethnic fractionalization as driving economic performance. But it is quite plausible that the overall economic performance in an economy might
affect ethnic antagonisms positively or negatively. Perhaps conditions of slow economic growth can inflame inter-ethnic tensions while prosperity may relax them? Developments in Malaysia and in Indonesia seem to connect economic crisis to increased communal violence (Darity, 2002, p. 135). The possibility of simultaneous causation has not informed research in this area to any significant degree, although Alesina and La Ferrara (2005, pp. 772–3) are well aware that any number of the ‘independent’ variables used in regressions of this type are subject to the endogeneity complaint.

Similarly, it can be asked whether the general level of inequality in a society drives the level of intergroup inequality, rather than the level of intergroup inequality shaping the general level of inequality (Darity and Deshpande, 2000). There may be no general answer; indeed, the particular answer is probably contingent on the particular structure and history of each country or region.

And what about the fundamental relationship between ethnic/racial conflict and economic inequality between groups? At the very start of his 1985 monograph, Racial Conflict and Economic Development, W. Arthur Lewis raised the following question:

Is economic equality necessary for social peace? From one standpoint every so-called racial conflict is sustained (or even initiated) by an economic conflict, covert or open. What poses as a conflict between a dominant and a subordinate group, we are told, is really only a way that an exploiting minority recruits supporters of its case from people with whom it has only racial ties. The proposition is somewhat doubtful. It is probably true that every dispute, racial or not, has, or acquires some economic edge, but this is not the same as saying that all disputes originate in economic conflict. If the economic conflict were mitigated by movement toward equality, would the racial conflict be lessened automatically?

The difficulty is that, far more often than not, effective movement toward economic equality requires disrupting the economically privileged position of the socially dominant group. Its members will resist or act to destabilize the policies that have been adopted to push the society toward greater intergroup equality, whether it is affirmative action, school desegregation or a program of reparations. The very effort to achieve greater intergroup equality will exacerbate racial conflict – a backlash effect – from the group that sees its position of privilege as being threatened. The threat arises precisely because the members of that group have a material benefit from maintaining their identity as a group. To eliminate racial conflict, the benefits associated with racial division would have to be removed. But to remove those benefits typically unleashes racial conflict. In that sense, economic equality is necessary for social peace, but the process of achieving economic equality between ethnic and racial groups is invariably far from socially peaceful.
References


PART VIII

THE STATE, INSTITUTIONS AND DEVELOPMENT
The proper roles of states and markets in fostering and sustaining economic prosperity comprise one of the oldest and most debated topics in the history of economic thought. A focus of Adam Smith’s attack on mercantilism in the late eighteenth century, the roles of states and markets were debated by the historical, institutional and early neoclassical schools of economics a century later. More recently, the issue enjoyed a place of prominence in late twentieth century debates over reasons for the economic growth of Japan and other East Asian economies, the most notable success stories of that period.

It should not surprise us, perhaps, to find disagreement over the roles of states and markets in the economy, because the two institutions have histories of both synergy and rivalry (Putterman and Rueschemeyer, 1992) stretching back to ancient times. Some 4000 years ago, populous societies marked by increasingly complex divisions of labor gave birth to the first city states and empires, and while those societies probably saw unprecedented expansions of market activity, their economies were also in some cases state-dominated, and all featured attempts by the new states to enrich themselves and to support larger armies and coteries of officials by controlling sources and flows of wealth. The interplay between traders and officials differed from one society to another, and in given societies across periods of time. While extractive, centralized states may have constrained economic prosperity in some instances, in others the absence of central authority constrained it. For example, the tenuousness of law, order and safe commercial routes in much of Europe following the collapse of Roman rule probably contributed to economic stagnation. It seems noteworthy that modern capitalism and the nation state arose in tandem in Europe after 1500. Since then, no prosperous modern economy has emerged in a society lacking a well-ordered state.

The roles of markets and states
Despite controversy, most economists agree that markets, permitting competition among independently managed enterprises free to select their products and methods of production, have been central to economic
progress and crucial to the growth of productivity, technological know-how and living standards since the Industrial Revolution. Independent enterprises can thrive, these economists would argue, only when free to interact with suppliers and customers in relatively uncontrolled labor, raw material, capital goods and product markets. At the same time, most economists also agree that market systems function poorly if at all without the protection of property rights, rule of law and availability of a stable currency, providing which have been central economic roles of states. Further, economists recognize domains in which competing enterprises cannot be expected to bring about optimal results, most prominently the cases of market power, public goods (including some key trade-facilitating infrastructures), and environmental externalities. Where debate exists is over the scope of the set of public goods (for example, whether they include health care and education), and over the degree to which government remedies (for instance, in the case of monopoly) tend to ameliorate rather than worsen unregulated outcomes. Full agreement is also lacking as to the macroeconomic responsibilities of governments, and whether inequalities in the distribution of income and wealth should be viewed as market failures, also calling for government interventions. A point of particular relevance to this chapter is whether governments can promote economic development by formulating strategies to promote growth, or whether the ideal role of government is simply to create a stable institutional environment and then allow the market to ‘work its magic’.

Although properly speaking markets are the loci of interactions between buyers and sellers, sometimes although not always associated with specific institutional frameworks or locations, economists use the phrase ‘the market’ to refer not only to the settings or sets of those interactions but also to the decentralized economic mechanism as a whole, including the presence of autonomous enterprises that purchase inputs and transform them into goods and services. A ‘market economy’ (sometimes called ‘the market’) in this broader sense is said to allocate scarce resources efficiently among competing needs and wants, and to stimulate technological progress, for several reasons. First, the market mechanism causes both final consumers and intermediate users of goods, services and resources to signal the values they attach to those goods, and so on, as a by-product of the exchange of offers to buy and sell. Second, the pressure that enterprise owners feel to maximize net returns, lest they be driven from business or at least suffer serious financial losses, induces them to attempt to produce goods of maximum value to consumers using the least-cost combinations and quantities of resources, while also striving to satisfy the needs of buyers in terms of quantity, quality and variety. Third, individual workers, including prospective managers and specialists of other kinds, are motivated to
invest in demanded skills due to the higher earnings which market competition assigns to those with scarce capabilities. Fourth, owners of non-labor resources can be expected to steer those inputs toward the uses most valued by society, responding to price signals which indicate, for example, the value of a parcel of land as an orchard versus its value as a grain field or a parking lot. Finally, the rents that accrue to innovation, and the fear of falling behind competitors, are viewed as major causes of the high rates of technological change observed in market capitalist economies.

But markets do not exist in social and political vacuums. Market interactions are embedded in social systems (Granovetter, 1985) and many trades would be impossible without social norms that increase trust and facilitate contract enforcement (Putnam, 1993; Greif, 1994). Although state power and the formal legal systems to which states lend their force may offer protections of final recourse, market interactions rely more directly and extensively on widespread adherence to such norms by ordinary members of society. Recent research suggests considerable cross-country variability in trust and social capital, and this variability correlates with differences in levels of economic development and growth (Fukuyama, 1995; Knack and Keefer, 1997). It can be argued that the existence of a stable political order, honest administration and a non-corrupt judiciary may be important facilitators of social trust and norm abidance (although the relationship probably goes both ways). State actions that contain the extremes of inequality, poverty and neglect of worker health and safety may, while intervening in the full expression of the logic of competition, end up helping the market by helping to stabilize the polity and society and to increase the social acceptability of leaving most economic coordination to the market.

Lessons from planned economies
While economists’ convictions about the efficacy of markets derive in part from classical and neoclassical theory as well as from observation of market economies, many draw lessons also from experience with alternative economic systems and policies. The premier experiment in operating large modern economies with almost no state-tolerated role for markets is the one that began in Russia in 1928 and lasted into the 1980s there and in numerous other countries eventually ruled by communist parties. In those centrally planned economies, prices were set administratively and the allocation of resources and determination of production plans was to take place not under the influence of market forces, but rather under the aegis of a planning bureaucracy directly weighing leaders’ political goals and perceptions of societal needs. Planners ignored notions of comparative advantage and were guided instead by the goal of building industrial
economies, which they hoped to achieve by directing huge investments into the capital goods sectors, ordering low-price crop deliveries from farmers, and exploiting natural resources without regard for opportunity cost and environmental impact. Planners paid only limited attention to consumer goods production, and even less to the provision of services other than health care and education.

According to observers (for example, Nove, 1983; Kornai, 1992), planners in the Soviet-type economies struggled with the problem of providing effective incentives to, and the need to elicit information from, enterprise managers. Quality, variety, spare parts and maintenance were perennial problems. Enterprises integrated vertically to avoid relying on the planning bureaucracy for the inputs they needed. Considerable amounts of resources were diverted into black market activities. Innovation proved difficult to engender at levels comparable to those of industrialized market economies, except perhaps in the military sector. Considerable activity took place outside of the approved plans, with some observers going so far as to argue that the claim that such economies were primarily plan-based is inaccurate.

For a time, rapid structural change and achievements in health and education sectors made the model attractive to some outsiders, especially in the developing world. The Soviet Union’s industrial output growth rate exceeded that of the United States during the 1930s and again from World War II until the late 1970s. Most estimates suggest that China achieved a higher rate of industrial growth than India from the late 1940s to the late 1970s, despite ideological excesses and political upheavals. Life expectancy in China exceeded that in India by some 12 years in 1978, and in general planned economies achieved higher life expectancy, lower infant mortality and higher literacy rates than non-communist countries at similar income levels. However, the curtailment of individual freedoms and comparisons with incomes and consumer good availability in neighboring countries like West Germany and Taiwan bred dissatisfaction with the system, ultimately leading to the system’s demise in both Europe and Asia. In addition, the planned economies’ growth was unbalanced, and much of the capacity put in place by their economic system had little value when the countries in question adopted market-oriented reforms and became more open to international trade.

State roles elsewhere
Less extreme in their departures from free market principles are the numerous cases in which market pricing and exchange were permitted but with key government interventions ‘distorting’ the price system. Most developing countries of the late twentieth century controlled foreign exchange transactions, usually overvaluing their country’s currency. By doing so,
they inadvertently discouraged exporting and necessitated combinations of import restrictions, tariffs and borrowing to deal with trade imbalances. Interest rates on bank loans were often subject to regulatory caps, and import licensing and tariff arrangements made some capital goods less expensive, creating artificially low prices for certain producer goods while the cost of capital to small-scale borrowers, including farmers, remained high. These and other interventions, described by some economists as ‘getting prices wrong’, led to limited growth of bank deposits and other forms of financial mediation (‘shallow finance’), excessive capital intensity and limited job creation in a few modern sector activities, and capital starvation and underemployment in other parts of the economy. By discouraging exports and domestic savings while encouraging imports, they also contributed to the growth of unsustainable burdens of debt at the national level.

Although these examples convince many of the virtue of markets, pure free market economies are textbook abstractions to which no modern national economy adheres in reality. After the 1930s, the ideal of a fully self-regulating economy was abandoned by most economists and politicians in industrialized countries, with macroeconomic stabilization being considered a responsibility of states. Other government roles were also growing. The proportion of national income used by governments to pay civil servants and to support various functions grew steadily until close to the end of the twentieth century. Governments were asked to respond to market failures, for example to set and police environmental standards, and to monitor the safety of workplaces, foods and pharmaceuticals. Governments invested in roads, bridges, maintenance of waterways and harbors, and rail lines. Due to some combination of market failure and distributive concerns, governments also provided unemployment benefits, food and health subsidies, pensions and other social benefits. The notion that the market and autonomous firms could be engines of production and technological progress, but that the distribution of social benefits could be partly separated from that of market rewards, was mainstream in the polities if not in the economics professions of the world’s most prosperous and technologically advanced nations after World War II.

At the same time as the role of the state was growing in industrialized mixed economies, economists were beginning to re-examine their depiction of government as a benevolent agent that could be counted upon to follow the prescriptions of normative economic theory regarding the correcting of market failures. Mainstream political economy viewed governments as being composed of individuals who might promote the well-being of citizens if imbued with social concerns or held accountable by an engaged public, but who might also be as motivated by self-interest as other
individuals. If monitoring by the public is costly and if formally democratic political institutions can be captured by groups with concentrated interests in particular areas, state officials and politicians might be poor servants of the public as a whole. Grievous policy errors could also result from simple misunderstanding of the effects of instruments like exchange controls and interest rate ceilings. The idea of ‘government failure’ entered the lexicon of economics, alongside the term ‘market failure’, and the possibility was raised that even when there exist imaginable state interventions that can increase social welfare if effected, actual government involvement might worsen rather than ameliorate some market failures. Efforts to help the poor might also have effects at odds with that aim in the long run if they resulted in reduced incentives to invest in human and physical capital and thus lower rates of growth.

The special problems of less-developed countries
Some arguably distinct aspects of the state–market relationship in those countries lagging far behind the most industrialized economies have been the subject of separate discussions at various points in time. In the nineteenth century, development strategies were proposed and to some degree adopted by national governments in then-lagging countries including the United States, Germany, Japan and Russia, usually including tariff protection against a range of importable manufactured goods and government assistance or active participation in the accumulation of investment funds. The so-called Great Industrialization Debate in the Bolshevik-ruled Russia of the 1920s would be re-examined by the advisors to leftist Third World governments like those of Mozambique and Tanzania as late as the 1970s. The post-World War II era in which the Bretton Woods institutions were formed to help manage the economic problems of less-developed and especially newly decolonized countries saw the growth of a new literature of development economics.

Early post-World War II writers such as Ragnar Nurkse (1953), W. Arthur Lewis (1954) and W.W. Rostow (1960) argued that the central problem of a developing economy was to raise the share of capital formation in national product to a level sufficient to fuel the growth of modern-sector activities and, in the formulation of Lewis and of Fei and Ranis (1964), to ‘drain off’ the pool of surplus labor underemployed in the traditional, mainly agricultural, sector by absorbing it into modern employment. As a result of such thinking, government development plans identifying the gap between domestic savings and investment targets were formulated and used as bases for seeking investment financing from international financial institutions and foreign governments. During the 1960s and 1970s, there was still considerable tolerance among Western development specialists and
advisors for using inflation as a tax to finance government investment, using tariffs to protect domestic industries deemed promising, and adopting other measures that would even then have been considered inadvisable in a developed-country context.

Governments were also thought to have a role to play as coordinators of the overall push for development. Although enterprises were mainly privately owned and prices determined by supply and demand, it was argued by some that national economic planning could still play a crucial coordinating role. The theory espoused by advocates of planning in post-World War II France, for example, was that firms might be reluctant to invest if they could not be confident that complementary investments were being made by others up and down the relevant production streams. Complementary investments in infrastructure and manpower training might also be called for. The market might be a suitable coordinator of short-term production decisions, the theory went, but the large-scale investment decisions needed to bring about true structural change might be too lumpy to overcome the hurdles of uncertainty in an unplanned economy. The argument was stretched by some to the point of suggesting that government itself had to do the investing, even in industrial enterprises and mines, to overcome private sector hesitation. But for the most part, it was used in favor of a government coordinating and facilitating role, not state ownership. The role of governments in Japan and later Korea were often understood in this light, but so too were the national plans typical in many other developing countries.

The ‘international division of labor’ had a central place in many discussions. Free trade among nations, it was argued, was more beneficial to rich than to poor countries. Europe’s poor former colonies in Asia, Africa and Latin America had been brought into the world economy for the benefit of their colonizers as sources of cheap raw materials and foodstuffs and as markets for European and North American manufactured goods. It was in the ‘core’ or ‘metropolitan’ countries’ interests that the ‘peripheral’ countries’ comparative advantage remain one based on unskilled labor and raw materials. ‘Free trade’ would perpetuate this because manufactured goods would remain less expensive for poor countries to import than to produce, so they would have no chance to learn by doing and to move towards the international frontier of industrial capability.

In response to such concerns, economic moderates called for the leveling of the playing field by reducing discrimination against developing-country agricultural exports, devising mechanisms to stabilize and maintain the prices of tropical commodities like sugar cane and coffee, and fully opening developed-country markets to Third World manufactures – a ‘New International Economic Order’. Believing that positive steps were required to foster structural change away from the old reliance on primary product
exports, many also believed that selected industries in poor countries should be protected from foreign competition by tariffs of sufficient magnitude to let domestic manufacturers obtain a foothold – the groundwork of the import substitution industrialization approach. More radical commentators called for de-linking developing economies from the world trading system, substituting (especially for smaller economies) links with neighboring countries and/or with Communist states (for a discussion, see Diaz-Alejandro, 1978). In either case, national strategies, which only governments could put forth and implement, were seen as requirements for escaping the self-perpetuating status of underdevelopment. Thus, a key role was assigned to the state; leaving things to market forces would only perpetuate underdevelopment and dependency.

More recent discussion
The 1980s were a watershed decade for policy and professional opinion on the problems of economic development. The decade was marked by slowdowns of growth in the industrialized market economies, a still more pronounced slowdown in productivity growth in Communist countries, the accumulation of unsustainable debt levels by many middle- and low-income developing countries, and increased international recognition of the remarkable growth achievements of a number of East Asian economies. These developments helped to fuel a conservative backlash against the ‘welfare state’ and state-owned industries in the West; the beginnings of radical economic reform in China and a last decade of reform experimentation in the Soviet bloc; initiation of structural adjustment programs in developing countries in the wake of their debt crises; and the fall from respectability of import substitution industrialization as a policy approach. Moved partly by the necessity of accepting International Monetary Fund (IMF) conditions for urgently needed loan programs, partly by the perception of the relative merit of East Asia’s more outward-foocussed orientation, most developing-country governments devalued their currencies, reduced spending, began reversing the trend of nationalization, and attempted to make their economies attractive to foreign investors. Going into the 1990s and the start of the twenty-first century, increasing flows of foreign direct investment and international bank lending, growth of trade volumes, and vigorous participation in world trade by China, India and other developing economies, became hallmarks of the intensification of international trade, investment, knowledge and cultural flows that was dubbed ‘globalization’.

Even though policies were trending in this period towards liberalization in comparison with prior decades, it would be incorrect to describe the approaches of most developing-country governments as laissez-faire. Nor would it be accurate to suggest that the prevailing view of the state’s role in
development had become the minimalist one favoring protection of property rights and of a stable currency as the only valid economic functions of government. The 1997 *World Development Report* of the World Bank stated that: ‘[a]n effective state is vital for the provision of the goods and services – and the rules and institutions – that allow markets to flourish and people to lead healthier, happier lives. Without it, sustainable development, both economic and social, is impossible.’ Government investments in the health and education sectors were viewed as important both to immediate well-being and to economic growth. The role of governance, especially the rule of law and the absence of corruption, was accorded considerable importance in World Bank publications and was the focus of a number of studies including Kaufmann et al., 2000.

China’s quarter-century growth spurt after 1979 occurred in an economy in which, until the mid-1990s, most industrial enterprises were owned by local or higher levels of governments. More importantly, even the China of the early 2000s, when the private sector had come to play a more important role, resembled more the Japan of the 1950s and 1960s with its government-guided industrial policy, including special subsidies and incentives to sectors accorded important roles, than it did laissez-faire. Until its admission to the World Trade Organization (WTO) in 2001, China maintained a dual trade regime, with substantial tariff protection of most domestic industry but also a more liberal regime for imported inputs applying only to export-oriented industries originally restricted to special economic zones and coastal ‘open cities’ (Naughton, 2007). These distinctions were only gradually dismantled during the early WTO years.

Opinion regarding the role of the state was also strongly influenced by the experience of countries in the former Soviet bloc transitioning from state socialism to market capitalism. Whereas initially the emphasis of economic advisors was on ending governments’ involvements in the economies of the countries concerned, it was soon recognized that healthy market economies could not exist without strong institutional supports, including legal protection of property rights, low tolerance for corruption and adequate monitoring of financial market institutions. Furthermore, the advantages exhibited by China since 1978, a little earlier by Taiwan, Korea and Singapore, and more recently by India, were seen to be in part the result of substantial public investments in education, transportation, communications and health. Comparative political stability has also been a major asset of these countries. Although the relatively closed nature of the Chinese and Indian economies before the 1980s has often been viewed as an error that retarded development, it is difficult to prove that those initial closed periods were not of some benefit to their economies, creating a breathing space during which indigenous capacities and skills could be incubated.
Efficient states, old states
With or without agreement on what governments need to do to facilitate development, there is evidence of a general correlation between more efficient and capable government and better economic outcomes. A number of studies have found a correlation between measures of government quality and rates of economic growth. Mauro (1995) found that countries with more corrupt governments had lower rates of investment and economic growth. Evans and Rauch (1999) found that developing countries whose government administrative structures exhibited more of the classic bureaucratic features of meritocratic recruitment and predictable long-term careers achieved higher rates of economic growth even after controlling for initial GDP and human capital. Kaufmann et al. (2000) found several governance measures to be correlated with rates of economic growth.

The fact that countries with relatively capable and stable governments have better growth records than others does not prove that good government is an independent cause of economic growth. The same propitious factors may give rise both to economic growth and to a well-ordered state. Interestingly, one of those factors may be a long history of large-scale political organization. Recent studies suggest that regions that saw earlier formation of kingdoms, states or empires, especially ones not subsequently disrupted by large population shifts due to colonization, enjoyed faster growth between 1960 and 2000 (Bockstette et al., 2002; Chanda and Putterman, 2007). The countries hosting old states also tend to have better scores on commonly used measures of institutional quality. Old states are associated with early transitions to agriculture, and one study suggests that the time of agricultural transition is a strong predictor of current level of development (Hibbs and Olsson, 2004).

Conclusion
Although most economists agree that private enterprises disciplined by market competition are more efficient providers of most goods and services than are government-owned enterprises, there is also general agreement that neither a prosperous national economy nor a sustained process of economic growth are achievable in the absence of a well-functioning state. States are needed to secure property rights, manage currencies and provide the civil order without which commerce and investment become excessively risky propositions. States help to create competitive markets by regulating and breaking up monopolies and promulgating rules for the operation of banking systems and financial markets. States help to address market failures in such areas as environmental quality and workplace safety. And states can pursue macroeconomic policies that contain fluctuations in price and employment levels.
More controversial is whether state guidance or strategic planning is either necessary or desirable for a country to be launched on a path of economic growth. Some governments, especially in Asia, appear to have enjoyed success with government activism in this respect, but there is as yet no consensus about these cases, and misguided interventions can be harmful. Building state capacity while focusing on education, health, transportation, and a legal and tax environment conducive to investment may be tall enough orders for some governments. These tasks, in any case, constitute a starting point that all governments should strive for, and that the people of every country should actively demand of their governments.

References

Any survey of the issues confronting monetary policy in developing countries must first address several basic questions. First, should the central bank target development as one of its objectives, or more broadly, can it contribute to development indirectly, through for instance maintaining financial stability, ensuring international competitiveness or delivering low inflation? Second, is there any scope for, and value to, monetary independence, either because monetary policy is under the thumb of fiscal policy, or because the country concerned is small and open, or because a foreign currency circulates domestically and currency substitution is rampant? Finally, assuming that there is some scope for an independent monetary policy, what should be the operational guide for policy-setting, and is it likely to be different depending on the degree of financial development or other structural features of the economy?

This chapter will be mainly about the third question, namely the way monetary policy should be set for the short to medium run, and the institutions and policy regimes that support and implement that policy, while recognizing that developing countries are very diverse and that ‘one size does not fit all’. In discussing the issue, the position will be taken that monetary policy cannot be separated from exchange rate policy – the two have to be considered together. This is not to say that there may not be instruments that under some conditions have differential effects on domestic monetary conditions and the exchange rate (open-market operations and sterilized intervention, for instance). However, for many countries – in particular the ‘emerging economies’ with access to world capital markets and few controls on capital flows – the scope for differential effects is small (for instance, because of constraints on the cost and effectiveness of sterilized intervention). Over time, the extent of capital controls has declined, as their effectiveness has been blunted and their distorting effects on economic decisions have become manifest. Thus, countries are ill advised to use monetary policy to target domestic objectives while targeting a fixed value for the nominal or real exchange rate.

As for the first question, the position taken here is that development should not be an explicit goal of monetary policy. Development is a longer-term, structural issue relating (primarily) to the real economy while monetary neutrality precludes long-run effects of the money supply on the level
of real output. Forty years ago, this position would not command a consensus; indeed many then advocated deliberate use of monetary policy to keep interest rates low (below world real rates), to channel credit to particular sectors, to undervalue the real exchange rate in order to stimulate exports, or to redistribute income (via inflation) to those with higher saving rates. However, a policy of deliberately keeping interest rates low to stimulate investment is likely in all but financially repressed economies to produce accelerating inflation rather than higher growth, and inflation quickly gets anticipated, blunting any favorable real effects. Following McKinnon (1973) and Shaw (1973), the pervasive inefficiency of financially repressed economies has been recognized. Long-run monetary neutrality does not preclude monetary policy from contributing to short-run stabilization, or the long-run level of inflation from having real effects, but the time horizons for the effects on development are so long and complicated that they make targeting development an inadequate guide for monetary policy setting. This leaves open which of the various possible intermediate targets for monetary policy (the exchange rate, the rate of inflation, nominal income or external competitiveness, among others) is most appropriate and will most contribute to development and welfare. This is issue three above.

Turning to the second question, for a discussion of monetary policy to be interesting, it must be assumed that there is some scope for monetary independence. At the same time, it will be noted below that in some developing economies the limits on that independence are tight indeed, and this may influence the choice of operational targets for monetary policy. In that respect, the size of countries and their structural characteristics matter in considering their appropriate monetary policy. Thus, monetary policy questions for developing countries are somewhat distinct from those for developed countries, for which there is an extensive literature. There are reasons to expect that countries with higher per capita gross domestic product (GDP), more developed financial sectors and stronger institutions face different policy trade-offs. In what follows we shall pay considerable attention to a policy regime – inflation targeting (IT) – that is increasingly finding favor among both industrial and emerging market economies, while considering whether the two sets of countries differ with regard to IT’s suit-ability as a monetary policy regime and its implementation.

**Interdependence of monetary and exchange rate policies**

The tight linkage between monetary and exchange rate policy is most clearly expressed in the doctrine of the ‘impossible trinity’: a country cannot at the same time maintain a fixed exchange rate and an independent monetary policy in a context of perfect capital mobility. Put another way,
a credible peg would not allow a country to run different interest rates from
those in the anchor country, provided arbitrage was free to operate and
unlimited in quantity. In practice, of course, capital is not perfectly mobile,
leaving some scope for monetary policy independence even with an
exchange rate target. Moreover, as Frankel (1999) has persuasively argued,
countries can trade off some exchange rate fixity for some monetary inde-
pendence, consistent with adopting an intermediate exchange rate regime.

The constraints on countries operating a pegged exchange rate have
increased as capital has become more mobile. Increased capital mobility
has occurred for essentially two sets of reasons. Increasing financial sophis-
tication, technological advances and increased financial wealth have vastly
increased the volume of capital flows and hence the resources that can be
marshaled to speculate against a currency peg. And a growing consensus
over the last three or four decades that liberalization increases economic
efficiency (subject to the cautions expressed by Diaz-Alejandro, 1985),
which were confirmed by the emerging market crises of the 1990s) has
meant that few countries attempt to maintain tight government regulation
of capital flows.

In contrast, during the early post-World War II Bretton Woods period of
fixed but adjustable exchange rates there were pervasive controls on capital
flows as well as extensive domestic financial regulation, or even ‘financial
repression’. In this context, monetary conditions could be set with little
concern for external consequences and interest rates were kept low to stim-
ulate investment. Some envisaged the deliberate use of inflation to raise the
relative price of capital goods, lower real wages, shift income to those with
higher saving propensities (that is, the rich), and call forth increased output
for development. For instance, a respected text on development, Higgins
(1959), discusses this strategy, advocated by Martin Bronfenbrenner; Hig-
gins accepts the argument in principle, but argues that the optimal rate
of inflation is likely to be less than 5 percent per year. Curiously, the incom-
patibility with the prevailing system of fixed exchange rates is not men-
tioned, nor does ‘exchange rate’ appear at all in the index of this
comprehensive text. At most, there is a reference to the fact that inflation
would make export industries increasingly unprofitable, aggravating
balance-of-payments difficulties (Higgins, 1959, p. 464). The context was
one where constraints on monetary policy came not from the fledgling
capital account but rather through the competitiveness of the trade
account.

**Targets for monetary policy**

Corden (2002) has provided a useful taxonomy of what he calls ‘approaches’ for exchange rate policy, which for the reasons above we will
adapt for our discussion of monetary policy: the ‘real targets’ approach, the
‘nominal anchor’ approach, and the ‘exchange rate stability’ approach. He
distinguishes these approaches from what he considers to be the three (not
two) polar regimes: an absolutely fixed exchange rate, pure floating, and the
fixed but adjustable exchange rate regime (FBAR), which was the regime
under the Bretton Woods period. Thus, his classification of regimes neatly
sidesteps the bipolar classification of proponents of a ‘hollowing-out’ of
intermediate regimes. But Corden maintains that the FBAR is not simply
a compromise between the other two regimes. Because of its credibility
problems – it involves an explicit or implicit commitment to a peg, without
a corresponding assignment of the instruments needed to ensure mainte-
nance of the peg – it differs from other intermediate regimes which do not
promise so much, such as managed floating or crawling pegs or bands.
The ‘real targets’ approach presupposes nominal wage rigidity, so that in
the short run, if not in the long run, expansionary monetary policy (or
nominal exchange rate depreciation) can affect real output, employment
and the real exchange rate. The value of using monetary policy in this
fashion is greatest when the economy faces negative real shocks and fiscal
policy is prevented from operating in a stabilizing way. However, the assign-
ment of monetary policy to real targets suffers the disadvantage that it does
not provide an anchor for the price level or the rate of inflation. Moreover,
it relies on a degree of money illusion that is endogenous, and is likely to
shrink drastically if monetary policy systematically tries to exploit it.
Hence the increasing emphasis, among central banks of the world, on
‘nominal anchors’ for monetary policy – a rigid link or a target for a
nominal quantity or price which is intended to prevent the economy’s
overall price level or rate of inflation from wandering off. Nominal anchors
can be divided into domestic variables (a monetary aggregate, nominal
GDP or the rate of inflation itself), and external anchors, in particular a
peg to another currency or to a world commodity (such as gold). In prin-
ciple, either a domestic or foreign nominal variable can anchor the price
level and produce long-run nominal stability, but different targets can yield
different short-run outcomes for output and inflation. They also differ in
terms of institutional requirements and expose the economy to a crisis in
varying degrees.
The ‘exchange rate stability’ approach is to be distinguished from the
desire to put in place a nominal anchor: it postulates that exchange rates left
to themselves simply add noise to the world economy, perhaps because of
self-fulfilling expectations and destabilizing speculation (Williamson, 2000).
Though in some cases exchange rate flexibility could facilitate adjustment,
movements in exchange rates are dominated by short-run volatility unrela-
ted to economic fundamentals and by medium-term misalignments. In this
view, a system of credibly fixed rates would clearly improve welfare compared to exchange rate flexibility. Williamson has argued that an intermediate regime with an explicit exchange rate target (for example, a band-basket-crawl – or BBC – regime) would have some of the same advantages of anchoring expectations and taming volatility.

The exchange rate as a nominal anchor
Exchange rate-based stabilization (ERBS) has been used, with varying degrees of success, in reducing a high initial rate of inflation. The advantage of the exchange rate as nominal anchor is that it is visible, easily explained to the public and requires little institutional credibility – hence its attraction for countries suffering from high or hyperinflation, as Chile in 1979, Argentina in 1991 and Brazil in 1994. In these and other similar cases, the only way to achieve a modicum of monetary policy credibility is to tie the hands of the central bank, since unlike the inflation rate or a monetary aggregate, the exchange rate is a variable which the public observes directly and continuously. However, the principal drawback of this strategy is that it requires an eventual exit from the peg unless the economy is to undergo a severe deflation to remove loss of competitiveness resulting from the accumulated inflation. Indeed, because of the stickiness of inflation, bringing it down to industrial-country levels in the space of even a few years still leaves embodied in the price level the integral of the inflation gaps incurred during that time. Eliminating them would require an extended period of deflation, involving output losses that few governments would willingly incur. Therefore, pegs associated with ERBS become increasingly non-credible over time, even in countries which are successful in achieving low inflation.

The problem is compounded in the presence of sufficient capital mobility that investors can take positions against the currency large enough to exhaust the authorities’ foreign exchange reserves. Thus, a speculative attack could force the authorities to devalue or float. Paradoxically, then, the attempt to gain credibility by using an external anchor sows the seeds of its own downfall. The trick is to ride the wave long enough to benefit from initial credibility gains without getting locked into a strategy that will eventually throw you up on the beach. Unfortunately, while the strategy is successful, there is little pressure on politicians to change, while when the peg is under attack, it is too late: exits in a crisis usually have dire consequences (Eichengreen et al., 1999).

There is a category of countries, however, for which monetary policy independence is of little use and hence a hard peg is credible. These are small, open countries with a high export concentration on a single commodity or service priced in an international currency, or a dominant trade
partner. In particular, countries in the Caribbean which rely heavily on tourism and banking services have long-standing and credible pegs to the US dollar. Other countries in this category are the small neighbors of South Africa – Lesotho, Namibia and Swaziland – which participate in the Common Monetary Area. This arrangement allows the countries to have their own currencies, exchangeable at par with the rand. Nepal is an example in Asia; that country pegs its currency to the Indian rupee. Finally, some countries – among which are Ecuador and Panama – have simply adopted a foreign currency, that is, have a regime of official dollarization.

It is relevant to examine the empirical determinants of changes in regimes. In Masson (2001) and Masson and Ruge-Murcia (2005), exchange rate regimes are divided into fixed, intermediate and flexible. Using data from 1975–97 for as many as 168 countries, the probability of changing regimes was related to macroeconomic determinants, foreign exchange reserves divided by GDP, and trade openness. The probability of abandoning any of the regimes was greater, the higher the rate of inflation and the lower real GDP growth (Masson and Ruge-Murcia, 2005). The intuition is clear: regimes are abandoned in bad economic times, not good ones. For instance, high inflation makes a peg increasingly precarious, but also makes it more likely that countries that are floating will use ERBS in order to reduce it.

Analysis of exchange rate regime transitions also permits testing formally the ‘hollowing-out’ hypothesis (Eichengreen, 1994). Hollowing-out requires transitions away from intermediate regimes, but not towards them from the poles of hard fixes and free floats. Using a constant transition matrix, that hypothesis can be rejected (Masson, 2001). Moreover, the existence of continuing negative shocks that produce high inflation and slow growth suggests that there will be continued cycling among regimes when the transition probabilities are endogenous, as described above. The idea that hard fixes such as currency boards are immune from crisis was decisively proved wrong by Argentina’s abandonment of its Convertibility Law in January 2002, and the floating of the peso. Despite having a credible ERBS based on institutional guarantees, the severe recession suffered by Argentina during 1998–2001 made maintaining the exchange rate straitjacket difficult. Argentina also illustrates the need for support from fiscal policy to make any monetary regime successful – but especially so for a fixed rate regime. Earlier fiscal adjustment would have allowed Argentina to avoid a debt crisis and would have helped maintain international competitiveness and current account balance.

**Domestic nominal anchors**

The principal choice of domestic nominal anchor is between a monetary aggregate and inflation targeting. Monetary aggregates have the advantage...
of being relatively easy to measure and of being a financial variable that, at least in the era of highly regulated financial systems, was relatively easy to control. It was postulated that there was a stable relationship between those financial and real variables, principally taking the form of a stable (and simple) money demand equation.

However, there is accumulating evidence that money demand is unstable. Greater access to other financial assets associated with liberalization has changed its nature, and probably increased the interest elasticity of demand for non-interest-bearing deposits and cash. Technological changes have also allowed greater opportunities for conserving on transactions balances. Finally, partial dollarization – the circulation of a foreign currency – provides another source of instability in the demand for the domestic currency. As a result of all these factors, targets for monetary aggregates, abandoned by almost all industrial countries, are now also increasingly being abandoned by developing countries.

Emerging-market countries have in a number of cases adopted inflation targeting. Chile was the precursor, announcing a target for consumer price index (CPI) inflation in 1991, albeit accompanied initially by an exchange rate target band. Other developing countries having forms of inflation targeting regimes include Brazil, Colombia, the Czech Republic, Israel, Korea, Mexico, Peru, Poland, South Africa and Thailand (see Mishkin and Schmidt-Hebbel, 2002, for details).

**IT in developing countries: prerequisites and experience**

Inflation targeting as a monetary policy regime has to be distinguished from a situation in which the central bank merely expresses a desire to lower the rate of inflation to a particular level or maintain it there. Price stability in some form is always part of a central bank’s mandate, but IT aims to enhance the credibility of the central bank’s commitment to price stability by improving its accountability. Announcing targets which are not met because they have no effect on policy or because they are over-ridden by other objectives does nothing to improve the credibility or effectiveness of monetary policy.

Two basic prerequisites for putting in place an IT regime are that the central bank, which is charged with implementing monetary policy, be given a reasonable degree of ‘instrument’ independence to carry out that task; and the absence of commitment to a target for another nominal variable (Masson et al., 1997). Hybrid regimes are of course possible (see below) in which countries have targets for both the rate of inflation and the money supply or the exchange rate; in practice, several developing countries have operated such a regime for a transitional period. In this sense, the advocates of IT who deride a ‘prerequisite approach’ to IT (for example,
Sterne, 2002) have a point: one can start doing ‘baby steps’ as a way of learning how to walk. However, from the standpoint of clarity it is important to be clear what the ultimate objective should be, so as to put in place the capabilities needed to achieve it.

Subject to the above two prerequisites, the IT regime needs to involve the following elements of a framework for monetary policy: quantitative targets for the rate of inflation, over a specified horizon; a commitment to those targets as overriding objectives for policy; a clear methodology for making inflation forecasts; and a transparent way of translating the possible expected deviations from target into changes in the instruments of monetary policy.

How successful are developing countries in meeting the prerequisites? First, IT is likely to be a candidate regime primarily for middle-sized or large middle-income countries – roughly speaking, the ‘emerging market economies’. Smaller, very open economies may well choose to peg – a credible monetary regime for them; and countries with low incomes would typically not have the financial development or institutional capacity to implement inflation targeting. Second, countries differ greatly as to the degree of central bank independence. In particular, where there is fiscal dominance, an independent monetary policy is impossible. Compared to industrial countries, seigniorage – used to finance fiscal deficits – in many developing countries is high. Third, de jure independence may not guarantee that the central bank is able to carry out its mandate in the face of lack of public support. A constituency in favor of low inflation has not developed in many emerging market economies. And de jure independence has sometimes been overridden, as in Argentina, where a central bank governor was summarily replaced for disagreeing with the Minister of Finance. Fourth, administered price changes and centralized negotiations (for example the Pacto in Mexico) that determine a large fraction of the economy’s annual rate of wage increase may interfere with the central bank’s ability to control inflation unless they are coordinated with the inflation target. Finally, forecasting inflation is difficult in many emerging-market countries because of highly unstable macroeconomies and lack of solid econometric relationships (due, for example, to an insufficiently long or homogeneous data sample).

These obstacles are not insurmountable, however. The experience of emerging-market economies that practice IT is generally favorable, since countries have typically met their inflation targets. Moreover, the inflation-targeting countries have had a better experience of avoiding the balance-of-payments crises than other emerging-market countries since 1994. Thus, taking as given that emerging-market countries face a more challenging environment for monetary policy, this does not imply that the choice of
regime should necessarily be biased away from IT. Instead, Fraga et al. (2003) suggest that IT be operated somewhat differently than for industrial countries, perhaps by using wider and ‘softer’ bands around the inflation target, higher targets so as to accommodate bigger shocks without risking deflation, and adjusting targets more flexibly in response to shocks (while making special efforts to communicate the reasons to the general public).

Another feature of the early years of inflation targeting in developing countries has been its association with exchange rate targets. Thus, Chile, Israel and Poland, for instance, for a time targeted a band for the exchange rate (crawling, with the band width also adjusted depending on circumstances). Such a policy had the advantage of easing into the IT framework, allowing experience with it to be built up while retaining a backstop which might prevent instability should the inflation target give the wrong signal. The disadvantage of combining the two targets, as described in Israel’s case by Bufman and Leiderman (2000), was lack of transparency and the danger of conflicting signals which could add to the public’s uncertainty about monetary policy. In practice, given the greater immediacy of the exchange rate variable (continuously observable and widely publicized), it tended to dominate the inflation target if they conflicted, and this then required an explicit change in the exchange rate band if inflation was to be given its proper weight. Over time, the bands for the exchange rate were widened, as inflation declined and confidence with the new regime increased. Israel now has no exchange rate target. A similar progression occurred in Chile’s case. At the time of the Russian crisis in August 1998 concern about external developments led the central bank to increase the short-term interest rate, causing the economy to go into a severe recession (Morandé, 2002). Sole emphasis on the inflation target would have allowed some easing of monetary policy. The exchange rate target was abandoned in September 1999, and the Bank of Chile at the time of writing in 2006 sets its monetary policy to maintain inflation within a 2–4 percent target range.

Unresolved questions concerning inflation targeting

While the verdict so far in emerging-market countries is positive, the track record is quite short. For most countries, it dates from the end of the 1990s. The environment has been relatively benign, with low inflation and low interest rates prevailing in the industrial world; this has permitted all developing countries, whether inflation targeters or not, to reduce inflation. Moreover, this period has not seen any major contagion from emerging-market currency or balance-of-payments crises. It remains to be seen whether a major world inflation shock would be weathered well by the IT regimes in place. As noted above, monetary and exchange rate
regimes have been periodically adopted and then abandoned – will the same be true of IT? Since IT is a less rigid regime – often characterized as ‘constrained discretion’ – it is likely to be less fragile than strict exchange rate targets. The danger remains however that the credibility gains that have accrued to IT central banks may be dissipated by persistent overshoots or indications that inflation does not provide an effective guide for month-to-month policy-setting. Thus, IT will face several challenges going forward.

An important question that should condition views of the advantages of IT is whether in fact it is subject to speculative attack. Kumhoff (2002) argues that the regime can be attacked, and that moreover it behaves much more like a fixed exchange rate regime than is often claimed. In the face of an unsustainable fiscal policy, it might have to be abandoned. Favero and Giavazzi (2004) provide a formal model in which such a fiscal policy could lead, through increasing default risk, to severe constraints on monetary policy’s ability to deliver on its inflation commitment. Contrary to the analysis of Fraga et al. (2003), they conclude that for at least a short period in 2002, Brazil’s economy might have tipped into a regime of fiscal dominance that, had it continued, would have doomed the IT regime. Thus, they are not convinced that the regime was ‘stress tested’.

The continued spread and popularity of IT may be affected by a parallel trend toward monetary unions, following in the wake of the successful creation of the Eurozone. Already, European integration has eliminated two inflation targeters, Finland and Spain, through their joining the Eurozone; and several among the new EU members, in particular the Czech Republic, Hungary and Poland, are expected to join the Eurozone within a few years. Sweden and the United Kingdom, prominent inflation targeters, could also conceivably join. While reducing the number of central banks implementing inflation targeting, of course should the European Central Bank (ECB) clearly adopt that regime, then the economic area it applied to might increase.

Turning to other continents, Africa, the Middle East and Asia are also considering regional monetary integration that might lead to a common currency. In Masson and Pattillo (2004) the argument is made in the context of Africa that the European example does not translate well to other regions with less broad-based integration projects and less strong regional solidarity. Thus, the success of an African single currency seems doubtful; instead, greater monetary integration could arise around regional poles such as South Africa, which already has a successful IT monetary regime. Since the Eurozone, the United States and Japan are likely to want to retain their monetary independence and exchange rate flexibility, regional currency blocs, if they are created, are unlikely to anchor their
exchange rate to any single reserve currency. On balance, then, inflation targeting among emerging market (EM) countries is likely to remain the regime of choice, whether or not regional integration proceeds.

When considering implementation of inflation targeting, a still unresolved issue is the weight to give to other variables, in addition to the inflation forecast. While of course other variables may influence the inflation forecast (for instance, the current output gap or the actual exchange rate) it seems that most countries give some additional weight to those variables – for instance, lowering interest rates if activity is weak, provided the inflation forecast is within the target range. Thus, the influence of other variables on policy may be asymmetric, but non-zero; this is consistent with Mervyn King’s dictum that central banks are not ‘inflation nutters’. Developing countries in particular are loath to ignore the exchange rate, and as Calvo and Reinhart (2002) show, exhibit a ‘fear of floating’. While intervention may smooth some exchange rate fluctuations, central banks may also want to use interest rate policy for that purpose – subject to the caveats evoked by Chile’s experience in 1998.

Finally, should the target for the (long-run) inflation rate be higher in emerging markets, as suggested by Fraga et al. (2003), because of larger shocks facing those countries? More empirical evidence is needed to resolve this issue. While the Balassa–Samuelson effect would suggest higher average inflation in faster-growing developing countries (provided one wants to avoid falling tradable goods prices in domestic currency), the argument for avoiding deflation really requires a greater understanding of the nature of downward rigidities. And the cross-country literature on costs of inflation is not very precise on when inflation becomes costly. These remain important issues for monetary policy in developing countries, whether they have adopted IT or not.

Notes
1. I am grateful to Max Corden, Frederic Mishkin and Miguel Savastano for comments on this chapter.
2. Even long-run neutrality does not command unanimity, however, and recent models suggest that transitory effects may last longer than previously thought (for example, Mankiw and Reis, 2001).
3. Programs supported by the International Monetary Fund typically included targets for the central bank’s net domestic assets.
4. However, Fraga et al. (2003) consider that Brazil ‘stress-tested’ its IT regime during 2002, when it faced a negative capital account swing of about 6 percent of GDP, missing its inflation target but not suffering a permanent loss of credibility for the regime.

References
Frankel, Jeffrey (1999), ‘No Single Currency Regime is Right for All Countries Or At All Times’, Essays in International Finance, No. 215, Princeton University.
Introduction
Fiscal policy plays an increasingly important role in many developing countries. Decisions on fiscal policy, especially if properly synchronized with monetary policy, can help smooth business cycles, ensure adequate public investment and redistribute incomes.

The four main components of fiscal policy are: (1) expenditure, budget reform; (2) revenue (particularly tax revenue) mobilization; (3) deficit containment and financing; and (4) determining fiscal transfers from higher to lower levels of government. Fiscal policy works through both aggregate demand and aggregate supply channels. Changes in total taxes and public expenditure affect the level of aggregate demand, whereas the structures of taxation and public expenditure affect, among others, the incentives to save and invest (at home and abroad), take risks, and export and import goods and services.

This chapter is organized as follows. It first analyses some basic fiscal issues confronting developing countries. Then it considers budgetary deficits and evaluates norms for tax and expenditure reforms. The final section concludes.

Tax and expenditure profiles of developing countries
Developing-country fiscal systems have three basic characteristics: (1) low tax–GDP and expenditure–GDP ratios compared to developed countries, even though developing countries need more public expenditure; (2) the fiscal stance is often procyclical; and (3) tax resources are more volatile than those of developed countries.

There are manifold pressures for high and growing government expenditure in developing countries. Because of their low per capita incomes and high poverty, developing countries face an urgency to raise growth rates. This places a heavy burden on policy whereas, concurrently, the limited efficacy of policy instruments and governance inadequacies constrain policy. Pressures for populism through price controls and the like are considerable. The state in many developing countries is politically weak and beset with lack of consensus on what constitutes a sound fiscal policy compared to most developed countries (Heady, 2004).
Table 55.1 shows government tax revenues in developed, transition and developing countries for two time periods, 1990 to 1995, and 1996 to 2002. In the median developing country the tax–GDP ratio was below 20 per cent whereas in the median transition economy it was 30 per cent and in developed countries 40 per cent. Unsurprisingly, on average, governments in many developing countries face a severe resource crunch.

Only three out of 21 developed countries and two out of 14 transition countries had revenues falling over the two time periods, whereas the corresponding magnitude for developing countries was five out of 13. The share of distortionary commodity and trade taxes in total central government revenue is higher in developing countries (Tables 55.2 and 55.3).

Table 55.3 shows that in the richest countries personal income taxes are the most significant and contribute more than half (54.3 per cent) of tax revenue. Next are various commodity taxes and then the corporation tax. Border taxes and seigniorage revenue are low, reflected in the low value for inflation. The informal sector is small. With falling gross domestic product (GDP) per capita tax revenue as a percentage of GDP drops and corporate taxes as a percentage of revenue rise. Income taxes remain at about 30 to 35 per cent of revenue and commodity taxes are high. High rates of inflation reflect high values of seigniorage revenue. The poorest among the developing countries raise almost a quarter of their revenue through seigniorage. The informal sector in developing countries is about twice that in developed countries.

### Table 55.1 Total government tax revenue as a percentage of GDP, 1990–95 and 1996–2002 (domestic prices, number of countries and median values of the simple averages)

<table>
<thead>
<tr>
<th></th>
<th>1990–95</th>
<th></th>
<th>1996–2002</th>
<th></th>
<th>Direction of change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Countries</td>
<td>Median</td>
<td>Countries</td>
<td>Median</td>
<td>Down</td>
</tr>
<tr>
<td>Separate samples</td>
<td>56</td>
<td>31.9</td>
<td>53</td>
<td>33.8</td>
<td></td>
</tr>
<tr>
<td>Combined sample</td>
<td>48</td>
<td>33.5</td>
<td>48</td>
<td>33.8</td>
<td>20</td>
</tr>
<tr>
<td>Developed countries</td>
<td>21</td>
<td>37.8</td>
<td>21</td>
<td>40.1</td>
<td>3</td>
</tr>
<tr>
<td>Transitional economies</td>
<td>14</td>
<td>34.7</td>
<td>14</td>
<td>31.4</td>
<td>12</td>
</tr>
<tr>
<td>Developing countries</td>
<td>13</td>
<td>18.7</td>
<td>13</td>
<td>19.2</td>
<td>5</td>
</tr>
</tbody>
</table>

**Note:** Tax revenue is computed as the sum of revenue accruing to central and local governments. For each country in each category average tax revenue as a percentage of GDP is computed. For each category of country the median value of this average tax revenue is reported in the chapter.

**Source:** UNPAN Statistics.

Table 55.1 shows government tax revenues in developed, transition and developing countries for two time periods, 1990 to 1995, and 1996 to 2002. In the median developing country the tax–GDP ratio was below 20 per cent whereas in the median transition economy it was 30 per cent and in developed countries 40 per cent. Unsurprisingly, on average, governments in many developing countries face a severe resource crunch.

Only three out of 21 developed countries and two out of 14 transition countries had revenues falling over the two time periods, whereas the corresponding magnitude for developing countries was five out of 13. The share of distortionary commodity and trade taxes in total central government revenue is higher in developing countries (Tables 55.2 and 55.3).

Table 55.3 shows that in the richest countries personal income taxes are the most significant and contribute more than half (54.3 per cent) of tax revenue. Next are various commodity taxes and then the corporation tax. Border taxes and seigniorage revenue are low, reflected in the low value for inflation. The informal sector is small. With falling gross domestic product (GDP) per capita tax revenue as a percentage of GDP drops and corporate taxes as a percentage of revenue rise. Income taxes remain at about 30 to 35 per cent of revenue and commodity taxes are high. High rates of inflation reflect high values of seigniorage revenue. The poorest among the developing countries raise almost a quarter of their revenue through seigniorage. The informal sector in developing countries is about twice that in developed countries.
Auriol and Warlters (2005) argue that the informal sector in developing countries is large because of the higher costs of entry into the formal economy. By keeping barriers to entry into the formal economy high, those firms and individuals who make it into the formal economy acquire large rents and hence may be easier to tax than a diffused set of small taxpayers. If this argument is correct then encouraging large formal sectors should be part of a government strategy to increase tax revenue. Data for 64 countries indicate that this is indeed the case, particularly in Africa, for example, 0.4 per cent of taxpayers account for 61 per cent of total domestic tax collection in Kenya and 57 per cent in Colombia. As general policy, Auriol and Warlters argue that developing countries should lower entry barriers and raise the size of the formal sector to raise tax revenues.

Further, rapid globalization, technological advancement and the accompanying movement of factors of production across national boundaries, and the emergence of multinational corporations as major actors have eroded many developing countries’ tax bases. Taxpayers can more easily raise income outside of conventional channels (Lao-Araya, 2003).

With inflexible public expenditures and low tax revenues government finances in developing countries are weak, with high deficits, debts and debt-servicing obligations. Consolidated figures for the finances of central and local governments together are not readily available but Table 55.4 presents these for central governments. Typically, developing countries’ revenues and expenses are lower and interest payments higher than in developed countries, although government consumption in developing countries is lower than that in developed countries (Table 55.5).

<table>
<thead>
<tr>
<th></th>
<th>No. of countries</th>
<th>Direct taxes</th>
<th>Payroll taxes</th>
<th>Sales taxes</th>
<th>Trade taxes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sample</td>
<td>139</td>
<td>27.1</td>
<td>5.9</td>
<td>34.8</td>
<td>14.6</td>
</tr>
<tr>
<td>Developed countries</td>
<td>24</td>
<td>34.8</td>
<td>28.5</td>
<td>28.1</td>
<td>0.5</td>
</tr>
<tr>
<td>Transitional economies</td>
<td>23</td>
<td>17.7</td>
<td>33.3</td>
<td>38.9</td>
<td>6.2</td>
</tr>
<tr>
<td>Developing countries</td>
<td>92</td>
<td>27.6</td>
<td>0.7</td>
<td>33.0</td>
<td>24.9</td>
</tr>
<tr>
<td>Africa</td>
<td>32</td>
<td>27.2</td>
<td>0.2</td>
<td>30.7</td>
<td>33.0</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>27</td>
<td>22.1</td>
<td>5.1</td>
<td>38.9</td>
<td>13.7</td>
</tr>
<tr>
<td>Asia &amp; Oceania</td>
<td>33</td>
<td>34.1</td>
<td>0.0</td>
<td>34.8</td>
<td>25.6</td>
</tr>
</tbody>
</table>

Source: UNPAN Statistics.
### Table 55.3 Sources of Government Revenue (1996–2001)

<table>
<thead>
<tr>
<th>GDP per capita</th>
<th>Tax revenue (% of GDP)</th>
<th>Income taxes (% of revenue)</th>
<th>Corporate income tax (% of income taxes)</th>
<th>Consumption &amp; Production taxes (% of revenue)</th>
<th>Border taxes (% of revenue)</th>
<th>Inflation rate (%)</th>
<th>Seigniorage Income (% of revenue)</th>
<th>Informal economy (% of GDP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;$745</td>
<td>14.1</td>
<td>35.9</td>
<td>53.7</td>
<td>43.5</td>
<td>16.4</td>
<td>10.6</td>
<td>21.8</td>
<td>26.4</td>
</tr>
<tr>
<td>$746–2975</td>
<td>16.7</td>
<td>31.5</td>
<td>49.1</td>
<td>51.8</td>
<td>9.3</td>
<td>15.7</td>
<td>24.9</td>
<td>29.5</td>
</tr>
<tr>
<td>$2976–$9205</td>
<td>20.2</td>
<td>29.4</td>
<td>30.3</td>
<td>53.1</td>
<td>5.4</td>
<td>7.4</td>
<td>6.0</td>
<td>32.5</td>
</tr>
<tr>
<td>All developing</td>
<td>17.6</td>
<td>31.2</td>
<td>42.3</td>
<td>51.2</td>
<td>8.6</td>
<td>11.8</td>
<td>16.3</td>
<td>30.1</td>
</tr>
<tr>
<td>&gt;$9206</td>
<td>25.0</td>
<td>54.3</td>
<td>17.8</td>
<td>32.9</td>
<td>0.7</td>
<td>2.2</td>
<td>1.7</td>
<td>14.0</td>
</tr>
</tbody>
</table>

**Source:** Gordon and Li (2005).
Table 55.4  Finances of central governments for country groups

<table>
<thead>
<tr>
<th>Country group</th>
<th>Revenue (% of GDP)</th>
<th>Expense (% of GDP)</th>
<th>Cash surplus or deficit (% of GDP)</th>
<th>Net incurrence of liabilities (% of GDP)</th>
<th>Debt and interest payments</th>
<th>Interest payment as % of revenue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income</td>
<td>13.5</td>
<td>13.0</td>
<td>15.5</td>
<td>15.5</td>
<td>-2.6</td>
<td>-3.2</td>
</tr>
<tr>
<td>Middle income</td>
<td>17.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower middle income</td>
<td>16.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper middle income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East Asia &amp; Pacific</td>
<td>8.4</td>
<td>11.5</td>
<td>12.0</td>
<td></td>
<td>-2.1</td>
<td></td>
</tr>
<tr>
<td>Europe &amp; Central Asia</td>
<td>31.0</td>
<td>31.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Latin America &amp;</td>
<td>20.9</td>
<td>23.0</td>
<td>-0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caribbean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle East &amp; North</td>
<td>28.3</td>
<td>23.5</td>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asia</td>
<td>13.2</td>
<td>12.4</td>
<td>15.4</td>
<td>15.1</td>
<td>-2.7</td>
<td>-3.1</td>
</tr>
<tr>
<td>High Income</td>
<td>26.0</td>
<td>28.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Europe EMU</td>
<td>36.3</td>
<td>35.7</td>
<td>38.8</td>
<td>38.6</td>
<td>-2.3</td>
<td>-2.3</td>
</tr>
</tbody>
</table>

Jha (2006) reports that the unweighted average of tax buoyancy (defined as \( \frac{\text{Percentage change in tax revenue}}{\text{Percentage change in tax base}} \)) for several developing countries is larger than one, indicating that an expansion of income would lead to an increase in the tax–GDP ratio. Gordon and Li (2005) argue that taxation, by its very nature, must depend on the formal economy since bank records are needed to identify taxable activity. In rich countries the intermediary services provided by the financial sector are considerable, so there is a high cost of abandoning it and conducting business in the informal sector. However, this is not the case in developing countries. Further, their tax base is likely to be narrow (biased towards capital income) and cover mostly capital-intensive firms that need the financial sector the most, and tariffs are used to protect the capital-intensive sectors and shortfalls in revenue (from public expenditures) would often be met through seigniorage.

Another important characteristic of fiscal variables in developing countries is their instability. Table 55.6 reports on key fiscal variables in 13 Latin American developing countries and 14 industrialized countries. In terms of all categories and in both nominal and real terms, computed coefficients of variation are much higher for Latin American developing countries than for industrialized countries.

Fiscal variables in many developing countries move in a procyclical fashion. Standard Keynesian models require that fiscal policy should be countercyclical, that is, during recessions taxes should be lowered and public expenditure hiked whereas during good times, taxes are raised and public expenditures lessened to reduce chances of overheating of the economy.

### Table 55.5 Government consumption as percentages of GDP, 1990, 1996, 2002 (domestic prices, median values)

<table>
<thead>
<tr>
<th>Number of Countries</th>
<th>1990</th>
<th>1996</th>
<th>2002¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete sample</td>
<td>114</td>
<td>15.3</td>
<td>15.7</td>
</tr>
<tr>
<td>less transitional economies</td>
<td>101</td>
<td>5.2</td>
<td>14.3</td>
</tr>
<tr>
<td>Developed countries</td>
<td>24</td>
<td>18.9</td>
<td>19.4</td>
</tr>
<tr>
<td>Transitional economies</td>
<td>13</td>
<td>20.0</td>
<td>18.0</td>
</tr>
<tr>
<td>Developing countries</td>
<td>77</td>
<td>14.2</td>
<td>12.7</td>
</tr>
<tr>
<td>Africa</td>
<td>26</td>
<td>15.1</td>
<td>12.8</td>
</tr>
<tr>
<td>Latin America &amp; Caribbean</td>
<td>25</td>
<td>12.9</td>
<td>13.4</td>
</tr>
<tr>
<td>Asia &amp; Oceania</td>
<td>26</td>
<td>12.2</td>
<td>11.7</td>
</tr>
</tbody>
</table>

*Note: ¹ Or latest data (2000, 2001).*

*Source: UNPAN Statistics.*
economy. In contrast the ‘Ricardian equivalence’ hypothesis suggested by Barro (1979) suggests that since rational economic agents make decisions based on perfectly anticipated tax and expenditure policies of the government, fiscal policy should remain neutral over the business cycle and respond only to unanticipated changes that affect the government’s budget constraint.

Using a sample of 56 countries (20 developed and 36 developing) Talvi and Vegh (2005) show that in G7 countries fiscal policy follows Barro, whereas for developing countries it has been procyclical. Two plausible explanations for this phenomenon exist. The first is that tax bases are so narrow and public expenditure so inelastic in developing countries that tax revenues and expenditures rise during expansions, whereas during recessions revenues and expenditures both decline for similar reasons. Second, as Talvi and Vegh (2005) argue, since fluctuations in the tax base are much larger in developing countries than in developed countries, full tax smoothing would require large surpluses during good times which is not possible since public expenditures are inelastic and resources may be wasted in enhanced public expenditures on public sector undertakings and subsidies, instead of retiring of debt as full tax smoothing would require.

Table 55.6 Coefficients of variation of key fiscal variables

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Industrialized Countries</td>
<td>Latin America</td>
</tr>
<tr>
<td>Total revenue</td>
<td>0.15</td>
<td>0.55</td>
</tr>
<tr>
<td>Current revenue</td>
<td>0.15</td>
<td>0.56</td>
</tr>
<tr>
<td>Non-tax revenue</td>
<td>0.19</td>
<td>0.58</td>
</tr>
<tr>
<td>Tax revenue</td>
<td>0.15</td>
<td>0.56</td>
</tr>
<tr>
<td>Total expenditure</td>
<td>0.16</td>
<td>0.55</td>
</tr>
<tr>
<td>Current expenditure</td>
<td>0.16</td>
<td>0.55</td>
</tr>
<tr>
<td>Government consumption</td>
<td>0.15</td>
<td>0.54</td>
</tr>
<tr>
<td>Interest payment</td>
<td>0.22</td>
<td>0.63</td>
</tr>
<tr>
<td>Transfers</td>
<td>0.17</td>
<td>0.58</td>
</tr>
<tr>
<td>Capital formation</td>
<td>0.17</td>
<td>0.57</td>
</tr>
</tbody>
</table>

Notes:
1. There are 13 Latin American and 14 industrialized countries.
2. In the sample the Gavin and Perotti (1997) database is used for the analysis.

Fiscal deficit issues
The exercise of fiscal policy in developing countries has its limits. The combination of low revenues and inelastic expenditures means that expenditures routinely, and even increasingly, outpace revenues. Jha (2004) argues that there is considerable heterogeneity in experience with respect to the fiscal deficit, between the middle- and low-income country categories and even within the low-income category countries. Indeed, the poorest among the least-developed countries are caught in an insidious resource trap and the average least-developed country economy has, since the 1970s, been exposed to adverse external trade shocks with an impact, in the worst years, approximately double the average of other developing countries (UNCTAD, 2000).

External finance is limited, especially for the poorest countries, although large, stable economies attract considerable capital inflows. Official aid has been falling and private equity flows go to the best-performing developing and transition economies. Private loans, as Harberger (1985) notes, are available at increasingly difficult terms since the domestic resource cost (often underestimated) of servicing these increases with additional borrowing. Other reasons for differences across developing countries include continuity and stability of policy regimes: Zambia, with several policy reversals, will be associated with greater risks than Mauritius, which has had a credible and stable policy regime.

Given financing constraints many developing countries have to opt for some non-bond (monetary) financing of the deficit. This establishes a direct link between fiscal policy and the monetary base of the central bank, blurs the distinction between fiscal and monetary policy, and compromises central bank independence.3 If bond financing is chosen, private investment may get crowded out.

Jha (2004), shows that in the long term public revenue and public expenditure are unrelated in many developing countries so that any excess of expenditure over revenue cannot be financed by generating budgetary surpluses over a long enough time horizon. Thus fiscal deficits are unsustainable in many developing countries. Mendoza and Ostry (2007) argue that whereas fiscal policy in most countries is responsive to budgetary deficits, high-debt countries do run a risk of having an unsustainable fiscal stance.4

Jha (2004) also shows that current account deficits are unsustainable in many developing countries. The fact that external sustainability conditions are hard to meet would imply the need for continual capital inflow in order to keep the balance of payments in equilibrium, necessitating the maintenance of a substantial rate-of-return wedge between domestic and foreign rates of return. This raises domestic interest rates substantially above global interest rates and acts as a drag on higher growth, making debt servicing harder, and exacerbates the fiscal deficit.
However, public expenditure could be productive, so whether public deficits impede or spur economic growth becomes an empirical question. In this context Adam and Bevan (2005) examine the relation between fiscal deficits and growth for a panel of 45 developing countries over 1970–99. Public expenditure is permitted to be both growth-enhancing as well as growth-inhibiting and distortionary taxes exist and fiscal deficits are permitted. They show that the impact of the deficit depends upon the mode of financing it. Deficits can be growth-enhancing if financed by limited seigniorage, growth-inhibiting if financed by domestic debt, and have opposite flow and stock effects if financed by external loans at market rates. These opposite effects define a threshold effect, before attaining which fiscal deficit has growth-enhancing effects and after which the effects of fiscal deficits are growth-inhibiting. Adam and Bevan find this threshold figure to be around 1.5 per cent of GDP after grants.

**Norms for tax and expenditure reforms in developing countries**

One of the principal aims of a meaningful tax and expenditure reforms policy would be to bolster the savings and investment rates in the economy in order to raise growth rates. A higher growth rate, it is widely accepted, is the best way to lower poverty over the medium term. Loayza et al. (2000a, 2000b) indicate that the most important determinant of savings, across both developed and developing countries, is the level of per capita income and the rate of economic growth. Thus the higher the rate of savings, the higher the economic growth rate and the higher the growth rate, the higher the rate of savings at least at low absolute levels of per capita income. Their results also point to the possibility of incomplete Ricardian equivalence, that is, a given rise in public savings is accompanied by a less than commensurate drop in private savings.

The gap between the real rate of return on savings and the discount rate is critical. Savers who are liquidity constrained may be more sensitive to such differentials compared to those who are not. As financial deepening takes place and fewer consumers remain liquidity constrained, this responsiveness may drop. However, as consumers become less liquidity constrained they might also become less risk averse and opt for investments with higher returns, boosting the savings rate. Thus the impact of the tax structure on savings is of critical importance, and distorting differences in effective tax rates across sectors and assets and tax-induced distortions that create inefficiencies and lower the potential rate of economic growth should be eliminated. This would be an important component of tax reform, the basic tenets of which are well known and briefly summarized below.

As an economy develops, reliance on indirect taxation for revenue should decline. This is because indirect taxes typically have an excess burden
associated with them (Jha, 1998). Furthermore, efficient indirect taxation (one that minimizes excess burden to the representative consumer) can be quite regressive.\(^5\) Indirect taxes can be made redistributive by sacrificing some efficiency, but the extent of this redistribution is limited (Sah, 1983).

If, however, indirect taxes can be levied on final consumption alone, tax-induced changes in relative prices that characterize production taxes such as excise duties could be avoided. Then, if consumer utility functions are weakly separable between consumption and leisure, a uniform tax on final consumption goods (say a value-added tax – VAT) would approximate a lump-sum tax.\(^6\) This tax, with only few exemptions (for items consumed in disproportionately large amounts by the poor), harmonized across levels of government in federal countries and few rates, is recommended. These could be supplemented with excise duties on environmental bads or ‘luxuries’. Peak tariff duties and effective rates of protection should be reduced gradually. If the tax base admits few exemptions and there are fewer rates, costs of compliance and monitoring will fall. But the VAT requires the netting out of input costs and the exemption of exports from the tax base. This, in turn, needs sophisticated account keeping which may be absent in many developing countries. The credibility of the tax regime is also important and tax reforms should aim for a stable tax environment and be well coordinated and, at all times, be simple. Tariff cuts should be accompanied by an upward revision of VAT rates to compensate for tax revenue.

However, Emran and Stiglitz (2005) show that the standard prescription of reducing trade taxes with revenue-compensating upward revision of the rate of a broad-based VAT is welfare-improving only in an economy with no informal sector, with all production and exchange activity in the tax net – conditions typically not satisfied in developing countries.\(^7\) When only the formal sector can be taxed, the introduction of a VAT (or a hike in its rate) may end up creating a distortion between the formal and informal sectors. Even broadening the VAT base to include more of the informal sector may reduce welfare (Piggott and Whalley, 2001). Similarly Bibi and Duclos (2007) show that for indirect tax reform to be poverty-reducing it must: (1) not remove all subsidies; (2) in some cases increase taxes on already taxed commodities and, concurrently, increase subsidies on already subsidized commodities; (3) not exclusively follow efficiency considerations, since redistribution may still play an important role in poverty reduction; and (4) concentrate on reform rather than removal of subsidy. Jha (2006) presents a taxonomy of the extant literature’s view on how to fine-tune the aforementioned tax reforms in order to make the resulting tax structures distributionally sensitive.

Tax structures in developing countries are not particularly progressive. Thus Chu et al. (2004) find that: (1) only 13 of the 36 overall tax systems
surveyed by them are progressive, seven are proportional, seven are regressive and the rest neutral or insignificant; (2) income taxes were progressive in 12 of the 14 cases studied whereas indirect taxes were broadly regressive. The progressivity of direct taxes declined over time in eight cases. This needs attention.

Another principle of tax reform is that the share of direct taxation in overall tax revenue should rise. Within direct taxation, reliance has to be shifted from corporate to income taxes. Since corporate profits are taxed at the level of personal income anyway, the rationale for separate corporate taxes is rather weak. There are only two arguments in favor of corporate taxes: (1) as a tax on foreigners’ incomes; and (2) as a tax on non-competitive profits. Within the sphere of income taxation, the rate and exemptions structures need to be rationalized. The number of tax brackets should be small, the degree of progression mild with the top marginal tax rate low. Tax reform theory advocates taxation of ‘full income’ the Haig–Simons definition of which is ‘all increases in human and physical capital during a period of time’. One cannot pick and choose the types of income one would like to tax.

Another area of importance for taxation is the conduct of commerce over the internet (e-commerce). Although e-commerce is a nascent industry it should be taxed since it would be inefficient as well as inequitable to tax goods traded through bricks and mortar stores and not tax e-commerce. A commodity that is sold in a bricks and mortar store and, therefore, subject to taxation would be deemed to be different if sold through e-commerce, and escape taxation. Further, those buying through e-commerce are likely to be rich. This exacerbates inequity. There is a rationale for zero customs duties on e-commerce in line with arguments for free trade, but not for zero taxes. A policy of not taxing e-commerce would provide another avenue for tax evasion as some US evidence shows. Further, given its projected phenomenal rate of growth, if e-commerce is not taxed there will be sharp erosion of the tax bases of governments that primarily levy sales taxes.

Another issue is the presence of tax havens. The Organisation for Economic Co-operation and Development (OECD) estimates that during 1985–94 foreign direct investment (FDI) by the G7 countries in some tax havens in the Caribbean and South Pacific increased more than fivefold to more than US$200 billion – an increase well in excess of the growth of total outbound FDI. These concerns extend to transition and developing economies and have probably worsened in recent years. ‘A race to the bottom’ may ensue with national and/or state governments using tax incentives competitively to attract FDI. Such incentives interact dynamically with the existing avenues for tax evasion (for example because some
incomes are not taxed) to reduce current tax revenues and prospects for higher future tax revenues. In the face of this tax reform, particularly direct tax reform, should have a considerable element of international cooperation.

A related issue is service taxation. Services have become the dominant sector in many developing countries but are hard to tax. Not taxing services is inefficient as well as inequitable: inequitable because it discriminates between providers of goods and services; inefficient because it has the potential of creating several distortions, thus increasing non-labour costs.

Expenditure reform
Tax reforms should be complemented with appropriate adjustment of government expenditures. Typically this calls for reduction of current subsidies and augmentation of subsidies for well-managed capital projects. The impact of public expenditure is usually ascertained through an ex post incidence analysis but we must evaluate not what does exist but what might exist – the theme of benefit incidence analysis. Such analysis is marginal (to capture differences from the status quo) and behavioural (to generate counterfactuals) and is difficult to conduct in many developing countries.

Delineating expenditure adjustments according to their effects on the poor cannot await the development of ex ante analysis. A good rule of thumb is to delay or reduce cuts in public expenditure on goods and services that are directly or indirectly of high importance in the poor’s budget, for example, coarser types of food, fuel and agricultural subsidies.

Within the broad category of basic services the selection of programmes needs to be sensitive to the type and severity of deprivation. If malnutrition is widespread, a programme of subsidized nutritional supplements would be more effective than an elementary education scheme. Rudra (2004) establishes that only the education component of public expenditure reduces income inequality in the face of globalization. Thus, when high inequality is a concern, expenditure on education should not be cut (van de Walle and Nead, 1995).

Conclusions
The role of fiscal policy in developing countries is as important as it is complex. Developing countries face the unenviable task of accelerating economic growth to reduce poverty in a short span of time even as they face greater uncertainty, in the face of globalization, about key elements of their fiscal policy such as the tax base. Furthermore, the exercise of fiscal policy is often circumscribed by increasing pressures from regulatory and exchange rate regimes in place, and subject to considerable pressure from external parameters such as competing countries’ tax rates; for example, it
would be difficult for a given developing country to have corporate tax rates very different from its competitors or to burden monetary policy with high fiscal deficits which could lead to sharp depreciation of the exchange rate.

This chapter has outlined some of the major challenges that developing countries face in some key areas of fiscal policy, particularly tax and expenditure. Even here the treatment has been selective (for example, there has been little discussion of corporate taxation and indirect tax harmonization) to provide an overview of the issues involved and an introduction to the literature on these topics.

Notes
1. I am grateful to Amitava Krishna Dutt and Jaime Ros for helpful comments on an earlier draft of this chapter. The usual caveat applies.
2. The rationale for the existence of multi-tiered governments owes much to the classic statement by Oates (1972), and has been extensively reviewed (for example, Jha, 1998). Intergovernmental fiscal relations are surveyed, among others, by Fjeldstad (2001) and Bird and Smart (2002). Fiscal federalism is not considered in this chapter.
3. This may also lead to an exacerbation of inflation as de Haan and Zelhorst (1990), Easterly and Schmidt-Hebbel (1993) and Buffie (1999) show.
4. They find these countries to be Malaysia, Hungary, Ecuador, Morocco, Panama, Philippines, Indonesia, Bulgaria, Côte d’Ivoire, Egypt, Israel, Jordan, Lebanon, Nigeria and Pakistan. Clearly both transition and developing economies belong to this group.
5. Efficient indirect taxation calls for tax rates to vary inversely with compensated elasticity of demand making them regressive.
6. Separability of the utility function between goods and leisure would indicate that taxation of goods would have no implications for the labour–leisure choice.
7. CSO (2000) notes that in 1999–2000 as much as 60 per cent of India’s GDP came from the unorganized sector and this sector employed 92 per cent of the labour force.

References
Introduction

Economic development strategy in open industrializing economies is dominated by stabilization policy and structural adjustment. Developing countries are faced by an exogenous and changing set of world prices and export demand schedules on the one hand, and rationed global credit markets whose dynamic is determined by financial cycles in the core economies on the other. In consequence, domestic demand stabilization in response to unexpected temporary shocks and supply adjustment to permanent shifts in global markets determine the growth path, rather than a process of intertemporal optimization in the stable and foreseeable world of textbook economic theory.

Global economic shocks are exacerbated by armed conflicts and natural disasters, which often affect entire regions, while accumulated debt positions affect not only fiscal and current account balances but also the future expectations (and thus current behaviour) of the private sector. The political economy issues arising from accompanying changes in employment levels, wage rates and sectoral output are complicated by the key role played by international institutions – particularly the International Monetary Fund (IMF) and the World Bank – both as providers of financial resources to governments and as arbiters of ‘sound’ economic management.

These orthodox macroeconomic management criteria are based on specific models of economic behaviour for small open economies as well as a bias against active intervention in markets. In consequence the disappointing record of orthodox stabilization and adjustment policies in Eastern Europe, Latin America and Asia in the 1990s has led to a renewed interest in ‘heterodox’ methods of macroeconomic management that correspond more closely to the structure and behaviour of the emerging market economies with domestic capital markets integrated to the international financial system. However, an alternative set of underlying macroeconomic models with the generality and power of the orthodox models has yet to be constructed.

This chapter opens with a brief survey of the analytical literature that reveals an excessive emphasis on inflation targeting in stabilization policy.
and a lack of attention to issues of investment and distribution in structural adjustment. It then outlines an alternative new-Keynesian approach to stabilization policy that explicitly contrasts with the standard IMF monetary programming framework. When capacity utilization and open capital account are included, inflation targeting is shown to have destabilizing consequences, requiring a return to active fiscal and monetary policy. The chapter then presents a new-Keynesian view of structural adjustment in explicit contrast to the standard ‘1-2-3’ open economy model that underpins World Bank analysis. It is shown that resource reallocation in the medium term can only take place through new investment, with significant distributional effects via employment and real wages. The chapter concludes by suggesting that emerging market authorities should and can engage in active macroeconomic intervention based on a more realistic analysis of the structure and behaviour of their economies.

Stabilization policy and structural adjustment: the analytical debates
I start by examining the macroeconomic programming models used by the Fund and the Bank in designing stabilization policy and structural adjustment, respectively.2 These two models are not entirely consistent as they are built on different assumptions as well as referring to different time horizons, despite earlier attempts to reconcile them (Khan et al., 1990), and thus must be treated separately.

The IMF Basic Financial Programming Framework (BFPF) is the standard model used by the Fund in designing stabilization programmes, the object of which is to reduce inflation to as near zero as possible and ensure debt service payments (IMF, 1987; Mussa and Savastano, 1999). The intellectual origins of the BFPF are Hicks’s interpretation of Keynes, expressed as the ‘absorption approach’ where excess domestic demand creates current account deficits and/or domestic inflation when imports are constrained by lack of external finance (Polak, 1957). The model subsequently metamorphosed into a version of the Chicago ‘monetary approach to the balance of payments’. It is concerned with the short run, where both real output and exports are taken as given, so the focus is on aggregate demand management. The budgetary balance plays a key role in this (particularly when financed by money supply) as does the nominal exchange rate because it sets import purchasing power.

Apart from the usual set of national accounting identities linking the macroeconomic and monetary variables, there are only two behavioural relationships in the BFPF: the demand for money and the demand for imports. A budget deficit beyond that warranted by output growth and the inflation target (that is, ‘seignorage’) is then reflected fully in the current account deficit if the exchange rate is fixed and foreign finance is available;
or fully in domestic inflation if the exchange rate is flexible and foreign finance constrained. Stabilization policy design then consists in the required fiscal adjustment\(^3\) in order to bring inflation down to target and restore external reserve levels to a prudent proportion of imports. This is supported by last-resort short-term lending from the Fund, which alleviates the fiscal adjustment required to meet these targets; and also provides powerful external leverage (‘conditionality’) to ensure compliance with this form of stabilization policy (Collier and Gunning, 1999). However, the behavioural relationships are clearly far too simplistic in the BFPF: for instance, interest rates and wealth (especially debt) stocks do not enter the model, while capital flows are exogenous, despite the central place of both in modern monetary theory.

The World Bank Revised Minimum Standard Model (RMSM) is used in designing structural adjustment programmes, the object of which is to restore current account stability and raise output growth (Addison, 1989 [1999]). The RMSM has Keynesian roots too: in the Harrod–Domar theory of growth constrained by savings, modified to include an external constraint reflecting the dependence of developing countries on imports of producer goods, leading to the ‘two-gap’ model (Chenery and Strout, 1966). However, it has since involved into the more neoclassical framework of a ‘computable general equilibrium’ model discussed below. The RMSM is concerned with the medium term, so aggregate supply is endogenous. Exports respond to the real exchange rate, which acts so as to allocate production factors between the traded and non-traded sectors: in other words, a relative price effect on supply instead of the income effect of the nominal exchange rate on demand in the Fund model. Investment is simply driven by the availability of savings: private saving (a constant proportion of disposable income) less the budget deficit plus external finance (‘foreign saving’).

In addition to the usual national accounting identities, the RMSM contains five behavioural relationships for the investment–growth linkage, import demand and export supply, fiscal income and private saving. Structural adjustment design seeks to relax the current account constraint on growth by raising exports through real exchange rate devaluation; and to raise the growth rate itself by reducing government expenditure and thus reversing the ‘crowding-out’ of private investment. Regulatory reforms follow the same logic, emphasizing trade and financial liberalization combined with extensive privatization to reduce the size and scope of the public sector. External finance in the RMSM plays three roles therefore: directly increasing public investment (for example in infrastructure) and output growth; reducing domestic borrowing to fund the budget deficit, and thus allowing private investment to rise; and funding more imports and thus
output. As in the case of the Fund, the Bank’s role as a leading provider of long-term official loans to poor countries, and its influence on other aid donors, ensures the adoption of this ‘sound’ approach to structural adjustment (Mosley et al., 1995).

As the RMSM is a one-sector model it is useful for macroeconomic programming, but not entirely appropriate for the analysis of structural adjustment, so World Bank policy design also has analytical foundations derived from the ‘dependent’ economy’ model set out by Dornbusch (1986) and Buiter (1988). This is disaggregated to generate what has now become the ‘industry standard’ with three products – exportables, importables, and non-tradable or ‘home’ goods and services – which we use as a framework in the fourth section below. This ‘1-2-3’ model has generated a wide range of applied computable general equilibrium models (Devarajan and Robinson, 1993) used by the World Bank to inform structural adjustment programmes and to link macroeconomic policy to poverty reduction strategies (Bourignon and Morrison, 1992).

A number of significant lessons are drawn from this simple yet powerful model. One of these is the well-known ‘Dutch disease’ interpretation of the effect of an unexpected increase in world primary commodity prices or a rise in external aid flows: the real exchange rate appreciates, the non-traded sector expands, other traded sectors contract, so imports rise and exports fall, which is unsustainable in the long run. Another lesson is the effect of fiscal expansion: as government expenditure is intensive in non-traded goods and services, the real exchange rate appreciates and traded export production falls as non-traded output rises, leading to unsustainable debt problems.

However, neither the RMSM nor the 1-2-3 model are dynamic and thus do not allow for intertemporal optimization by economic agents: that is, the fact that households, firms and governments take investment, saving and borrowing decisions looking forward over many years. This is the basis of modern neoclassical macroeconomics and allows resource allocation behaviour to be endogenized. Further, they fail to reflect the elements of modern growth theory in general and the role of public expenditure in physical and human capital formation in particular. Last but not least, the simplistic view of the negative effects of budget deficits (on inflation for the Fund and on private investment for the Bank) ignores the modern macroeconomic theory of intertemporal budgetary and financial policy. Indeed from a strictly neoclassical viewpoint this persistence of the ‘financing gap’ tradition can be seen as invalidating the proposals from the Bank and the Fund on additional lending and debt forgiveness (Easterly, 1999).

These orthodox models have also been subjected to a much broader critique from non-neoclassical standpoints. Four such lines of argument are:
(1) the Keynesian critique of the failure to understand the exogenous nature of cycles in open economies; (2) the structuralist critique of the neglect of supply constraints in developing countries; (3) the Kaleckian critique of implausible assumptions on investment and savings; and (4) the Fabian critique of the exclusion of poverty reduction from macroeconomic strategy. These critical theoretical views have been fuelled by the evident failure in most cases of stabilization policy to get beyond inflation reduction, and of structural adjustment to achieve sustained growth (Williamson, 1997).

The underlying assumption in the IMF model that output is unaffected by demand and indeed that the economy operates at full factor employment is clearly implausible. Excess capacity in the Keynesian sense is often present in practice, as well as chronic underemployment in the Lewis sense. Moreover, the central issue in monetary policy for most developing countries today is not inflation as such but rather countering the effects of externally generated cycles exacerbated by inherited debt positions (Ocampo, 2000). In the upswing of a cycle the interest rate declines and the exchange rate appreciates, but any attempt to counter the boom attracts still more funds and the exchange rate appreciates still further. In the downswing, markets push for devaluation but this forces up interest rates and exacerbates production declines, promoting further capital flight and debt default. Moreover, the budgetary dependence on foreign borrowing makes the fiscal stance automatically procyclical. The application of the standard IMF policy model during these financial crises worsens economic recessions and further destabilizes capital flows (Stiglitz and Greenwald, 2003).

A central feature of the standard theory of structural adjustment is that any imbalance between traded and non-traded sectors in the dependent open economy is a result of distorted domestic relative prices (that is, differing from ‘world’ prices): thus the emphasis on real exchange rate correction and trade liberalization. However, this analysis rests on the twin assumptions of full employment of labour and capital on the one hand and the perfect substitution of existing factors between sectors in response to relative prices on the other; which is clearly unrealistic. Indeed, excess capacity and immobile factors explain much of the lack of supply response to structural adjustment (Taylor, 1993). Moreover, the assumption that domestic prices are not affected by exchange rates (and thus that devaluation is not ‘passed through’ into inflation) is similarly implausible for small open economies (Taylor, 1988). These supply response failures are exacerbated by the lack of business liquidity caused by restrictive monetary policies, because even under normal circumstances credit rationing prevails and output as well as prices are affected by interest rates (Blinder, 1987).

More generally, there is a clear parallel between orthodox structural adjustment theory and neoclassical trade theory because the internalization
of world prices is intended to bring about an intersectoral resource reallocation in line with comparative advantage. Trade liberalization raises the return to the abundant factor of production (assumed to be unskilled labour in developing countries); and because primary exports are taken to be more labour-intensive than tariff-protected industry there should also be a net employment creation (Obstfeld and Rogoff, 1997). However this does not often occur in practice: either because exports are based on natural resources in which case rents rise and little unskilled employment is generated; or because skilled labour is the scarce resource and export expansion opens up wage differentials (Wood, 1994). Moreover, there is in consequence no theoretical reason to believe that income distribution will necessarily improve with structural adjustment.

Central to the theoretical approach of both Bank and Fund is that private saving is a fixed proportion of disposable private income, and that private investment (and thus growth) is determined by private savings less the budget deficit plus external finance. However the large fluctuations observed in the savings rate for developing countries and the empirical evidence of the influence on private investment of other factors such as profit rates, credit conditions, public infrastructure, debt overhang, regulatory change and political stability all suggest that in developing countries at least investment is not constrained by private saving (FitzGerald, 2003). Indeed, the policy uncertainty caused by violent and unpredictable stabilization and adjustment episodes is among the most depressive influences on investment in developing countries (Rodrick, 1991). Nonetheless, successful structural adjustment and sustained growth require high rates of investment so that production capacity can change and thus the desired structural adjustment takes place.

This process cannot simply be considered as an overall proportionate expansion based on a fixed savings rate (augmented as necessary by external funds) once domestic resources have been reallocated, as the RMSM does; nor as a smooth process of reallocation of labour and capital between sectors in response to changing relative prices as the 1-2-3 model does. Both modern intertemporal macroeconomics and traditional Keynesian theory tell us that the investment process has its own dynamic based on future profitability, and this has profound implications for adjustment policy. Further, the financing gap theory used in both Bretton Woods models assumes that extra external finance always contributes to growth, by simply and directly adding to investment funds: but it is well established that capital inflows often lead to increased consumption (Jansen and Vos, 1997).

Last, but far from least, the neglect of distributional considerations in both the Bank and Fund models is not only inconsistent with their institutional commitment to poverty reduction but also leads them to
underestimate the political economy constraints on macroeconomic policy. There exists a long-standing critique of adjustment policy in this respect in terms of the negative effect on social service provision of fiscal expenditure cuts as the central macroeconomic policy tool (Cornia et al., 1987). Targeted poverty reduction programmes, while desirable in themselves, do not redress the effects of macroeconomic policy design on employment and wages, which are more significant in determining the welfare of the majority of the population. These effects in turn determine social support for economic policy, and thus its political sustainability.

**Stabilization policy, inflation targeting and monetary autonomy**

We have seen that the open developing macro-economy works in a different way from that which the Fund model supposes—the role of domestic credit rationing and external capital flows being crucial in the short-run context. This section sets out, therefore, a model with a formal framework similar to that in IMF (1987) except that: (1) output can be below capacity and is determined by the level and costs of credit; and (2) the interest rate and exchange rate are related though arbitrage across the capital account.

The standard inflation-targeting model can be set out as follows (IMF, 1987). As this is a short-run model exports ($X$) and real output ($Q$) are exogenous, as are the capital flows, net of debt service, that determine the net change in external liabilities ($\hat{F}$). The level of domestic debt ($D$) and foreign exchange reserves ($R$) are set according to fixed prudential rules. The endogenous variables are thus domestic aggregate income ($Y$), the level of imports ($M$) and the demand for money ($B$) and for credit ($H$) from the private sector. The nominal exchange rate ($E$) floats under the current Fund doctrine, and thus is also endogenous. The target variable is the price level ($P$) and the policy instrument is the interest rate ($i$).

I start with three accounting identities. Nominal income ($Y$) and inflation ($p$) are:

\[
Y = Q \cdot P \\
p = \frac{\hat{P}}{P} \quad (56.1)
\]

The balance of payments (denominated in foreign currency) is the familiar:

\[
X - M = \hat{R} - \hat{F} \quad (56.2)
\]

and the domestic monetary balance (Khan et al., 1990, p. 158) is:

\[
B = D + H + E \cdot R \quad (56.3)
\]
There are three behavioural equations in this standard model, each of which reflect a key aspect of aggregate private sector macroeconomic behaviour. Import demand in nominal terms is a proportion \( m \) of aggregate income \( Y \):\(^{14}\)

\[
M \cdot E = mY
\] (56.4)

Deposits in the banking system (that is, ‘demand for money’) depend on income \( Y \) and the interest rate \( i \) for a given velocity of circulation \( v \) and positive interest ‘elasticity’ coefficient \( \alpha \):

\[
B = Yv + \alpha i
\] (56.5)

The credit (and cash) requirements of the private sector – that is, the supply of money – have a similar form because the Fund model assumes that the authorities always accommodate the monetary needs of the market (that is, passive rather than active monetary stance) and that the impact \( \beta \) of the interest rate on this demand is of course negative:

\[
H = Yu - \beta i
\] (56.6)

The ‘prudential rules’ for domestic debt \( D \) and reserves \( R \) are:

\[
\dot{D} = \lambda Y \\
R = \theta M
\] (56.7)

The reserves rule \( \theta \) is based on a specific degree of ‘liquidity’ in the form of import coverage,\(^{15} \) while the domestic debt rule \( \lambda \) effectively constrains the fiscal deficit as a proportion of GDP.\(^{16} \)

This model is simple to solve because it can be distilled down to two reduced-form equations based on (56.2) and (56.3). The domestic price level \( P \) is determined from the domestic monetary balance by substituting (56.5), (56.6) and (56.7) into (56.3) using (56.1) and (56.4) to yield:

\[
P = \frac{D_{-1} - (\alpha + \beta)i}{Q(v - u - \lambda - m\theta)}
\] (56.8)

from which it is clear not only that higher interest rates \( i \) reduce the price level and thus inflation, but also that a key determinant of inflationary pressure is the domestic debt overhang \( D_{-1} \), as indeed is the prudent fiscal deficit \( \lambda \), thus the emphasis on fiscal retrenchment in Fund stabilization programmes.
An alternative formulation of the reserves rule that has recently found favour in the Fund with the spread of full currency convertibility is that there should be maintained a constant proportion ($\kappa$) of the money supply ($H$). This gives a similar result:

$$ER = \kappa H$$  \hspace{1cm} (56.7a)

$$P = \frac{D_{-1} - \{\alpha + \beta(1 + \kappa)\}i}{Q\{\nu - u(1 + \kappa) - \lambda\}}$$  \hspace{1cm} (56.8a)

The external foreign currency balance found by substituting (56.1), (56.4) and (56.7) into (56.2) and using (56.8) serves to determine the nominal exchange rate ($E$) for a given domestic price level under the floating exchange rate regime espoused by the Fund:

$$E = \frac{mQP(1 + \theta)}{X + (\theta M_{-1} + \hat{F})} = \frac{\{D_{-1} - (\alpha + \beta)i\}}{(v - u - \lambda - m\theta)} \frac{m(1 + \theta)}{\theta M_{-1} + \hat{F}}$$  \hspace{1cm} (56.9)

This in turn implies that the real exchange rate ($e$) using world prices as numeraire$^{17}$ – and thus export competitiveness in the medium term – is endogenous and appreciates (that is, $e$ falls) with positive external shocks such as capital inflows ($\hat{F}$) or increased commodity export income ($X$) because:

$$e \approx \frac{E}{\hat{F}} = \frac{mQP(1 + \theta)}{X + (\theta M_{-1} + \hat{F})}$$  \hspace{1cm} (56.10)

However, note also that although from (56.9) inflation targeting will affect the nominal exchange rate (higher interest rates leading to appreciation) the real exchange rate in (56.10) remains unaltered. As a whole, therefore, the policy stance applied by the Fund model is procyclical because not only is the impact of external shocks on the economy unmitigated by active domestic fiscal or monetary policy, but also any unexpected decrease in output ($Q$) must be met by higher interest rates to keep $P$ (in 56.8) stable and vice versa in the inflation targeting approach.

I now adapt the model to allow for the two characteristics noted at the beginning of this section, which are essential in order adequately to describe middle-income ‘emerging market’ economies and the larger low-income countries with a domestic capital market open to foreign investment. I retain the same basic modelling framework in order to facilitate comparison between our new-Keynesian approach and the orthodox model.

The first modification is to the private credit channel. The relationship in (56.11) appears superficially similar in form to (56.6) but in fact causality
has been reversed: in a credit-rationed economy monetary policy determines the level of real output \( (Q) \) as this responds to credit supply \( (H) \) within the limit of capacity \( (AK) \). This response comes about from both consumer credit expansion affecting demand and working capital availability affecting supply, and is a more realistic representation of emerging market economies than the Fund model. The authorities can alter the supply of money by straightforward monetary emission as an alternative to debt issue for fiscal deficit finance, by varying reserve requirements on banks or by changing the mode of financing foreign exchange reserve holdings. Note that this relationship means that raising interest rates \( (i) \) will reduce output. Only with full capacity utilization is the demand effect felt on prices \( (P) \) rather than output, and only then will reductions in money supply or higher interest rates reduce inflation. In contrast, my bank deposit function \( (B) \) function is similar to that of the Fund model:

\[
\begin{align*}
Y &= H(\phi - \beta i) \\
Y &= QP \\
Q &= \frac{H}{P}(\phi - \beta i) \text{ when } Q < AK \\
P &= \frac{H}{K}(\phi - \beta i) \text{ when } Q = AK \\
B &= Y(v + \alpha i)
\end{align*}
\]  

(56.11)

The second modification is to open up the capital account of the balance of payments in (56.2) by expressing capital flows – changes in external liabilities \( (F) \) – as a function of domestic interest rates \( (i) \) and changes in the nominal exchange rate \( (E) \). This contrasts with the Fund model where capital flows (limited to aid and foreign direct investment, FDI) are entirely exogenous:

\[
\dot{F} = \gamma \left( i - \frac{E}{\dot{E}} \right)
\]  

(56.12)

Note also that foreign investor risk appetite or world interest rates (both reflected in the parameter \( \gamma \)) can shift suddenly in practice, and that a sufficient imbalance between interest rates and exchange rate changes can lead to a capital outflow \( (\dot{F} < 0) \).

Reflecting observed practice in emerging market economies, the ‘prudential rules’ for the management of external reserves and domestic debt in our model need to be adapted to these structural characteristics. Thus instead of the import coverage rule in (56.7), the central bank maintains a reserve level adapted to the external debt position \( (F) \) as a form of insurance against external capital account shocks.

\[
B = Y(v + \alpha i)
\]  

(56.11)
And the debt solvency rule in (56.7) is applied in an intertemporal context, changing with the primary fiscal deficit ($Z$) and limited to a given ratio ($\lambda$) of output at full capacity:

$$D = Z + (1 + i)D_{-1}$$

$$\frac{D}{PK} \leq \lambda$$  (56.14)

Finally, although the emerging market economy is exposed to external shocks even if the nominal exchange rate is allowed to float, the real exchange rate in our model is no longer determinate, unlike the Fund model (56.10), because it is affected by the domestic interest rate through the capital account. This endows the monetary authorities with a degree of freedom both to ensure that exports remain competitive and to respond adequately to external shocks. The primary deficit ($Z$) thus becomes the policy instrument affecting inflation, while interest rates set the real exchange rate and credit levels determine capacity utilization.

We can see how this policy framework works in practice by condensing the model into three reduced-form equations. The first is that real credit supply ($H/P$) be set so as to ensure full capacity utilization. From (56.11) we have this condition as:

$$\frac{Q}{K} = \frac{H}{P} = \frac{K}{\phi - \beta i}$$  (56.15)

The second is the balance of payments identity (56.2), re-expressed in terms of the exchange rate, interest rates and output by substituting in equations (56.1), (56.4), (56.12) and (56.13). Assuming that the target of stabilizing the real exchange rate is in fact achieved (that is, $\pi = 0$ and thus $p = E/E$), inserting expressions for the real exchange rate (56.10) and inflation (56.1) yields the real exchange rate ($e$):

$$X - \frac{mpK}{E} = (\pi - 1)\gamma \left(1 - \frac{E}{E}\right)$$

$$e = \frac{E}{P} = \frac{mK}{X + \gamma(1 - \pi)(i - p)}$$  (56.16)

In other words, the policy instrument that determines the real exchange rate in an emerging market economy with an open capital account is the real interest rate ($i - p$). The higher the real interest rate, the more the real
exchange rate will appreciate (that is, $e$ falls). The desirable policy target is thus clearly to maintain the real exchange rate at a stable and competitive level, keeping real interest rates low and adjusting them actively according to world market conditions.

The third reduced-form equation is derived from the monetary balance (56.3): substituting in (56.5), (56.13), (56.14) and (56.15) yields an equation for the price level ($P$) in terms of the other targets and instruments:

$$P = \frac{Z + (1 + i)D_{-1}}{K\left[(v + \alpha i) - \frac{1}{\phi - Bi}\right] - e\pi\{F_{-1} + \gamma(i - p)\}} \quad (56.17)$$

Because the interest rate ($i$) instrument is already employed to stabilize the real exchange rate ($e$), and the credit level ($H$) is already set so as to stabilize output ($Q = K$), the budget deficit ($Z$) plays the key role of price stabilization. However from (56.17) it is clear that a low level of inflation does not mean that that the budget should always be in balance ($Z = 0$) or even a fixed proportion of aggregate income ($Z = \lambda Y$), but rather should compensate for exogenous shocks to international asset demand ($\gamma$) or production capacity ($K$) even when international price fluctuations are smoothed by the real exchange rate.

In sum, in this type of open economy integrated – albeit asymmetrically – to international capital markets, a decision by the central bank to raise the interest rate in order to curb inflation from (56.17) will actually have three undesirable effects: inflation will rise due to the effect of debt service on the budget deficit; the exchange rate will appreciate through the capital account effect; and real output will fall from the credit channel effect. My more realistic new-Keynesian model for the emerging market economy thus shows not only that this is a misguided stabilization policy, but also that a wider range of policy instruments should be used to achieve multiple stabilization targets, with particular emphasis on trade competitiveness and full employment. These instruments include low real rates of interest, a balanced budget over the cycle and above all strong prudential control of bank credit.

**Structural adjustment, sectoral investment and income distribution**

We have also seen that emerging market economies work in different ways from that which the World Bank model supposes – the existence of surplus labour and sector-specific installed capital being crucial. This section sets out, therefore, a medium-term model with a framework similar to the 1-2-3 model used as a formal analytical basis for the RMSM, except that: (1) although production capacity may be fully utilized, this does not involve
the full employment of the labour force; and (2) shifts in production patterns are not possible except through new investment.

In the canonical dependent economy model there are two sectors producing traded \((T)\) and non-traded \((N)\) goods from homogeneous production functions with sector-specific labour \((L)\) that is intersectorally mobile, so that the nominal wage \((W)\) is equalized across sectors. Total labour supply is fixed and fully employed at the equilibrium wage. For the two sectors \((j)\):

\[
Q_j = Q_j(L_j) \quad (56.19)
\]

and the real exchange rate \((e)\) is now defined as the ratio of traded \((P_T)\) to non-traded prices \((P_N)\):

\[
e = \frac{P_T}{P_N} \quad (56.20)
\]

Each sector employs labour up to the point where the marginal product of labour is equal to the single intersectoral wage, which is thus equalized between the two sectors. Defining the real wage \((w)\) in terms of non-traded prices \((P_N)\) we thus get:

\[
w = \frac{W}{P_N}, \quad Q'_T(L_T) = \frac{w}{e}, \quad Q'_N = w \quad (56.21)
\]

It follows that the ratio of the marginal products of labour in the two sectors is equal to the real exchange rate \((e)\) at equilibrium:

\[
\frac{Q'_N(L_N)}{Q'_T(L_T)} = e \quad (56.22)
\]

Demand for labour is the inverse function of the product wage and there is full employment of the total labour force \((\bar{L})\):

\[
L_T(ew) + L_N(w) = \bar{L} \quad (56.23)
\]

The real exchange rate thus drives labour allocation across the two sectors, and hence output. The sectoral supply functions become:

\[
Q_j = Q_j(e/w), \quad Q'_T > 0, \quad Q'_N < 0 \quad (56.24)
\]
And generally an inverse relationship between the real wage \( (w) \) and the real exchange rate \( (e) \) is also implied by this result, because with full employment in (56.23) and exploiting the implicit function rule:\(^2^6\)

\[
w' = - \frac{wL'_N}{L'_T + eL'_N} < 0 \quad (56.25)
\]

Disaggregating traded goods into exportables \((x)\), importables \((m)\) and non-tradable or ‘home’ goods and services \((h)\) generates the 1-2-3 model (Devarajan and Robinson, 1993), which we use as a framework in order to facilitate comparison. The prices for exportables \((P_x)\) and importables \((P_m)\) are determined by world prices multiplied by the nominal exchange rate \((E)\), while home goods prices \((P_h)\) depend upon supply conditions, as the domestic market must clear to ensure full capacity utilization.

However, in addition to the unwarranted assumption of full employment of labour, the notion in the 1-2-3 model that capital can simply be moved between sectors with a constant elasticity of transformation (CET) production function is obviously implausible. In fact the key issue in structural adjustment is investment behaviour because installed capital cannot be shifted between traded and non-traded sectors, while labour cannot easily be substituted for capital when technology is largely imported. The sectoral production functions are thus better seen as separate, limited by the installed capital stock in each.\(^2^7\)

World prices for exports and imports \((P_f, P_m)\) and thus terms of trade \((\tau)\) are exogenous, while the unit import \((m)\) and labour \((l)\) input coefficients are technologically fixed. As before there is a single nominal wage rate \((W)\), but in our model it is set either institutionally or by the reserve price of labour from the household sector, rather than by labour market clearing, and unemployment persists. Sectoral employment \((L_j)\) is thus determined by output and there is excess labour supply, due to the Leontief fixed technical coefficients:

\[
Q_j \leq A_j K_j \\
G_j = Q_j (P_j - W_l - m_j EP_m) \\
L_j = l_j Q_j, \quad \sum_j L_j < \bar{L} \\
\tau = P_f / P_m \quad (56.26)
\]

Traded exportable prices \((P_x)\) are as before, but home goods prices \((P_h)\) are formed by a mark-up \((g)\) on production costs, where the nominal wage as well as the exchange rate plays a central role. I use home goods prices \((P_h)\) as the numeraire in order to define the real exchange rate \((e)\) and the real wage rate \((w)\):
I assume as before that the export sector produces at capacity because world demand is infinitely elastic to supply by our country. The demand for home goods is a proportion \(a\) of the aggregate factor income, but the domestic market clearing condition will not now determine price as in the standard model, but will rather determine output within the capacity constraint. Substituting the income (56.26) and price (56.27) equations into the demand function for home goods gives the output level, therefore:

\[
P_x = EP_f \\
P_h = (1 + g)(Wh + Em_hP_m) \\
e = \frac{EP_f}{P_h} \quad w = \frac{W}{P_h} \quad (56.27)
\]

Note the so-called ‘Dutch disease’ effect of an improvement in the terms of trade \(\tau\) leading to an expansion of the home goods sector as incomes rise.

Taking the home good price \(P_h\) as the numeraire and rearranging (56.27), we now have real sectoral profit levels \(\bar{R}_j\) in terms of the real exchange rate \(e\), the real wage rate \(w\) and sectoral output \(Q_j\):

\[
\bar{G}_x = Q_x \left[ e \left(1 - \frac{m_x}{\tau} \right) - w\right] \\
\bar{G}_h = Q_h \left[ 1 - \frac{em_h}{\tau} - wh \right] \\
\bar{G}_{x}(e) > 0, \quad \bar{G}_h'(e), < 0 \quad (56.29)
\]

The channel through which the real exchange rate and the real wage rate affect profits is now evident and the implications for investment decisions and income distribution can be explored properly – something that is not done in the standard model.
I now examine in detail the investment decision for the case when \( e \) rises (that is, real depreciation) but the obverse is simple to work out; as are the consequences of other shocks such as changes in labour productivity \((f)\). I assume for convenience that in previous periods firms have been able to adjust their capital stocks to the desired level (that is, \( Q_j = A_j \)). From (56.29) real depreciation will raise real profits in the traded export sector and reduce them in the non-traded home goods sector; but any shift in output depends on investment, which is irreversible. The two-period \((0, 1)\) problem for firms in each sector is whether to invest or not. If they do not invest, then capacity (and thus production) falls by the amortization rate \((\delta)\). If they do invest then they must do so at the level that maximizes its present value \((V)\) discounted at the interest rate \((i)\) where the installation cost \((J)\) expressed in home goods prices is an increasing function\(^{29}\) of real investment \((I)\) using imported equipment at the world price \((P_m)\) and the corresponding import coefficient \((m_k)\):

\[
V_{j,0} = \bar{G}_{j,0} + \frac{\bar{G}_{j,1}}{1 + i} - J_j
\]

\[
J_j = I_j (em_k P_m + \frac{1}{2} \sigma I_j)
\]

\[
I_j \geq 0
\]

\[
K_{j,1} = K_{j,0} (1 - \delta) + I_j
\]

The optimal investment level \((\tilde{I})\) is simply found by differentiating \(V\) with respect to \(I\) in order to maximize the present value of the firm:

\[
\frac{dV}{dI} = \frac{1}{1 + i} \frac{dG}{dI} - m_k e^{-\frac{I}{i}} - \sigma I = 0
\]

\[
\tilde{I} = \delta \left\{ \frac{1}{1 + i} \frac{dG}{dK} - e^{-\frac{I}{i}} \right\}
\]

(56.31)

Because the capital stock in the home goods sector \((K_h)\) was adjusted to the previous real exchange rate so as to maximize profits, from (56.29) no investment takes place in the sector (that is, \(I_h = 0\)) and capacity declines by the amortization rate:

\[
K_{h,1} = K_{h,0} (1 - \delta)
\]

(56.32)

In contrast, the export firms do invest as real profits have risen and:
\[
\tilde{I}_x = \frac{1}{\sigma} \left\{ \frac{A_x}{1+\tau} \left[ e \left( 1 - \frac{m_x}{\tau} \right) - w \right] - e^{-\frac{m_x}{\tau}} \right\}
\]

\[
K_{x,1} = K_{x,0}(1 - \delta) + \tilde{I}_x
\]

\[
\dot{I}_x'(e) > 0, \dot{I}_x' (\tau) > 0, \dot{I}_x' (w) < 0, \dot{I}_x' (i) < 0 \quad (56.33)
\]

Note that improved terms of trade or real depreciation raise export sector investment, while higher wages or interest rates reduce it.

Traded output capacity thus rises while non-traded output capacity falls, due to the changes in their respective capital stocks. Aggregate real output \(Y\) only rises if the net output capacity shift is large enough:

\[
Y = eQ_x + Q_h
\]

\[
Y_1 > Y_0 \text{ if } eA_x (K_{x,1} - K_{x,0}) > \delta A_h K_{h,0} \quad (56.34)
\]

Moreover, even if the adjustment is sufficient to cause aggregate real output to rise, there is no necessary reason why net employment should do so as well. The general condition for this to happen is found by substituting the employment functions from (56.27) into (56.33) to yield:

\[
L_1 > L_0 \text{ when } Y_1 > Y_0 \text{ if } l_x > e_l_h \quad (56.35)
\]

This result has three interesting characteristics. First, there is no guarantee that total employment will increase with real devaluation because this depends on the labour intensities \(l\) of each sector as well as the investment outcome: clearly only if the export sector is the more labour-intensive \(l_x > l_h\) is this likely to occur. Second, the ratio of the real wage rate to the real exchange rate \((w/e)\) is clearly critical to the outcome. If real wages fall then employment will rise, although this trade-off is not the result of factor substitution along the constant elasticity of transformation (CET) curve, but rather of investment incentives. Third, an increase in the interest rate \(i\) will reduce the employment gain due to the investment disincentive (56.33): this is the reverse of the factor substitution effect textbook theory would predict. Moreover, for investment to take place, private investors must have confidence in future profits and be provided with sufficient credit and infrastructure.

As we have seen in (56.25), the dependent economy model implies that depreciation of the real exchange rate will reduce the real wage rate, which with full employment implies a deterioration of the overall distribution of income. The new-Keynesian model reveals a more complex relationship. Rearranging (56.27) we have:
which has interesting characteristics: there is again a negative relationship between real wages and the real exchange rate; but the home goods profit mark-up also plays an important part in income distribution, and productivity in the home goods sector (the inverse of the labour input coefficient \( l_h \)) is also a key determinant of real wages. This last point should remind us that while traded investment is the central focus of adjustment policy, improved living standards require a greater supply of wage-goods.

Further, if I define workers’ living standards \( \omega \) as the nominal wage deflated by the cost of living \( P_c \) defined by the mean domestic product prices weighted by home goods consumption propensity \( a \) in (56.28):

\[
\omega = \frac{W}{P_c}
\]

\[
P_c = aP_h + (1 - a)P_x
\]

Substituting the relevant definitions from (56.29) into (56.37) then gives:

\[
\omega = \frac{W}{a + (1 - a)e}
\]

Here the welfare consequence of the fall in the \( w/e \) ratio in (56.33) required to raise investment and permit structural adjustment is clearly revealed as a deterioration in workers’ living standards.

In other words, the targeting of the real exchange rate necessary in order to maintain export competitiveness is in fact an ‘incomes policy’. If employment expands then overall income distribution can improve, but this requires active intervention in order to raise investment rates as we have shown. Monetary policy should be geared to low real interest rates and producer credit provision on the one hand, and an active fiscal stance geared to damp exogenous macroeconomic shocks on the other. This desirable outcome can be reinforced by a system of dividend taxation (designed to stimulate investment) and social spending specifically aimed at raising the living standards of employees’ families.30

Conclusions

In this chapter we have seen how the macroeconomic models conventionally used to analyse stabilization policy and structural adjustment are open to a wide range of criticism from both neoclassical and Keynesian standpoints. Specifically, the existence of excess productive capacity, mark-up pricing by firms, credit rationing by banks and open capital accounts on the
one hand; and the central role of investment in determining sectoral output and the impact of the real exchange rate on both export levels and real wages on the other; mean that the standard IMF and World Bank models can become seriously misleading as the basis for macroeconomic policy.

By including more realistic formulations of credit supply and external capital flows for emerging markets in the standard stabilization model, I have shown not only that inflation targeting using the interest rate can have serious procyclical consequences, but also that a more active monetary policy based on fiscal and credit instruments can make for higher and more stable output solutions – that is, sustainable stabilization. Similarly, by including the pricing and investment behaviour of firms in the standard adjustment model, I have shown that effective structural adjustment will not take place unless active exchange rate and monetary policies are implemented in support of traded production.

Finally, this new-Keynesian approach also differs from that of the Bretton Woods institutions in its attitude to macroeconomic intervention. The Bank and the Fund insist that inflation-reduced public expenditure and balanced budgets are essential in order to promote growth. This passive ‘rules-bound’ approach is central to their lending conditionality, and is built into significant institutional reforms such as central bank independence. In marked contrast, this chapter has shown how emerging market authorities can combine active fiscal and credit management with real exchange rate targeting in order to cope with exogenous shocks and promote longer-term export-led growth in a more purposive stabilization policy. Finally, it has demonstrated that active intervention to maintain both a competitive real exchange rate and a low real interest rate is necessary in order to promote sufficient investment in the traded sector and thus ensure not only export growth but also employment expansion so that real wage constraints do not lead to a worsening income distribution as a consequence of structural adjustment.

Notes
1. This chapter does not address issues such as privatization of public enterprise or government expenditure reform that, while central to structural adjustment and stabilization policy in practice, are covered in other chapters of this Handbook.
2. These two well-known models can be found in the official sources referenced below and are summarized clearly in Khan et al. (1990) and Agénor (2000).
3. Usually government expenditure cuts rather than increases direct tax pressure – which would depress saving, assumed to be a constant proportion of private disposable income in this model – with obvious distributional consequences.
4. In consequence, it might better be described as a ‘three gap model’ (Bacha, 1990).
5. In the sense a small open economy that is a price-taker in world markets, popularized by Dornbusch (1986), rather than the wider Latin American notion of dependencia.
6. This model is lucidly set out in Montiel (2003, Part V).
7. Because there are one country, two sectors and three products in the model.
8. See Corden (1984). In fact ‘British disease’ would be more appropriate. The Netherlands did indeed experience real exchange rate appreciation in the 1980s due to North Sea gas finds, which rendered manufactured exports uncompetitive; but the fiscal resources were reinvested in infrastructure and skills, with subsequent growth based on advanced services exports. The Thatcher administration used the North Sea royalties to reduce UK profit taxes, stimulating consumption and depreciating the real exchange rate; but the consequences are now visible in deteriorating public transport and education.


12. As opposed to the justifiable concern with hyperinflation in the 1980s.

13. Strictly speaking, bank deposits in the Fund’s BFPF model.

14. This implies unitary price and income elasticities of import demand, which are much higher than those found empirically and means that import contraction through reductions in real demand \((Y/P)\) become the main channel for stabilizing the current account of the balance of payments.

15. The conventional rule of thumb is three months’ cover (that is, \(\theta = 0.25\)).

16. As in the ‘Maastricht Criteria’, where \(\lambda = 0.03\).

17. This corresponds to the IMF definition of the ‘effective real exchange rate’ as the nominal rate divided by the ratio of the domestic price level \((P)\) to the weighted mean of the price levels in trading partners (unity in our case). The alternative definition of the real exchange rate in terms of the ratio of traded to non-traded prices is discussed below. On both definitions, see Dornbusch and Helmers (1988) and Montiel (2003).

18. Which in turn is given in the short run but depends on investment in the medium term, a point taken up below.

19. This is a simplified form of the full international demand function for emerging-market assets, which itself can be derived from standard portfolio theory (FitzGerald, 2006).

20. The most cautious position would be to maintain reserves equal to short-term external liabilities, commonly known as the ‘Greenspan rule’. The value of \(\pi\) will then depend on the maturity structure of external liabilities \((F)\).

21. See Missale (1999) for a discussion of optimal debt models in a full intertemporal context, from which this familiar rule is derived.

22. Formally, Walras’ Condition is satisfied because the model has 11 equations and 11 variables (eight endogenous and three targets); while Tinbergen’s criterion is met by having three instruments \((H, i\) and \(Z)\) with which to hit the three targets \((Q, e\) and \(P)\).

23. The model is very well explained in Chapter 2 of Agenor and Montiel (1999).

24. That is, \(PQ(L_j) = W\).

25. Note that this is the formulation used to derive the Harrod–Belassa–Samuelson model of real exchange rate trends in the long run.

26. if \(F(x, y) = \text{constant}\), then \(dx/\gamma y = -F_x/\gamma F_y\).

27. In the standard 1-2-3 model set out above, there is a domestic production function with a constant elasticity of transformation \((\sigma)\)

\[
Q = A[\gamma Q_x^\sigma + (1 - \gamma) Q_y^\sigma]^\frac{1}{\sigma} \quad 0 < \sigma < \infty
\]

where the convention is to adopt a Cobb–Douglas unitary elasticity \((\sigma = 1)\). Below I adopt the more realistic Leontief form \((\sigma = 0)\) to reflect the fact that once installed, capital is entirely immobile.

28. This form is in fact the constant elasticity of substitution (CES) consumption function used in the canonical 1-2-3 model discussed above, but with unitary own-price elasticity and zero (Leontief) substitution between home and export goods.

29. This is a standard representation of the intertemporal optimisation process for the firm, which as part of the ‘AK’ model underpins endogenous growth theory – see Heijdra and van der Ploeg (2000, Chapter 2). In developing countries this can also be seen as reflecting limited local project implementation capacity.
30. This topic lies beyond the scope of this chapter, but see FitzGerald (1993) for a further discussion of such a policy based on dividend taxation and social expenditure; and FitzGerald (2002) for the derivation of an optimal profits tax to fund infrastructure provision.

References
Bourguignon, F. and C. Morrison (1992), Adjustment with Equity in Developing Countries, Paris: OECD.
57 Economic planning in developing economies

*Bill Gibson*¹

**Introduction**

‘Planning’ is a term that generally has fallen into disuse. Todaro defines development planning as ‘the conscious effort of a central organization to influence, direct and in some cases even control changes in the principal economic variables (such as GDP, consumption, investment, savings, etc) of a certain country or region, over the course of time in accordance with a predetermined set of objectives’ (Todaro, 1971, p. 1). Planning connotes, but does not logically imply, command-and-control mechanisms by which authorities issue directives for which compliance becomes a matter of administrative law.

Development planning was attempted in the Soviet Union and Eastern Europe and to a degree in India, Cuba, Egypt and Tanzania and other countries in the immediate post-World War II period. Indeed, it was largely the success of the Soviet Union in raising per capita incomes in the first half of the twentieth century that demonstrated the existence of a practical alternative to market allocation. Soviet performance impressed policy-makers in developing economies who had come to see the market as inadequate to the task of industrialization. In non-communist countries, planning without enforceable command-and-control mechanisms was widespread in the immediate post-World War II period. The United Nations and other sources even withheld development aid unless a plan was in place and as a result, planning ministries became commonplace throughout the developing world. Planning models that demonstrated how foreign aid could be coordinated to achieve maximum impact on growth and development were especially popular. Despite its increasing technical sophistication and theoretical appeal, planning in the post-World War II period led to widespread disillusion and rejection by even formerly ardent supporters. By the end of the 1970s, Chowdhury and Kirkpatrick noted that many economists were talking openly about the failure of planning, and as early as 1965, Waterson had concluded on the basis of a study of 55 country experiences that ‘the majority of countries have failed to realize even modest income and output targets’ (Chowdhury and Kirkpatrick, 1994, p. 2).

Since the 1970s explicit plans in developing countries have largely been abandoned. Many of the problems planning was designed to confront are
still present, of course, and the need for some kinds of planning persists. As a result, planning has re-emerged in a more market-friendly variant, development policy management, with emphasis on the price mechanism, incentives and schemes such as ‘cap-and-trade’, that rely heavily on decentralized implementation.

This chapter critically reviews planning as applied to developing countries. The next section addresses the general question of the relationship of planning to the market. Economy-wide planning models and techniques are discussed in the following section, while the subsequent section turns to microeconomic planning and cost–benefit analysis. A concluding section discusses the uses of surviving planning models in the current context.

**Plan versus market**

Plans may either be economy-wide or partial. Heal reviews the theory underlying the economy-wide planning procedures (Heal, 1973). He notes that much of the early writing on planning, by distinguished economists such as Lange, Lerner, Arrow and Hurwicz, sought to establish that an efficient centrally planned system would employ the same marginal equalities as in the Walrasian system, with the central planning board playing the role of the auctioneer. Plans in which individual preferences are constitutive of the social objective function, therefore, yield the same pattern of resource allocation as would a competitive market. In other words, there is nothing inherently inefficient about planning. This conclusion is today widely accepted inasmuch as planners’ preferences often proxy a social welfare function under the assumption that a freely functioning competitive market mechanism would produce an identical allocation of scarce resources.

In partially planned economies, planning is generally conceived as a response to market failure, including externalities, informational asymmetries and public goods. If market failure is widespread then it follows that central planning can serve as a substitute for the market; if not, then planning can, in principle, resolve allocational issues related to market failure.

In addition to concerns about market inefficiencies, equity was also considered a legitimate objective. The Coase theorem holds that efficiency and equity are separable, but the distinction in the early days of planning was not clearly recognized. Force-draft industrialization had achieved rapid modernization in the Soviet Union, but at the great expense of a debilitated agricultural sector. The First Five Year Plan under Nehru in India in the early 1950s explicitly prioritized reducing unemployment and poverty over maximizing the rate of economic growth. The principle that income could be redistributed without disturbing the price-guided marginal equality of social costs and benefits was ignored.
There is significant disagreement as to the extent to which government can improve outcomes by realigning social and private costs. In standard theory, a properly tuned set of taxes and subsidies could repair markets that failed and public sector institutions could fill in when markets were missing altogether. In developing-country practice, however, public policy often did not improve outcomes and the term ‘government failure’ gained currency to describe counterproductive intervention by states. The necessity of a one-to-one relationship between policy objectives and policy instruments, originally due to Tinbergen, shows how precarious is the entire mission. The collapse of earlier planning initiatives was in part due to a mismatch in this relationship, with goals grossly exceeding the number of instruments, other than command-and-control, available for implementation.

Killick (1976) provides a comprehensive discussion of government failure in development planning. He argues that the plans failed because their creators assumed that politicians see the planning problem ‘essentially as economists do’. The assumption that governments are composed of ‘public spirited, knowledgeable and goal oriented politicians . . . clear and united in their objectives, choosing policies which will achieve the optimal results for the national interests’, is unwarranted. Anticipating much of the subsequent public choice literature, Killick argues that politicians should be seen as rational, self-interested, acting to maximize the short-term probability that they will be re-elected. The implicit assumption of the existence of a ‘benevolent despot’ was at variance with both reality and the ‘liberal-individualist’ tradition of Western civilization.

Modern public choice theory does indeed suggest that planning will be undertaken for the benefit of the planners themselves or their clients, and that command-and-control directives will give rise to rent-seeking behavior and other principal–agent problems that deprive a country of needed resources and talents. In democratically organized societies, a major problem arises when costs of a directive are widely distributed, while benefits accrue to a smaller set of individuals. Significant pressure to change course can develop as a result, with powerful groups lobbying effectively to push the economy away from its socially optimal path.

Planning, *ipso facto*, could never have resolved these deeper issues of policy-making. Problems of coordination, incentives and the trade-off between efficiency and equity are at the root of the problem of underdevelopment and were beyond the reach of technocratic planners and their tools. Planning was abandoned within a broader current of change that involved rethinking the role of the state generally. Much of the planning literature reviewed above is now seen as archaic, and to say that planning is ‘out of fashion’ is an understatement. As demand management, automatic stabilizers, incomes policy and the Phillips curve gave way to rational expectations,
new public choice theory and inflation targeting, planning became caught up in a generalized retreat from *dirigisme* and ascendancy of the market mechanism. Killick, who had so thoroughly excoriated the planning process earlier on, came later to wonder if there had not been ‘a reaction too far’ in moving so decisively away from planning toward the market. We shall return to this issue in the concluding section.

**Economy-wide planning models**

Planning models can be classified in several different categories: aggregate, main sector, multisectoral, regional and project-specific models. Economy-wide models include the first three categories, but not the last two, and may be static or dynamic. They typically reflect the accounting regularities and conventions of national income and product accounts, balance of payments, and income and expenditure balances of the public sector (Taylor, 1979). These can be simulation models or more traditional econometric constructs. The former employ informal calibration procedures, while the latter are estimated formally, using statistical theory under the usual assumptions. The simulation approach does not rely on statistical theory, but rather on whether the model captures salient features of the economy (Gibson, 2003).

**Input–output models**

Economy-wide planning models have their roots in the model first described by the young Harvard economist W. Leontief in the 1930s and 1940s. The inter-industry or input–output approach pioneered by Leontief, and first implemented with his help in the Soviet Union, served a means by which consistent intersectoral plans could be drawn up. Input–output models have their roots in Quesnay’s *Tableau Economique*, a Physiocratic device that was the first effectively to separate real from nominal resources flows.

Input–output models are used to analyze the impact of a change in final demand on the levels of production. Let $A = \{a_{ij}\}$ be the coefficient matrix such that each $a_{ij}$ describes the use of input $i$ for the production of one unit of output $j$, and $X = \{X_j\}$ be a column vector of gross outputs, including intermediate goods. Factors of production, labor, $L$, and capital, $K$, are treated separately, usually with fixed coefficients under the assumption that factor prices remain unchanged. Together with the labor and capital coefficients, the $A$ matrix represents the technique by which goods are produced.

Final demand is denoted by $F = \{F_j\}$, a column vector of outputs, and may be disaggregated into consumption, government spending, exports and imports as needed. The essential equation of input–output analysis, known as the material balance, is then:
In a one-commodity world, say corn, let the output of six ears require the input in the form of seed corn of one ear. In order to consume 500 ears of corn, we must then produce $X = (1/6)X + 500$ or $X = 600$ ears to make sure that there is enough for both final, 500 and intermediate, or seed, of 100. Since we could just as easily produce 1200 ears with 200 ears of seed, the model evidently assumes constant returns to scale.

So-called dual variables can also be defined and interpreted as prices, denoted here by row vector $P = \{p_j\}$. The equation dual to the material balance is then:

$$P = PA + V_A \tag{57.2}$$

where $V_A$ is a row vector of value added, and may be disaggregated into wages, profits, imports, taxes and rents as needed.

As a result of a linear production technology, input–output models are relatively inexpensive and easy to formulate and run. Since prices were often administered and incentives less relevant, the absence of a functioning price mechanism in the model was unimportant. They were, consequently, enormously popular in early development planning.

The model can be made dynamic if investment, $I$, is first disaggregated from final demand, $F$, and then used to determine the time path of capital stock, $K$. This is done by way of the 'stock-flow equation':

$$K_t = K_{t-1}(1 - \delta) + I$$

where $\delta$ is the depreciation rate. Consistent forecasts of intermediate demand, labor and foreign exchange requirements, for example, could then be made, contingent on a forecast for investment.

The framework just presented is the 'open Leontief model’, but a closed version is available in which all elements of final demand and value added are made dependent on $X$. It was left to von Neumann to show that a maximum rate of sustainable growth is well defined by the model. Despite the elegance of the von Neumann model itself, there were limited direct practical implications of the closed Leontief model for planning. One reason was that prices played virtually no role whatsoever; the technical coefficients, whether for capital or feeding labor, determined the entire balanced growth path. ‘Turnpike optimality’, as exhibited by the von Neumann model, was intellectually appealing, but offered little insight into the nature of the far bumpier road on which developing countries were traveling.
Linear programming models
The chief limitation of all linear models is that they do not allow for substitution in response to changing prices of goods and factors. A partial solution to this problem is provided by the linear programming approach. Introduced by Dantzig for the US Air Force in 1947 and popularized in a classic text by Dorfman, Samuelson and Solow, linear programming models allowed for prices to have a limited impact on the allocation of resources (Dorfman et al., 1958). Unlike their more rigid input–output counterpart, linear programming models could be set, for example, to maximize employment by choosing a sectoral pattern of output consistent with a foreign exchange constraint or some other supply-side limitation (Blitzer et al., 1975).

In a typical linear programming model, there is usually more than one feasible solution. The feasible solutions are then ranked according to an explicit objective function that depends on prices of goods and factors, or some other methods of valuation. An optimal primal solution satisfies all constraints and provides a maximum of the objective function. Like input–output models, there is a dual solution which minimizes the value of the dual objective function. A powerful and fundamental duality theorem of linear programming establishes complementary slackness, which holds that if a constraint in the primal solution of a linear program does not bind, that is, it is satisfied only as an inequality, then the corresponding dual variable is zero. In less formal language, an additional unit of a resource that was already in excess supply could have no effect on social welfare.

The impact of complementary slackness on development planning cannot be overestimated, since at once the notion clarified the relationship of a ‘social optimum’, however planners wished to define it, to factor abundance and the related production technology. From linear programming and complementary slackness, practitioners derived the idea of a ‘shadow price’, or the change in the value of the social objective function with respect to a change in the quantity of a specific binding resource. Now the size of the wedge between the social and private cost of resources could be computed. The application was immediate: in economies with surplus labor, the shadow value of unskilled labor was effectively zero and thus planners would be justified in substituting a lower than market wage when computing the social cost of any particular project or policy intervention.

While linear programming models allow for choice of technique, they do not allow for smooth substitution and infinite divisibility between discrete techniques. This may in practice be more realistic but does give rise to jumps in the values of the solution variables. Data permitting, smooth substitution can always be approximated to any degree of accuracy by
increasing the number of available techniques of production. Moreover, since linear programming models are special cases of non-linear programming models, computer software available for the solution of the latter, for example, General Algebraic Modeling System (GAMS), Matlab, Mathematica, and so on, also compute solutions for the former. Specialized packages exist for linear programming problems, such as Lindo, that are fast, efficient and give highly detailed computational results.

That linear programming would show only how one analyzes resource constraints given the objective function, rather than the deeper problem of how social objectives are themselves to be defined, would ultimately lead to its undoing. But for a while, the technique enjoyed immense popularity, and still does in many specialized applications. Moreover, that it could neatly separate the role of the policy-maker, who determined the coefficients in the objective function, from that of the economist or planner, who designed, built and ran the model, only enhanced its scientific patina.

**SAMs and CGE models**

Social accounting matrices (SAMs) extend the usual conceptual categories of input–output frameworks to account for more detailed expenditure and distributional categories. The constructs are not properly referred to as ‘models’, but rather serve as a database to which behavior equations can be calibrated.

Just as linear programming generalized input–output analysis, so computable general equilibrium (CGE) models take the next step in integrating price signals in more fundamental ways (Gunning and Keyzer, 1995). They are usually multisectoral, economy-wide models, calibrated to SAMs. They may be static or dynamic with short-run coverage of one to three years, three to seven for the medium run, and long-term models that extends beyond a decade. Static models compare two points in time without explicit attention to the path connecting these points, while dynamic models trace out a locus of points with explicit stock-flow adjustment processes. The models may exhibit a wide range of adjustment mechanisms, from closed, purely competitive, Walrasian models to macro structuralist models in which foreign exchange availability determines the level of output in some key sectors.

The structure of a typical CGE model can be briefly sketched as follows. Beginning with the material balance, in equation (57.1), the model links the various elements of final demand to goods prices and incomes. Factor demand equations determine factor prices when supply is binding, but this need not be the case and some other mechanism might be introduced to determine nominal factor prices. CGE models (CGEs) can be constructed
in real or nominal terms, but it is a characteristic feature of structuralist CGEs that equations are given for nominal quantities, which are then converted into real terms by the price vector that results from general equilibrium. This implies that money or some other nominal quantity be fixed exogenously and thus, inflation can be modeled in dynamic systems.

There is, of course, no need to specify supply and demand equations in CGEs since the underlying determinants are modeled directly. Production functions combine labor, $L$, and capital, $K$, so that equation (57.1) can be expressed as:

$$X(K,L) = AX + F(P,Y)$$

with final demand written as function of income, $Y$, given by:

$$Y = V_A X.$$  

Since value added depends on factor supplies, equation (57.2) should be re-expressed as:

$$P = PA + V_A(K,L).$$

Unlike the input–output or linear-programming models, both $X$ and $P$ must be solved for simultaneously. Prices appear throughout the model in a more integral way, causing substitution of both goods and factors and determining incomes.

It follows that the elasticities of substitution must be carefully calibrated for each application. Overestimating these elasticities implies a failure to recognize structural rigidities that may be present in the actual economy. Models in which response elasticities are too high underestimate the effect of policies, since the model allows adjustment in both production and consumption to be smoothly and easily accomplished. In the real economy, there may be significant transactions costs associated with substitution, and thus policies may be more effective in the real economy than in the model.

Dynamic CGEs are more cumbersome, and to the extent that they are designed to reflect Walrasian dynamic adjustment mechanisms, with perfect foresight, are less realistic than models which depend on an explicit investment function. The latter can employ parameters that are econometrically estimated to enhance realism. Dynamic CGEs can be calibrated to historical time series in the same way as large econometric models can be, and provide much more detailed and consistent information than typical time-series models.
Environmental planning models
Computable general equilibrium models in theory can be extended to address a range of related policy problems, such as environmental components. So long as stable contaminant coefficients can be found and linked to production and consumption levels, the models can generate an endogenously determined estimate of environmental quality along with its forecasts for production, consumption, investment and international trade. There are several important problems of implementation, however, the first of which is that contaminant levels can vary significantly between two industries that have been aggregated into a larger category, and even within an industry pollution levels can vary between two firms. Moreover, the coefficients presently in use are derived from studies of US manufacturing firms and one can only guess how these coefficients would need to be adjusted to conform to conditions in developing countries.

Environmental policy analysis thus requires considerable sophistication. Without some detailed microeconomic analysis built into the model, it might become difficult or impossible to judge how firms would react to the introduction of tradable emissions permits, that is, pollution rights that can be bought or sold in a specified market. Earlier planning models could adequately capture a command-and-control system that targeted output levels, but would fail to capture more nuanced response to cap-and-trade policies, such as the time-phasing of investment in compliant technologies. Moreover, models that do not include a feedback loop from the toxic contaminants to price or output levels would also fail to capture reality. While environmentally augmented CGEs have been employed in a small subset of developing countries, they are in their infancy.

Growth and long-term planning models
Even if resources are efficiently allocated statically, a sequence of Pareto-optimal states need not be Pareto optimal when viewed as a sequence (Dorfman et al., 1958). Hence markets may function well to allocate resources over space, yet do a poor job over time. This is especially difficult when the allocation problem stretches over generations, some of which are not yet born. Heal has argued recently that markets systematically err in valuing the future. Thus, inadequate capital accumulation due to uninsurable risk, credit rationing, asymmetric information and other imperfections is related to, though not the same thing as, imperfections that block trades between agents who happen to be alive at the same time. In this limited but important regard, the coefficients in the planner’s objective function may be more accurate than market-determined weights.

Growth models have a distinguished history in planning, stretching back to the 1920s when issues of capital accumulation were first addressed in the
Soviet Union. Following Fel’dman’s work in the USSR, Indian statistician P.C. Mahalanobis in 1953 developed a two-sector model that examined the allocation of investment between capital-good-producing and consumption-good-producing sectors and implied that investment needs to be allocated to the former sector to increase the rate of growth of the economy. By implicitly ignoring the agricultural sector (under the assumption that cheap food could be extracted from the agricultural sector through favorable terms of trade), it was a real failure of planning inasmuch as agriculture stagnated and ‘cheap food’ became expensive, sometimes prohibitively so. The models eventually fell out of fashion.

Following the emergence of the one-sector Solow model in the 1950s, gap models, essentially aggregate growth models with both savings and trade constraints, became popular planning tools. Gap models continue to be used to resolve issues of whether faster growth will be self-canceling by stimulating imports to the point that a balance-of-payments crisis develops.

Under very restrictive conditions, dynamic planning models can be used to determine optimal accumulation paths far into the future. One of the most well-known early models in economics, due to Ramsey, employs the calculus of variations to find the optimal savings rate, the one that maximizes the discounted value of future consumption. Despite their technical sophistication, these optimal growth models, like the von Neumann model, never guided real planning exercises in any important or practical way. Similarly, endogenous growth models have been current since the 1990s, but neither have they gained much traction for development planning.

Regional models
Regional models comprise a final subcategory of planning models. Since data requirements are hefty and data availability is sometimes scanty, regional models have lagged in application. The exception was in Eastern Europe, where data were more abundantly available, even if fabricated out of legal necessity. It is clear, however, that in the case of India and China, which together comprise almost half the developing world, regional models are not merely desirable, but unquestionably necessary. Combining regions in China could be as misleading as aggregating North and South America, and therefore aggregate models could grossly distort the true state of economic activity.

Micro-level planning
In the 1970s there was an explicit attempt to integrate micro planning into comprehensive models that were used to check consistency and direct and indirect effects of policies. The two best known were the Organisation for Economic Co-operation and Development (OECD) manual written
by Little and Mirrlees (1974) and the UNIDO Guidelines for Project Evaluation by Dasgupta et al. (1972).

Cost–benefit analysis
Public sector projects for electrification, hydrological development or transportation and communications infrastructure are key components of any development plan. Costs and benefits of projects are optimally evaluated using an hierarchical methodology in which the project is sequentially evaluated at ever higher levels of aggregation. Eventually, of course, the model may not ‘see’ the project, simply because the project is too small to matter at the aggregate level.

The private sector criterion for project acceptance is either that the present discounted value of costs and benefits as they are distributed over time should be positive, or that the internal rate of return of these same costs and benefits exceeds the cost of capital to the firm. Because externalities are so prominent in developing countries, however, the private project selection procedure has long been considered inadequate for use by development planners. While the present value template itself is appropriate, it is social – rather than private – costs and benefits that must be reconciled. Shadow, rather than market, prices are then used to evaluate project costs and benefits.

As discussed above, shadow prices are intended to reflect the marginal social benefit of available resources. Computation of these shadow prices, however, is fraught with controversy due to the large number of assumptions required for their determination. Projects that would utterly fail a private screening can, perhaps, be accepted using one method of computing shadow prices, but not another. Since shadow prices purportedly measure the marginal impact of aggregate welfare, the whole procedure had now to be vetted by the political process. This dulled the technocratic gloss that project evaluation had acquired under the direction of the authors cited above.

As noted, the linear programming approach imputes a shadow value of zero to factors of production in excess supply. Much of the early literature was devoted to calculating shadow prices in specific markets: labor, both skilled and unskilled, foreign exchange, and capital markets. The United Nations Industrial Development Organization (UNIDO) guidelines developed an extensive analysis consistent with optimal accumulation paths in surplus labor economies, all done in an analytically rigorous fashion (Dasgupta et al., 1972). Recently, more elaborate economy-wide simulation models have been used to calculate shadow values, but have not escaped intense methodological criticism since so much depends on the objectivity of the price scheme.
In retrospect, it is hardly surprising that the less analytically demanding scheme of Little–Mirrlees became dominant. It is the approach to shadow pricing most widely accepted today (Little and Mirrlees, 1974). The economy is divided into traded and non-traded goods markets and there is a competitive primary factor market as well. The shadow price of traded goods is simply the border price, since the import border price is the clearest measure of what the country is willing to give up in order to secure an additional unit of a good. Similarly, if foreigners are willing to pay the border price for our exports, that stands as the next-best alternative to any domestic use. It is a straightforward application of the basic principle of opportunity cost and requires no political justification, defense or intervention.

Non-traded goods are still difficult to shadow price. If there happens to be a separate factor of production for every traded good, and input–output relationships are known, it would be possible to solve for the shadow prices of non-traded goods and factors as a function of the known traded goods prices. If the number of factors is greater than the number of tradables, then the indeterminacy must be removed by additional information. If, for example, it is possible to deduce the foregone output of a traded good upon removing a unit of unskilled labor, then we would have a measure of the shadow value of unskilled labor that could be used to reduce the number of unknowns. If the number of factors were less than the number of tradables, the system would be overdetermined and there would exist two shadow prices for the same good.

Eventually shadow prices would be calculated directly from computable general equilibrium models, but this did not fully resolve the problem either. Model structure clearly matters and moreover, shadow prices are sensitive in general equilibrium models to how projects are financed. If a project is offset by an increase in lump-sum taxes, then the effect on aggregate welfare is the simplest to calculate. But since these tax vehicles are not usually available in developing countries, one immediately has to contend with distortionary mechanisms like income or sales taxes, which add another assumption-laden level of complexity to the analysis. Other complications include economies with segmented goods (traded and non-traded) and labor markets (which may also be regulated), large informal sectors, credit rationing, an inadequately developed or captured regulatory apparatus, and the like (Squire, 1996).

Projects that do not represent Pareto improvements, since they may easily imply a loss of welfare to some members of society while others gain, can be accepted. Income distribution need not, however, be taken into account in project appraisal if an appropriate scheme of taxes and subsidies is available to compensate losers. This is a big ‘if’ however, and some
authors have tried to incorporate distributional concerns directly into
the procedures for project evaluation. Government policy-makers may
choose to redistribute income from current to future generations or
within the current generation from one class of households to another. As
Chowdhury and Kirkpatrick note, distributional weights applied to utility
representations of individual households is an explicitly subjective exercise,
which varies across both time and space (Chowdhury and Kirkpatrick,
1994, p. 2). Efficiency calculations are rarely of such magnitude that they
cannot be reversed by small changes in weights in the aggregate welfare
function. For this reason, planners have been reluctant to mix concerns of
equity and efficiency.

Public investment in infrastructure projects including electrification,
telecommunications, transportation and marketing facilities would seem to
address problems of static and dynamic market failure. Oddly, it has been
argued that there in fact has been too much investment in infrastructure.
Project evaluation techniques – even when undertaken by competent eco-
nomists, such as the staff of the World Bank – fail to account properly for
the welfare loss in cost recovery. On the other hand, welfare losses per
dollar of public revenue raised are typically calculated using static com-
putable general equilibrium models and therefore cannot account for the
dynamic market failure of the underproduction of public goods. Getting
prices wrong ultimately means they will not be used for any politically sen-
sitive decision. Planning succumbed in large measure because, in demo-
cratically organized societies, only the market has been able to claim
objectivity in determining shadow values.

Current uses of planning models
Planning and planning models may be out of fashion, but they can still
serve a useful purpose. The most obvious use is that they allow policy-
makers to form quantitative estimates of the various trade-offs in prepar-
ing development policies. They can be used to comb out inconsistencies in
the ways in which policy-makers believe the economy is working. The
models also enhance internal communication, adding clarity to discussions
within the policy establishment as well as between these individuals
and politicians, the public and other interested parties, such as non-
governmental organizations (NGOs). Planning models also serve as a
means of external communication. The models communicate the thinking
about how resources are employed and the explicit assumptions (behavioral
parameters, elasticities and the like) underlying the model can be reviewed
and evaluated by outsiders. Models can signal to donors that contributed
resources will be used wisely and in ways consistent with broad develop-
ment objectives. Finally, planning models with sufficient structural detail
also can be used to counterbalance any undue influence of generic, one-size-fits-all models.

Proper incentives were often ignored in early planning and this was reflected in the models themselves. More recently, CGE models explicitly incorporate the incentive structure. They derive their strength from the comprehensive picture they paint of the economy and can account for the combined effects of numerous simultaneous policies, from labor markets to exchange rates, taxes and transfers. Planners can conduct realistic ‘what if’ experiments, refining their understanding of the various channels by which adjustment processes unfold. Some, although not all, unintended consequences are likely to be anticipated, allowing for corrective policies to be put in place.

Planning, as an institution throughout the developing world, has not entirely disappeared but rather has changed forms in significant ways. Policies often have unintended consequences, most often when they are blind to the implicit incentive structures they erect. Consequently, planning ministries have given way to development policy management offices. The latter explicitly strive to enhance market outcomes. Rather than having to anticipate the various ways in which the private sector may try to evade the planners’ directives, modern theory suggests that a market-driven approach can yield more satisfactory results. Planners set broad overall planning objectives and then encourage the private sector to maximize their own interests subject to these imposed constraints. Decision-making is decentralized and the social cost of compliance is minimized.

This enlightened approach takes much of the conflict out of planning and the negative connotation associated with command-and-control is thereby lessened. As states abandoned coercive methods, fewer trades were blocked, and economic efficiency automatically increased. This is not planning ‘lite’, but rather a different approach that tries to exploit fully the informational content of prices rather than issue legally binding directives.

Note
1. Thanks to Diane Flaherty and the editors of the Handbook for many constructive comments and criticisms.

References


State-owned enterprises and privatization

Anindya Sen

Introduction
The 1960s and 1970s were characterized by the rapid expansion of the public sector in both developed and developing economies. But from the 1970s there was a breakdown of the social-democratic Keynesian consensus of the preceding 30 years about the strong social and economic role of the state. Ideas of market failure came to be replaced by ideas of government failure. The 1980s therefore witnessed widespread attempts by policymakers to curb the state’s economic role. Privatization was a significant element of these attempts. Moreover, the fiscal crises faced by many developing countries in the 1980s considerably curtailed the capacity of the state to invest in state-owned enterprises (SOEs). The consequent attempts at reforms were sometimes part and parcel of structural adjustment programmes that emphasized speedy privatization. In particular, countries facing foreign exchange constraints which had to approach the international financial institutions – the IMF and the World Bank – for assistance were given such assistance subject to certain conditionalities, including privatization of SOEs.

It is interesting to note that the term ‘privatization’ was originally coined by Peter Drucker and replaced the term ‘denationalization’. The first large-scale ‘denationalization’ programme of the post-World War II era was launched by the Adenauer government in the Federal Republic of Germany. In 1961, the German government sold a majority stake in Volkswagen in a public share offering. The offering was heavily weighed in favour of small investors. Four years later another larger offering took place. Both offerings were initially received favourably. However, a later cyclical downturn in share prices forced the government to bail out many small shareholders.

The next major European initiative came 20 years later with the successful British Telecom initial public offering in November 1984. This was followed by increasingly massive share issue privatizations in the late 1980s to early 1990s. As a result, the share of SOEs in British gross domestic product (GDP) reduced from around 10 per cent to almost 0 per cent in the space of 18 years. France, Italy, Germany and Spain followed with their own programmes. Typically these were public share offerings, often launched by socialist governments.
Among Asian countries, Japan has sold only a relatively few SOEs, mostly via share issue privatization. Some of these have been truly enormous. The $40 billion Nippon Telegraph and Telephone offer in November 1987 remains the single largest security offering in history. In China, numerous small privatizations have taken place, but relatively few outright sales of SOEs. The reason may be that most Chinese SOEs are burdened with social welfare responsibilities. It will be difficult to implement a very large privatization programme since that will seriously undermine the state’s economic role.

India has undertaken a ‘disinvestment programme’ since 1991, but the targets have consistently exceeded the actual proceeds realized, and major controversies have erupted over some of the SOEs privatized. In Latin America, many countries have undertaken large-scale privatization – Chile, Mexico and Brazil being the most prominent. In sub-Saharan Africa, privatization can be characterized as a ‘stealth economic policy’ (Megginson and Netter, 2001) since few governments have openly adopted an explicit divestment strategy. But there has been more privatization than is commonly believed. For example, in South Africa the African National Congress came to power on the planks of nationalization and redistribution of wealth, but the Mandela and Mbeki governments have almost totally refrained from nationalization and have even sold off several SOEs.

In Central and Eastern Europe, privatization efforts are part of a broader effort to transform from command to market economies. Two imperatives facing these countries are the lack of financial savings and the infeasibility of foreign purchases of divested assets because of political considerations. These two imperatives compelled the launch of ‘mass privatization’ programmes that involved the distribution of vouchers to the population. The vouchers could be used to bid for shares in companies being privatized. Such mass privatizations resulted in massive reduction in state ownership. Later this technique became very unpopular, especially in Russia where a perception grew that privatization had led to robbery by the old elites and new oligarchs.

It has been estimated that cumulative value of proceeds raised from privatization exceeded $1 trillion in mid-1990. Annual proceeds peaked at $160 billion in 1997.

Definition and techniques
Privatization involves a transfer of ownership and control from the public to the private sector. Privatization can be accomplished in two ways. The government can sell its assets to private buyers. Or the government can stop providing a service directly and rely on the private sector to deliver the
service. Given that a government often does not unload 100 per cent of its holdings in a company, the distinction between public and private entities can become nebulous. The problem is compounded by our inability to identify precisely the point at which control passes to private agents.

The public enterprises relevant here are revenue-generating entities originally owned or controlled by the state. An example of a revenue-generating entity is a municipal corporation that gets property taxes.

There are a number of techniques of privatization (Lopez-Calva, 1998). The most well known among these is the sale of equity to the general public. This is called ‘divestiture’ (divestment or disinvestment), and may be complete or partial. Divestiture can be done through both direct sales and equity offerings. While the developed countries have often utilized equity offerings as a technique for privatization, developing countries have found the process to be difficult for a number of reasons. The inadequacy of national stock markets and the lack of domestic capital in these countries have sometimes led to a shortage of local buyers, while foreign investors, unable to obtain sound information on the enterprises offered, often lacked sufficient interest. Even in developed countries, the direct sales approach may be costly and slow, owing to the complexity of preparing each state asset for sale individually, and then ensuring that buyers observed all contract provisions.

One special type of direct sales is a management–employee buyout. Shares of an enterprise are sold or given to some combination of managers and other employees. Well-structured management–employee buyouts can sometimes lead to efficient results, since the people who know best about an enterprise, that is, the employees and managers, become the owners. It is also rapid and easy to implement. Nevertheless, experience shows that these buyouts suffer serious disadvantages. Yielding to insider interests often entails large costs in inefficiency and poor management. Insiders may also lack many of the skills necessary to function in a market-oriented economy. Further, the process is seen to be inequitable, handing employees, rather than the population at large, most of the benefits.

In a number of former socialist countries, restitution has been employed to privatize SOEs. Restitution refers to the return of state assets to their former private owners in situations where the government’s original acquisition is seen as unjust, such as uncompensated seizure. Restitution, in such cases, it is argued, is essential on moral grounds. Opponents of restitution counter that the process is necessarily selective, and therefore an unsatisfactory way of achieving justice retroactively. Moreover, private claims can often be complicated and drawn out, bogging down privatization unnecessarily. In practice, the transition countries have seldom used restitution, except for Estonia and, to a lesser extent, the Czech Republic.
In such economies, another important technique has been mass privatization. In mass, or equal-access, voucher privatization, the government generally gives away, or sells for a nominal fee, vouchers that can be used to purchase shares in enterprises. This technique has proved to be popular particularly in the Czech Republic. Voucher privatizations can not only help to overcome the shortage of domestic capital, but they are also politically popular because they address the perceived unfairness of other approaches and avoid the charges of a sell-out of national assets to foreigners. The main risk is that a dispersed ownership structure will lack the focus and power to direct effective corporate management. These problems have been partly addressed by pooling ownership interests in investment or mutual funds. The funds, however, do not always have adequate management, control and supervisory powers. In such cases, voucher privatization becomes merely ineffective absentee ownership.

Contracting out or leasing out of government services can be another technique of privatization. For example, a municipal corporation can contract out the task of garbage collection to a private party.

Associated with privatization usually are processes of liberalization and deregulation. Liberalization refers to the introduction or promotion of competition in a traditionally monopolized industry. Deregulation refers to the abolition of statutory barriers to the operation of market forces. For example, the government of India controls the prices of many commodities through the administered pricing mechanism. If some commodity is taken out of the purview of this mechanism, then this is a deregulatory measure, because the price will now be determined by market forces.

**Reasons for the establishment of SOEs**

State owned enterprises (SOEs) were created for a number of reasons. It was believed that nationalization of private sector enterprises and establishment of SOEs would provide governments with access to much-needed revenues. These profits or surpluses could then be channelized to develop the priority sectors of the economy. Implicit in this line of thinking was the assumption that the private sector would not help in the rapid and sustained development of the economy if left to itself. There was a need to control the ‘commanding heights’ of the economy, that is, the strategic industries. If the government controlled these industries, it would be able to steer the economy in the right direction and overcome critical bottlenecks. National security reasons were sometimes added as justifications to the above, particularly in the context of heavy industries.

In many developing countries, lack of private entrepreneurs was also a major concern and forced the state to take an active role in the process of industrialization. Local private entrepreneurs often were in short supply.
Even if they did exist, they might not have access to adequate capital, partly because stock markets were not well developed. In some countries, private entrepreneurs came from unpopular minorities or were linked to foreign powers.

Distributional considerations also played a role. In India, SOEs were set up in backward regions to lessen regional inequalities. SOEs were also used to increase employment generation.

In political terms, SOEs constituted important resources for state elites – politicians and bureaucrats. They could be used to provide jobs to potential voters and service constituencies (for example, a railway minister could order railway officials to provide better links to his or her constituency).

**Why privatize?**

Over time, in many countries, the performance of SOEs turned out to be, by and large, unsatisfactory. They incurred losses, or did not make as much profit as they should have, given that they had privileged access to capital, various subsidies and protection from domestic and foreign competition. The main reason for this failure seems to have been the problems of gathering appropriate information and devising appropriate incentives for the pursuit of public interest. The incentives for serving customer interests and controlling costs were usually weak. Management was given ill-defined objectives and these frequently conflicted with political objectives. ‘The accounting systems were not appropriate to the information needs required for efficient pricing, i.e. setting prices equal to long-run marginal costs and investment projects were often appraised against technical criteria rather than economic hurdle rates of return’ (Jackson and Price, 1994, p. 2) As a result, many governments started considering privatization of SOEs as a solution to these problems because it was felt that markets provide better incentives to participants and use information more efficiently. Privatization would provide greater incentives for cost minimization, encourage more effective managerial supervision and stimulate greater employee effort.

We next examine the possible impacts of privatization in economic terms. In addition to efficiency, distribution and stability factors also need to be taken into consideration.

**Fiscal impact**

When a SOE is sold off to the private sector, the government gets the sales proceeds. Further, if the SOE had been making losses and was being subsidized, then these subsidies come to an end, which further helps the government. Thus the immediate generation of revenues is supplemented by reduction in recurrent expenditures.
But does the government really gain? In the simplest case, the buyer will be willing to pay only so much as the SOE is expected to bring in the future. The discounted sum of the future stream of returns from the SOEs is what a buyer will pay (Van de Walle, 1989). The government would have got the same revenue had it not sold the SOE. Therefore, it would seem that privatization does not have any real impact on the government’s finances.

There are two reasons why privatization might still make a difference. First, a privatized firm might be expected to be more efficient than a SOE. Hence, the sum of discounted returns will be higher than that under government ownership. Secondly, the government, when it privatizes, is getting funds immediately. This added liquidity might be desirable for a number of reasons: for example, because the government might want to spend on education or infrastructure.

It is interesting to note that in theory, for a loss-making SOE the price might be negative. This is not very far-fetched. Governments have sometimes given so many concessions to the buyer to induce them to buy loss-making concerns that in effect the price has turned out to be negative.

One pertinent question here is whether selling bonds is a better means of raising revenues than equity sales. Suppose that privatization via equity sales does not change the earnings prospects of the firm in question. In countries with liquid bond markets, selling bonds might involve lower transaction costs than privatization. Moreover to garner popular support for privatization, equity is often underpriced. However, countries constrained in their ability to sell bonds, for example those facing debt crises, or those that are trying to limit borrowings to commit to an anti-inflationary policy, might be compelled to sell equity. In addition, the perceived risk of default on bonds might be higher than equity risk.

Efficiency gains

The argument for privatization often rests on the supposed superiority of the private sector in attaining the goal of economic efficiency – both allocative efficiency and productive efficiency. Proponents of privatization have argued that a change in ownership can have an important effect on economic efficiency. In SOEs, prices sometimes did not reflect scarcities properly. For example, if the government gives a subsidy for an input used by an SOE, the SOE would tend to overuse that resource. Or, if a SOE is a monopoly, then it can set its own price. SOEs therefore would not attain allocative efficiency.

It has also been argued that SOEs are likely to exhibit greater internal inefficiencies than private firms for various reasons. Public managers are given numerous and inconsistent objectives. Instead of control by
shareholders who are interested in profit-maximization, there is bureaucratic control which puts more emphasis on ‘playing it safe’. Suppose that productive efficiency requires use of an input that is not available in a competitive market. The manager in an SOE is required to obtain competitive quotations for almost everything, and hence may have no way of using this particular input because there is only one seller.

Whether such efficiency will indeed be achieved depends in turn on the goal(s) that managers pursue in private enterprises vis-à-vis public enterprises. In theoretical literature, it is often assumed that a manager of a private enterprise maximizes economic profit, while the manager of a public sector has the liberty or is forced to pursue a more diffused agenda. If this assumption is correct, then of course ownership per se becomes a determinant of efficiency, and there is no dearth of empirical research looking into the ownership issue. However, there are reasons to believe that the monitoring system and incentive system in private enterprises may not always work perfectly, and to a large extent, the efficacy of these systems depends on the market structure and the regulatory policy. Then not only is it necessary to turn our attention to a comparative evaluation of the incentive structure in these two types of enterprises, but we are forced to consider other alternatives for achieving efficiency.

In many economies today, the incidence of owner-managers is declining and most large firms are coming to be characterized by the separation of ownership from control. Even if the shareholders can be assumed to have profit-maximization as the overriding objective, managers can pursue their own goals because of the existence of asymmetric information: in general the managers have much better knowledge about market conditions, technology and their own effort levels than shareholders. Since shareholders cannot control managers, various instruments for curbing managerial discretion have been suggested. These include: product market competition which imposes a Darwinian survival requirement of profit maximization; labour market signalling which leads to diminution of market value of non-performing managers; the threat of hostile takeover and use of incentives like employee stock option plans to align the interests of managers with those of shareholders.

All these instruments have their own requirements to be successful. In many markets, competition is muted. The Old Boy network bypasses the signalling effect of non-performance. In response to the threat of hostile takeovers, managers have developed a host of takeover defences to discourage or foil such bids. Moreover, as Grossman and Hart (1980) have pointed out, a free-riding consideration which makes shareholders refuse to part with their shares in the anticipation of an imminent rise in share prices can abort attempted takeovers. The experiences of offering stock

Amitava Krishna Dutt and Jaime Ros - 9781848442818
Downloaded from Elgar Online at 07/19/2019 02:48:07AM via free access
option plans have been mixed, mainly because the design of these plans often insures managers against the downside risk.

Of course, in addition, market failures from externalities and public goods nature of certain commodities may mean that profit-maximization does not lead to efficiency.

For public sector enterprises, on the other hand, there is no market for shares and hence no market for corporate control. How does government monitoring compare with monitoring in private enterprises? In theory, the government has the ability to correct for all types of deviations between social and private returns in goods and factor markets (Yarrow, 1986). The dissatisfaction with government control stems from the fact that the ‘market’ for political control is highly imperfect. The voting public controls the government, but its knowledge about specific enterprises may be very weak, and normally it cannot vote separately on the issue of running public sector enterprises (though when the public sector constitutes a very large part of the economy, poor performance by such enterprises can become a significant political issue in elections). The resources at the government’s command can also lead to the soft budget constraint – continuing support of even non-viable and inefficient enterprises to keep them going. If it is argued that it is necessary to keep these enterprises alive to attain distributional goals (for example, provision of cheap products to poorer sections), then one is faced with the question whether there are alternative and better ways of attaining these goals. Again, sometimes, a fraction of the shares of an SOE are sold to the public, and then outsiders can monitor the enterprise’s performance to some extent. However, it still remains unclear how much divestment must take place before the outsiders can exert a significant oversight on the running of such enterprises.

As already mentioned, product market competition can be one instrument for checking managerial discretion. Product market competition forces firms to minimize costs and maximize profit for long-term survival. Moreover the observance of the performance of competing firms provides shareholders and governments with additional information about managerial inputs and the firm’s true opportunities. Thus in any year, profitability will depend on the levels and the quality of managerial inputs as well as a host of other factors outside of the manager’s control. It then becomes difficult to employ profitability as a correct indicator of managerial inputs and base rewards on such a measure. However, when the profitability of other firms can be observed, this to some extent aids shareholders in disentangling managerial contribution from random factors in a firm’s performance.

It is clear that for SOEs operating in competitive markets, prices would better reflect scarcities and therefore allocative inefficiency would be less. Then the gains from privatization would also be less. On the other hand,
transforming a public sector monopoly into a private sector monopoly would also not lead to increases in allocative efficiency. We can expect large increases in allocative efficiency to be achieved when a public sector monopoly is privatized and the market opened up to other players.

**Distributional impact**

Privatization can have rather significant effects on income distribution. One can discuss distributional issues either by looking at the sources of distributional changes or by identifying the major groups who might be affected by privatization. Perhaps the most important sources of redistribution effects are the changes in the prices of privatized assets and the pricing of commodities after privatization. If, for example, water and electricity supplies are privatized and water and electricity charges go up after privatization, this can affect large segments of the population adversely. If privatized assets are sold at discounted prices, there is a transfer of wealth to the new owners from the wider public and taxpayers in particular (Vickers and Yarrow, 1988). Such sales at discounts are sometimes politically attractive because the risk of shares being unsold is minimized, and ‘because the beneficiaries tend to be more aware of their gains than the losers feel the losses’ (Yarrow, 1986). In all these instances, it is clear that different groups are affected differently. If privatization is followed by layoffs, employees are affected. If privatization leads to higher profits, shareholders gain.

Sometimes the goods and services made available by the SOE to the poor may become less accessible after privatization. For example, a privatized airline may choose not to fly on unprofitable routes. On the other hand, if privatization is accompanied by a more competitive environment, then greater varieties of commodities may be available at lower prices. Especially in the telecommunications sector, privatization has given access to new and cheaper services to the population at large.

While privatization can have some adverse distributional implications, one can also argue that the SOEs have not had a very successful record of reaching the poor and the disadvantaged sections of the population. For example, in India, it has been repeatedly shown that the Public Distribution System does not do a good job of enabling essential commodities to reach the poorest sections; it is more useful to the richer sections.

**Some key issues**

**Implementation issues**

Worldwide experience shows that implementation of privatization programmes has lagged well behind stated intentions. Barring a few countries,
privatization has been limited to small SOEs of the manufacturing and the services sector. There are quite a few problems faced by countries trying to privatize SOEs. Firstly, in some developing countries, there is a lack of well-established, competent management consulting groups, accounting firms and investment bankers. These are needed to provide technical advice and valuation of SOEs. As a result, in some instances, foreign experts have been brought in.

Secondly, a valuation of the SOE has to be carried out before it can be offered for sale and the valuation exercise has faced severe problems. Valuation is a sensitive subject politically, because governments want to get high sales prices and at the same time the valuation process might raise questions about past public management and investment decisions. There have been inordinate delays in valuation. The problem is aggravated when poor records are maintained by SOEs.

Thirdly, once the valuation has taken place, administrative capacity is needed to assess buyers’ bids, arrange finance and insurance, and deal with a host of complex legal issues. Sometimes, a comprehensive rehabilitation plan for the SOE has to be designed, evaluated and financed before privatization is possible. Moreover, appropriate regulatory structures may not exist and may have to be set up, particularly when privatization leads to the creation of a monopoly.

Fourthly, capital markets in many developing countries are typically weak and poorly regulated. Large investments in equity are quite unusual. SOEs are some of the largest firms in the country and the private sector may not be in a position to fund the purchase of large assets. The private sector may also be suspicious about the government’s intentions, given the record of nationalization in the past. On the other hand, the government may not be willing to sell assets to foreign investors.

**Political constraints**

Generally, the costs of privatization are borne by a small group of people, for example the workers of the enterprise who may lose their jobs, or the suppliers who may lose favoured contracts. The benefits, however, are spread out over a large number of people, sometimes a very large section of the population. Public choice theory suggests that in such situations, it will be easier to organize opposition to the privatization programme than support. Experience tells us that in many countries, privatization programmes fail to mobilize popular support and in fact give rise to strong opposition.

Trade unions, in particular, tend to react strongly against privatization. Trade union power is often concentrated in the public sector and the public sector provides a base for such power. Unions oppose privatization, not
only because of the direct effect on employment, but also because of a fear that trade union power will be reduced in the private sector. The restructuring process generally involves laying off part of the workforce. Usually, forced dismissals are politically infeasible and only generate more opposition to privatization. Governments therefore try to adopt some kind of voluntary approach. Components of voluntary approaches that have been tried out include monetary compensation (for example through voluntary retirement schemes), retraining and redeployment. Cash-strapped governments may find it difficult to cover the cost of laying off workers. Sometimes the government agrees to accept a lower price for the enterprise in return for an assurance from the new owner that employees will be retained even after privatization. In the East German privatization programme, there is an instance where an enterprise was sold for 1 Deutschmark, because the bidder promised to retain all the workers.

Should restructuring occur before or after sale of the unit?
Most SOEs will not fetch a good price if they are sold in their current condition. For historical reasons, many have excess workers, are burdened with obsolete machinery and technology, and often are run bureaucratically. One option before the government is to restructure these enterprises before placing them on the market, for example, by laying off excess workers, by inducting new workers with appropriate skills, by selling off non-strategic parts of the business, computerization of operations, and so on. These enterprises will then become attractive to private investors who would be willing to pay high prices for them. On the other hand, proponents of speedy privatization (the ‘big bang approach’) argue that the attempt to restructure these enterprises before sales will inevitably lead to delays and the entire momentum for privatization will be lost. Further, it is doubtful whether the governments are at all adept at restructuring.

The evidence
Historically, it appears that SOEs have contributed quite significantly to the gross fixed capital formation of many economies. They have played an important role even in the highly successful East Asian newly industrializing countries. Moreover, there is no clear evidence of a negative correlation between the size of the SOE sector in an economy and its economic performance in terms of the rate of growth. There are also acknowledged outstanding cases of efficient SOEs, for example the giant Korean state-owned steel enterprise POSCO.

There are two separate empirical issues that must be kept in mind. One is the question of whether SOEs are necessarily less efficient than comparable private sector enterprises. The other is the question of the success of
privatization programmes. The first question can be summarized as ‘ownership matters’, that is, regardless of all other factors, the mere act of transferring a SOE to the private sector will increase its efficiency. The discussion in the section on the reasons for privatization has demonstrated that there is no a priori theoretical reason to believe that ownership alone matters.

Empirical investigations of these two questions run into several problems. For one, using profitability as a performance indicator for comparing SOEs with private enterprises misses out the point that SOEs are often established for reasons other than making profits. Therefore, if this measure is used it will tend to flatter privatization if under state ownership non-profit goals had been pursued. Even studies using profitability do not establish that SOEs are invariably inefficient. Moreover it is difficult to control for the effects of factors other than ownership which might affect a firm’s performance. Chang (2003) notes that there might be country-specific, industry-specific and firm-specific factors that are the determinants of performance differentials rather than ownership per se. Thus, countries with successful records of privatization sometimes underwent substantial macroeconomic changes that created a climate conducive for realizing microeconomic efficiency gains. So far as industry-specific results are concerned, the evidence of successful privatization in the telecommunications sector, to take one example, cannot be generalized to other sectors. Rapid technological changes in this sector have increased competition and reduced regulatory problems.

Parker and Kirkpatrick (2005) note that to assess the impact of privatization, in addition to using the correct performance measure, there is the problem of taking into account relative price changes with spillovers into other sectors of the economy, and redistribution effects on different socio-economic groups. Their review of the empirical literature leads them to conclude that: ‘The studies vary in terms of the financial and economic performance measures and show that privatization measures can lead to widely differing results.’

**Conclusion**

Privatization in essence represents a reduction in the role of the government in the economic activities of the nation. The process of privatization has been impelled by ideology as much as by necessity. Even if we restrict ourselves to efficiency concerns, there seem to be no convincing theoretical reasons for arguing that privatization per se will improve efficiency. This is especially important in developing countries where a number of preconditions needed to carry out successful privatization programmes are often missing. It is therefore no wonder that the empirical evidence is also mixed, particularly in the context of developing countries.
References


The attention given to corruption and its causes and consequences for economic development has expanded enormously in recent years. While in the 1970s and 1980s corruption was often treated in the economics literature as a benign redistribution of economic rents, since 1990 there has been increasing recognition in both academic and policy circles that high levels of corruption can have a large detrimental impact on economic growth and development. The change in thinking reflects the increasing focus of the economics profession on the role of institutions in economic development, as a rapidly growing body of economic research in the 1990s examined the implications of different institutional arrangements on the functioning of markets and the supply of public goods.¹ This change also partly reflects political developments – most notably the end of the Cold War, the transition from communism in Central and Eastern Europe, and the break-up of the Soviet Union – as the softening of superpower competition created conditions in which international organizations and donor agencies could be more forthright in challenging corrupt regimes in developing countries. The opening up of political regimes in other parts of the world, including increasing democratization in Africa, Asia and Latin America, also provided a window for citizens to organize and put increasing pressure on their governments to fight corruption.

Corruption can be defined as ‘the use of public office for private gain’, or it can be defined more broadly also to encompass the abuse of positions of trust in the private sector for personal gain. For the purposes of this chapter the term is limited to the narrower definition pertaining to the public sector, around which most research and policy work has focused. Fraud and abuse in the private sector is most often addressed by literature on corporate governance.

Measuring corruption
Before the mid-1990s it was generally assumed that corruption could not be measured – both because it was difficult to define and because perpetrators wanted to keep it secret – and thus it was a phenomenon that could only be discussed in generalities. Since the mid-1990s, however, survey tools have been developed to measure levels and patterns of corruption. While significant inaccuracies no doubt remain, these survey tools have added
immensely to economists’ ability to analyze the phenomenon and its impact on other economic variables.

There are two general types of survey tools: polls of ‘experts’ and surveys of economic actors. In each case the questions can be focused either on opinions and perceptions or on actual experiences. The first type of tool asks questions of people who are considered to have expert knowledge on a particular country, and tends to focus primarily on their opinions and perceptions on various issues, including the incidence and severity of corruption.2 The second type of tool surveys economic actors – whether enterprise managers, citizens or public officials. While such surveys may also contain questions about opinions and perceptions, surveys of economic actors often try to gather information about actual experiences in dealing with government – for example, the amount of bribes paid in obtaining a business license, getting access to medical care or interacting with tax officials.3 A third type of tool, pioneered in Transparency International's Corruption Perceptions Index, is an attempt to combine all surveys on an individual country into one composite indicator that facilitates rankings and comparisons among countries.4

Economics literature and applied policy analyses have utilized these various survey tools extensively since the mid-1990s to understand the extent and patterns of corruption and their implications for economic development. Aid donors have also relied increasingly on these tools to help focus aid in countries with lower levels of corruption.5 As the number of types of measurement tools have increased and begun to have greater practical impact in recent years, it has also become increasingly important to understand the pros and cons of different approaches and how they compare to each other.6

Levels and types of corruption

The surveys undertaken since the mid-1990s have proven what casual observers already know – that levels and patterns of corruption vary widely among countries. It is indeed not really possible to talk about ‘corruption’ as one phenomenon, as public office can be abused for private gain in a myriad of different ways. Some literature distinguishes between ‘petty’ and ‘grand’ corruption, with the difference between the two being largely a function of the size of the bribe and the status of the briber. Large bribes paid to senior officials for major public contracts are seen as ‘grand’ corruption, while small bribes paid to the traffic police, for example, are classified as ‘petty’. More recent literature distinguishes between ‘state capture’ and ‘administrative corruption’, with the former referring to corruption in the formation of laws and regulations (for example bribes to legislators or regulators to tilt lawmaking in certain directions) and the latter...
referring to corruption in their implementation (for example bribes to get goods through Customs or enroll children in favored schools). State capture is often seen as the more pernicious, because it affects the rules by which markets function. Both state capture and administrative corruption come in various forms, depending on who the parties to the corruption transaction are and what is being bought, and both can occur on a large or small scale. ‘Unbundling’ corruption into its various types can give analysts and policy makers a more nuanced sense of its economic and political causes and consequences.

The costs of corruption

High levels of corruption can have devastating impacts on an economy and a society. Among the most pernicious are bribes that allow people to evade laws that protect public safety. Importers, for example, may bribe customs officials to allow dangerous or ineffective drugs into the market, or builders may bribe regulatory agencies to erect buildings that do not meet safety codes. Also costly to public welfare is corruption that affects access to public services, as when parents must bribe doctors or teachers for medical care or education for their children. In all cases, widespread corruption undermines citizen trust and rule of law, and thereby impedes the arm’s-length transactions among strangers that are so fundamental to a market economy.

A large body of academic analysis has been undertaken since the mid-1990s to try to measure the economic impact of corruption more precisely, with various dimensions of economic impact being studied. One of the first studies was the 1995 analysis by Mauro of investment in a cross-section of 67 countries, which found that corruption has a significant negative impact on the level of investment in relation to GDP. Further studies in the late 1990s using a variety of corruption indicators reinforced this overall finding. One such study differentiated among types of corruption and found that in settings in which bribes had less predictable outcomes – that is, where bribers were less confident about getting what was paid for – corruption had a stronger negative impact on investment. Another study differentiated between ‘centralized’ (or coordinated) and ‘decentralized’ (or uncoordinated) bribe-taking, arguing that the economic impact of the latter is likely to be more severe, and using the model to illustrate the increased cost of corruption in post-communist Russia as compared with the centralized monopolistic corruption imposed by the Communist Party. Analytic work has also focused on the negative impact of corruption on foreign direct investment (FDI). Wei found that corruption at the level found in Mexico was equivalent to a 20 percent tax as compared to that found in Singapore.
While corruption and gross domestic product (GDP) per capita are highly correlated, the direction of causation is more difficult to untangle. On the one hand, corruption can hamper growth by reducing the efficiency of public spending and the effectiveness of public service delivery. On the other hand, poorer countries have a more difficult time tackling corruption, both because bribes may be more tempting when public sector salaries are low, and because it takes resources to fund ‘watchdog’ groups needed to prevent corruption, such as the press, accounting and auditing services, and police and other investigative and law enforcement agencies. A large body of recent literature attempts to unravel the effects of corruption on either the level or the rate of economic growth. While many methodological difficulties make this line of research difficult and skepticism remains about the findings of individual studies, the broad consensus supports the view that corruption – rather than ‘greasing the wheels’ of commerce – has a substantial negative impact on economic growth.

Other economic costs and consequences have also been identified through research in recent years. Corruption has been found to increase inequality, and the reverse – that higher inequality leads to more corruption – has also been shown. Corruption reduces public revenues, leads to lower quality in public investments and public services, and tends to skew public spending away from education to other types of spending – such as large infrastructure projects or military procurement – where bribes are more lucrative. Corruption has detrimental effects on the environment by reducing the effectiveness of environmental regulation, and it is positively correlated with a country’s rates of inflation and crime.

Most fundamentally, corruption reduces citizens’ trust in government and the political system, which undermines the ability of governments to cooperate with the citizenry in the formulation of policies and enforcement of laws. A dysfunctional and mutually reinforcing equilibrium of citizen distrust, poor government performance and corruption can result, undermining economic growth and development in the many ways outlined above.

**Tackling corruption**

In its simplest form, corruption is facilitated by the ability of a public official to exercise discretion in the carrying out of his or her duties without having to answer to formal organs of accountability. Hence the formulation that $C = M + D - A$ (corruption equals monopoly plus discretion minus accountability). This simple formulation points to several variables that can affect the level of corruption in an economy. A large role for government and greater discretion for individual public officials opens avenues for potential corruption, and such corruption is more likely to materialize...
if formal institutions of government accountability are weak. Corruption is also more likely when valuable and sought-after assets are under government control, as, for example, in countries with abundant state-owned natural resources (for example, oil- and gas-producers) or in transition countries with large-scale privatization programs. The simple lesson would be to minimize government’s role and public officials’ discretion while strengthening oversight institutions in an economy.

Reality is much more complex, however. First of all, although people may disagree at the margin about the appropriate roles for government, such roles will never be eliminated entirely, as there are certain public goods (for example, defense, law and order, basic infrastructure and education, environmental protection) that governments have an important role in supplying. Furthermore, giving discretion to public officials is often critical to getting good results, particularly in more complex areas of public policy. Building in mechanisms to ensure transparency and accountability in public decision-making – for example through public hearings on draft laws, ‘freedom of information’ and publication of government decisions, or internal or external audit procedures – is always desirable, but these can be costly or difficult to implement, particularly in poor countries with few public resources and a severe shortage of skills. The challenge in any particular case is to understand these trade-offs and try to tailor the role of government, the discretion granted to officials, and efforts to strengthen accountability to the particular needs and characteristics of each country situation.

Controlling corruption also requires the active commitment and involvement of the citizenry. Formal public law enforcement is expensive and necessarily operates only at the margin in any country; most laws are enforceable in practice because citizens willingly obey them and are willing to take action when someone else breaks them. If, in contrast, people lack confidence that laws can be enforced, this can become a self-fulfilling prophecy, as they will neither respect and follow the laws themselves, nor bother to report others who break them. One can envision two situations of equilibrium: one where most people follow the law and expect others to follow it, and thus where transgressions are few and can be managed by formal law enforcement; and the other where most people do not follow the law nor expect others to, and transgressions are too numerous for formal law enforcement to handle. Both equilibrium conditions are common in today’s world: corruption is the rare exception in some countries, but it is systemic and widespread in others. The challenge that many systemically corrupt countries face is how to move from one equilibrium state to the other.

In practice, governments that are strongly motivated to reduce corruption have many policy and institutional levers they can use. A multi-pronged
approach to tackling corruption would address many institutional dimensions, as shown in Figure 59.1.23

State capture tends to thrive when the private sector is monopolized, and economic conditions and policies that enhance private sector competition and transparency will tend to reduce incentives and opportunities for corruption. These include, for example, policies that promote international trade,24 low entry barriers for new firms, a stable macroeconomic framework with low inflation, and well-designed regulatory, corporate governance and anti-monopoly laws.25 Competition and transparency in the political realm can also help, including clear and binding rules for political party financing or asset declaration and conflict of interest rules for senior government officials.26 More fundamentally, government structures that limit power and create horizontal or vertical checks and balances – for example, through legislative or judicial oversight of government decisions or through multi-tiered layers of government28 – can greatly enhance political competition and accountability. On a day-to-day basis governments can take steps to improve public sector management by adopting rules and enhancing incentives and skills for meritocratic civil service staffing,29 honest tax and customs administration, and transparent public procurement and budgeting procedures. As there are almost always losers as well as winners in these reforms, governments must make intensive efforts to build public support and publicize early ‘wins’ to overcome countervailing pressures. Governments can also restrict opportunities for corruption by taking steps to reduce their role in an economy – for example, by privatizing commercial firms and reducing regulations on business activity. While it is true that cross-country regressions do not show a statistically significant correlation between the overall size of government and the level of corruption,30 it is also true that opportunities for corruption can be reduced in any particular country by shedding unnecessary activities in the public sector. Finally, governments can help create the legal basis and political openness for a vibrant and independent media31 and a strong civil society to function. Both are critical counterweights to government power and are essential to the control of corruption.

Reforming the political system to increase transparency and accountability is part of the challenge, as noted above. However, the presence or absence of democracy per se is not necessarily the determining factor. Young democratic systems with limited transparency and accountability can be highly corrupt, and pay-offs often increase during election years as politicians trade bribes for votes or private parties buy jobs or favors from new governments. However, political reformers and the international community need to persist in promoting democratic values, because there is clear evidence that a longer exposure to democracy lowers corruption.32
Figure 59.1  Addressing corruption requires action across many fronts

- **Structure of government**
  - Horizontal distribution of powers (executive, legislative, judiciary)
  - Vertical distribution of power among levels of government.
  - International dimensions: FDI, trade, donor rules/controls

- **Political accountability**
  - Political competition, credible political parties
  - Transparency in party financing
  - Disclosure of parliamentary votes
  - Asset declaration, conflict-of-interest rules

- **Civil society voice & participation**
  - Freedom of information
  - Public hearings on draft laws
  - Media/NGOs
  - Community input/monitoring

- **Competitive private sector**
  - Economic policies
  - Restructuring of monopolies
  - Regulatory simplification
  - Transparency in corporate governance
  - Collective business associations

- **Public sector management**
  - Meritocratic civil service with adequate pay
  - Public expenditure management
  - Decentralization with accountability
  - Frontline service delivery (health, education, infrastructure)
  - Tax and customs
And when there is a window of opportunity to change the design of the democratic system, reformers should note that parliamentary systems appear to do a better job than presidential systems in controlling corruption, at least as long as political parties are generally oriented toward the public interest. The Westminster system of democracy found in the UK – with a long and unbroken history, direct election of candidates in lieu of proportional representation, strong and accountable political parties, high electoral participation, and voting districts that are not unduly small – is considered by many to be the model of democracy most likely to control corruption.

The question of whether decentralization of government power reduces corruption is a subject of active debate, and many aid donors are pressing for further decentralization and community participation in countries with weak governance. Decentralization advocates argue that moving government closer to the citizenry will force greater transparency and accountability, while more cautious observers question the capacity of local governments (particularly in poor countries) and the likelihood of ‘capture’ of decentralized power by local elites. On the one hand, it is clear that some of the best-governed countries in the world (Denmark, for example) are also among the most decentralized. On the other hand, it is also clear that state and local governments in many settings (the United States, for example) have been more prone to corruption scandals than national governments. The extensive research on this issue is inconclusive and suggests that the details of each country’s historical, cultural, political and economic situation are likely to be more important than the extent of decentralization per se in determining how decentralization will affect governance and corruption.

Progress is possible
Experience around the world shows that strong and committed leaders can indeed reduce corruption in government. Among the more advanced countries, for example, both the United States and the United Kingdom undertook far-reaching civil service reforms in the nineteenth century that helped to reduce corruption. More recently, corruption in state governments in the USA is widely thought to have been far worse 50 years ago than it is today. And, more recently still, many of the former communist countries of Central and Eastern Europe appear to have made significant strides in reducing corruption since the mid-1990s. In-depth surveys of businesses undertaken in 1999, 2002 and 2005 indicate that the frequency of bribery in many types of public services and the constraint that corruption imposes on doing business declined significantly in many countries over this period. A combination of better economic policies, stronger
economic and political institutions, faster economic growth and a more vibrant civil society combined to put pressure on governments for greater honesty and accountability.

Eliminating all corruption is not feasible. It exists and will continue to exist in every government in every country in the world. However, a much more feasible – and in the end more important – goal in every country should be to reduce the level of corruption and its negative social and economic impacts. For developing countries with widespread and systemic corruption, the goal is to move to an equilibrium where corruption is the exception rather than the norm. There is little doubt that declining corruption, enhanced government effectiveness, greater citizen trust and increased private investment are mutually reinforcing phenomena that together spur economic growth and development.

Notes

1. See, for example, North (1990), Bardhan (1997a) and World Bank (2002).
4. Examples include the Transparency International (TI) indicators (www.transparency.org) and indicators put together by the World Bank Institute (www.worldbank.org/wbi/governance).
5. For example, both the level of resources made available to poor countries by the World Bank’s International Development Association (IDA) and the eligibility of countries to access grants from the US Millennium Challenge Account (MCA) are dependent in part on countries’ scores on various governance and corruption indicators.
7. For further explanation of the concept of state capture, see World Bank (2000) and Hellman et al. (2006).
8. For a synopsis of some of the vast literature on corruption and development, see Bardhan (1997b) and Lambsdorff (2006).
13. Some recent research that attempts to use instrumental variables to address these simultaneity problems finds that corruption leads to lower GDP per capita. See, for example, Hall and Jones (1999) and Kaufmann et al. (1999).
17. Tanzi and Davoodi (1997) and Gupta et al. (2002).
21. La Porta et al. (1997).
24. Research points to the value both of low international trade barriers and of a uniform (rather than highly diversified) tariff structure in reducing opportunities for corruption: Gatti (1999) and Ades and Di Tella (1999). Wei (2000a) argues that a country’s ‘natural openness’ (that is, small size and non-remote location) is more important than trade policy per se in increasing competition from trade and thereby reducing corruption.
26. With regard to electoral rules, there is some support for the view that corruption is less likely in countries where voting districts are larger (allowing greater political competition) and where citizens vote for individual candidates rather than party lists (which increases individual accountability of politicians to the electorate). Persson et al. (2003).
27. For an analysis of the importance of an independent judiciary in controlling corruption, see World Bank (1997) and Ades and Di Tella (1997).
28. See fuller discussion on decentralization below.
29. Evans and Rausch (2000). Raising civil service salaries can also help to reduce corruption (van Rijckegehem and Weder, 2001), but, above a certain reasonable minimum, raising salaries is not as important as promoting merit-based hiring and promotion.
30. It is not surprising that cross-country regressions show no significant correlation between government size and the level of corruption, given the two-way nature of causation. While an overextended public sector may create more opportunities for corruption, corruption (and the low level of trust and public sector effectiveness it causes) may also make it more difficult for governments to collect public revenue. Thus only the more effective governments may have the capacity and citizen trust to grow and remain large. Indeed, many of the world’s least corrupt countries – most notably in Scandinavia – are also among the countries with the largest public sectors. When the Scandinavian countries were omitted in one study, size of government was found to be positively correlated with the level of corruption. LaPalombara (1994).
34. With regard to culture, Husted (1999), drawing on Hofstede (1997), finds that more hierarchical, more materialistic, and more risk-averse cultures are likely to be more corrupt.
35. A large number of studies have been undertaken in recent years to try to assess the links between decentralization and corruption, with some supporting the view that greater decentralization is correlated with lower corruption and some refuting that view (generally finding that the correlations disappear when the sample changes or when decentralization is measured in a different way). Another strand of this literature looks at federal and unitary states, with some studies finding that federalism is correlated with higher levels of corruption and others showing no effect. See Lambsdorff (2006) for further discussion.

References


Rule of law
In his manuscript *Lectures on Jurisprudence*, based on his lectures at Glasgow University in the early 1760s, Adam Smith stated that a factor that ‘greatly retarded commerce was the imperfection of the law and the uncertainty in its application’. This is still one of the main messages of the Law and Economics literature as it pertains to development. Law and Economics is a thriving subject in the USA, and it is now being widely adopted in other countries, including in law schools of developing countries. But its Chicago origins and the general American mold may have given a particular slant to the development of the subject, which is not always quite appropriate for these countries. In this chapter I shall focus very generally on some of the special issues that arise in the context of developing countries that the literature on Law and Economics needs to address if it is to be applicable there. These special issues arise primarily because the institutional, political and behavioral context in these countries is different from the usual context of the literature.

In this literature as well as that of recent Institutional Economics the major emphasis is on contract law and security of property rights. In the pervasive context of incomplete contracts the emphasis is rightly on the residual rights of control, and the security and predictability of these property rights are crucial for economic performance and long-term investment. Throughout history in any time-separated activity – for example, if the seed planter cannot be secure in reaping the harvest, if a trap-setter cannot claim the trapped game, or a lender is uncertain of being repaid – economic life is hampered by insecurity of property rights. North and Weingast (1989) trace the success story of development in English history to the King giving up royal prerogatives and increasing the powers of the Parliament in 1688, thus securing private property rights against state predation and allowing private enterprise and capital markets to flourish. The more recent empirical literature has tried to quantify the effect of these property rights institutions – or what they call in this literature the ‘rule of law’ variable (one standard measure combines indices of effectiveness and predictability of judiciary, enforceability of contracts and incidence of crime) – on economic performance from cross-country aggregative data. Since these institutions may be endogenous (that is, economically better-off countries may have
more of those institutions, rather than the other way round), the literature
tries to resolve the identification problem by finding exogenous sources of
variations in those institutions. See Acemoglu et al. (2001, 2002). Rodrik
et al. (2002) use similar data to show that once the property rights institu-
tions are accounted for, the role of other factors like geography or openness
to trade in explaining cross-country variations in per capita income is
minimal.

What is often ignored in this literature is that the ‘rule of law’ actually
involves a whole bundle of rights, and we need to ‘unbundle’ it. Even for
security of property rights, different social groups may be interested in
different aspects of these rights. For example, the poor may be interested
primarily in very simple rights like land titles, and also, to a very important
extent, in protection against venal government inspectors or local mafia; to
them that is the most salient aspect of security of property rights. For the
richer investors, however, a whole range of other issues like protection of
the minority shareholders in corporations, oversight of capital markets
against insider abuse, bankruptcy laws, and so on, loom large; these are
what investors emphasize when they talk about security of property rights.
As different groups are thus interested in different aspects of security of
property rights, these rights may have differential political sustainability,
depending on how politically influential the corresponding groups are in a
given polity.

‘Rule of law’ should also include other rights, some quite different from
mere security of property rights. For example, one part may involve various
democratic rights of political participation, association, mobilization and
expression of ‘voice’. An analysis of cross-country variations in human
development indicators (which includes education or health variables like
mass literacy or life expectation) shows that an institutional variable mea-
suring ‘voice’ or participation rights is just as important as that measuring
security of property rights as an explanatory variable; – see Bardhan
(2005), Chapter 1. In other words, the part of ‘rule of law’ that refers to
democratic participation rights explains a significant amount of variation
in human development indices across countries. Those who emphasize
property rights often ignore the effects of participatory rights, and there is
some obvious tension between these two types of rights included in the
standard package of ‘rule of law’.

The idea of security of property rights has been extended to the case of
intellectual property rights for the preservation of incentives for innov-
ation. Since innovations are the main source of economic growth, laxity in
the enforcement of international patents and copyrights in developing
countries for products that are knowledge-intensive or require expensive
investments in research and development is often regarded as harmful for
long-term economic growth. This has been the rationale for the incorporation of TRIPS (Trade-Related Intellectual Property Rights) in World Trade Organization (WTO) rules, when developing countries accepted these rules under some pressure from rich countries. While keeping incentives alive for new research and innovations is extremely important, the question from the point of view of a developing country is usually whether the enormous costs (including the often exorbitant monopoly prices charged by the patent holder for a prolonged period) are always worth the benefits, and whether there are better alternative ways of encouraging research. It is recognized now by many scientific researchers that existing patents often act as an obstacle to further research that tries to build on earlier findings (in developing countries this includes research for adapting new technology to the special conditions there). This is linked with the question of the optimal patent breadth, which is about how broadly the protection of existing innovations ought to extend to related innovations in the future. The alternative method of subsidizing research inputs (rather than rewarding research output with temporary monopoly) has the advantage of encouraging information-sharing and collaborative research. Of course, upfront funding carries with it the moral hazard problem that researchers, once having secured funding, may be tempted to pursue activities or lines of research other than those most desired by the public sponsor. This problem may be mitigated if researchers expect to apply for public funding in future.

The problem of international patents in life-saving drugs in poor countries recently caught public attention in connection with the controversies about the prices of anti-retroviral drugs for AIDS patients in Africa. The major problem in corporate drug research is that only a tiny fraction of what the companies spend on finding new diet pills or anti-wrinkle creams is spent on drugs or vaccines against major killer diseases of the world’s poor, like malaria or TB, and the situation has not changed with the onset of TRIPS and is not expected to change even with a more stringent enforcement of TRIPS in poor countries. So alternative avenues of encouraging such research have to be sought. There are now the beginnings of some international attempts to make credible arrangements on the part of international organizations like the World Health Organization (WHO) in collaboration with non-governmental organizations (NGOs) like Médecins sans Frontières, private foundations (like the Gates Foundation) and donor agencies and governments to a commitment to purchase vaccines to be developed by pharmaceutical companies against some of these diseases. For a discussion of the incentive issues in vaccine purchase commitments, see Kremer (2001). For other diseases (like diabetes or cancer) which kill large numbers of people in both rich and poor countries, the incentive argument for enforcing patents in poor countries is weak, since that research will
be carried out by the transnational drug companies in any case as the market in rich countries is large enough (provided resale can be limited).

We have earlier commented upon the different kinds of security of property rights being relevant for different social groups. In the case of intellectual property rights as well, the transaction costs may limit the symmetry of access of different groups to those rights. Khan and Sokoloff (1998), in a historical comparison of the patent systems in the USA and Britain in the first half of the nineteenth century, show that while the British system used to effectively limit access to intellectual property rights to the relatively wealthy and well-connected, access in the American system was much more broad-based, and this contributed to a much more vigorous and wider spread of patenting activity in the USA in that period.

**Beyond formal legal institutions**

While nobody will deny the importance of innovations in the process of economic growth, in the case of manufacturing technology in most developing countries the problem is really in adaptation of technology theoretically available elsewhere. Much of the effective use of that technology particularly in these alien circumstances is not codified, but implicit or tacit, and cannot be just transplanted from abroad. Learning by doing and domestic efforts to adapt and assimilate are critical, costly and time-consuming, and in this, government investment in market-supporting infrastructure and in research and training and extension are quite important. Just putting in place a legal system facilitating private efforts may not be enough. As Pack (2003) points out, in recent years many developing countries have liberalized domestic and international trade regulations but have not realized high total factor productivity, in the absence of a set of institutions constituting a national innovation system and extension services that facilitate appropriate training and technology absorption.

There are also corresponding implications for the inadequacy of just a legal framework in developing credit and equity markets or the requisite financial infrastructure in general. Investment in learning by doing is not easily collateralizable and is therefore particularly subject to the high costs of ‘imperfect information’. At an early stage (which can be prolonged in poor countries) when firms are not yet ready for the securities market (with its demands for codifiable and court-verifiable information), there is often a need for some support and underwriting of risks by some centralized authority (with, of course, its attendant dangers of political abuse). There is also the problem of interdependence of investment decisions with externalities of information and the need for a network of proximate suppliers of components, services and infrastructural facilities with large economies of scale. Private financiers willing and able to internalize the externalities
of complementary projects and raise large enough capital from the market for a critical mass of firms are often absent in the early stage of industrialization. Historically, the state has played an important role in resolving this kind of coordination failure by facilitating and complementing private sector coordination – as the examples of state-supported development banks in nineteenth-century France, Belgium and Germany, and more recently in Japan, Korea, Taiwan and China, suggest. There are, of course, many examples of state failures in this respect and politicization of financial markets in other developing countries. In much of the literature on Law and Economics, as in Institutional Economics, the importance of the state is recognized only in the narrow context of how to use its power in the enforcement of contracts and property rights, and at the same time how to establish its credibility in not making confiscatory demands on the private owner of those rights. The history of the successful as well as failed cases of the state as a coordinator of technology assimilation and financial market development has lessons which should be analyzed in a framework that goes beyond this narrow context.

Why does a society not always adapt its legal and institutional set-up to facilitate productivity-enhancing innovations? Such innovations have gainers and losers, but in most cases the gainers could potentially compensate the losers. The problem is that it is politically difficult for the gainers from a change to commit credibly to compensate the losers ex post. As Acemoglu (2003a) puts it, there may not be any political Coase Theorem, whereby politicians and powerful social groups could make a deal with the rest of society, give up some of their control on existing rules and institutions that are inefficient, allow others to choose policies and institutions that bring about improvements in productivity, and then redistribute part of the gains to those politicians and groups. Such deals have severe commitment problems; those in power cannot credibly commit to not using this power in the process, and others cannot credibly commit to redistribute once the formerly powerful really give up their power for the sake of bringing about new rules and institutions.

A central issue of development economics is thus the persistence of dysfunctional regulations and institutions over long periods of time, as we discuss in Bardhan (2005), Chapter 2. In particular, the history of underdevelopment is littered with cases of formidable institutional impediments appearing as strategic outcomes of distributive conflicts. Acemoglu and Robinson (2002) develop a theory where incumbent elites may want to block the introduction of new and efficient technologies because this will reduce their future political power; they give the example from nineteenth-century history when in Russia and Austria-Hungary the monarchy and aristocracy controlled the political system but feared replacement and so
blocked the establishment of rules and institutions that would have facilitated industrialization. These replacement threats are, of course, often driven by extreme inequality in society.

In explaining the divergent development paths in North and South America since the early colonial times, Engerman and Sokoloff (2002) have provided a great deal of evidence of how in societies with high inequality at the outset of colonization rules and institutions evolved in ways that restricted to a narrow elite access to political power and opportunities for economic advancement. Initial unequal conditions had long-lingering effects, and through their influence on public policies (in distribution of public land and other natural resources, the right to vote and to vote in secret, primary education, patent law, corporate and banking law, and so on) tended to perpetuate those institutions and policies that atrophied development. Even in countries where initially some oligarchic entrepreneurs are successful in creating conditions (including securing their own property rights) for their own economic performance, as long as that oligarchy remains powerful, they usually get away with regulations that raise entry barriers for new or future entrepreneurs, and this blocks challenges to their incumbency and thus sometimes new technological breakthroughs. See Acemoglu (2003b) for a theoretical analysis of this kind of dynamic distortion in oligarchic societies even when property rights are protected for the initial producers. The classic example of inefficient rules and institutions persisting as the lopsided outcome of distributive struggles relates to the historical evolution of land rights in developing countries. In most of these countries the empirical evidence suggests that economies of scale in farm production are insignificant (except in some plantation crops), and the small family farm is often the most efficient unit of production. Yet the violent and tortuous history of land reform in many countries suggests that there are numerous roadblocks on the way to a more efficient reallocation of land rights, put up by vested interests for generations.

Inequality in power distribution in society also influences the social legitimacy of laws enacted or decreed by the powerful, and the degree of commitment of the general population to the rule of law. When the state is captured by a narrow clique, or when the state is weak so that there is an ‘oligopoly’ of coercion and authority (as opposed to the ‘monopoly of violence’ that Max Weber attributed to the state) shared by various protection rackets and corrupt officials (police, judges, bureaucrats), there is usually a big gulf between laws that are in the statute books and their enforcement, and, most importantly, a deficiency in every citizen’s expectations about others’ compliance, which form the foundation of the rule of law. Along with the underlying power distribution and enforcement mechanisms in society, some overarching social norms and political commitments provide
the main structure within the confines of which the formal legal system operates, and compared to the former the latter – which is the focus of much of the Law and Economics literature – is often in a secondary role.

These important elements of the institutional, political and social framework are ignored in a recent burgeoning of empirical literature on the effects of legal origins of a system. La Porta et al. (1997, 1999) have called attention to the superior effects, across countries, of the Anglo-Saxon common law system based on judicial precedents over the civil law system based on formal codes, on the corporate business environment both in terms of more flexibility with changing needs of business and in terms of better protection for external suppliers of finance to a company (whether shareholders or creditors). Apart from some doubts about the establishment of causality in these cross-national studies, one can also question the historical evidence even in the rich countries themselves. Lamoreaux and Rosenthal (2005) have done a comparative study of the constraints imposed by their respective legal systems on organizational choices of business in the USA (with its common law system) and France (with its civil law codes) during the middle of the nineteenth century around the time when both countries were beginning to industrialize. They conclude that there was nothing inherent in the French legal regime that created either a lack of flexibility or a lack of attention to the rights of creditors or small stakeholders. Many of the rules in the USA for minority shareholder rights actually came after the insider scandals of the Great Depression period. Franks et al. (2003) point out that in the UK it was not until as late as 1948 that the Parliament began to enact limited legislation to protect minority shareholder rights. Rosenthal and Berglof (2003) also question the primacy of legal origin in explaining institutions of investor protection; drawing upon the legislative history of US bankruptcy law they show how the USA, with an English common law legal origin, ended up with a bankruptcy regime quite different from that in the UK, and how political and ideological forces shaped financial development.

Several legal scholars – see, for example, Roe (2003) – have pointed out how the nature of corporate governance even in American large firms depends more on socio-political factors than on the form of corporate laws.

In any case, as we have indicated earlier, the importance of the legacy of the formal legal system is rather moot where much too frequently in developing countries the enforcement of whatever the laws are in the statute books is quite weak, and the courts are hopelessly clogged and corrupt. Take the two largest developing countries, China and India. India has inherited the English common law system, and being a democracy, legal rights there are more well defined and the legal system is less subject to political discretion than in China under the monopoly control by a Communist Party. And yet, according to the World Bank Report on Doing
Business in 2005 (World Bank, 2005), it is China which seems less disadvantaged in most indicators of regulatory and judicial effectiveness in business matters. For example, registering property requires 67 days and costs about 14 per cent of property value in India, whereas in China it is 32 days and 3 per cent of property value. In enforcing debt contracts it requires 425 days and costs about 43 per cent of debt value in India, whereas in China it is 241 days and 26 per cent of debt value. On closing an insolvent business it takes about ten years in India, in China 2.4 years.

In many developing countries the efficiency of courts as mechanisms of resolving disputes or enforcing contracts is shaped by a rather warped system of incentives: judges, even when they are not corrupt, do not care about delays, lawyers earn more when court proceedings are prolonged, appeals are too easy and some defendants deliberately seek continual delay in judgment. Courts are congested because of too-lengthy procedures and built-in incentives for over-litigation, apart from administrative delays in appointments of judges. Such low judicial effectiveness in commercial law, apart from raising transaction costs all around, has important effects on the size and structure of firms. This is because the more effective the judicial process, the more you can have relatively complex contracts, larger firms can thrive and more complex goods be produced.

Social and behavioral presumptions
Finally, I am going to comment on some of the broad presumptions of the Law and Economics literature which may need to be changed or made more flexible if it is to be applied to developing countries. One relates to the scale of economic activity. In small peasant communities where the scale of economic activity is not large, informal relational contracts may be more efficient than rule-based contracts supported by elaborate legal–juridical procedures. Breaches of relational contracts are often observable by other community members even when not verifiable by courts, and punishment is usually through social sanctions and reputation mechanisms. Another advantage is flexibility and ease of renegotiation. But as the scale of economic activity expands, as the need for external finance becomes imperative, and as large sunk investments increase the temptation of one party to renege (and as increased mobility and integration with the outside world improve exit options), relational contracts and reputational incentives become weaker. As Li (2003) points out, relation-based systems of governance may have low fixed costs (given the pre-existing social relationships among the parties and the avoidance of legal–juridical and public information and verification costs of rule-based systems), but high and rising marginal costs (particularly of private monitoring) as business expansion involves successively weaker relational links.
Of course the transaction costs of legal–juridical systems are asymmetric in their incidence on the rich and the poor as they try to get legal remedies, and it is not surprising that the legally handicapped poor often feel that the law is just another ‘stick’ with which the resourceful rich can beat them. In small face-to-face communities what anthropologists call the ‘politics of reputation’ may provide some modest measure of protection for the weak against the strong; as long as all parties belong to what is perceived to be the same ‘moral community’ in terms of which reputation is defined, there are some accepted limits and symbolic sanctions against the kind of ruthless exercises of power that sometimes accompany the cut-throat impersonality of the legal system enforced by the gendarmerie of the state.

It also needs to be recognized that in a world of highly imperfect information and the interconnected and multiplex nature of traditional informal contracts, the establishment of market relations enforced by the legal system in one market can crowd out implicit contracts in other related markets. Kranton and Swamy (1999) show in a study of the impact of the introduction of civil courts in British India on the agricultural credit markets of the Bombay Deccan that while it led to increased competition in the credit market, it reduced lenders’ incentives to subsidize farmers’ investments in times of crisis, leaving them more vulnerable in bad times, with insurance markets largely absent. In the context of environmental management of the village commons, Seabright (1993) has pointed out that as contracts are necessarily incomplete, attempts to enforce private property rights may weaken the mechanisms of cooperation that previously existed among the resource users, who may have shared implicit non-contractual rights in the common property resource.

The Law and Economics literature has inherited from mainstream economics the latter’s behavioral postulate of rational self-interested individuals. This postulate is being increasingly questioned in the branch of economics that is now called ‘behavioral economics’, but there may be special reasons for questioning it in the context of poor countries. In traditional communities where your conformity to community norms is at a special premium, we may have to pay particular attention to social preferences (‘other-regarding’ as opposed to self-centered, or ‘process-regarding’ as opposed to simply outcome-oriented) which may go beyond the narrow interpretation of self-interested behavior. For example, social reciprocity (individuals going out of their way to reward helpful actions by other members of the community, or taking revenge for perceived unfair or nasty behavior on the part of others at some considerable cost to the revenge-taker – ‘honor killings’ in many traditional societies being the extreme but not uncommon case) is often a foundation stone of community norms,
which define the informal institutional framework within which particular legal rules can be implemented.

It is also a questionable presumption of the Law and Economics literature that individuals always behave in their best interests. Common observations of myopic, weak-willed, procrastinating and time-inconsistent behavior fly in the face of the inexorably rational economic man of our textbooks. This may be a special problem in poor countries where public information media are weak, many people are uneducated and superstitious, and there is a surfeit of touts, middlemen and operators trying to manipulate people to make hasty uninformed decisions. The innate psychological characteristics of people may not be different in poor countries, but their circumstances and information sources are often quite different, and capacity for complex calculations is an acquired trait, honed only as transactions become more complex. Also, people often internalize their constraints and by all accounts the constraints are much more severe in the case of poor people. All this may sometimes call for more paternalistic regulations than are admitted in the rational-choice framework of Law and Economics. For example, consumer protection regulations in food labeling and health warnings, publicizing of information about often the exorbitant implicit interest rates charged in instalment purchases of durables from retailers and pawnbrokers, publicizing the odds of winning lotteries (which are very popular, as most people systematically overestimate their chance), are all instances of paternalistic regulations that are particularly important in poor countries. One, of course, has to be wary of the slippery slope here that may easily end up in heavy-handed regulations or regulatory capture, but one cannot deny that the sovereignty of the rational consumer is a particularly egregious myth in such contexts.

Furthermore, the Law and Economics literature, particularly through its Chicago origins, has inherited a presumption about voluntary contracts that one may have to be careful about. Milton Friedman and others have repeatedly asserted that if parties enter into a transaction voluntarily (without adverse effects on third parties), legal rules should not interfere; they should play only an enabling or facilitating role in that transaction. There are, however, many cases, particularly in poor countries, where it is possible to show that one party in this transaction would have been actually better off if the law intervened to take out certain options from the choice set. Take the case of ‘bonded labor’. Genicot (2002), in describing what she calls ‘the paradox of voluntary choice’, constructs a case where the strategic interaction between the landlord and the local credit institutions can constrain the poor peasant to ‘choose’ a bonded labor contract, whereas if bonded labour were banned it would have resulted in welfare-enhancing credit opportunities for the peasant. Basu (2000) models a
somewhat similar case of a woman choosing a ‘sexual harassment contract’ where she would have otherwise been better off if such contracts were disallowed. Similar cases can be argued for legally taking out the option for a poor worker to work in unsafe or hazardous conditions. These are all cases for interventionist regulations in the context of extremely unequal but ‘voluntary’ contracts.

Let us end with a comment on a fashionable attitude to the rule of law in the context of development that is sometimes expressed at the opposite end of the political spectrum. We have indicated earlier in this section as well at the beginning of this chapter that the rule of law is often an instrument in the hands of the propertied, ruling over and restricting the activities of the propertyless. This undoubted fact sometimes leads commentators to dismiss the rule of law merely as an instrument of class oppression or as part of a modernizing elitist project that rides roughshod over the ‘subaltern’. In the face of such tendentious simplifications we can do no better than to quote here from the far more nuanced historical analysis of E.P. Thompson. At the conclusion of his 1975 book, *Whigs and Hunters* (which shows how a political oligarchy in eighteenth-century England invented callous and oppressive laws to serve its own interests) Thompson writes:

> We reach, then, not a simple conclusion (law = class power) but a complex and contradictory one. On the one hand, it is true that the law did mediate existent class relations to the advantage of the rulers . . . On the other hand, the law mediated these class relations through legal forms, which imposed, again and again, inhibitions upon the actions of the rulers . . . In a context of gross class inequalities, the equity of the law must always be in some part sham . . . We ought to expose the shams and inequities which may be concealed beneath this law. But the rule of law itself, the imposing of effective inhibitions upon power and the defence of the citizen from power’s all-intrusive claims, seems to me to be an unqualified human good. To deny or belittle this good is . . . a desperate error of intellectual abstraction. (pp. 264–6)

**Notes**

1. For a discussion of the limitations of such exercises see Bardhan (2005, Chapter 1).
2. Even when the original patent is about to run out, the transnational company holding the patent often has various ways of effectively extending it: by slightly changing the composition of ingredients in the product and then taking out a new patent, bribing or intimidating the potential producers of the generic substitute, and through high-pressure advertisement keeping many of the customers hooked on to the original brand.
3. For a discussion of some of these issues see the papers by Gallini and Scotchmer (2001) and Kremer (2001).
4. For a review of the theoretical political economy literature on credibility of commitment see Bardhan (2005, Chapter 4).
5. For example, among developing countries many French legal origin countries are in Africa or Latin America and it may be standing as a proxy for other (unmeasured) deficiencies in state capacity in several of these countries.
6. Some of the pros and cons of relational contracting are empirically studied in the case of Vietnam's emerging private sector by McMillan and Woodruff (1999).

7. For a formal treatment of the subject see Dixit (2003).

8. The standard argument that 'irrational' behavior is weeded out in the evolutionary process is much too limited. Other-regarding cooperative behavior may be more successful in many cases. Evolutionary success in replication and the economist's narrow conception of efficiency may not go together if pay-offs to adherence to particular behavioral rules depend on adherence by others, or if there are positive and negative interactions of different behavioral rules.

References


Introduction
The New Institutional Economics (NIE), which has expanded rapidly since the mid-nineteenth century, has placed the issue of property rights at the forefront of its intellectual priorities. This is not surprising since, in the absence of well-established and well-enforced property rights, trade transactions may involve considerable transaction costs which have the effect of slowing market development. Perhaps to the dismay of those who believe in the absolute superiority of private ownership, a transaction-cost analysis aimed at assessing the relative efficiency of different property regimes does not necessarily point to the desirability of private property. In this chapter, we elucidate the kinds of circumstances under which private property is likely to dominate alternative property regimes, and this is done by limiting our attention to land and other natural resources (forests, lakes, pasturages, and so on). The demonstration proceeds in two steps. First, we compare a regime in which there are no property rights (the so-called open access regime) to alternative regimes in which such rights exist. Thereafter, we discuss the advantages and disadvantages of a regime of collective or community ownership vis-à-vis a regime of private ownership. The final section offers concluding remarks.

Open access versus property rights
A simple but basic principle taught by economics since Adam Smith is that a commodity, whether a good, service or asset, carries a low value if its supply is abundant relative to the demand expressed for it. The value can even be nil if the commodity is so plentiful that there is actually no competition between potential users. The next step is to recognize that, if a resource has no value, users have no incentive to seek to establish and acquire property rights which would assure them exclusive and free disposal of a portion of that abundant resource. The above proposition applies very well to land and other natural resources. Thus, in an environment characterized by low population density, access to land and natural resources is so easy that a portion of them can be occupied and exploited by individuals or collective entities without there being any need for them to defend against possible encroachments from outside.
Problems arise when an increasing number of individuals want to use a resource that is protected by no exclusive property rights acting as barriers to entry. More exactly, the regime under which no property rights exist (known as *res nullius*) begins to be problematic as soon as the entry of an additional user in the resource domain causes a damage to those users who were exploiting it before, and there is no penalty that this additional entrant is required to pay for such negative externalities. There are in fact two different sorts of externalities involved here.

The first type is known as congestion externalities. They manifest themselves in all the cases where the use of a natural resource necessitates that it is divided into distinct portions individually exploited by the users. When pressure on this resource increases, the activity of an individual user then unavoidably entails ecological costs for the other users, especially those located in the immediate neighbourhood. For example, a farmer who has cut off trees in the upper part of a hill in order to open a new field can cause erosion that will result in the silting of the lower parts.

The second type of externalities are rent-dissipating. They are observed in conditions where a natural resource is jointly, rather than individually exploited. And the problem arises only insofar as the resource is subject to decreasing returns, which implies that there is some degree of population pressure. In these conditions, through his additional harvesting effort, a new entrant causes a fall in the average productivity that is felt by all the previous users. A rational individual will enter into a resource domain, or continue to increase its activity therein, as long as the benefit he obtains exceeds the cost. What needs to be emphasized is that he will so act even though his income, or part of his income, is earned at the expense of previous users of the resource whose incomes are diminished as a result of his additional efforts. It is even possible that the additional effort of the new entrant does not give rise to any increase of output (the marginal productivity of effort is nil), yet the new entrant finds it individually profitable to apply this effort.

From the social viewpoint, such a decision is evidently undesirable since the costs thereby incurred are borne in vain, that is, they are entirely unproductive. A social waste of scarce production factors (labour, capital, fuel, and so on) occurs to the extent that the same aggregate output could have been obtained without applying additional amounts of these factors. The natural resource can then be said to be inefficiently managed. If the unproductive factors are used, it is because they enable the individual who owns them to participate in the exploitation of the resource and to obtain the average product which decreases as the number of harvesters rises. In the open access equilibrium, the resource rent is entirely dissipated (the costs eat the product), an absurd outcome since the natural resource is scarce.
(subject to decreasing returns), and should therefore have a positive value reflected in the rent.

In fact, the root cause of this inefficient use of a natural resource lies in the absence of property rights. Indeed, had property rights prevailed, the addition of an effort unit to be applied to the resource would have been decided by the owner only if the expected additional product exceeded the cost. In other words, it is the marginal product instead of the average product that would have been compared to the cost in order to decide whether an additional unit of effort must be applied. The marginalist rule guarantees the efficient use of a resource and, which amounts to the same thing, the maximization of the rent associated with it. It has this property because it allows a perfect internalization of externalities (see Demsetz, 1967; Alchian and Demsetz, 1973, for a first formulation of this well-known principle).

Dynamic losses must also be taken into consideration when assessing the pros and contras of a given property regime. More precisely, in the same way that he has no reason to be concerned with the damage caused to other users by causing a fall in their incomes, an individual harvester is not induced to forsake present benefits in order to ensure better the long-term conservation of the resource. For example, a fisherman has no incentive to throw back juveniles into the sea so that they can grow to mature size and be caught in their adult state at a later time. His reasoning is, indeed, that if he does not keep those juveniles when they are caught in his net, other fishermen will not hesitate to kill them with the result that they will anyway be unable to reach mature size. Had private property rights existed, the owner would have seen to it that conservation measures are adopted, since the future benefits of present sacrifices would have accrued to him rather than to other users.

Private property versus communal property

General considerations
In the absence of transaction costs, communal property allows rightsholders to internalize externalities as effectively as private property. Under this condition of zero transaction costs, the two property regimes are thus strictly equivalent (Platteau, 2000, Chapter 3). As a matter of fact, a group of people who own a resource will make exactly the same decisions as an individual owner regarding the amount of effort to be applied. Such an outcome results from the fact that, like an individual, a group seeks to maximize the rent or the surplus, that is, the difference between the value of the flow of produce extracted from the resource and the capital and labour costs. In both cases, the marginalist decision rule is applied.
But the assumption of zero transaction costs is totally unrealistic: it is useful to set a theoretical reference point, but not to describe reality. Once transaction costs are taken into account, the members of the property rights school contend, private property appears superior to any system of collective or communal property. The underlying argument can be summarized as follows: while an individual owner forms an organically integrated decision unit (he need not discuss with anyone else than himself in order to reach decisions regarding the use of the resource), so as to create a one-to-one relationship between individual actions and their effects, a collective owner must achieve an agreement between its members to decide how to use it. An agreement will not necessarily emerge from intra-group discussions. And, even if an agreement can be eventually reached, the negotiation process will necessarily entail non-trivial costs, such as the opportunity cost of the time spent in meetings, transport expenses, the costs of communicating the time and place of the meeting, and so on (Baland and Platteau, 1998b; Platteau, 2000, Chapter 3). Lastly, assuming that an agreement has been found and that the details of its modus operandi have been worked out, there remains the delicate question as to how it will be effectively enforced. At the very least, one may fear that its implementation will not remove all the possible sources of inefficiency, something which single private ownership is apparently capable of achieving. In the following, all the costs plaguing group ownership will be referred to as governance costs.

The determinants of governance costs
The feasibility of an agreement regulating the use of a natural resource at village level is itself dependent on a certain number of factors, among which the size of the user group and its degree of heterogeneity stand foremost. Regarding the first factor, it is evident enough that the smaller the number of rightsholders the lower the negotiation costs involved in the process of devising the regulatory agreement. If their number is too high, on the contrary, one may fear that no regulatory mechanism will be put into place for lack of an internal agreement. The impact of heterogeneity is also evident. Thus, it is easy to understand that divergences between group members regarding the intended uses of a natural resource will make an agreement more difficult to reach. Illustrations are numerous, as attested by the prolonged conflicts between farmers and herdsmen around land areas claimed by each category for their own specific purpose. Herdsmen want to maintain their customary rights to large grazing areas at least during a part of the year (so that animals can feed themselves on crop residues after harvest time), while farmers are increasingly eager to win exclusive rights over well-delineated zones so as
to be able to practice more intensive forms of agriculture requiring continuous cultivation and long-term land improvements.

Income or wealth inequality constitutes another form of heterogeneity that tends to make regulation of the use of natural resources difficult to achieve. It has indeed been shown that the more unequal the distribution of income between members of a village community or any user group, the harder it is to find a regulatory scheme that satisfies all the people concerned. Moreover, if such a scheme exists, the efficiency gains that it will yield compared to a situation with no regulation diminish as income distribution becomes more unequal (Baland and Platteau, 1998a; 2003). The underlying intuition is simple: when there is a need to regulate the use of a resource, the group involved must not only determine the extent to which the intensity of use must be reduced to approach efficiency, but also the manner in which the effort reduction will be shared among the various users. If users are relatively identical, the latter problem is unlikely to be serious: a uniform reduction of individual effort levels appears as the natural solution. If, on the contrary, the users are different in terms of wealth or income and these differences are reflected in different rates of resource use, the problem of sharing the burden of effort reduction obviously becomes more complicated. In particular, the efficient solution might well imply that the larger part of this burden be borne by the less productive or the more impatient (those more preoccupied with subsistence constraints) users who often are the poorer members of the community.

Let us nonetheless assume that an agreement can be found that entails efficiency gains and allows each resource user to improve their situation compared to the status quo state of no regulation (only access rights exist). There remains the tricky issue of the enforcement of the regulatory scheme: once the rules are decided and agreed upon, each user has an incentive to violate them. The problem is especially serious because, by rendering the resource more valuable than before, effort restriction has the effect of increasing the benefits which can be obtained by exceeding one’s allowed quota while other users follow the rule. To put it in another way, the outcome of the agreement is to enhance the temptation for individual users to free-ride on the sacrifices incurred by fellow users.

In addition, again assuming that an agreement is feasible, there is the question as to who will bear the costs of formation of collective action, understood as the costs involved by the very process of creating collective mechanisms for both decision-making and enforcement. Again, the characteristics of the user group – its size and heterogeneity, in particular – influence the extent to which this problem can be surmounted (Baland and Platteau 1997; 2003). To begin with, the impact of group size is identical to
the one observed when the problem consists of devising a regulatory scheme: the smaller the size of the community, the more likely the costs of formation of collective action will be actually incurred and, therefore, the more likely the agreed rules will be designed and applied. The main argument here is known as the incentive dilution argument (Olson, 1965). Each individual makes a personal calculation when he decides whether or not to contribute to the production of a local public good (or, in our context, to the creation of a collective mechanism, on the one hand, and to the implementation of a regulatory scheme, on the other hand). The individual compares the gains from abstaining from contributing to the collective effort with the cost. When a group is smaller, the cost of withdrawing participation obviously rises relative to the benefit.

In addition, when a group is smaller, members tend to know each other better and, therefore, reputation effects are more important. More attention will be paid to the future consequences of opportunistic behaviour in order to avoid punishment in the form of exclusion from the group or denial of the right of access to the resource. Furthermore, not only is communication facilitated within a smaller group, but the formation of collective identity feelings is also easier and, as a consequence, individuals are more induced to take into account the effects of their decisions on the other members (Baland and Platteau, 1996, pp. 75–8).

The impact of heterogeneity is more ambiguous than the impact of group size. In fact, it cannot easily be predicted a priori. Let us examine, in particular, the impact of inequality of income or wealth on efficiency in the production of a public good, such as the formation of a regulatory body or framework. Two effects are at work which run into opposite directions. On the one hand, a great inequality creates a situation in which the rich guy internalizes a large part of the externalities created by his particular contribution to the public good, thereby inducing him to apply the required effort. Yet, on the other hand, those who are at the lower tail of the income distribution find themselves in exactly the opposite situation: they will draw only minor benefits from the collective good and hence they will have weak incentives to contribute.

It is thus impossible to predict in a general manner whether a higher degree of income inequality will actually result in an increase or a decrease of the aggregate amount of contributions to the local public good. The aggregate amount will rise only if the increased contributions of the village elite (who better internalize the externalities) exceed the reduced contributions of the common people. What is certain, however, is that an extreme inequality corresponding to a total concentration of all the wealth in the hands of a single villager will lead to an efficient provision of the public good (see Baland and Platteau, 2007).
The evolution of governance costs
As is evident from the above discussion, communal property gives rise to serious incentive problems, especially when it involves the regulation of the use of natural resources besides the setting of access rules. A reasonable position therefore consists of admitting that inefficiencies are bound to persist under this ownership regime, whether in a static form (the dissipation of part of the resource rent) or in a dynamic form (the lack of investment to conserve the resource, including actions to fight against predators who threaten its long-term stock).

A straightforward consequence of such a situation is the following: the people’s ability to cooperate in the management of common access resources determines the profitability of jointly held resources compared to their profitability when they have been individualized, and it also influences the allocation of resources between various uses. Inasmuch as this cooperation ability varies from one area to another, one must expect to observe geographic variations in the uses and rates of profitability of local-level natural resources. For example, it has been shown that in Mexico, when cooperation fails in the management of collectively grazed pastures, more land is allocated to crops than under successful cooperation and less to pastures, while the stocking rate on pastures is increased. This results in too much land in extensive crops and too many animals per hectare of pasture (McCarthy et al., 1998).

The point that we want to make now is that efficiency losses caused by externalities are likely to grow with the value of the resource, hence the frequent emphasis in the literature on the unit value of natural resources as one of the main determinants of its privatization (division). To illustrate, in his classical study of the Swiss Alps, Netting contrasts the lowlands of the valley which are fertile and therefore tend to be privately appropriated with the more arid highlands which are used as communal (summer) pastures under the authority of the village council (Netting, 1976; 1981).

Population pressure bears upon the efficiency gains of division in two different ways. For one thing, by increasing the number of users per unit area, it creates more room for external effects and, thereby, the governance costs and the inefficiencies involved in the joint exploitation of the resource rise. For another thing, by making the resource increasingly scarce, population pressure enhances its value and therefore makes for increased aggregate losses from collective exploitation. In other words, the amount of the rents foregone by not dividing the resource tends to increase with population. This is especially true when population pressure involves a transformation of the pattern of resource use, such as a shift from extensive to intensive agricultural or grazing practices, since intensive practices have the
effect of increasing the potential value of the resource per unit area and thus enhance the gains of private property.

Market penetration and the ensuing commercialization of products from primary activities is another critical determinant of the relative profitability of private property. It is by enhancing the realizable value of natural resources that growing integration of rural communities into developing market networks increases the benefits of resource division. Thus, in many developing countries, dramatic increases in prices for fuelwood or fresh fish as a result of a rapid expansion of urban markets have prompted rural inhabitants to intensify the exploitation of many forest and fish resources during the post-independence period. Significant efficiency losses have resulted from the growing pressure on these resources and the rising incidence of negative external effects that have accompanied it (see, for example, Baland and Platteau, 1996, pp. 262–70).

At this stage of our analysis, it appears that two series of factors bear upon the efficiency of regulated communal property compared to that of private property. On the one hand, there is the cooperation ability of the resource users, which is itself determined by the size and the degree of heterogeneity of the group or community to which they belong. On the other hand, there is the degree of scarcity of the resource as reflected in its value, which is determined by: (1) the intensity of population pressure in the area; and (2) the extent of market integration. A third series of factors influences the extent of efficiency losses resulting from the collective exploitation of a resource or the extent of the potential efficiency gains of its privatization. This last set of factors bears upon the costs of privatization, understood as a process of division of a jointly held resource accompanied by its partitioning into individually held portions. In the remainder of this chapter, we focus our attention on the two main costs of privatization, namely direct transaction costs and opportunity costs.

*The role of direct transaction costs*

When the problem of choice of ownership regime is considered from the exclusive standpoint of governance costs, the balance sheet is unmistakably favourable to the division and private appropriation of local-level natural resources. But this is only one side of the balance sheet. So far, indeed, we have implicitly assumed that the establishment and protection of property rights are costless operations. Such an assumption is manifestly unrealistic and we need to remove it now. What appears then is that, compared to communal property, private property is costlier from the viewpoint of direct transaction costs, which primarily include set-up and protection expenses. There thus exists a trade-off between two kinds of costs that have a different impact according to the ownership regime considered: the governance costs
that affect communal but not private property, on the one hand, and the
direct transaction costs that are smaller under the former than under the
latter regime, on the other hand.

The latter conclusion follows from the fact that it is less costly to fence,
demarcate and protect a territory of a given size than to do those things for
divided portions of that territory. As a corollary, when the surface area of
a resource domain is larger, the per capita direct transaction cost of priva-
tizing it increases. To put it in another way, direct transaction costs increase
with the physical base of the resource: the more spread the resource base
(or the less concentrated the resource) the higher the costs of delimiting and
defending the resource territory. Other things being equal, therefore, the
more spread a resource is, the less profitable it is to privatize it, and the more
compact the resource, the more attractive is its privatization.

A consequence of the above is that we expect private property rights to
be established over high-density resources and communal property rights
over resources with the opposite characteristic. It is good news for eco-
nomic theory that such a prediction is systematically verified in reality. Yet,
at the same time, one must reckon that practically it is often difficult to dis-
entangle the impact of the density of a resource from that of its value on
the probability of privatization. As a matter of fact, high-value resources –
for example, fertile lands that are susceptible to being irrigated owing to
their favourable location – tend to be divided more often than low-value
resources – for example, semi-arid lands that are hardly suitable for any
other purpose than extensive grazing. In other words, there exists a strong
correlation between the density or compactness of a resource and its value.
In fact, in many instances the second characteristic causally determines the
first one. What empirical evidence reveals is that compact resources with a
high value (for example, intensively cultivated lands or fertile lands located
near an important market town) are generally held under private property
while resources that stretch over large areas and carry a low value (for
example, the immense low-quality grazing areas in Mongolia or the Maasai
Mara in Kenya) are jointly held by a local user group or community. Our
analytical argument simply shows that these two correlated characteristics
of a natural resource – its high value per unit area and its high density –
tend to make its privatization relatively profitable.

Two remarks are in order. First, there exist natural resources, the division
of which would entail prohibitively high direct transaction costs under the
present state of technology. For example, the open sea – or, more exactly,
the fish stock contained in it – presents insuperable difficulties for private
appropriation. The enforcement of exclusive property rights to individual
patches carved up in the ocean would, indeed, be infinitely costly. This is
especially evident when fish species are mobile and move over large water
spaces, since exclusive rights are too costly to establish and enforce whether over the resource or the territory in which it moves. The example of wildlife reserves also comes to mind.

Interestingly, even in the case of maritime fisheries, privatization may sometimes be a viable solution. This tends to happen when species are rather sedentary (for example, lobsters, shellfish, molluscs, seaweeds) and live in relatively compact and well-delimited spaces, such as when the fish are found around islands (the Pacific islands, the Shetlands in Scotland, and so on) or in relatively well-sheltered aquatic zones (for example, in deltas or in backwaters) where fishing locations can be easily demarcated and protected against external encroachments. In these conditions, fishing spots are frequently assigned to individuals or families for their exclusive use and these private rights can generally be inherited by future generations as long as they are used effectively (see Platteau, 2000, p. 85, for references to the anthropological literature).

Second, the direct costs of resource division are not exogenously fixed. In the above, we have considered that these costs are determined by the inherent characteristics of each resource, it being understood that a resource may take on various forms and characteristics depending on the precise location and environment in which it is found. Here, we want to point out that, in fact, direct transaction costs may fall with technological progress. One well-known example is the discovery of the barbed wire which proved to be a decisive step in the reduction of the cost of protecting property rights through cheaper fencing of agricultural fields (North, 1981). Another example is the introduction of modern borehole drilling facilities in arid and semi-arid areas where this has the effect of facilitating the privatization of common grazing areas. Before this invention, in a country like Botswana, water extraction was subject to important scale economies as a result of which grazing lands were always the collective property of herders’ communities (Peters, 1994). Private appropriation is thus not only facilitated by factors which contribute to enhance the value of a resource, but also by factors which have the effect of reducing the direct cost of partitioning.

The role of opportunity costs

Two types of opportunity costs appear to play an important role in this respect: scale economies and insurance benefits associated with collective ownership.

Let us first consider the impact of scale economies. Resources offering multiple products tend to be subject to scale economies to the extent that they form part of an overall ecosystem. This multiple product character of the resource is a reason often mentioned to argue against the parcelling out
of forest areas into individual holdings. In the case of hunting, on the other hand, wild animals require large territories to survive and reproduce, so that division of a hunting domain into smaller parcels would imply the destruction of the resource.

When discussing the role of scale economies, it is important to bear in mind that they may be present not in the resource itself but in complementary factors. The obvious advantage of coordinating the herding of animals so as to economize on shepherd labour in extensive grazing activities is probably the best illustration of the way scale economies in a complementary factor may prevent the division of a resource domain. Another illustration can be taken from fisheries (maritime or inland). In many cases, indeed, the guarding of privately apportioned fishing spaces is subject to strong scale economies. As a consequence, it may make sense for several individuals or families to get together to enforce a de facto right of collective property over a given fishing ground. This is actually what many traditional fishing communities have done in the past when competition around scarce fish resources began to develop, particularly in inland fisheries.

Opportunity costs of privatization also come into the picture when returns to a resource are highly variable across time and space. The need to insure against such variability is then a consideration that militates against resource division (McCloskey, 1976; Dahlman, 1980). When a resource has a low predictability (that is, when the variance in its value per unit of time per unit area is high), indeed, users are generally reluctant to divide it into smaller portions because they would thereby lose the insurance benefits provided by the resource kept whole. In the words of Nugent and Sanchez: ‘the lower the quality of land or the more variable the weather, the more important it is that the land be held in communal, that is, tribal form’ (Nugent and Sanchez, 1993, p. 107).

The example of extensive grazing and also that of maritime fishing again provide us with good illustrations of the above. Herders (fishermen) typically need to have access to a wide portfolio of pasture lands (fishing spots) insofar as, at any given time, wide spatial variations in yields result from climatic or other environmental factors. Assuming that the probability distributions are not correlated too much across spatial groupings of land or water and that they are not overly correlated over time, a system offering access to a large area within which rightsholding users can move freely appears as highly desirable from a risk-reducing perspective.

It must nevertheless be pointed out that private property rights over portions of the resource could apparently solve the predictability problem. Consider the case of extensive grazing again. Due to the unpredictability of rain-induced growth of grasses within any small region, what is important for herders is to be able to move over large ranges of land and rapidly
change location when the need arises. By holding exclusive ownership rights over widely dispersed patches of pasture lands, they would therefore achieve their objective of risk reduction. Clearly, to account for the maintenance of communal property, one needs to appeal to transaction-cost considerations and the high exclusion costs of a spread resource base (Platteau, 2000, p. 88). Because it would be prohibitively costly to enforce exclusive rights over widely dispersed and infrequently visited ranch patches, the division of the resource domain turns out to be infeasible. The same situation actually obtains in many fisheries.

The additional (direct) transaction costs implied by the necessity to insure against income fluctuations in the event of division or privatization of a resource increase with the variability of incomes and the surface area of the domain of this resource. As a matter of fact, the higher the variability of incomes the larger the number of resource portions that a particular user needs to insure himself and, hence, the higher the costs of establishing and protecting private property rights. On the other hand, the more stretched is the resource base the higher the (direct) transaction costs caused by the demarcation and the guarding of a given portion of the resource.

**Conclusion and final considerations about the evolutionary approach to institutions**

Two central conclusions emerge from our analysis. First, the transaction-cost economic theory of institutions leads us to expect that an increasing number of village-level natural resources will be divided and individually held as they acquire more value under the combined impact of population growth and market penetration. The important role of governance costs associated with collective ownership goes a long way toward explaining this gradual shift from corporate to private forms of ownership. Nonetheless, and this is our second conclusion, certain resources possess characteristics that make their division and their private appropriation especially costly. The costs involved are those required to establish and protect private property rights (direct transaction costs), or opportunity costs resulting from the loss of benefits provided by communal property.

The second conclusion raises a thorny issue. Indeed, in order that communal property be viable in conditions where private property is infeasible owing to prohibitively high direct transaction costs, or undesirable owing to high opportunity costs, it is essential that governance costs remain within tolerable limits. If this condition is violated, the natural resource concerned will not be regulated in a satisfactory manner and efficiency losses will be significant, perhaps considerable. It will either become a non-regulated common property (meaning that it is characterized by access
rules while rules of use are absent), or it will fall under the open access regime. If the number of users is large, these last two regimes will produce more or less equivalent results in the form of grave inefficiencies both in the static (rent dissipation) and in the dynamic senses (destruction of the stock following a lack of conservation investments or a lack of control of the extraction efforts applied by users).

**Note**
1. ‘A primary function of property rights is that of guiding incentives to achieve a greater internalization of externalities.’ (Demsetz, 1967, p. 348).

**References**
Introduction
Are some cultures more prone to development than others? Does economic development require cultural change? Which is more important, economic development or cultural integrity? These are among the core questions that swirl around the debate on the relationship between culture and development. The debate was initiated in its modern form by Max Weber’s classic work, *The Protestant Ethic and the Spirit of Capitalism* (1958), and it has continued ever since. Social scientists past and present have staked out positions on all sides of the issues. Traditional modernization theorists take it for granted that economic development is primary, but they differ on whether cultural change leads development or vice versa. Some see culture functioning as either an instrument of or an obstacle to development, while still others see culture as irrelevant. Post-development localization theorists, on the other hand, disparage modern economic growth and argue that culture is paramount. They are highly resistant to Western-style corporate-led globalization, which they believe destroys local culture. Inbetween are many intermediate positions, like those who hope to maintain cultural integrity while also allowing for the development of market economies, the expansion of trade and economic growth.

Historical background
To begin the exploration of the role of culture in development, it is worthwhile to review briefly the rise of the modern economy in the Western world. Two facts stand out from an examination of the history of modern capitalist development. First, capitalism has succeeded in producing quantities of goods and services unprecedented in history; second, it has done so in a temporally and spatially uneven manner. The economies of some nations take off into a self-sustaining growth, with other nations eventually catching up. Still others, the loosely called ‘developing countries’, seem to be left hopelessly behind. This pattern occurs across nations as well as across regions within nations.

One of the great economists of the twentieth century, Joseph Schumpeter, captures this dynamic process in his concept of ‘creative destruction’: ‘The fundamental impulse that sets and keeps the capitalist engine in motion comes from the new consumers’ goods, the new methods
of production or transportation, the new markets, the new forms of industrial organization that capitalist enterprise creates . . . [These developments] incessantly revolutionize the economic structure from within, incessantly destroying the old one, incessantly creating a new one’ (Schumpeter, 1950, p. 83).

In both industrial and poor countries this creative-destructive process of development has created socio-political tensions both because of its uneven nature and because of its challenge to traditional values and ways of life. Today this is true particularly in countries with strong Islamic roots and in countries where readily identifiable minorities control the wealth of the society.

A simple policy of free markets and free elections may not advance the welfare of the poor in the world, and it must be remembered that the Western world did not develop that way itself. Development was slow, its major transformation spread over more than a century. For much of that period democracy was limited and countervailing policies were gradually enacted to curb the divisive effects of growing inequality.

Most historians, whether their orientation is political, cultural or economic, recognize that the eighteenth century was a turning point in the nature of the Western world that saw momentous movements and events – intellectual, political, military, social, cultural and economic. The Enlightenment, with its emphasis on reason, natural law and progress, and its avant-garde – the philosophes and physiocrats – opened new vistas, even though most of the population of Europe and the remainder of the world scarcely glimpsed those vistas.

In economic affairs, the eighteenth century began with Francois Quesnay’s campaign against mercantilism and ended with the completion of the campaign by Adam Smith. In the process, the classical school of economics, a new social science, came into being. Finally, in the course of the century, the agricultural and commercial revolutions of the previous two centuries initiated the Industrial Revolution in England that formed the basis of our modern economies.

Now that self-regulated market capitalism had arrived, what were its characteristics? In its textbook purity, a capitalist market economy is controlled, regulated and directed by markets alone. Socially beneficial outcomes in the production and distribution of goods are entrusted to this self-regulating mechanism, based on the expectation that human beings behave so as to achieve money gains.

In contrast, during preceding historical periods (and even today in many of the poor countries), markets were never more than accessories of economic life. Instead, the economic system was embedded in the social-cultural system. In places like Babylonia and Greece the local markets
(trading centers) were compatible with the established social way of life; markets did not expand at the expense of the society. Even under the mercantile system of the previous two centuries, where markets had expanded to involve a large part of the nation, they were not free markets, for they were subjected to centralized administration. Karl Mannheim argued that the move to self-regulating markets entailed a transformation from a regulated and socially controlled mechanism into the very organizing principle of society itself (Mannheim, 1950, p. 191).

Moreover, such a market-first institutional pattern cannot function unless other aspects of a society’s life are subordinated to its requirements, which is what happened over time in today’s developed economies. A market economy can only exist in a market society, and the requisite process of social, cultural and institutional change evolved in conjunction with the transition to a market economy. Nations wanting to catch up economically are thus naturally led to the question of how such a market society can be created in countries far removed from Western culture and in a time span shorter than the century or two that was required in Europe. On the heels of this question, another follows close behind, which is whether the creation of such a market society should even be a goal of development. Needless to say, there is substantial disagreement over these questions.

Nailing down a working definition of culture is itself problematic, in part because culture is easier to identify in others than it is in oneself and one’s own society, just as most people think others speak with accents, not themselves. That in itself makes talking about, and especially evaluating, culture potentially invidious, because in discussing cultures people are inevitably discussing the ways of life of others. Going way beyond people’s culinary, musical preferences, dress and traditions, Rao and Walton understand culture to be:

about relationality – the relationships among individuals within groups, among groups, and between ideas and perspectives. Culture is concerned with identity, aspiration, symbolic exchange, coordination, and structures and practices that serve relational ends, such as ethnicity, ritual heritage, norms, meanings, and beliefs. It is not a set of primordial phenomena permanently embedded within national or religious or other groups, but rather a set of contested attributes, constantly in flux, both shaping and being shaped by social and economic aspects of human interaction. (Rao and Walton, 2004, p. 4)

Culture is thus about our deepest beliefs, values, sense of identity, ways of life and longings, which makes it unsurprising that the discussion of culture and its significance for development generates controversy.

For economists the interest in culture has centered on its support of traits that contribute to economic growth, that is, thrift, hard work and
reinvestment by the middle class; hard work, obedience and contentment for the working class. The key is to discover the historic role of culture in generating the capitalist spirit of entrepreneurship, which for Schumpeter is so basic. For example, what were the psychological conditions – the capitalist spirit – that accompanied and aided the development of a thriving capitalist economy? Greed and the pursuit of riches are nothing new. Money-lending, commercial trading, piracy, plunder and other forms of unrestrained avarice are as old as history. But a way of life based on the rational, calculated pursuit of pecuniary profit through Smith’s ‘truck, barter and exchange’ and its organization into an economic system using free wage labor is a modern phenomenon.

It was only after centuries of struggle that capitalism established its claim to legitimacy, for it involved a code of economic behavior and a system of human relations sharply at variance with traditional religious customs and values. Originality, self-confidence and tenacity of purpose were required to initiate and carry on this struggle. This was the role of entrepreneurs. They emerged partly because changing economic conditions helped the Reformation succeed and helped shape the development of new theologies and creeds. In turn, the emerging religious beliefs helped direct and shape the subsequent economic development. Economic reasons alone are insufficient to account for the extraordinary power of entrepreneurship and rational profit-seeking in the modern world.

Alternative views on the role of culture in development

How does this historical experience of the rise of capitalist economies in the now developed world impact upon our theory and practice of development today? Mainstream views of development today continue to follow in the modernization tradition that arose in the aftermath of World War II. On this view the primary measure of development continues to be a self-reinforcing tendency to economic growth and material progress. It is from this basic developmental reality that other positive features of a good society spring, like better health care, improved education and democratic governance. As Benjamin Friedman says at the close of his recent moral defense of economic growth: ‘Only with sustained economic growth, and the sense of confident progress that follows from the advance of living standards for most of its citizens’, can a nation hope to achieve ‘an open, tolerant, and democratic society’ (Friedman, 2005, p. 436). Over time it has become clear that an equitably growing economy requires financial capital, the adoption of efficient technologies, investment, human capital, entrepreneurship and market-promoting policies and institutions, which begs the question of how to get nations to build and use these necessary attributes. For many theorists who see a connection between culture and development,
inquiries into this question lead directly to issues of how cultural adaptation similar to that which occurred historically in the West can be facilitated so as to accommodate the needs of modern market economies throughout the world.

On the other hand, it needs to be said that many economists believe that culture is essentially irrelevant to economic development. Instead, the road to development is paved with market-promoting economic policies and institutions. Hernando De Soto, for example, argues that efforts ‘to explain why capitalism fails outside the West remain mired in a mass of unexamined and largely untestable assumptions labeled “culture”, whose main effect is to allow too many of those who live in the privileged enclaves of this world to enjoy feeling superior’ (de Soto, 2000, p. 225). Along with economist Mancur Olson, De Soto argues that people of all cultures respond similarly when property rights and contract-enforcing institutions are established (Olson, 2000). In his high-profile treatment of twenty-first-century development challenges, Jeffrey Sachs calls the cultural thesis a myth, saying that cultures often follow rather than lead economic change and that culture-based arguments ‘are usually made on the basis of prejudice rather than measurable evidence’ (Sachs, 2005, p. 317). In his critique of Lawrence Harrison’s emphasis on child-rearing, for example, Sachs points out that children are taught the value of hard work more consistently in Nigeria, South Africa and Tanzania than they are in the United States (Sachs, 2005, p. 318). For William Easterly too, the issue is not culture, but the proper structuring of incentives (Easterly, 2006). Summarizing these lines of thought, David Throsby points out that: ‘mainstream texts in economic development have no time for culture; taking three such texts more or less at random, an inquisitive reader can find no reference to culture in the subject indexes of any of them’ (Throsby, 2001, p. 67). For such theorists, it is no surprise that nations of greatly diverse cultures, like Ireland, South Korea, France, the United States, and now China and India can all enjoy the fruits of economic progress without fundamental changes to their cultures.

And yet there are perhaps a greater number of development theorists, like Weber, who find culture to be of central importance. For many of the early thinkers in this tradition, like economist Bert Hoselitz and sociologist Talcott Parsons, transitions from traditional to modern patterns of life required nothing short of major cultural overhaul. Major efforts to discover potential sources for such cultural change led David McClelland to identify the significance of a people’s psychological ‘need for achievement’, what he calls ‘n-achievement’. McClelland argues that modern societies have been built by innovative entrepreneurial types with high n-achievement. N-achievement, however, is so deeply embedded in people’s psyches, which develop gradually
during child-rearing and other social and cultural nurturing practices, that he despairs of offering practical advice for how actually to promote development. He notices, for example, that religious and ideological changes, like the rise of Protestantism in some Mexican communities, were associated with a rise in n-achievement. But he is doubtful this awareness of the relationship will itself lead to acceptable development policies. And though he believes education might be of some help, he doubts it can make a large contribution, because people's personalities are too deeply formed by prior and ongoing child-rearing practices. Ultimately, the best he can do is to encourage policies that facilitate the interaction among entrepreneurs from developed countries with scarce but vital entrepreneurial types engaged in business in ‘underdeveloped countries’ (McClelland, 1961).3

In a similar vein, Everett Hagen, recognizing both the importance of entrepreneurship and the role of psychological formation and traditional cultural patterns in creating people resistant to progressive change, identifies one major source of change to be a socially deviant group that finds a psychological outlet in violating traditional patterns of life and thus creates its own identity through entrepreneurial change and success in business. Subdominant or threatened minorities, or marginalized but progressive immigrant communities, might be able to provide an impetus to change that breaks through traditional cultural patterns and points the way to modernity (Hagen, 1962).

Still today, many contemporary theorists within the modernization tradition continue to focus on the fundamental importance of cultural change. Perhaps most well known among these is Lawrence Harrison, who after long experience in development work in Latin America has developed his version of the cultural thesis in a series of books written over a twenty year period starting in the mid-1980s.4 According to Harrison, the basic thesis is that ‘values, beliefs, and attitudes are a key but neglected factor in understanding the evolution of societies and that the neglect of cultural factors may go a long way toward explaining the agonizingly slow progress toward democratic governance, social justice, and prosperity in so many countries’ (Harrison, 2006, xiii). In Harrison's view, progress in such basic areas as life, health, liberty, prosperity, education and justice depends on the adoption of a democratic capitalist way of life, which in turn depends on cultural orientations.

Following many of his forbears who have developed lists of the contrasting cultural characteristics of traditional and modern societies, Harrison too identifies cultural traits that either inhibit or advance progress. Over the years, his list has expanded to 25 core cultural traits that make societies either progress-prone or progress-resistant (Harrison, 2006, pp. 36–7). These include religious orientations (for example favorable or
non-favorable attitudes toward material pursuits), values (for example how flexible a society’s ethical code is), economic behavior (for example whether people have entrepreneurial inclinations) and social behavior (for example the radius of trust). Societies that value competition instead of fearing it as a threat to equality, for example, are more likely to progress, as are those that focus on success in this world over their place in the next. If progress is to come about in poor countries, then it is culture that must ultimately be changed.

Harrison is the name most often associated with the cultural thesis, but he is not alone. David Landes, in a sweeping study of development patterns across the world, both historical and contemporary, is drawn toward cultural explanations, ultimately assigning a major causal role to culture. Landes argues that: ‘just because markets give signals does not mean that people will respond timely or well. Some people do this better than others, and culture can make all the difference’ (Landes, 1999, p. 522).

Recent studies lend the cultural thesis some degree of empirical support. In a study of various factors that influence growth, David Weil finds that openness to new ideas, an inclination to work hard and to save, and the level of trust play a significant role in explaining economic growth (Weil, 2005, p. 427). Other studies analyze World Values Survey data with a special focus on the role of religion and find that religious faith and traditions definitely matter. Luigi Guiso et al. conclude that: ‘on average . . . religion is good for the development of attitudes that are conducive to economic growth’, and that ‘on average, Christian religions are more positively associated with attitudes that are conducive to economic growth, while Islam is negatively associated’, and that Protestants and Catholics have different mixes of positive and negative factors’ (Guiso et al., 2003, p. 280). In neither the Weil nor the Guiso study is the direction of causality firmly established. In an attempt to address this shortcoming, Robert Barro and Rachel McCleary find that religious beliefs, especially as regards the existence of heaven and hell, seem to play a causal role in the achievement of higher economic growth. Like Weber’s theories about the role that Calvinism played in Europe’s drive toward capitalism and industrialization, Barro and McCleary conjecture that such ‘religious beliefs stimulate growth because they help to sustain aspects of individual behavior that enhance productivity’, like thrift and a greater work ethic.5

**Social capital**

Another line of thought that has arisen since the 1990s is focused on the causal role played by social capital and civil society. In the economics literature, social capital has come to mean that ‘social relations’ are important factors in the economy. Francis Fukuyama and Robert Putnam have each
studied the extent to which social relations promote group cooperation, civil society, good governance, trust and productive economic activity. Fukuyama emphasizes interpersonal trust as a key cultural aspect, arguing that: ‘one of the most important lessons we can learn from an examination of economic life is that a nation’s well-being, as well as its ability to compete, is conditioned by a single, pervasive cultural characteristic: the level of trust inherent in society’ (Fukuyama, 1995, p. 7). In a similar vein, Putnam shows that the cultural predilection to work together cooperatively in civic groups is a major well-spring of democratic governance and economic well-being (Putnam, 1993). Fukuyama and Putnam both argue that cultures that foster trusting working relationships outside of narrow family interests, referred to by Fukuyama as ‘weak ties’ and by Putnam as ‘bridging capital’, are much more prone to establish successful democratic capitalist societies.

The recent attention to social capital, combined with the renewed focus on religious beliefs, has given rise to the concept of spiritual capital, which refers to the spiritual or religious resources that contribute to a well-functioning community. Putnam argues that religion is by far the largest generator of social capital in the United States, contributing to more than half of the social capital in the country. For developing countries, where religious commitments are generally stronger than they are in the economically developed countries, religious sources of social capital may be even more important. Theodore Malloch claims that: ‘In the ultimate sense spiritual capital is the missing leg in the stool of economic development, which includes its better known relatives, social and human capital’ (Malloch, 2003, p. 2).

Social capital studies have become common in development economics due, at least in part, to the World Bank which has been working on the concept since the 1990s. On one of its websites it says that: ‘Social capital refers to the norms and networks that enable collective action. Increasing evidence shows that social cohesion – social capital – is critical for poverty alleviation and sustainable human and economic development.’ Still, the extent to which social capital and spiritual capital are intellectually valid conceptual categories for economic development is carefully explored in a World Bank-supported volume edited by Anthony Bebbington et al. The presumption throughout the book is that the character of a society’s social capital is a culturally defined reality that is central to the empowerment, participation and inclusion of people in the development process. The purpose of the book is to study efforts at the World Bank, populated as it is with technically trained economists, to figure out ways to include social capital in their analysis and in their programming (Bebbington et al., 2006).
In the hands of economists, social capital is typically integrated into neoclassical microeconomic theory, making it a behavioral property of individual actors. The impact of social capital is thus seen to flow from the rational choices of those individuals. The most common ways to model social capital are: (1) as a preference in a utility function; (2) as an individual resource owned by persons or firms; and (3) as an instrument to reduce risk.

Against the instrumentalist view of culture
The neoclassical tendency to consider social capital as a resource for development has generated some strong criticism. Using the term ‘capital’ is seen as misleading and even ideological, and the empirical support is also seen as weak and circular. As Van Staveren and Knorringa argue: ‘one of the most central flaws . . . is a circular explanation of social capital: a group’s success is attributed to its social capital, but social capital is measured by group success’ (Van Staveren and Knorringa, 2007, p. 110). Moreover, if social capital is a resource for development, then social relations are stripped of their own value and become nothing more than instruments in the service of economic growth. One of the troublesome implications of such a view is that cultures can thus be judged worthy or unworthy, good or bad, based on their ability to generate economic growth. Not surprisingly, such invidious views are not readily accepted by people who have learned to value their own cultures for more than their pecuniary potential.

Amartya Sen, for example, argues that a focus on whether cultures are ‘good or bad’ fosters prejudicial attitudes, sometimes leading to a ‘blame the victim’ mentality that can cause great harm. English responses to Irish famines in the nineteenth century, for example, were thought by the British to be the result of Irish cultural deficiencies, which, if true, meant that direct assistance would only exacerbate the problem. What the Irish really needed were civilizing influences, a cultural makeover. Thousands died. By contrast, economic downturns in England were seen by the British as resulting from events beyond the control of the citizenry. Fast forward to the present day and one finds Catholic Ireland’s economy growing much faster than Protestant England’s. The same is true for India, which with its caste system and other-worldly religious views, was always assumed to have a culture resistant to development. Like China and Ireland, India is now among the fastest-growing countries in the world. Did their cultures change? Or was something else at work? (Sen, 2004, pp. 37–58).

Even some neoclassical economists have trouble with this tendency to instrumentalize culture. Oliver Williamson, responding to the definition of trust as rational expectations of the behavior of others, says ‘calculativeness will devalue the [social] relations’ because it ‘may well be destructive of
atmosphere and lead to a net loss of satisfaction between the parties’ (Williamson, 1993, p. 481).

Such disrespect for local culture is even more vehemently rejected by the post-development school, which, with leaders like anthropologist Arturo Escobar, is much to the political and philosophical left of Sen. Escobar, in a postmodern vein, argues that the discourse dominating development thought and practice today is of Western origin and thus embeds the superiority of its cultural orientations in everything it says and does. When the international development discourse emanates from rich-country governments, universities, non-governmental organizations (NGOs) and multilateral organizations like the World Bank, all dominated by the institutions of Western and/or Northern culture, is it any surprise that its democratic capitalist ways of life and thought are privileged over others? When international economic, political and social institutions are structured on the basis of powerful democratic capitalist countries, is it any wonder that other cultures fail to compete successfully? In almost diametrical opposition to cultural modernizers, who believe Western-style cultural reforms are the way to developmental salvation for the poor world, Escobar believes they will bring destruction, for in disrespecting and ultimately undercutting their own cultures, the requisite cultural changes will cause local cultures to lose their identities, meaning systems and control over their own lives, resources and communities. There is in post-development a strong environmental orientation as well, for often people are dispossessed of their property and their livelihoods to make way for dams, agro-export products and mass production techniques that destroy traditional ways of life. Modernizers like Harrison might think these to be progressive developments, but Escobar believes they lead to domination, dispossession, violence, cultural chaos and poverty.

Unlike modernization theorists like Harrison, whose prescription is to criticize local culture and to enact policies to bring people into the modern world, Escobar exalts local culture, insisting on local solutions, discovered and implemented through locally developed institutions, and understood in terms of local languages and ways of life. Escobar and the post-development movement are thus among the main theoretical supports to the contemporary localization movement, of which David Korten is one of the most prominent supporters, but which also includes popular authors like Wendell Berry and novelist Barbara Kingsolver (Korten, 2001; Kingsolver, 2003).

In addition to the modernizers who discount culture entirely or treat it like one of any number of instruments, and post-development scholars who have a tendency to sacralize local cultures, there are a good number of scholars who take culture seriously while not being so judgmental about it. In these perspectives, culture is seen as a vital and, sometimes, revered
aspect of a people’s identity, and as representing the integral and holistic nature of a society that changes in response to internal and external pressures. Culture is not apart from or outside of economic life, but is instead integral to and interactive with it. Lourdes Arizpe, in reference to the UN Commission on Culture and Development, says: ‘it is not culture that is embedded in development; it is development that is embedded in culture’ (Arizpe, 2004). Gunnar Myrdal was among the early economists to consider culture in this way, emphasizing that the development of a society needed to be fundamentally based on the people’s own choices, which would in turn be based on their own values (Myrdal, 1968).

Another leader of this third way was Denis Goulet, who advocated an understanding of development that respected local cultures while at the same time recognizing the need for cultures to change. Goulet thus tried to find a way out of what he termed ‘The Cruel Choice’, which forced cultures outside the Western mainstream to choose between keeping their local cultural traditions and staying poor, or opting to join modernizing trends and losing their identity and sense of meaning (Goulet, 1971, 1980). His solution was to work within cultures and to find the ‘latent dynamisms’ that allowed cultural groups to respond constructively to the challenges of modernization. His approach coincides with the views of anthropologist Mary Douglas, who thinks the question of which cultures are ‘better’ is misguided and dangerous (Douglas, 2004). The common practice of evaluating which religious traditions are more prone to progress, for example, is the wrong way to approach the cultural question. Instead, she promotes a theory that sees every culture as a mixture of four groups and tendencies: the hierarchical, the entrepreneurial, the dissenting and the apathetic. Each of these first three plays crucial roles in the maintenance, protection and growth of cultures in their ever-changing environments. The hierarchical types, often government and religious leaders, want to keep traditions as they are and thus ensure social stability. Entrepreneurs, often from the world of business, are change agents who test the limits of their cultures by trying out and promoting new ways of doing things. Dissenters are typically idealists and visionaries who provide checks and balances on both groups. The apathetic tendency arises out of practices and patterns of life that marginalize whole groups of people or isolate them from the circles of power and decision-making. As the size of the apathetic group grows, so too does cultural distress. In many societies where poverty is prevalent, a study of the interaction of these four groups will reveal much about who holds power, how it is used and how whole groups are consigned to poverty. The analysis is not so much about whether a culture is good or bad as a whole, but about the way power is held and used by the different groups within the culture.
Interaction of development and culture

The notion that economics and culture make up integral parts of a large whole has led David Throsby to hint at the possibility of a new development paradigm that manages to bring these ‘two disparate fields closer together’, for ‘culture is in fact central to and inextricable from the development process, providing both the context within which economic progress occurs and the very object of development from the perspective of individual needs’ (Throsby, 2001, pp. 164, 165). This is the project of the two World Bank-spawned volumes cited earlier, and it seems also to be the project of Amartya Sen, who has been a highly influential voice in the development of the Human Development Index that is reported in the United Nations’ annual *Human Development Reports*.

As an economist, Amartya Sen is oriented toward matters of efficiency, growth and distribution, but, like Goulet, believes that any decisions to change culture must arise from the people themselves, in processes that ensure and facilitate their active and informed participation. Sen has been in the forefront of a new perspective on development known as the ‘capabilities approach’. Based on the idea of individual freedom within cultural contexts, this perspective respects culture by insisting on the legitimacy and inclusion of every voice by allowing the people themselves to evaluate their own cultural institutions and ways of life. Understanding culture as varied and complex, Sen affirms that culture and economic development are tightly interwoven, arguing also that both cultures and economies evolve in mutually reinforcing patterns. Whether and how economies and cultures should change, however, can only legitimately be determined by the people themselves, which is why Sen is so insistent that the capabilities to make such decisions be widespread throughout society. The capabilities approach thus assesses more whether people have both the individual capabilities (for example, education) and the necessary participatory and inclusive social structures for making informed choices about their own development path, than it does their achievement of predetermined ends, like high gross domestic profit (GDP) growth rates. While respectful of culture, Sen also points out that culturally based arguments are often improperly employed by powerful leaders who invoke the culture argument to defend oppressive systems. He is not persuaded, for example, by the argument that Asian values are more oriented to authoritarian styles of governance than European or other sets of values. Instead, he finds it is the leaders and holders of authoritarian power rather than the masses who support so-called Asian values. He argues that the real purpose of the Asian values argument is not to support local culture, but to legitimize the leaders’ hold on power.

For Sen, generalized poverty is often explained by the fact that the poor are disempowered and marginalized and lack key individual and social
capabilities. He is famous for noting the fact that famines do not occur in
democratic countries, which is a consequence of the breadth and depth of
participation in social, political and economic institutions. Where wealth is
widespread, on the other hand, there will also tend to be a widespread
diffusion of education, opportunity and voice, which combine to define
capability.

Sen’s influence at the UN is especially transparent in the 2004 Human
Development Report entitled Cultural Liberty in Today’s Diverse World
(UNDP, 2004). It respects culture, recognizes that cultures can and should
change in the development process and identifies widespread participation
as the legitimizing principle for choosing. The report begins with the asser-
tion that: ‘cultural liberty is a vital part of human development because
being able to choose one’s identity – who one is – without losing the respect
of others or being excluded from other choices is important to leading a
full life’. In contrast to some of the studies cited above, the report argues
that ‘there is no evidence from statistical analysis or historical studies of a
causal relationship between culture and economic progress or democracy’,
thus rejecting the Weber hypothesis. The report goes on to argue in favor of
multicultural democracies that ensure the broadest possible participation
in the social choices that affect people, identifying such choices as funda-
mental human rights. Furthermore, the report does not accept the reality
of Goulet’s ‘cruel choice’, for participation guarantees that there need be
no trade-offs between cultural integrity and human rights, democracy and
economic improvements.9

Yet many would argue that Sen and the Bank are too optimistic.
Modernizing development strategies have always meant conflict with tradi-
tional cultural institutions. Goulet pointed out in one of his last works:
‘Under the banner of development, powerful standardizing forces dilute
cultures and relegate them to purely ornamental, vestigial or marginal
positions in society’ (Goulet, 2005, p. 23). He highlights three of these
standardizing forces: (1) technology, especially media technology, which
spreads the values of individualism, instant gratification and consumerism;
(2) the modern state which centralizes everything, including ideas and
values; and (3) the managerial ethos which spreads cost–benefit analysis to
every sphere of activity as the best way to make decisions.

Moreover, the pervasiveness and the creative destructiveness of these
standardizing forces have fomented cultural resistance in many parts of the
world, probably best seen in Muslim countries. The present style of devel-
lopment and globalization threatens to generate a whirlwind of cultural and
political backlash.

Culture can both hinder and/or aid development, and development can
in turn both harm and/or help culture. What drives the change, and whether
development institutions like the World Bank, national aid agencies and NGOs ought to promote such change, are fundamental and controversial questions. Regardless, cultures must and will change. How they will change becomes the issue. Will the changes be destructive, resulting in people losing what gives them meaning in life? Or will people become part of a change process that allows them to adapt gradually and intentionally with minimal social and cultural cost, and without violent and regressive backlashes? Over 30 years ago Jim Lamb pointed out the road that needed to be traveled if that cultural and political backlash is to be avoided: ‘Development should be a struggle to create criteria, goals, and means for self-liberation from misery, inequity, and dependency in all forms. Crucially, it should be the process a people choose, which heals them from historical trauma, and enables them to achieve a newness on their own terms’ (Lamb, 1973, p. 20).

Notes
1. The basic sources for this section are: Tawney (1926) and Weber (1958).
3. See especially the final chapter.
5. See Barro and McCleary (2003, p. 37). The paper can also be found at http://www.economics.harvard.edu/faculty/barro/papers/Religion_and_Economic_Growth.pdf. Interestingly, Barro and McCleary also find that though certain religious beliefs have a causal impact on economic development, economic development itself also generates a decline in overall religiosity, thus supporting to some degree the secularization thesis.
8. See Sen (1999) for his most complete explanation of the theory. Another main voice in developing the capabilities approach is philosopher Martha Nussbaum. See Nussbaum (2000) for a discussion of how she integrates the legitimacy of local cultures with an attempt to discover universal values.

References
Easterly, William (2006), The White Man’s Burden: Why the West’s Efforts to Aid the Rest have Done So Much Ill and So Little Good, New York: Penguin Press.


Introduction
The incidence of violent conflict among poor countries is high: seven out of ten of the poorest countries have recently experienced some sort of civil war. Since conflict has a serious adverse impact on development, as well as causing massive human suffering, efforts to promote development and reduce poverty must include policies to prevent conflict and to protect populations during conflict. Moreover, as poorer economies are more likely to experience conflict and conflict inhibits development, a vicious cycle can ensue – of underdevelopment–war–underdevelopment – which it is essential to break if either peace or development is to be sustained, yet to do so is very difficult. This review covers both sides of this cycle: the following section analyses economic causes of contemporary conflicts; the subsequent section explores economic and social consequences.

Economic explanations of violent conflict
While some attribute contemporary conflicts to fundamental differences arising from ethnicity or religion (for example Huntington, 1993), such differences are evidently insufficient as an explanation since many multi-ethnic or multi-religious societies live peacefully while others are at peace for decades before experiencing conflict. In fact, the vast majority of multi-ethnic societies are at peace (Fearon and Laitin, 1996). Therefore, we need to look beyond ethnicity to issues of power and economics to understand conflict (Cohen, 1974). Below we consider four explanations that have dominated recent economic analysis of conflict: group motivation and inequalities; private motivations; a failed ‘social contract’; and environmental pressures (‘greenwar’).

Group motivation
Political conflicts consist in fighting between groups – groups that wish to gain independence or take over the state, and others that resist this (Horowitz, 1985). Such groups bring individuals together with a common purpose. While individual motivation is also important, this perspective argues that group motivation and mobilization underlie most political conflicts.

Groups engaged in internal conflict are often united by a common ethnic
or religious identity. Since 1945, the proportion of conflicts attributable to ethnic violence has been steadily increasing (Figure 63.1). While such conflicts are generally presented in religious or ethnic terms, and such identities provide a powerful source of mobilization and unity, underlying differences in access to economic or political resources are generally also present, providing both leaders and followers with a strong motive to fight. Gurr (1970, 1993) terms such group differences 'relative deprivation' and Stewart (2000, 2008) defines differences in groups’ access to economic, social and political resources as ‘horizontal inequalities’. Horizontal inequalities consist of inequalities in access to resources between groups differentiated by racial, ethnic, linguistic or religious characteristics, in contrast to vertical inequality which measures inequality among individuals or households. The horizontal inequalities explanation of conflict is based on the view that when such cultural differences coincide with economic and political differences between groups, this can cause deep group resentments that may lead to violent struggles.

Note: The data from which these figures were calculated – the list of ‘major episodes of political violence’ compiled by Monty G. Marshall of the Centre for Systemic Peace – provide a ‘magnitude’ score for each episode, ranging from 1 (mildest) to 10 (severest). Total magnitude here is calculated simply as incidence weighted by magnitude.


Figure 63.1 Ethnic violence as a proportion of ‘major political violence’, 1946–2004
Empirical evidence is accumulating that horizontal inequalities constitute a significant cause of violent conflict. Cross-sectional quantitative analyses have shown a significant relationship between various dimensions of socio-economic inequality and conflict. Mancini (2008) shows that differences in infant mortality rates – a broad proxy for levels of socio-economic deprivation – between ethnic or religious groups among districts in Indonesia help explain the location of the communal conflicts that occurred after 1998. A similar relationship – between ‘spatial’ horizontal inequalities and the intensity of insurgency – has been found in Nepal (Gates and Murshed, 2005). Although multi-country studies have been hampered by poor data, supporting evidence has been found by both Østby (2004) and Barrows (1976). There is also substantial case study evidence: Stewart’s (2002) review of the experiences of nine countries shows not only that severe socio-economic horizontal inequalities preceded the emergence of violent conflict, but that reductions in socio-economic horizontal inequalities – such as occurred in Northern Ireland during the 1980s – may contribute to the conditions for a peaceful resolution of such conflicts. However, some societies show severe horizontal inequalities without experiencing conflict. Political inclusiveness is one reason that some societies avoid conflict despite severe economic horizontal inequalities; other reasons are lack of unity among the deprived groups, and state repression.

Political horizontal inequalities – the exclusion or under-representation of groups within the political structure of a state – can provoke violent conflict, especially when they change abruptly. In Côte d’Ivoire, three decades of post-independence rule by Félix Houphouët-Boigny avoided significant conflict, largely due to the policy of balancing representatives of the major groups in positions of importance in the government and bureaucracy. Following Houphouët-Boigny’s death and the introduction of multi-party elections in the early 1990s, political leaders sought to mobilize ethnic sentiments to enforce their grip on power and thus undermined Houphouët-Boigny’s careful balancing act, leading to a spiral of ethnicization, xenophobia and, ultimately, civil war (Langer, 2005).

It is important to note that relatively rich groups may instigate conflict, as well as the relatively poor. The relatively rich do so to preserve their riches (and/or power), while the relatively poor do so out of a sense of injustice with the intention of achieving some redistribution.

Private motivation
People who fight are, of course, individuals with their own private motivation as well as being members of a group. War confers benefits as well as costs on some individuals. Political sociologists (Keen, 1998; Duffield, 1994), and economists (for example Collier and Hoeffler, 2001), have
emphasized private or individual motivation as the fundamental cause of conflict, arguing that the net economic advantages to individuals motivate them to fight. In this approach, which has its basis in ‘rational choice’ assumptions, group identities are not an independent factor but are instruments, created to help fulfil the private motives of those who fight, especially leaders (Hirshleifer, 1994).

Keen lists many ways in which war confers individual benefit on particular categories of people: it permits people, especially uneducated young men, to gain employment as soldiers; it offers opportunities to loot, to profiteer from shortages and from aid, to trade arms and to carry out illicit production and trade. Where alternative opportunities are few, and the possibilities of enrichment by war are considerable, wars are likely to be more numerous and longer. Conflicts may persist because some powerful actors benefit through the manipulation of scarcity, smuggling, and so forth and have no interest in resolving the conflict. An oft-cited case used to support this view is the role of ‘conflict diamonds’ in the prolongation of the civil war in Sierra Leone (Collier, 2000, p. 5).

However, case studies suggest that even where natural resources are abundant, private maximizing motives are rarely the full explanation. A study of seven countries in conflict concluded:

very few contemporary conflicts can be adequately captured as pure instances of ‘resource wars’ . . . Economic incentives have not been the only or even the primary causes of these conflicts. (Ballentine and Sherman, 2003, pp. 259–60)

In most cases of conflict, the risk of death or debilitating injury are high, so that ‘rational’ actors might be likely to choose another option before engaging in rebellion. This argument may not apply to leaders, who are less often killed or injured, while followers may be coerced into fighting, or persuaded to fight by leaders playing up religious or ethnic differences and grievances:

Grievance is to a rebel organization what image is to a business . . . [A] sense of grievance is deliberately generated by rebel organizations . . . [rebel supporters] are gulled into believing the discourse which self-interested rebel leaders promote. (Collier, 2000, p. 5)

At this point the group explanation and individual explanation of conflict come together. Grievances are hard to sell to the extent of people risking their lives if they are not genuine (that is, unless there is some sort of exclusion or economic horizontal inequalities), while leaders are motivated by political exclusion (that is, political horizontal inequalities) which denies them access to resources and power.
While generally not a sufficient explanation of conflict, it is clear that expected rewards often play a role in the decision to rebel. Econometric evidence confirms that conflict incidence is higher in resource-rich areas (Humphreys and Varshney, 2004). The gains (and motivation) in resource-rich areas may be individual or group, or both. As Collier notes, citing the cases of Aceh (Indonesia), Biafra (Nigeria) and Katanga (Zaire), separatist rebellion often emerges in resource-rich areas of a country (Collier, 2000, p. 10). Yet all these conflicts were framed in ethnic terms. Moreover, in many cases the leaders of the rebellions left lucrative and safe positions to instigate rebellion. For example, Hassan di Tiro left a secure position at the United Nations to instigate the Acehnese uprising. In the case of Colombia, often depicted as a ‘greed’-motivated conflict, interviews with both leaders and those mobilized to fight show that generally their economic position worsened as a result of participating in the conflict – most put forward ideological reasons for their actions, especially the issue of land reform (Gutierrez Sanin, 2004).

There are also examples of separatist movements in regions with poor resource endowment such as Eritrea, Bangladesh (then East Pakistan) and the Tamil rebellion in Sri Lanka. Moreover, it is not possible to create an identity out of nothing (Smith, 1991). A common history, language, culture or religion is generally required to generate felt identities powerful enough to mobilize people for conflict.

**Failure of the social contract**

A third theory of violent conflict derives from the view that social stability is implicitly premised on a social contract between the people and the government. According to this hypothetical contract, people accept state authority so long as the state delivers services and provides reasonable economic conditions in terms of employment and incomes. With economic stagnation, or decline, and worsening state services, the social contract breaks down and violence results. Hence high (and rising) levels of poverty and a decline in state services would be expected to cause conflict (Nafziger and Auvinen, 2000). High vertical inequality might also be associated with such a failure, unless accompanied by populist measures to compensate the deprived. Conversely, political institutions that are able to channel and respond to socio-economic discontents strengthen the social contract, thus reducing the risk of conflict.

Considerable evidence from econometric studies shows that conflict incidence is higher among countries with lower per capita incomes, life expectancy and economic growth (Elbadawi and Sambanis, 2000; Nafziger and Auvinen, 2000; Collier and Hoefler, 2001). Many analyses have found an inverted U-curve relationship between the extent of democratization in
a country and the risk of conflict (for example Ellingsen, 2000), with the usual interpretation being that ‘stable’ democracies are able to avert violent conflict through a strong social contract, while strongly authoritarian regimes are able to suppress conflict. However, Reynal-Querol (2002) has argued that it is the particular type of democracy – whether majoritarian, presidential or proportional representation – that affects propensity to conflict, rather than the level of ‘democracy’ per se.

‘Greenwar’ and environmental scarcity
The fourth explanation of violent conflict, associated with the work of Homer-Dixon and the ‘Toronto Group’ (for example Homer-Dixon, 1994; Percival and Homer-Dixon, 1998), is the ‘greenwar’ or ‘environmental scarcity’ argument. The essence of this perspective is that contest for control over declining natural resources, often intensified by population pressures, is a major cause of violent conflict around the world. Poorer societies are more at risk because they will be ‘less able to buffer themselves’ from environmental pressures (Homer-Dixon, 1994, p. 6). Three dimensions of environmental scarcity are identified which may lead to conflict: ‘supply-induced scarcity’, linked to the ‘depletion and degradation of an environmental resource’; ‘demand-induced scarcity’, linked to population growth and the consequent extra pressures on existing resources; and ‘structural scarcity’, which ‘arises from an unequal distribution of a resource that concentrates it in the hands of a relatively few people’ (Percival and Homer-Dixon, 1998, p. 280). Homer-Dixon thus predicts ‘an upsurge of violence in the coming decades that will be induced or aggravated by scarcity’ (Homer-Dixon, 1994, p. 6).

However, while it is clear that pressures arising from environmental scarcity may play an important role in many conflicts, the environmental scarcity hypothesis is – and really does not claim to be more than – a partial theory that contributes towards our understanding of a set of conflicts, but not the general conditions under which conflict is more likely to arise.

The environmental scarcity hypothesis overlaps substantially with the other hypotheses discussed here. It overlaps with the social contract hypothesis in viewing poverty as the root cause of conflict, although it points to specific environmental causes of such poverty. It also often overlaps with the group motivation approach, as environmental pressures usually lead to conflict where there are ‘groups with strong collective identities that can coherently challenge state authority’ (Percival and Homer-Dixon, 1998, p. 280). Indeed, the ‘structural scarcity’ dimension of the greenwar approach is very similar to the group motivation hypothesis, albeit restricted to a particular dimension of inequality.
The environment scarcity view has been criticized by Fairhead (2000) who argues that it is environmental riches, not scarcity, that is associated with conflict because people fight to control lucrative natural resources, as in the Congo. This view fits well into the private motivation or greed hypothesis. In fact both environmental poverty and environmental riches may cause conflict, for different reasons and in different circumstances.

The theories outlined above appear, in their extreme formulations, to be diametrically opposed – as manifest in the ‘greed versus grievance’ debate (Ballantyne and Sherman, 2003). But, as we have seen, proponents of one perspective usually accept in part the insights of other perspectives. Some conflicts fit neatly into one of the explanations, some into others, and some clearly have multiple causes. One rather simple conclusion, therefore – that qualitative analysts of conflict are mostly aware of, but that quantitative analysts tend to overlook – is that each of the broad causal theories discussed above involves a degree of oversimplification and generalization. The causes and dynamics of any single conflict are typically complex, sometimes contradictory, and involve aspects of many, if not all, of the perspectives discussed above.

Yet it is important to understand which explanation dominates in a particular case, since this has important implications for appropriate policy prescriptions for the prevention and resolution of the conflict. There is not space here to discuss policies in detail. In brief, where group motivation is a fundamental cause, policies need to be inclusive and correct large horizontal inequalities (Stewart, 2008). To the extent that private motivation is key, policies need to reduce the profitability of illicit war-sanctioned activities (like drug production and smuggling), and to offer combatants income-earning opportunities as an alternative to fighting. If a failed social contract is the fundamental cause, then the aim should be to improve the functioning of government in relation to security, the economy and the provision of social services. Greenwar conflicts require an attack on the fundamental causes of environmental pressure, both from demand and supply perspectives. Such policies should be applied to all societies at risk of war, which includes all low-income countries, any country which has experienced conflict in the past few decades, and any economy suffering severe horizontal inequalities.

Two points are worth making about this policy set. Firstly, none of these policies are a central part of the current development agenda of the international financial institutions; secondly, they are all desirable in themselves, quite apart from their impact on conflict prevention.

**The economic consequences of war**

War in general, and civil war in particular, is one of the main causes of human suffering and economic underdevelopment. Most of the human and
economic costs of war do not result directly from battle deaths and injuries, but indirectly from the loss of livelihoods caused by the dislocation of economy and society. An important implication of the substantial indirect costs is that policy might be able to reduce costs substantially if appropriately designed, even while conflict is ongoing.

The economic effects of war are the result of a complex interaction between the particular processes of war and the economy in which it takes place. Figure 63.2 provides an overview of the main relationships likely to be affected, with arrows indicating the direction of causality and expected direction of impact. It is helpful to distinguish between the direct economic consequences of the conflict, and the compensating behaviour of economic agents in their attempt to moderate or offset the negative impacts of war.

Direct effects include: output loss as people move from their place of work because they join the fighting, are killed or flee; the destruction of capital and consequent loss of output; disruption of transport links due to physical destruction; a loss of trust among economic agents, reducing market transactions; disruption of international markets due to frontier closure or embargoes; reduced foreign investment and the diversion of foreign exchange from economic and social needs to military uses.

These effects will tend to reduce aggregate levels of output. Labour markets will be disrupted as many of the unskilled of prime working age become military recruits and some suffer violent deaths, while much skilled labour is likely to leave the country. Reduced agricultural output, and disrupted internal and international markets, are likely to affect exports particularly heavily. Reduced foreign exchange availability for productive inputs results, leading to a shortage of imported inputs and to a further fall in output and exports.

Compensating behaviours which can moderate the negative impact of the effects of war include: increased capacity utilization and import substitution; the provision of international credit or aid, offsetting loss of foreign exchange; the emergence of new forms of social capital compensating for loss of trust in formal institutions – such as enhanced cooperation and trust among members of a group on the same side of the war; and rapid government or community action to reconstruct facilities destroyed by belligerents.

To understand the total impact of these mechanisms on individuals one needs to go beyond money incomes to explore how they affect different types of individual entitlements, including market entitlements (acquired largely through employment and self-employment), direct entitlements (subsistence production), social entitlements (provided by the state), civic entitlements (provided by the community and NGOs) and extra-legal entitlements (see Stewart et al., 2001a, Chapter 1).
Figure 63.2  Predicting changes in the economy and entitlements during conflict
There are serious methodological problems in estimating the costs of war (Stewart, FitzGerald and Associates, 2001a, Vol. 1, Chapter 1). Below we summarize results adopting a variety of methods.

**Macroeconomic consequences**

All studies find negative effects on gross domestic product (GDP) growth. One study found a negative impact on GDP per capita growth in 13 out of 14 countries suffering the worst conflicts between 1975 and 1995, with considerable variability in magnitude. The worst losses occurred in long and pervasive wars (Stewart et al., 2001a). Regression analysis of 92 countries, 1960–89, showed an annual loss of 2.2 per cent during the war and in the immediately following years, compared with a no-war situation (Collier, 1999). Other cross-country regression analysis for 1960–99 came to similar conclusions, with an average loss of growth of 2.4 per cent per annum (Hoeffler and Reynal-Querol, 2003), although Imai and Weinstein (2000) suggest somewhat lower costs. The wide range of estimates indicates how dependent they are on the methodology adopted. Nonetheless, all give negative results. Evidence shows that wars with more widespread geographic coverage have a more negative impact (Stewart, Fitzgerald and Associates, 2001; Imai and Weinstein, 2000).

Reduced economic growth is the result of capital destruction, lower investment and disrupted markets. All types of capital stock are eroded or destroyed in war. Physical facilities suffer direct attack – roads, ports and energy plants are often targeted. In Mozambique, Brück estimates that there was a two-thirds reduction in operational dams and plant nurseries, with 40 per cent of rural facilities destroyed or eroded. Social infrastructure is also commonly destroyed – again in Mozambique, almost 60 per cent of primary schools were closed or destroyed (Brück, 2001, pp. 64–7). Human capital is killed, or flees. About half the doctors and 80 per cent of the pharmacists left Uganda in the late 1970s (Dodge and Wiebe, 1985). The spread of AIDs that results from the sexual activities of combatants further reduces human capital. Institutions are destroyed – in Uganda, the agricultural extension system virtually disappeared (Matovu and Stewart, 2001). Social capital is weakened, with a severe loss in trust, particularly across groups. Yet new forms of institution and social capital emerge, for example informal banking systems.

Rates of investment and savings fall due to increased uncertainty. Evidence shows falling domestic saving rates (Stewart et al., 2001a; Brück, 2001), while capital flight accelerates. The proportion of private wealth held abroad rose from 9 per cent to 20 per cent in the course of civil wars according to Collier et al. (2004). Voluntary private lending from abroad tends to fall with increased uncertainty, but changes in official foreign lending
depend on political factors. ‘Forced’ foreign savings may occur, as countries renego on debt servicing obligations. In fact, aggregate foreign savings seem to hold up more than might be expected, evidenced by the huge accumulation of foreign debt during conflict (Stewart et al., 2001c).

Private (domestic and foreign) investment is adversely affected by uncertainty, rising costs of transport and difficulties in securing finance. Foreign investors are likely to be concerned about the safety of their personnel and equipment, and the increased foreign exchange risk. Country studies show a fall in foreign direct investment as expected – indeed this was the most important macroeconomic cost of Sri Lanka’s conflict (O’Sullivan, 2001). Imai and Weinstein (2000) show a strong negative impact of civil war on private investment. Government investment is likely to be negatively affected by reduced revenue, and diversion of expenditure to military uses – Ra and Singh (2005) estimate that development expenditure in Nepal fell by one-third, 2001–04. Country studies show that aggregate investment on balance does not fall as much as domestic savings, possibly due to buoyant small-scale investment as the informal sector expands.

Exports are also negatively affected, as a result of the general fall in production, a shift towards domestic markets and disruptions in international markets. In most cases, imports hold up much better than exports, financed by foreign debt. The share of imports going to military items and food imports rises, however, leaving a much smaller share for inputs into the productive sector.

Government revenue is likely to fall absolutely and as a proportion of gross national product (GNP) as the government finds it more difficult to collect taxes and major sources of revenue (for example, from export taxes) fall away. There are sharp divergences across countries. In Uganda (1979–80), revenue as a share of GDP fell dramatically, but in both Mozambique and Nicaragua in the 1980s, the revenue ratio rose. Budget deficits increase as government expenditure rises faster than revenue (Stewart et al., 2001c).

Inflation is expected to accelerate, as governments resort to deficit financing to finance the conflict and other essential services and public confidence in the currency declines. In recent wars, there seems to have been only minor acceleration in price inflation (Stewart et al., 2001c).

Meso-economic consequences
In general, there is a shift from tradeable to non-tradeable sectors, as a consequence of market disruptions, including undermining of formal organizations such as banks, reduced trust and failures of the transport system. One aspect of this is a switch to subsistence and informal activities, including simple production (even arms) and trading (particularly smuggling).
For example, Mozambique experienced rapid growth in the urban informal sector; while the ability to shift from producing marketed crops to subsistence agriculture also helps protect food production and nutritional standards – it did so, for example, during the Amin era in Uganda.¹

The share of government expenditure going to military items invariably increases, making it difficult to sustain social and economic expenditure. On average, it is estimated that military expenditure rises from 2.8 per cent of GDP (average for developing countries in peacetime, 1995) to 5.0 per cent of GDP (Collier et al., 2003). One consequence is likely to be a fall in the share of social expenditure. In Uganda, social expenditure fell by roughly 9 per cent annually during conflict. In exceptional cases such as Nicaragua in the 1980s, countries give increased priority to social expenditure and succeed in increasing public entitlements. Where government revenue collapses there may be a dramatic decline in public entitlements: this was true of Uganda in the late 1970s and mid-1980s and Afghanistan in the early 1990s.

Civic entitlements can substitute for reduced government provision. There was extensive foreign non-governmental organization (NGO) provision of social and economic services, for example, in Afghanistan in the early 1990s (Marsden and Samman, 2001). In Sri Lanka, local NGOs and communities and the Tamil rebel forces provided significant support (O’Sullivan, 2001). But in the worst-affected areas, communities disintegrate as people flee, and NGOs are able to do little – examples are Uganda, Southern Sudan and Sierra Leone.

**Human costs**

In addition to deaths and injuries, flight and ensuing psychological trauma, human costs result from the changing economy, with worsening entitlements of most types:

- Market entitlements decline as household incomes fall with worsening employment conditions, while the main earners may leave the household to fight or flee. However, extra-legal entitlements rise, with big gains for some households profiting from types of illegal production, but losses for others, subject to theft and looting.

- Worsening nutrition arises from reduced incomes and agricultural output. Calorie consumption fell in over 70 per cent of the countries worst affected by conflict, in 1970–95, falling to below 1700 calories per person per day in Ethiopia, Liberia, Mozambique, Sierra Leone, Somalia and Afghanistan (Stewart et al., 2001c, p. 90). The agricultural sector is typically badly hit in civil wars, as people are forced to move. Rising food prices can have a devastating impact on access to food. Millions of deaths in the Bengal famine of the 1940s have been...
attributed to war expenditures leading to food price increases (Sen, 1981). Speculative traders can also contribute to rising food prices (Ravallion, 1987). The impact on nutrition may be offset by food subsidies, food distribution and rations (including food aid). In Nicaragua, nutrition actually improved in war-affected zones as a result of such measures (Utting, 1987), while in Mozambique, food aid seems to have prevented mass starvation (Stewart and Samman, 2001).

- Social entitlements, including health and education services, worsen as a result of destruction of facilities, reduced government resources and flight of personnel. Government expenditure on health and education is estimated to have fallen in ten of 14 war-affected countries (1970–95), by over 40 per cent in Angola, Liberia, Uganda, El Salvador and Iran (Stewart et al., 2001c, p. 87). Primary school enrolment fell substantially in some countries – notably Angola and Mozambique. Doctors per person also fell significantly in about half the war-affected countries. Civil society in some situations is able to substitute for government services but it too disintegrates in the worst cases.

- Health is affected by increased infection rates associated with the mass migration that often accompanies war – as many as one-third of the people in Mozambique were forced to move, while one-third of the Afghanistan population left the country in the 1990s. The spread of AIDS has also become a particular feature of war, as soldiers are notably highly infected and infect others, including sometimes through mass rape. HIV prevalence in the military was estimated to be 40–60 per cent in Angola, the Democratic Republic of Congo and Sierra Leone in the late 1990s (Collier et al., 2003). The extent of health costs is indicated by rising infant mortality rates: in Uganda additional infant deaths, compared with non-war regional trends, amounted to over 2 per cent of the 1995 population. Econometric estimates across countries show an increase in infant mortality rate (IMR) of 13 per cent during conflict (Hoeffler and Reynal-Querol, 2003). Moreover, some case study evidence suggests increases in adult mortality rates sometimes exceed increases in IMR (Guha-Sapir and Van Panhuis, 2002). World Health Organization (WHO) estimates of disability-adjusted life years show a major loss due to increased infectious diseases (Ghobarah et al., 2003), which persists in the post-conflict era.

In summary, while the direction of impact on most variables is demonstrably negative, and most types of entitlement worsen, there are large variations in the magnitude of costs across countries and on the burden of
entitlement loss across groups within a country. Variations in the social and economic impacts of war arise from differences in the nature of the war, the structure of the economy, the character of the government, the response of the international community and people’s own actions. Moreover, while the evidence on worsening GDP and average consumption levels suggests that the net situation is one of loss, there are also gains from conflict for particular groups which can form one of the private motives for war or its prolongation as discussed above.

The economic consequences are obviously highly dependent on the nature of the war itself. Firstly, and above all, its duration is important. In a long war, reserves become exhausted, so vulnerability is increased. The geographic spread of the war is also relevant. When confined to one part of the country the war may have only small direct effects on the economy as a whole – conflict in Northern Uganda, which persisted over decades, had much smaller economic effects than the conflict in the mid-1980s which was centrally located. The extent of foreign involvement in the war is another factor affecting vulnerability, since external support may compensate for lost export earnings; it may also, however, contribute to prolonging the conflict.

Secondly, the structure of the economy helps determine the costs. An economy heavily dependent on the agricultural sector will be especially badly affected by widespread disruption of the sector, but may be less affected by reduced imports, since it is possible to retreat into subsistence if markets break down; an inflexible economy with a sizeable industrial sector may be particularly vulnerable to foreign exchange loss, although this can be offset by external grants or credit. An economy with a flexible industrial sector operating at less than full capacity may suffer little, as it is able to make up for loss of imports by domestic production – the UK in World War Two is a classic example.

A third critical factor is the nature of the government: a government that is or becomes very weak loses the ability to provide essential services and relief. Strong governments can sustain services, but they will only do so if they are relatively benevolent, wishing to provide for all the people, despite the war. Some strong governments may deliberately reduce food and basic services to ‘enemy’ territory, as was the case in Sudan.

Fourthly, international actions contribute to variations in costs. The large supplies of food aid in Mozambique undoubtedly reduced death rates, but in Sudan in the 1980s food aid was too little, too late and its delivery and use were distorted by government policy (Keen, 1994). In Afghanistan in the 1990s, massive support for the refugees in Iran and Pakistan greatly lessened the human costs. In contrast, the international community did little to offset costs in Sierra Leone in the 1990s, while the
USA increased the burden of civil war for vulnerable groups in Nicaragua by trade and aid embargoes.

Finally, people’s own actions are important in moderating both human and economic costs. In almost all cases, people found new economic possibilities – many created by the war – which enabled them to survive. The burgeoning of the informal sector in Mozambique is one example (Chingono, 2001). People also protect themselves by fleeing, relocating within the country, or emigrating. People are rarely completely passive victims, but in the worst situations, there is little they can do to protect themselves.

**Policies towards economies in conflict**

The analysis above suggests that economic and social policies of both governments and donors can be designed to reduce the economic and human costs of conflict, even during the conflict.

An overriding aim of economic and social policies towards countries in conflict should be to maintain entitlements of the vulnerable, especially to food and health services, if possible in a self-sustaining way. In addition, the policies should also aim to tackle the causes of war, following the analysis above.

There can be no generalization about policy irrespective of the actual situation, since it is essential to understand the major cause of entitlement collapse, whether it is loss of employment opportunities, escalating inflation or destruction of key assets; and to understand the nature of the authorities, whether they are so weak that for short-run action they must be bypassed, or are strong but ill-intentioned so resources channelled through them will not reach those in need, and whether there are alternative structures (for example local authorities or rebel authorities) that can handle projects.

Monitoring is essential to identify appropriate actions before the situation becomes critical. Official monitoring is often weak (and politically biased). Greater use could be made of NGOs, local and foreign, who can be well situated to monitor developments and provide early warning of impending adverse changes in human conditions, for example, distress selling of assets, small movements of people or adverse changes in nutrition. In the Sudan, a major reason for the huge rise in death rates in the early 1980s was the failure to take early action, and this partly stemmed from the development community being slow to note early warning signs (Keen, 1994).

Successful macro-policies are generally much more effective in maintaining essential entitlements than direct relief. The aim should be to sustain the economy – which may require external aid and support for export markets – and to prevent escalating inflation. It is vital to sustain revenue to support public entitlements, compensating for declines in the
normal revenue base by devising taxes on war-related activities. The sale of food aid, for example, can be an important source of revenue.

Meso-policies need to be directed towards sustaining the share of aggregate output going to supporting entitlements of the vulnerable. On the social side, the objective is to ensure that everyone has access to adequate food and to sustain public entitlements to basic health and education. Preventive health measures, notably immunization, are particularly vital in war because unusual movement of people causes infections to spread rapidly. Expenditure on basic health and education accounts for only a fraction of social expenditure, so that strong prioritization of these services can ensure their maintenance even if the total is being cut. Yet the problem, of course, is not only one of money. Teachers and doctors may flee, and facilities can be destroyed. A flexible approach is needed. For example, Mozambique introduced mobile clinics and classrooms when Renamo was targeting health and education buildings.

Ensuring food security requires that food prices are monitored and escalating prices prevented, through some combination of increasing supplies (via food aid), controlling prices and rationing – policies which advanced countries have adopted when themselves at war. For the rural population, a combination of ensuring adequate agricultural support (seeds, fertilizer and so on), employment schemes, and the provision of food in schools and clinics, can achieve wide food access.

As well as domestic policies, there are important spheres of international action which can help (or worsen) the situation – including long-term policies to support development (such as improved terms of trade and aid flows), and short-term policies to reduce human costs (such as welcoming refugees, and providing food aid), as well as policies to reduce the financing of conflict (including policies towards trade in conflict commodities: for example, diamonds). International policies need to take into account the impact on the livelihoods and survival of poor populations. Economic sanctions, for example, are frequently harmful to the poor and often ineffective in achieving political objectives (Clark, 1996; O’Sullivan, 2003).

**Conclusion**

This overview of the economics of war and development suggests three major conclusions. Firstly, economic analysis of causes and consequences of conflict is essential to design appropriate policies. Secondly, both causes and consequences are diverse, varying across countries, so in-depth knowledge of the particular case is vital. Thirdly, there is a rich menu of appropriate policies which can help reduce the incidence and costs of conflict – yet in most cases these policies currently form a part of neither the normal development agenda nor the normal relief agenda.
Note
1. The burgeoning of subsistence and informal activities means that official statistics can greatly understate production, so that the aggregate costs of conflict may not be as great as they appear from official data.

References


PART IX

ECONOMIC DEVELOPMENT EXPERIENCE
Introduction
According to conventional economic logic the surest way for poorer countries and communities to achieve lasting gains in economic welfare is by getting rid of state-created ‘distortions’ that impede market forces in allocating existing, and mobilizing future, resources. In the recommended reform package, opening up to international markets and firms offers the surest way of bringing about the right set of incentives to help realize those gains. The liberalization of international trade will enable countries to exploit their comparative advantages better, with big gains expected in the South given their legacy of protectionist policies. Where finance is a constraint on growth, the liberalization of international capital markets will ensure that investment funds flow from the capital-abundant North to the capital-scarce developing countries. Attracting foreign direct investment (FDI), including through the sale of state-owned assets, will help gain quick access to new technologies and management practices as well as providing ready-made export opportunities. Entrusting development to these international market forces is seen as having the additional advantage of their being less vulnerable to capture by local interest groups and rent-seeking coalitions. A plethora of econometric studies have backed up this case for hitching development to open markets, and a string of popular economic pundits have been ready at hand to explain just how their ‘win–win’ logic has already begun to flatten out the global economy, raising expectations of a swift eradication of extreme poverty, a narrowing of income gaps, and the emergence of a truly global middle class.

There is little doubting that, in the wake of the debt and development crisis of the 1980s, such thinking, oftentimes with the strong backing of the international financial institutions, affected a radical shift in policy-making in many developing countries. During the 1990s, new technologies and international business practices (closely identified, if not synonymous, with globalization) were added to the reform mix, ending, on many accounts, any further debate on development strategy. Efforts to direct globalization were deemed futile and resistance would result only in marginalization, or perhaps worse.1 Developing countries were, accordingly, advised to adapt
to this new reality by fully relinquishing economic sovereignty to mobile capital and the forces of international competition.

This chapter questions such advice. It argues that economic trends during the past quarter-century do not support the utopian pronouncements of many globalization enthusiasts; that the destructive impulses released by the radical turn in policy after the debt crisis have, in many countries, outweighed the creative impulses; and that expunging the historical detail and structural diversity from the catch-up process is unlikely to provide the direction on policy advice and institutional reform needed to narrow income gaps worldwide.

Openness, accumulation and structural change in an interdependent world

Most development economists agree that a strong productivity performance is essential to any successful catch-up growth path, not only because it translates (though not always directly) into rising living standards, but also because it enables poorer countries to manage better the various adjustments, trade-offs and distributional conflicts that taking such a path will generate. Most would also agree that this performance is more about galvanizing dynamic economic forces than it is about maximizing the static gains from an improved allocation of existing resources.

Certainly, recent efforts to revive the idea of market-driven convergence for a globalizing world have tried, by introducing a broader conception of capital (including human capital and other less tangible wealth-creating assets, and an expanded role for foreign capital), more variegated technological trajectories and additional behavioural parameters, to accommodate more dynamic forces in their blend of (old) trade and (new) growth models. The empirics of this new convergence literature have already provoked much debate and controversy and methodological questions surround the idea of 'conditional' convergence. However, what is more revealing about this literature is its enduring commitment to a set of standardized policies derived from the identification of distortionary market impediments within mathematically tractable equilibrium models. This approach continues to trump historical detail and structural differences in the design of development strategy (Kenny and Williams, 2001).

An alternative place to begin thinking about development strategy is with the empirical regularities identified by economic historians and classical development economists linking industrialization, and more particularly its manufacturing component, to strong productivity and income growth. This leading role is due less to any uniquely intrinsic qualities of industrial activity and more to the confluence of growth impulses that accompany its evolution, including the presence of increasing returns whether at the plant, firm or industry levels, high elasticities of supply and demand for
manufactured goods, strong externalities linked to a high rate of technological innovation, and complementarities between production and consumption. Successful industrialization paths have, moreover, been closely identified with a ‘well-filled input–output matrix’ with an increasingly dense set of links between sectors (a high level of sectoral articulation between, for example, rural and urban, and consumer goods and intermediate goods), and a structure of demand such that a high proportion of domestic production is sold to domestic wage earners (Wade, 2003, p. xlviii).

In the interplay of elements making up this strong growth regime, capital accumulation plays a pivotal role. A given pace of accumulation can of course generate different growth rates, depending on its nature and composition, as well as the efficiency with which production capacity is utilized. This is one of the main reasons why econometric studies have failed to establish a one-to-one relation between the rate of investment and economic growth. Unfortunately, systematic discussion of the forces that govern the process of capital accumulation has long been a stumbling block in the development literature (Hirschman, 1958, p. 35), and the recent growth literature is no exception. In particular, the neglected role of profits for financing investment in developed and developing countries alike, cuts off conventional economic analysis from a careful examination of how dynamic industrial activity can provide abundant opportunities to create rents whose reinvestment is key to perpetuating a dynamic growth regime. This profit–investment nexus has been identified in the recent fast-growth experiences in East Asia (Akyüz and Gore, 1996; Singh, 1999; Ros, 2000) and provides an initial guide to why and how policy interventions might make a difference in stimulating catch-up growth (UNCTAD, 1997).

While industrialization can generate the kinds of cumulative impulses that describe a successful growth process, these are not automatically self-sustaining. Various constraints, traps and coordination failures can upset the process, making it generally impossible to rely on market forces to establish the linkages associated with internal integration and to move economies through the various stages of industrialization. This provides further rationale for policy intervention and institutional learning. The variety of linkages and their local specificity are now much better appreciated. Among these, insufficient domestic demand to absorb the growing industrial output can be just as important as failures on the supply side. The close links between industrialization and ‘external’ integration have, of course, been a familiar feature of the development process since Adam Smith saw the size of the market as a constraint on the division of labour. By broadening the size of the market, exports allow scale economies to be exploited; they also provide the foreign exchange needed for capital accumulation, in view of the dependence of most developing countries on
imported capital and intermediate goods. At the same time, investment improves export potential by adding to production capacity and improving competitiveness through productivity growth. As such, a successful industrialization path is usually characterized by rising investment, exports and manufacturing value added, both in absolute terms and as shares of gross domestic product (GDP).

The historical evidence provides little support for an independent role for rapid liberalization of foreign trade and finance and the deregulation of domestic markets in stimulating and sustaining catch-up growth. In most cases, the state has provided a necessary complement to, and sometimes a corrective influence on, the market, particularly by promoting a rapid pace of capital accumulation and technological progress linked to expanding industrial output, employment and exports. Accordingly, establishing a robust nexus between investment, profits and exports remains key to the design of development strategy, helping identify the cumulative links that describe a successful industrialization path, as well as providing a framework in which to explore the institutional and policy challenges thrown up by the catch-up process.

The international economic environment: open for business

Following the debt crisis of the early 1980s, deregulation of domestic economic activity and its opening up to international firms and market forces became the leitmotif of economic policy design in many developing countries, more often than not with the overt support of the international financial institutions. Success has tended to be measured in terms of monetary and fiscal discipline, an increasing volume of international trade and capital flows, and rising ratios of trade and FDI to GDP. On these measures, many poorer countries, and the world economy more generally, have since the late 1980s registered a good deal of success. Trade has consistently outpaced global output, with the pace of expansion much faster in the 1990s, and with developing countries in the vanguard (UNCTAD, 2003, pp. 41–4). As a result there has been a rapid and ubiquitous rise in the share of exports and imports in GDP in developing countries, as well as a rapid increase in the share of these countries in global trade – from about 24 per cent of total exports in 1980 to 28 per cent in 1995 and 34 per cent in 2004.

The rise in capital flows has been even more dramatic. The global stock of financial assets rose more than elevenfold between 1980 and the end of 2004, from $12 trillion to $136 trillion; in 1980 they were roughly equal to global GDP but by the end of 2004 were more than three times higher. Much of the increase in flows has been among developed countries, but the 1990s saw a strong surge of financial flows to developing countries, following a sharp dip
in the 1980s (UNCTAD, 2003, pp. 23–31). The greater part of these flows consisted of short-term investments: bank loans, equities and short-dated government securities, inter-bank and other deposits. However, beginning in the early 1990s, FDI in developing countries also rose sharply, more than doubling between 1995 and 2000, when it reached over $250 billion, rising from a quarter to close to one-third of global flows.

These numbers do not fully capture the changes in the workings of the international economy. In a world where a good deal of trade takes place between affiliates, where technology transfer is tightly controlled from corporate headquarters and where credit is extended by firms to their customers, corporate governance has become a much more prominent factor coordinating international economic relations. Innovation has also characterized financial institutions as a growing share of their activity is distributed across multiple locations and these institutions tend to offer a variety of services and undertake multiple activities to an extent that breaks with past practice.

Multilateral institutions, rules and arrangements have, over this same period, also gone through a series of mutations in line with a more open and liberal system of trade and capital movements. The International Monetary Fund (IMF), since the mid-1970s, has abandoned its objective of seeking to ensure stable exchange rates in an orderly international financial system and, instead, openly encouraged the opening up of capital accounts as a way of attracting international financial flows. The World Bank has also given up its emphasis on longer-term infrastructure project lending, concentrating instead on adjustment lending and poverty reduction. With the Uruguay Round negotiations, the governance of international trade has moved towards a single-tier system of rights and obligations, in which trade liberalization has been given priority over economic growth and full employment, and linked to a series of ‘related issues’ which take the multilateral rule-making process much deeper into the workings of national economies.

Trade liberalization and the investment climate

Conventional trade theory does not distinguish between different activities in terms of their differential impact on economic growth. In practice, however, most economic policy-makers worry about the composition of trade for this reason. There have been some positive trends in recent years. In particular, the sharp rise in the share of manufactures in exports from developing countries, from one-fifth in the early 1980s to around 70 per cent by 2004, points to an important change in global trade dynamics and one that most proponents of market-led globalization have been quick to highlight as a further measure of its success. However, this is hardly an
exhaustive picture of the changes in the trading system. In the first place, the rise of manufacturing exports is down to just a handful of countries, largely from East Asia. Indeed, simply taking out the first-tier newly industrialized economies (NIEs) from the developing-country group, their share of world trade drops from around 22 per cent in the early 1980s to below 15 per cent by the end of that decade, but has now (in 2005) climbed back to the earlier figure; however, taking out China leaves the developing countries with less than one-sixth of total world trade. By implication, other developing regions have exhibited a stagnant or declining share since 1980, with exports very heavily skewed towards natural resources; these account for more than 30 per cent of export earnings in over 80 countries, often from just one or two products. Together these trends imply that many countries remain stuck with undynamic trading regimes.

Another, and less reported, trend in the trading system is that while many countries have been trading more, they have been earning less from doing so. In a recent study of 127 developed and developing countries, Dowrick and Golley (2004) found that between 1960 and 1980, increased trade helped productivity to grow in poorer countries at double the rate in richer countries, but that this gain was reversed in the period of more open trade between 1980 and 2000, when the marginal impact of trade on productivity growth favoured the richer countries, and indeed turned negative for poorer countries. Wacziarg and Welch (2003), using the Sachs–Warner methodology for determining openness, found that its links to faster growth were period-sensitive, with much weaker links in the 1990s than in earlier decades, and with more open economies if anything benefiting less than relatively closed economies. Finally, increases in the developing countries’ share of world manufactured exports since 1980 have not been matched by a corresponding rise in their share of global value added, and in a number of cases shares of global manufacturing income have actually fallen over the past decade or so, even as their share of world exports of manufactures was rising, while for others it has risen by much less than that share.

One possible explanation for these trends is biases in the liberalization process which have prejudiced growth prospects in developing countries by discriminating against sectors where they had or could build strong export sectors, even as domestic producers were forced to confront strong competitive rivals on home markets (UNCTAD, 1999; ILO, 2004). However, the fact that many countries were trading more but earning less during the 1990s suggests more deep-seated biases in the operations of the trading system (UNCTAD, 2002). These seem likely to be linked to the lopsided reliance on external demand as the basis of sustained growth (Heintz, 2003).
The risk of falling export prices, resulting from too many producers chasing too few markets, has of course been familiar to commodity exporters, where returns have long been asymmetrically skewed through organized markets in rich countries and more disorganized and fragmented markets in poorer countries. However, the structure of some key markets for developing-country manufactured exports seems to point to similar pressures emerging even where trade expansion has been rapid (UNCTAD, 2002, pp. 121–4). In this respect, the suggestion that in the new era of globalization developing-country exporters of some high-technology products have been able to bypass technological constraints is often misleading. Even when the final product has been classified as high-tech, many developing countries have in reality only been involved in low-skill assembly activities, using imported capital and intermediate goods, and whose contribution to value added is determined by the cost of the least-scarce and weakest factor, namely unskilled labour.

The confusion over what constitutes a dynamic export is linked, in part, to the increase in FDI flows through international production networks (IPNs), in which corporations slice up their value chain, relocating or outsourcing the different parts, from product design to final delivery, in a way that enhances overall profits. In some cases, production is organized by large transnational corporations (TNCs) producing a standardized set of goods in several locations (as in electronics and transport industries). In others, production involves groups of small and medium-sized enterprises located in different countries and linked through international subcontracting (as in clothing). Under these conditions, FDI could bring macroeconomic benefits to the host through a positive impact on its investment climate and on its balance-of-payments position. Certainly the expectation after the debt crisis, and particularly following the Brady Plan, was that liberalization, particularly when export-oriented FDI was attracted, would bring such benefits (Camdessus, 1997).

Although the bulk of FDI continues to flow between advanced countries, developing countries have, indeed, experienced a fifteenfold rise in average annual inflows of FDI since the mid-1980s, which has translated into it taking a significantly higher share of capital accumulation in most developing countries and regions. Moreover, between 1990 and 2003, the share of manufacturing in the FDI stock of the group of developing countries rose from 25 per cent to 37 per cent while the share of developing countries in the global stock of manufacturing FDI increased from one-fifth to close to one-third during the same period. However, FDI surges have often failed to stimulate a domestic investment recovery and there is evidence that it has even crowded out local investment (Gosh, 2004). The extent to which these flows have eased the balance of payments constraint has varied
unpredictably with the share of TNC profits in value added, the degree of import dependence, and the proportion of the final good sold in domestic markets (Akyüz, 2004). Certainly where the propensity to import of foreign firms has been much higher than that of domestic firms, and their export propensities similar, increased FDI has added to balance-of-payments pressures (Chudnovsky and López, 2002). In fact, the broad body of evidence points to this constraint remaining a tight bind on economic growth in many developing countries (UNCTAD, 1999).

In the absence of favourable macroeconomic effects from attracting FDI, much depends on technological and other spillovers that might strengthen internal integration. The broad body of evidence suggests that such effects are plant-, sector- and country-specific, though generally significant effects seem to depend on local absorptive capacity already being in place. As such, FDI tends to lag rather than lead the growth process (UNDESA, 2006). From this perspective, production networks may well increase the risk of TNCs being attracted into enclaves with limited linkages to the domestic economy; taking account of the different methodologies employed, there is little evidence to date of positive spillovers from the recent surge in FDI to developing countries, and virtually none on a significant scale.

Financial liberalization and capital flows
Financial markets have undergone a dramatic transformation since the early 1980s, thanks to a combination of deregulation, internationalization and innovation. While the impulse for this transformation came from the advanced countries, economic logic promised much for the world’s poorest countries. Deregulated and open financial markets would not only increase the availability of investment finance, both domestic and foreign, but they would also help create a more stable and disciplined investment climate, and free deficit countries from the unpredictable politics of ODA flows (Camdessus, 1997; Mishkin, 2006).

An upsurge in flows began in the 1990s, albeit in part a return to trend after the blighted years of the 1980s. Still, a plethora of new financial instruments promised to mitigate risk, particularly in those emerging markets that greatly excited investors after the Berlin Wall collapsed, providing arbitraging opportunities but also encouraging herding behaviour to become a more significant influence on the direction of flows. Consequently, these flows were very unevenly distributed, increasingly concentrated in a small group of 20 or so emerging markets which received over 90 per cent of total inflows of capital in the 1990s, compared to some 50 per cent before the outbreak of the debt crisis. Thus, the share of low-income countries in net private capital flows to developing countries has been steadily declining
since the second half of the 1980s, from 20 per cent (around $6 billion) to just 6 per cent in the second half of the 1990s (around $14 billion), with an attendant rise in the share of middle-income emerging markets. In recent years more than three-quarters of bond issues have been accounted for by less than ten borrowers in Latin America and Asia and much of the syndicated bank lending has gone to half a dozen countries in Asia. These same countries were also the main recipients of international equity investment.

Despite the ever louder beating of the financial liberalization drum during the 1990s, including efforts to put capital account openness in the IMF’s Articles of Agreement, these flows proved increasingly difficult to manage in a way consistent with faster and more inclusive economic growth. For countries seeking re-entry into international financial markets after the debt crisis, higher real interest rates and a stable exchange rate were prerequisites for attracting new inflows. While the resulting financial stringency could prove attractive to foreign investors, a tight monetary and fiscal stance, oftentimes with an appreciating currency, did little to stimulate domestic investment or to improve export prospects. Indeed, increased debt-servicing obligations resulting from higher interest rates, along with import surges, ran the risk of reproducing an unsustainable debt burden. In many cases, a combination of capital outflows, profit remittances and the accumulation of exchange reserves greatly reduced net inflows, and of these a growing proportion was absorbed by activities which added little to productive capacity (UNDESA, 2005). Particularly in the form of short-term loans and portfolio equity, these inflows, often highly leveraged through derivative contracts and hedge funds, could be very unstable and an unreliable source of development finance.

As the workings of financial markets became increasingly disconnected from the longer-term demands of productive investment and industrialization, unregulated financial flows triggered boom–bust cycles, which became a recurrent feature of the developing world during the 1990s. The precise circumstances in which the vulnerability to the reversal of capital inflows arose, and the subsequent impact on growth, varied from region to region. An early warning was given by the Mexican peso crisis of 1994. However, the full force of unregulated financial flows was revealed by the financial crises in East Asia, a region with a long-standing record of strong growth and fiscal discipline. As in other episodes of financial crisis and currency turmoil, the crisis in East Asia was preceded by financial liberalization and deregulation which, in some cases, constituted a major break with past practice. Moreover, the extremes of collapse were amplified by unnecessarily tight monetary policies which deepened the debt deflation process, served to depress output and employment, and caused serious dislocations in the corporate and financial sectors (Stiglitz, 2002).
Global slowdown and instability
The debt crisis and its aftermath generated a lost decade for many developing countries in the 1980s, with incomes contracting in many cases. On some assessments this has given way to a much more benign macroeconomic climate with lower real interest rates, less-volatile growth in the leading markets and diminished inflationary threats. Despite this, the slowing global trend persisted for much of the 1990s. Indeed, even as recovery in the USA turned into more sustained growth from the mid-1990s and growth in China and India accelerated, the average global growth rate for the decade was still not above that of the 1980s, and remained below that of the 1960s and 1970s.

This slowing trend has been associated with a good deal less stability in the growth performance of many developing countries. Some of this (as in the case of transition economies) can be traced to political shocks. But, if history is any guide, it should not come as a surprise that a relaxation of regulations put in place in response to earlier excesses generates ‘manias, panics and crashes’ (Kindelberger, 1984), or that these would hit weaker economies the hardest. On one estimate, during the 1990s the financial system was in crisis for 40 out of 120 months (Plender, 2003, p. 57). According to Barry Eichengreen (2002), the period since the collapse of Bretton Woods has seen a sharp increase in the incidence of financial crises, principally in the form of currency crises but also in conjunction with banking crises. Most of these have been in the developing world.

The global slowdown and increased incidence of financial crises has coincided with a declining frequency of strong growth episodes and a rising frequency of negative growth episodes. According to Ocampo and Parra (2005), in the 1960s and 1970s about 40 per cent of developing countries had successful growth episodes (with annual average rates of per capita GDP growth greater than 3 per cent over at least a five-year period) but this proportion fell to less than 20 per cent through most of the past quarter-century, while in 40 per cent of countries there were negative growth episodes compared with just 15 per cent in the earlier period. Similarly, Hausmann et al. (2004) searched for episodes of rapid acceleration in economic growth that were sustained for at least eight years and found that while there were 23 and 30 such episodes in the 1960s and 1970s respectively, there were only 14 in each decade of the 1980s and 1990s.

Fragmented development
While the influence of radical market-based reforms on policy-making in advanced countries was heavily qualified by checks and balances in their political systems (Krugman, 2007), these, thanks in no small part to the support of the international financial institutions, quickly became a fully
fledged development blueprint showing how poor countries could extricate themselves from the problems of the debt crisis and establish a new growth path. It promised to remove structural and institutional impediments to growth, improve productive capacity and trade performance, and put an end to stop-go development associated with excessive indebtedness and periodic payments crises. Above all it promised strong convergence in the global economy as growth in poorer countries outpaced that in richer countries and income gaps across the world economy narrowed sharply. The previous section has raised some initial doubts about blindly trusting in international market forces and firms to achieve this outcome. This section looks in more detail at how the interplay of external and internal integration forces has played out across the developing world.

**Growth, gaps and international inequality**

Contrary to much contemporary rhetoric, the growth in cross-border trade and financial flows since the mid-1980s has failed to stop the growing gap in real per capita incomes between rich and poor countries. Economic divergence is, here, the simple product of the higher starting income of richer countries and their faster average annual per capita growth rates (Milanovic, 2005). This has happened despite the sharp slowdown in the advanced countries since the mid-1970s, a slowdown which, as noted earlier, has resulted in a general slowing of the world economy over the same period.

Such a high level of aggregation can, however, be misleading and there are important differences across regions within the developing world. Asia has persistently maintained a much stronger growth performance than other developing regions, enjoying catch-up growth with the advanced countries since the early 1990s, and for a good deal longer in East Asia. By contrast Africa, particularly south of the Sahara, and Latin America have continued to fall further behind. Given that, among developing countries, Latin America was the richest developing region by some margin at the end of the 1960s, subsequent growth trends have generated convergence across the developing world. There are also some noteworthy differences in growth performance over time, with developing countries outpacing advanced countries in the 1970s and again over the decade since the mid-1990s, with a lost decade sandwiched between in the 1980s when developing-country growth performance was below its own historical average and that of the more advanced economies.

In both popular and scholarly accounts, the actual and projected tilting of the world’s economic axis towards Asia is often taken as synonymous with globalization. In fact, that shift has been ongoing since the 1950s, when Japan entered a period of rapid and sustained growth, joined soon
after by the small Asian tigers, notably Korea and Taiwan, China.\textsuperscript{15} A second tier (of more variegated economies) from South-East Asia joined in during the 1980s. However, it has been the emergence of China and India, particularly the former, that has added a new dimension to the Asian catch-up story.\textsuperscript{16} While both are still a long way from being middle-income countries, they have a large number of citizens who would qualify as middle class. Their emergence has already had a significant impact on global growth dynamics as well as on the wider regional performance, with growth in South Asia, since 1980, improving sharply over the preceding two decades, and high growth rates maintained in East Asia in the 1990s despite the onset of a severe financial crisis in 1997, though the pace did not match that of the 1970s. There is also little doubt that the impact is being felt beyond the region, though just how the balance between their growing demand for natural resources and their addition to the world’s unskilled labour force will play out elsewhere is a subject of much dispute.\textsuperscript{17}

After the lost decade of the 1980s, Latin American economies enjoyed a brief renaissance in the early 1990s when the intensification of structural reforms enabled them to return to the international capital markets; but after 1997 developments again turned sour and produced another ‘lost half-decade’ (Ocampo, 2002). Growth performance in the 1990s was similar to that in the 1960s but well below that in the 1970s. Sub-Saharan Africa, like Latin America, also suffered a ‘lost decade’ of development in the 1980s, but with a weaker (if less erratic) recovery in the 1990s; there was generally little or no inflow of private capital into the region, and no abrupt reversal at the end of the decade.\textsuperscript{18}

Against the backdrop of a global slowdown, tight macroeconomic policies and the increasing frequency of financial crises, many developing countries have seen a weakening of growth relative to their own past performance. Of a total of 124 developing countries, growth in 95 of them (that is, over three-quarters) was faster in the period 1960–78 than between 1978 and 1998 (Milanovic, 2002), and only a handful of countries have been able to hit the kind of growth targets needed to address their economic and social deficits:\textsuperscript{19} between 1980 and 2000, of 140 developing countries only 20 grew at annual rates above 5 per cent, a number that rises to 30 for the period 1990–2000, but if a rate of 7 per cent is taken as the benchmark, the number of success stories falls to just five and six respectively.\textsuperscript{20}

Countries at the very bottom of the income scale appear to have lost most ground. This has lead some to talk of a ‘twin peaks’ global income distribution (Quah, 1996), with a ‘hollowing-out’ of the middle-income range of countries (UNDESA, 2006b, p. 8). Milanovic and Yitzhaki (2001), for example, estimate that just 8 per cent of the world’s population
Polarization can, in large part, be explained by the fact that the general slowing of global growth since the 1980s has hit poorer countries particularly hard. According to Milanovic (2005, p. 5), the average annual per capita growth rate of the group of LDCs was just 0.1 per cent between 1980 and 2002, compared to 1.9 per cent in the ‘old’ Organisation for Economic Co-operation and Development (OECD) economies, and while the former figure hides a very wide dispersion of performance, there were no stellar growth performance of the kind found in East Asia. Growth collapses were, not surprisingly, much more frequent among this group of countries than other developing-country groupings.

Given the close (though not direct) connection between income growth and poverty reduction, there have been some clear differences in poverty trends across regions over the past two decades. Overall the number of people living on less than $1 a day has dropped since the early 1980s by around 400 million. However, this headline figure hides very large falls in some countries, notably China, which pulls down the regional figure for Asia, along with little change or increases in almost all other regions. Indeed, the improvement in the world poverty figure can be attributed almost exclusively to China.

Just how income and poverty trends translate into a picture of global inequality is an ongoing source of controversy among economists. The Gini coefficient is often used as a more integral measure that aims to take all members of a chosen set into account, although measurement problems abound. On this measure, a number of studies have reported a reduction in international inequality since 1980, linking this in turn to a more open world economy. However, these findings are open to serious questioning on both methodological and empirical grounds. Indeed, as Sutcliffe (2006) has noted in his balanced assessment of the debate, there is little agreement on how significant the change has actually been or just when the reversal might have begun. More importantly still, the result hinges on the performance of a single outlier; removing China from the country set reverses the trend to one of rising international inequality, and even more sharply than before 1980. The Theil decomposition of international inequality shows a similar result (UNDESA, 2006, p. 14). Given its size, China’s performance is obviously central to the analysis of global trends. However, from a comparative perspective what happens in a single country (however large) should not be used to obscure what is in fact a highly variegated picture across the global economy (Berry and Serieux, 2004).

The variety of development experiences in the 1990s
A recent assessment by the World Bank (2005, p. 30) of its own policy advice during the 1990s has acknowledged that it persistently overestimated...
growth prospects in regions implementing adjustment programmes, and underestimated growth performance in those that were not. This is a belated acknowledgment of a good deal of academic research which finds little evidence that adjustment programmes have had a positive impact on growth, employment or poverty. However, there is a reluctance to accept that the adjustment path that these policies helped fashion has actually set back development prospects, in many cases by distorting the process of internal integration, focusing instead on what might have been left out of the package in terms of social policies, good governance, institutional detail, and so on.

In fact, regions that pursued extensive adjustment programmes in the 1980s found it particularly difficult to reverse the sharp drop in the share of investment triggered by the debt crisis. In some countries, this dipped to below the levels needed to replace depreciated capital, and where adjustment programmes persisted, the investment cycle remained volatile, even after the immediate disturbances of the crisis subsided. The downsizing of public investment has been a prominent part of this story, and in many cases this has crowded out domestic private investment. But, as important, has been the approach to macroeconomic fundamentals guiding policy design, which has ignored the mutually reinforcing links between aggregate demand, capacity utilization and investment. Tight monetary conditions and accompanying currency gyrations have further discouraged growth based on capital accumulation, adding to the difficulties encountered by domestic industry in introducing the kind of restructuring that would help raise productivity, even as it faced stiff competition from lower trade barriers. In part as a consequence of these difficulties in the productive sectors of the economy, and in part as a result of ongoing liberalization of the financial sector, ‘rentier’ investments (including in government bonds) and ‘capital flight’ have become much more attractive options. In many cases this financialization of the domestic investment regime has coincided with and reinforced commodity dependence, often linked to increased FDI inflows (Magalhães Prates and Paulani, 2007).

Under these conditions, the cumulative nature of industrial development can very easily go into reverse gear. Certainly, where industrial stagnation was the norm in the 1980s, as in Latin America and Africa, most countries found it difficult to reverse the trend in the 1990s, with premature deindustrialization a visible trend in some cases (UNCTAD, 2003, pp. 92–9). Even where this has not been a dominant trend, there is little sign of the diversification and upgrading which describes a healthy internal integration dynamic. This is often associated with enclaves of specialized development. In the case of many Latin American economies these enclaves have been in the primary sector, though in some cases manufacturing activities
have been able to carve out niche markets. Where productive capacities are even more rudimentary, as in much of sub-Saharan Africa (SSA), the likelihood of enclave development in the primary sector generating a natural resource curse seems to be even greater.

Moreover, in economies with declining shares of investment and manufacturing value added, a stagnant or falling share of manufactures in total exports was often the norm, even as the overall composition of developing-country exports was shifting rapidly towards manufactures, including more skill- and technology-intensive goods. Labour market performance, key to tackling poverty, has also been caught up in these divergent developments. The impact of liberalization on labour market performance is a complex matter, with the effects differing between countries depending on the domestic and international conditions prevailing when liberalization occurred, as well as on the broader development strategy being pursued by policy-makers. However, a good deal of evidence from Latin America and sub-Saharan Africa, where neoliberal policies have been pursued most intensively, suggests that increases in unemployment and/or increasing wage inequality have often accompanied rapid trade liberalization. In many cases, adjustment policies and the downsizing of the public sector have led to a hollowing-out of the middle class. The counterpart of these trends has been an expanding informal economy, which by the end of the 1990s accounted for anywhere between one-third and three-fifths of the labour force in Africa and Latin America (Schneider, 2002).

By contrast, the countries in East and South Asia that bucked the declining investment trend in the 1980s were able to maintain a reasonably stable or even rising pace of capital accumulation for most of the 1990s, with the share of public investment holding up or rising further (UNCTAD, 2003, pp. 65–73). But just as importantly, strong investment has supported structural change. In countries where investment and industrial output expanded in the 1980s, this continued in the 1990s. In these cases rising and fully utilized capacity levels have had a knock-on effect on productivity growth through both the demand and supply sides. Indeed, Asian development since the 1970s confirms the key role of industrialization in establishing a rapid growth path, with positive cumulative effects through strong export drives linked to rising levels of productivity (UNDESA, 2006). During the 1990s, just eight East Asian countries accounted for 70 per cent of developing countries’ trade in manufactures. Outside this group, export strategies relied on low (and in some cases falling) wages or currency depreciation rather than strong productivity growth, and while this stimulated recoveries in some countries, few were able to reach a threshold level of exports consistent with a vibrant industrialization path (UNCTAD, 2003, pp. 99–102).
Taking stock of trends since the debt crisis of the early 1980s, it is clear that by itself the degree of exposure to global market forces is not what distinguishes ‘winners’ and ‘losers’. Rather, the main difference, particularly between the East Asian NIEs and most other developing countries, was that liberalization followed the successful implementation of industrial and trade policies; protection and support were removed in large part because they were no longer needed. In the latter, on the contrary, liberalization has largely been triggered by the failure to establish efficient, competitive industries in labour- and/or skill-intensive sectors. From this perspective, the different ways in which internal and external integration have taken shape in developing countries have failed to trigger convergence since the early 1980s, but have instead created a very uneven economic landscape of diverse experiences.

**Mature industrializers** This group includes the first-tier NIEs, notably the Republic of Korea and Taiwan Province of China, which achieved industrial maturity through rapid and sustained accumulation of capital, and growth in industrial employment, productivity and output, as well as manufactured exports. In the 1990s these economies enjoyed a share of industrial output in GDP above the levels of advanced countries, exports had shifted to more capital- and technology-intensive goods, and industrial growth was starting to slow down as resources shifted towards the service sector.

**Rapid industrializers** A number of countries saw a rising share of manufactures in total output, employment and exports, based on strong investment in resource-based and labour-intensive activities, and were beginning to upgrade to middle-range technology products. This group included the second-tier Asian NIEs, but also isolated success stories from other regions, as well as the waking giants of China and (albeit more cautiously) India.

**Enclave industrializers** Some countries moved away from dependence on commodity exports by linking to international production chains, often by attracting large amounts of FDI and with heavy reliance on imported inputs and machinery. Export growth was often very fast, as in the Philippines, Mexico and, to some extent, Morocco. However, overall performance in terms of investment, value added and productivity growth was often quite weak.

**Premature deindustrializers** This group included most countries in Latin America, which had achieved a certain degree of industrialization but were unable to sustain a dynamic process of structural change through rapid
accumulation and growth. In a context of rapid liberalization, declining shares of manufacturing employment and output and a downgrading to less technology-intensive activities were common trends.

**Commodity-dependent exporters** Many poorer economies, particularly in sub-Saharan Africa, remained heavily dependent on one or two commodity exports. In the face of relatively stagnant markets, volatile prices and declining terms of trade, investment dropped further, diversification stalled and productivity remained stagnant. In some cases enclaves of faster export growth emerged in the extractive sectors, usually tied to FDI, but with weak linkages to the rest of the economy. However, some wealthier developing countries, notably Chile, did achieve a faster pace of investment and growth based on their natural resource endowments.

**Conclusion**
The historical experience of advanced countries, including recent graduates in East Asia, establishes that a broad and robust domestic industrial base remains key to successful development, because of its potential for strong productivity and income growth. Success hinges on building such a base from local conditions and overcoming local constraints on its expansion and evolution. This is still the big policy challenge for most developing countries. The experience of the 1980s and 1990s suggests that the policy direction launched in many parts of the developing world after the debt crisis failed to establish flourishing alternatives. Despite the widely shared belief that a more open economic environment would demonstrate the benefits of unrestricted capital mobility and the superiority of markets over government intervention, the period since the collapse of the Bretton Woods system has instead been marked by very heavily concentrated flows to comparatively wealthy countries and by an increasing incidence of financial crises, and their growing virulence in terms of lost output and jobs. Recent efforts to salvage the strategy have emphasized omissions: not enough good governance, not enough market access, not enough FDI or aid. Doing so continues to overlook the damage to macroeconomic growth fundamentals and to the direction of structural change that have accompanied the turn to neoliberal policy reform.

By contrast, success stories in the 1990s built steadily on improving performance established in the 1980s or earlier. In most cases they followed what Birdsall et al. (2005) have called ‘heterodox gradualism’, using an array of policy options to manage integration into the global economy and ensure that more of the value added linked to trade stayed at home. Doing so involved experimenting with a range of more strategic measures to encourage strong capital formation, expand domestic markets and support
technological upgrading. These measures have been rooted in specific institutional settings reflecting national political and social cultures and consistent with the bureaucratic and entrepreneurial capacities of local elites. In their different ways, all have eschewed a softening of the state and instead premised their structural transformation on a harder development state, that exhibits both ‘adaptive efficiency’ and the effective creation and utilization of ‘policy space’ (Kozul-Wright and Rayment, 2007). These same features will need to be recovered in countries that have fallen back since the debt crisis, if catch-up growth is to be reignited. This will require active policies, particularly on such matters as industrial support, technological progress and public infrastructure, all of which will have to be tailored to the particular circumstances of the countries concerned.

Notes
1. In an article deploring the slow progress in the Doha round of trade negotiations, the Financial Times journalist Martin Wolf described French President Chirac’s scepticism as to the virtues of unlimited free trade as ‘foolish, even depraved’ since if the negotiations ‘should fail, disorder alone should triumph’, Martin Wolf, ‘The World has Everything to Lose if Trade Liberalisation Fails’, Financial Times, 2 November 2005.
3. The Nobel Prize-winning economist Douglas North (1994) has candidly admitted that the aversion of neoclassical economics to historical processes and structural discontinuities precludes a proper understanding of the development process.
4. These regularities are associated with the empirical work of Verdoorn, Lewis and Kaldor. For a review see Toner (1999). See also Rodrik (2006) for a recent assessment of why manufacturing still matters to the development process.
6. This conclusion applies, we believe, to much of the recent discussion of the importance of the ‘investment climate’ in developing countries.
7. Dowrick and Golley call for further research to understand these findings. Their own tentative suggestions are that the nature of technology transfer through MNCs has changed in the latter period and that the range of complementary policies that supported successful liberalization in the earlier period have gone missing in the ‘one policy fits all’ approach of the latter period. Both suggestions are in line with the arguments presented in this and subsequent chapters.
8. Certainly, the period of rapid export expansion in the advanced countries in the three decades after World War II coincided with strong wage growth and the development of deeper domestic consumer markets (Armstrong et al., 1984). However, intra-industry trade, strong productivity growth and high rates of capital formation were not the automatic outcome of market forces in these countries but were closely interrelated components of a politically fashioned socio-economic compromise.
10. Such networks are not a new development, dating back to the 1960s in parts of East Asia, and becoming a more prominent feature of the international division of labour in the 1970s, see Helleiner (1973) and Henderson (1991).
12. The annual capital inflow in the 1990s was around 5 per cent of GNP, which was roughly the level prevailing in 1975–82. If China is excluded, the ratio is actually lower than in the earlier period by one percentage point.

13. Although the call for financial liberalization was heeded across the developing world in the 1990s, the majority of countries, and particularly those in Africa, attracted little private flows, and certainly not enough to offset declining aid flows during the 1990s. FDI was also very unevenly distributed: three-quarters of the total in the 1990s went to just ten emerging market economies, and China, Brazil and Mexico together took nearly one-half. Indeed, China accounted for about one-third of all FDI in the developing countries in the 1990s and about two-thirds of the total went to East Asia. Other developing regions – sub-Saharan Africa, South Asia, North Africa and the Middle East – have only marginal shares of the total and in the case of sub-Saharan Africa a falling one, mainly in extractive sectors or in public utilities.


15. The city states of Hong Kong and Singapore are fascinating stories in their own right. However, it should be noted that their per capita incomes in the early 1950s were comparable to Japan, and higher than much of the European periphery. Recognizing this goes a long way in dissociating their transformation from the current phase of globalization.

16. China’s break with its earlier, highly erratic growth path occurred in the mid-1970s and has been building strength since the early 1980s. There are some significant differences between these two awakening giants, notably the much slower pace of urbanization in India, and with this a much slower pace of industrialization and capital accumulation. There is a good deal of dispute among economists as to which of the two development paths is likely to be the most sustainable.


18. The recent pick-up in African growth rates is largely a result of an increase in commodity prices linked in no small part to growing demand in China. According to a recent IMF Working Paper, a dozen middle-income and oil-producing African countries have been the principal beneficiaries of a stronger growth performance since 1997, see Tahari et al. (2004).

19. Obviously in light of what has already been said it would be wrong to identify any hard and fast benchmark for all developing countries. Still, many regard a 6–8 per cent target as what is needed to tackle those deficits and visibly to close income gaps with the more developed countries.

20. Identifying potential growth rates, and the related challenge of suggesting growth targets, must be seen as a somewhat hit-and-miss business. But assuming that a country’s labour force is growing on average at 2–3 per cent per year and that productivity growth needs to match that rate to maintain internal and external balance, then 5–6 per cent would be a minimum growth rate for GDP. Many, including ourselves, would regard a 7–8 per cent target as more appropriate if the policy objective is to make real progress in tackling the massive social deficits that have built up in most developing countries, and to start closing the income gaps with the more developed countries.

21. The Gini coefficient derives from the Lorenz curve which shows the cumulative share of the income received by the cumulative shares of the population, starting from the poorest income-receiving units. The coefficient measures the area between this curve and the diagonal of perfect equality and varies from 0 (maximum equality) to 1 (maximum inequality or from 0 to 100 when expressed in percentages. Milanovic’s review of the literature notes that there is little disagreement on the degree of international inequality, with most studies finding a Gini coefficient in the range of 63 to 68, a figure that is higher than the coefficient in highly unequal countries such as Brazil or South Africa (Milanovic, 2006, pp. 140–42).

23. See further UNCTAD (2003, pp. 74–6), also Ramirez and Namzi (2003).

Bibliography

Agosin, M. and R. Mayer (2000), ‘Foreign Investment in Developing Countries: Does it 
Crowd in Domestic Investment?’ UNCTAD discussion paper no. 146, United Nations 
Conference on Trade and Development, Geneva.
International Monetary and Financial Issues for the 1990s, Vol. 11. United Nations publication, 
New York and Geneva: UNCTAD.
Akyüz, Y. (2004), ‘Trade, Growth and Industrialisation in Developing Countries: Issues, 
Experience and Policy Challenges’, Penang, Malaysia, UNDP Asia-Pacific Conference on 
Trade.
Series, no. 38, Geneva, UNCTAD.
World Development, 24 (3).
Akyüz, Y., H. Flasebeck and R. Kozul-Wright (2005), ‘Globalization, Inequality and the 
Amsden, Alice (2001), The Rise of the ‘Rest’: Challenges to the West from Late Industrializing 
Economies, New York: Oxford University Press.
Arbache, Jorge S., Andy Dickerson and Francis Green (2004), ‘Trade Liberalisation and 
Wages in Developing Countries’, The Economic Journal, 114.
223–51.
University Press.
Berry, A. and J. Serieux (2004), ‘All about the Giants: Probing the Influences of World Growth 
Advanced Countries’, Economic Series Working Papers 99144, University of Oxford, 
Department of Economics.
Birdsall, Nancy, Dani Rodrik and Arvind Subramanian (2005), ‘How to Help Poor 
Countries’, Foreign Affairs, July/August.
Institute, Washington, September.
Braunstein, E. and G. Epstein (2004), ‘Bargaining Power and Foreign Direct Investment in 
China: Can 1.3 billion Consumers Tame the Multinationals?’ in Will Milberg (ed.), Labor 
and the Globalization of Production: Causes and Consequences of Industrial Upgrading, New 
speech to Los Angeles World Affairs Council, 17 June.


ILO (2004), A Fair Globalization: Creating Opportunities for All, Geneva: ILO.


Introduction

Economic development in Latin America, finally, seems on the verge of beginning again. After close to 25 years of poor performance, despite dramatic policy changes in fiscal policy, and hence inflation, and expansion in international trade, the region seems to be expanding. Results both in 2004 and 2005 have been better than initially foreseen. The key question is whether this process will continue over the next decades, or whether the current advance will soon collapse – as it has done in the past – with declining commodity prices, limited foreign demand for exports, less foreign investment and more inflationary domestic policy.

Much has altered within the region’s economies since the 1990s. Domestic politics have also changed profoundly, as democracy has continuously spread and deepened. Yet a high degree of political dissatisfaction has evolved, and successive elections have accurately reflected the population’s unhappiness. Although inflation rates are now virtually at a record low, benefiting substantially the bottom third of the income distribution, people do not vote for those that promise fiscal discipline. The Left has run well, and won, virtually everywhere since the beginning of the 21st century. But, with the exception of Hugo Chavez, President of Venezuela, and possibly Nestor Kirchner, President of Argentina, policies have remained responsible and even responsive to the international markets. Evo Morales’s election in Bolivia introduces a new element, control over the trade in drugs, while recapitulating once more the question of state control over energy resources that has grown in significance in recent years with the rising price of petroleum in world markets.

These contemporary events, both economic as well as political, are better understood within a historical context. Latin America was a latecomer to continuous economic expansion. Only since the last decades of the nineteenth century did economic growth become a regular process. That was the era when the region began its increasing participation in external markets, taking advantage of its natural resources, and simultaneously sought to implement, late, its own industrial revolution. That is where we start the story in the next section. It extends through the Great Depression, with its dramatic decline in external demand, up to the recovery of the world economy at the end of World War II.
Next, we analyze the rise and decline of import substitution in the three following decades until the debt crisis that began in 1982. That history simultaneously comprehends a surge of military dictatorships within virtually the entire region. At the same time, that experience encompasses a period of above-average growth in Latin America. Finally, this interval also represents growing dependence on external finance to finance the increased domestic investment that drove expansion. This combination proved too fragile to survive the successive oil shocks of 1973 and 1979.

In the next section we take up the economic performance of the region during a subsequent period of profound retrenchment, altering the previous large role of the state almost everywhere. Fiscal deficits were sharply curtailed, much privatization of state-owned activities occurred, and protection against international imports was sharply curtailed. Latin America in the 1990s gave signs of joining Asia and Eastern Europe in rapid expansion. But economic development soon proceeded much less rapidly than had been hoped, giving rise to increasing criticism of the Washington Consensus and sparking a clear move to the political Left that continues today.

A final part takes up the central problems and policy issues that confront virtually all the countries of the region today, in the midst of a record number of presidential elections. Politics and economics are intrinsically linked. That combination sometimes leads to an inability to focus effectively upon the longer run. But Latin America needs to continue to change if it is to keep up with global advance. It requires more and better educational and health outlays, both to improve the quality of the labor force as well as to improve the lamentable degree of inequality that plagues almost all countries. Macroeconomic reform has to continue, both in governmental tax collection as well as more efficient expenditure. And that subject necessarily includes the indispensable increase in domestic savings and investment. A third matter is the continuing increase, as well as diversification, of export earnings and thus reduced sensitivity to variation in external demand. Finally, the region will have to recognize the continuity of present economic policy as a virtue rather than a vice, even as the electoral process becomes more inclusive and civil society flourishes. Institutional change, making permanent these structural advances, requires continuing attention and additional major effort.

A brief conclusion extends the lessons of the Latin American and Caribbean experience to the broader agenda of contemporary development economics.

**Boom, bust and war recovery: 1900 to 1945**

Latin America, despite the achievement of political independence early in the nineteenth century, missed out on the initial Industrial Revolution. As
a consequence, overall, its standard of living probably improved little between 1820 and 1870. Those were years of political instability and modest engagement in international trade. Only during the period 1870 to 1913, when an initial era of globalization fueled economic activity, did the region expand its relative export position, helped by an increasing number of immigrants and much greater foreign investment. That investment underwrote great expansion of the rail infrastructure and construction of port facilities to facilitate export of wheat and meat, coffee, sugar, rubber, nitrates and other primary products demanded by the United States and Europe.

Such were the mounting receipts from exports after 1900 that not only Argentina returned to the gold standard, but Brazil and Chile also joined, then still a signal of successful developing-country adaptation within the expanding world economy. The region impressively focused on exporting primary products – something of which about 97 percent of exports were classified as such in 1913 – even while high levels of tariff protection assured a growing market for domestic industrial production. Import substitution generally began in the larger countries before the Great Depression of the 1930s, focusing on the elemental, but substantial, areas of foodstuff and textile and clothing production.

What is important to note is the great difference in incomes among countries of the region just before World War I. Data are presented in Table 65.1. Per capita income in 1913 in Argentina was comparable with that of Western Europe as a whole, and some three-quarters of that in the United States. Chile followed, about a third below Argentina’s pace-setting level. The rest of the region trailed much farther behind. At the very bottom was Brazil, burdened by the poor north-east that offset the rising, and more diversified, economy to the south.

These regional income differences reflected a variable commitment to exports, as can easily be seen in the same table. Argentina and Chile were the leaders in their share of income generated by exports. They were, as well, producers of non-tropical exports more in competition with developed countries than the products from other parts of the region, and hence, as W.A. Lewis has shown, able to gain higher incomes. Both of these countries also benefited from high levels of capital inflow from abroad, contributing to higher levels of domestic investment. Finally, the labor force in the case of both countries was more literate than elsewhere, and the movement toward local non-agricultural activities was equally more substantial. Bulmer-Thomas runs a regression for 13 countries, relating economic growth to exports over this interval, and finds a high degree of explanation; all seven of the countries separated correspond well.
The World War in 1914 had variable effects within the region.¹ No longer is it viewed as an unmixed blessing. The sudden reduction of imports that the conflict imposed did offer opportunities for domestic producers in several countries in the industrial sector. Yet capacity for production of capital goods was still rudimentary, and that constraint limited possibilities for significant output gains. Foreign investment from Europe ceased, and that of the United States was still directed primarily to the north of the region. Argentina, in particular, disappointed. With lower imports almost everywhere, governmental receipts flagged, and inflation began to expand in many countries. What helped was the relative brevity of the conflict, allowing many of the countries to go back to what they had been doing before: exporting their primary products.

Thus the 1920s featured a return to previous history. Imports of manufactured products went up, as a consequence of rising primary exports. Even Chile again managed to expand its sales of nitrates by 1929 to three-quarters its 1913 real level. National performance, as before, was related closely to success in international trade. There was a new entrant with much larger income upon the stage, Venezuela, whose petroleum exports found an expanding market within the United States. Colombia and Peru also improved their positions.

### Table 65.1 Latin America before World War I

<table>
<thead>
<tr>
<th>Country</th>
<th>Per capita income 1913</th>
<th>Growth per capita 1900–1913 in %</th>
<th>Exports as % of income</th>
<th>Literacy %</th>
<th>% Labor force agriculture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>3797</td>
<td>38</td>
<td>0.36</td>
<td>63</td>
<td>34.2</td>
</tr>
<tr>
<td>Brazil</td>
<td>839</td>
<td>19</td>
<td>0.3</td>
<td>35</td>
<td>66.7</td>
</tr>
<tr>
<td>Chile</td>
<td>2653</td>
<td>36</td>
<td>0.54</td>
<td>56</td>
<td>37.7</td>
</tr>
<tr>
<td>Colombia</td>
<td>1236</td>
<td>27</td>
<td>0.13</td>
<td>41</td>
<td>70.5</td>
</tr>
<tr>
<td>Mexico</td>
<td>1467</td>
<td>27</td>
<td>0.17</td>
<td>63</td>
<td>63.7</td>
</tr>
<tr>
<td>Peru</td>
<td>1037</td>
<td>27</td>
<td>0.24</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Venezuela</td>
<td>1104</td>
<td>34</td>
<td></td>
<td>30</td>
<td>72</td>
</tr>
<tr>
<td>Seven countries</td>
<td>1134</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relative to USA</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sources:** Per capita income and growth: Angus Maddison, *Monitoring the World Economy*, OECD Development Centre, 1995
Exports as % of income, literacy and % labor force: *An Economic History of Twentieth Century Latin America*, ed. Rosemary Thorp, Enrique Cardenas and Jose Antonio Ocampo 2000, Vol 1, p 26

The World War in 1914 had variable effects within the region.¹ No longer is it viewed as an unmixed blessing. The sudden reduction of imports that the conflict imposed did offer opportunities for domestic producers in several countries in the industrial sector. Yet capacity for production of capital goods was still rudimentary, and that constraint limited possibilities for significant output gains. Foreign investment from Europe ceased, and that of the United States was still directed primarily to the north of the region. Argentina, in particular, disappointed. With lower imports almost everywhere, governmental receipts flagged, and inflation began to expand in many countries. What helped was the relative brevity of the conflict, allowing many of the countries to go back to what they had been doing before: exporting their primary products.
But there were also some important differences. Two stand out. First was the reality of major external change. Globalization, and British leadership, was a phenomenon of the past. There was increasing financial dominance of the United States in the world, and also at the regional level. Much of the renewed surge in investment flows went to local and state sources instead of, as earlier, to national governments. As it would soon turn out, the new American investment banks competing for business were more effective in finding a supply of finance than identifying productive applications of the money. The weakened position of Europe had consequences not only for investment, but also for international trade. The gold standard had vanished, leaving individual countries confronting hyperinflation and internal political disruption. The League of Nations, an attempt at supra-national economic guidance, provided little assistance in the midst of these new circumstances.

Second, the renewed rise in export markets in the 1920s was weaker than it had earlier been. Countries varied in their ability to cope. Brazil, for example, despite losing out in exports of rubber and facing a slowing market for coffee, managed to continue a growth predicated upon continuing diversification of its industrial structure. Others, like Argentina and Chile, saw lesser rises in expansion in the period from 1913 to 1929 than they had previously. This modest deviation from the historic export-oriented model of growth was, of course, soon to become the beginning of a new commitment to import substitution after the start of the Great Depression.

Much of Latin America definitively changed its development style in the 1930s. There was little option. Prices of the region’s raw material exports plunged with declining demand. Quantities also fell. This time, unlike the earlier decline after World War I, recovery did not occur within the industrial center. As a result, it was necessary to cease full payment of interest on the public debt, which had much risen as a proportion of export earnings. Exchange rate controls were applied to limit imports of consumption goods capable of being supplied internally. Real exchange rate devaluations occurred almost everywhere. They most frequently took the form of multiple exchange rates enhanced by direct controls.

The deterioration of the balance of payments meant a simultaneous decline in government revenues: some countries were still dependent upon import duties and export taxes for as much as half of their total receipts. As a natural consequence government deficits showed a tendency to rise, financed through increases in the money supply, and thereby stimulating domestic demand. Inflation did not increase much as domestic supply instead responded well to the lack of import competition. Special circumstances, such as the potential Brazilian civil war with a seceding Sao Paulo
in 1932, or the Leticia War between Colombia and Peru in 1933–34, occasionally helped in establishing such proto-Keynesian policies and contributed to earlier and stronger recovery.

With a rise in export earnings after 1932 for most countries, rising imports of intermediate and capital goods became available to satisfy increasing demand for such inputs as could not be supplied locally. These sometimes consisted of used machinery rather than that newly produced. Labor migrated to the cities to meet the demands. For this period, unlike the later decade of import substitution in the 1950s, production increases were dependent upon increased labor, rather than capital, input. In some instances, extra shifts were employed to satisfy an expanding market. Noteworthy is the greater rise in industrial output than in gross domestic product that occurred virtually everywhere in the region.6

Such expansion was not without a negative side. Military leaders emerged in many countries in response to the new economic challenges being confronted. Constitutions were rewritten, or reinterpreted. Authoritarianism had a preview before its re-enactment in the 1960s and 1970s. There was an increasing degree of state intervention and regulation, not always efficiently. Markets were not again to operate in an unfettered fashion as they had somewhat done during the years before 1930.

With the approach of a new war in the late 1930s, economic conditions began to worsen slowly. Current dollar trade growth reversed after 1937, giving further stimulus to domestic production where it was possible. When the European war emerged in 1939, Latin America, like the United States, stayed out. With Pearl Harbor in December 1941, and United States engagement, several of the countries declared war. But only Brazil was actively engaged in the conflict, with a detail of some 25,000 troops. Indeed, the southern cone countries, with a long and strong German influence upon their military, were late in their formal espousal of the Allied side. That indecision did not affect their trade flows much. Between 1940 and 1945 there was a major reversal: all countries suddenly found their exports dominantly flowing northward to the United States, with portions going to other neighboring countries within the region.

The war saw slower expansion in Latin America as a result of much lesser imports. Reserves accumulated, substantially in some countries. Inflation too accelerated domestically, leading to real appreciation of unchanged nominal exchange rates. When peace returned, repressed demand and favorable prices led to substantial import flows. There was now a domestic sector in many countries resistant to such foreign competition and a return to the experience of the 1920s. Not surprisingly, within the region a new group of young economists emerged calling for a conscious strategy of import substitution. They soon took form around the Economic
Commission for Latin America (CEPAL) in Santiago under the leadership of Raul Prebisch. Their views were given powerful reinforcement by the failure of the Latin American countries to obtain their own Marshall Plan, as had the European countries, thereby assuring integration into the wider global economy.

Table 65.2 sets out the growth experience that the principal regional countries had achieved since 1913. What these data reveal are three characteristics. First is the extraordinary rise in income per capita of Venezuela: it moves from a third of the Argentine level in 1913 to one that exceeds it by half as much again. Note as well that the second-largest rate of advance is attained by Peru. That provides the basis for those countries’ choice not to move to import substitution. Second is the significant slowing in the performance of Argentina and Chile, whose leadership had been so predominant earlier on. Third is the relative rise of Brazil, Colombia and Mexico as a result of the move toward domestic industrial production.

But, ultimately, and alas, there is the unfortunate regional relative aggregate. It moves from 29 percent of the per capita level of the United States in 1913 to 27 percent in 1950. Even with the better performance of Latin America during the 1930s, the rapid recovery of the United States during the war years and immediately thereafter gave proof to the continuing circumstance of regional backwardness.

**Post-war growth**

Latin America, or at least a goodly number of countries in the region, launched itself toward a continuation of industrialization, impelled by a
more active state policy in the 1950s. Helped by the rise in primary product prices in the wake of the Korean War, and thus sufficient export receipts, countries – both large and small – imposed high tariffs, quotas and even multiple exchange rates, to permit giving substantial incentives, and subsidies, to the intermediate and capital goods sectors where imports retained a large percentage of the domestic market. As the world gradually moved away from the limitations to world commerce imposed during the Depression, Latin American countries largely ignored the opportunities afforded for greater diversification as well as market expansion of their traditional export products.

In this attempt, most countries badly failed, including the two regional leaders, Argentina and Chile. Three countries, Mexico, Brazil and Colombia, managed the process better during the 1950s. They had started farther behind, with large internal markets, and managed to focus on less-inefficient undertakings. They also benefited from lesser balance-of-payments shocks during the expansion of the 1950s, assisted by more foreign investment. Two countries, Venezuela and Peru, continued their past emphasis on petroleum and mineral exports, respectively, and emerged well.

What characterized the CEPAL model were three marked deviations from standard economic theory. First, the state role in the process of economic performance was much enhanced. Priorities were a matter of public policy: planning groups were established, production was directly encouraged, special tariff protection was granted, subsidies appeared, and so on. Matters during the 1950s in Latin America were less formal than the Indian Mahalanobis scheme of targeted expansion, or the planned efforts of the Soviet Union and its followers in Eastern Europe. Nonetheless, they represented new and important interventions, then and subsequently, designed to achieve desired results within the industrial sector.

Second, at the macroeconomic level, there was clear understanding of the recurrent need for external support of the balance of payments. International markets would not by themselves, in the short run, absorb a sufficient quantity of primary exports to support the large quantity of imports of intermediate and capital goods required to allow import substitution to proceed. Neither, in the midst of the Marshall Plan and World Bank efforts directed toward Western Europe, were there public funds available. Foreign direct investment, although frequently criticized as part of the greater degree of nationalism characteristic of the period, helped in two ways: it provided the needed finance, but also a direct access to the newer technology needed for more advanced stages of industrialization to proceed. This tension was to continue through subsequent years.

Third, at the microeconomic level, all efforts to enhance industrial expansion were provided. Public investment in urban centers occurred to
meet the needs of a rapidly expanding population, many now coming from the rural areas. New highways were built, as was occurring in the United States and Europe, to extend the transportation network. Older railways, established to facilitate primary exports, languished. The national political balance was frequently transformed in a variety of countries as this new effort was launched: labor unions became more relevant, and wage policy frequently surged to the national level for decision. At this time, minimum wages were a new instrument of government policy.

These characteristics, and their strong element of directing the market, rather than reacting to it, stood in sharp contrast to the capitalist style of earlier export orientation. But the import-substitution model also gave rise to important internal contradictions that very soon – almost within a decade – represented the beginning of the end.

One of the problems with the strategy was exactly its focus upon encouraging domestic production to substitute for continuing imports. Alas, that necessarily required still other imports of capital goods and other inputs in order to accomplish it. So net savings of foreign exchange were not as great as initially were conceived. Frequently, imports were initially permitted as an incentive to establish the industry domestically. Moreover, to ensure that those imported inputs were inexpensive, an overvalued exchange rate was the rule. In turn, this ensured reliance only on export of the basic primary commodities that had been present from the start. There was no motivation, nor was there initial efficiency, enabling exports to emerge from the new manufacturing firms that had been established. Instead, there was great emphasis on import protection, equivalent to a tax on the few agricultural exports able to compete. What saved the day, as noted above, was foreign direct investment and thus greater dependence on the outside world instead of the independence so much vaunted.

Another difficulty encountered was a rising government deficit, soon resulting in higher rates of inflation. Government expenditures and employment rose, while taxes lagged behind. An increased money supply was the solution. A debate ensued in most countries, between the structuralists – seen to be progressive – and the monetarists – conservative and crotchety. It is true that one can transfer some resources to the government sector through an inflation tax of limited amount. At that time, advocates of a bolder governmental strategy were speaking only of something like price increases of 10 to 15 percent, a rate that was soon exceeded in most cases. The problem was that the process of inflation acceleration was auto-generated. Continuously larger price increases – to make possible the needed transfer – were necessary as the public reduced their holdings of money in order to limit the loss of income. Another, and important, consequence was deterioration of the income distribution. Those at the bottom
of the income distribution, half and even more of the population, were affected negatively; they had no alternative but to use cash whose value began to deteriorate more rapidly.

Additionally, the very success in establishing industry had its negative counterpart in the lack of expansion of traditional sectors. Agriculture and mining lagged behind, unaccompanied by application of new technology or capital applied to enhance productivity. Because industry was capital-intensive, the growing labor force could not find increased employment there, but rather in urban services and governmental employment. Additionally there were pressures placed on agricultural producers to keep prices of foodstuffs low for the expanding urban areas. A sectoral imbalance emerged, one that prejudiced policy in many countries. Argentina is perhaps the best example of this, with political appeals to the military to prevent relative deterioration of agriculture, whose large exports of foodstuffs were regularly impeded and reallocated to domestic consumption. Everywhere, internal relative prices favored industry and penalized agriculture.

By the end of this first decade, even CEPAL could see that its initial hopes were unrealized, and that some change in policy was necessary. Raul Prebisch opted for a slightly altered strategy, focusing this time upon import substitution at the regional level. This, by allowing trade with neighbors, extended the basis for a viable industrialization through economies of scale. Despite the creation of a Latin American Free Trade Area in 1961, that undertaking made scant progress at the time. There were ambitious, but fortunately unrealized, plans for new geographic specialization in a variety of industries. Later, in 1967, the United States even accepted waiver of the ‘most favored nation’ principle: greater protection could be legitimately applied against its exports than those of Latin American neighbors.

Another option, emerging from the Left, was for much more fundamental internal reform, and less reliance on trade with the outside world. This view, in a few years to emerge more fully in the guise of dependency theory, saw the problem as a continuing commitment to the market system. There was too little focus on wide-ranging national reforms. Deficiencies of the import-substitution period began to be recognized. But the solution was now to come through greater centralization and state engagement, not exactly of the Soviet type, but more sympathetic to planning as well as more committed to elimination of rank income inequality.

A third alternative, and the one actually pursued, but only for a short time, was the Alliance for Progress, joined by the new Inter-American Development Bank. Both were to be new sources of public foreign capital, capable, together with private investment, of carrying Latin America to a stage of higher rates of economic growth and internal reform: land
redistribution, a new more equitable tax code, commitment to wider education at the basic level, and so on. Planning agencies were made universal, and approval of national plans was necessary to receive resources. Large groups of technical personnel made their way to Latin America to offer assistance. The region, atypically and suddenly, was momentarily at the center of attention.

This bold notion of democratic reform throughout the hemisphere was a United States response to the Cuban Revolution. It had barely begun its operation before another option soon gained force and spread through much of the region. That was explicit military intervention, first apparent in the case of Brazil in 1964, but rapidly spreading to other countries thereafter. The Alliance continued for additional years, but largely in name alone.

The military role was now somewhat a repeat of the Depression decade, although the economic circumstances were far different. The major objective was to preserve capitalism in the midst of the Cold War, and to eliminate the more radical options that threatened. Not all countries in the region were affected: Colombia and Venezuela in South America, Costa Rica in Central America, and Mexico retained civilian leadership. But the decade of the 1970s was later to be recalled as a decade of infamy for the violations of human rights that occurred in the Southern Cone.

Complicating matters still more, the petroleum crisis in 1973 exposed a major weakness in the model of Latin American development. Countries were substantially dependent upon petroleum imports, even those with national oil companies. Venezuela alone was able to benefit greatly from the fourfold rise in price from $3 to $12 a barrel in that year. Others coped by relying on a great surge of borrowing. At first, because interest rates remained low while commodity prices rose, such debt appeared a sensible way to finance the increased trade deficits. Unfortunately, debt-led development soon gave way to debt-led debt, a cumulating problem whose magnitude further multiplied with the outbreak of the war between Iraq and Iran in 1979, and a doubling once more of the price of petroleum.

Some in the region initially benefited. On this positive side, Venezuela was joined by Mexico, which had, during the 1970s, expanded its production considerably. Others, like Brazil, Chile and Argentina, had immediate adverse effects, the more so as international interest rates rose to new highs, for Paul Volcker’s Federal Reserve now sought much more actively to restrain inflation in the United States. As events would unfold, it was one of the oil beneficiaries, Mexico, that first signaled an inability to pay. That happened in July 1982, when an appreciated exchange rate and lack of international reserves forced it to seek help from the International Monetary Fund (IMF) and the United States. Thereafter, soon came many
others. By the mid-1980s, more than a dozen countries in Latin America were receiving financial assistance, and advice, from the IMF. Even with that help, imports had to be drastically curtailed. The international banks that had been so eager to lend in the 1970s had completely withdrawn. Debt-led debt had come to a disastrous end.

At the virtual peak in 1981, as Table 65.3 reveals, many of the countries in the region had shown a rapid expansion from 1950: per capita income had doubled, and for Brazil, almost tripled. Venezuela retained its position as the leader in per capita income, but failed to alter its dependence on oil exports alone as the source of wealth. Argentina and Chile only grew modestly, indicative of the failures of import substitution, compared to the relative successes of Brazil, Colombia and Mexico. More impressively, average Latin American income per capita had finally gained on the United States. It is no wonder that many look back to those three decades with continuing longing for the levels of high tariff protection, for substantial government assistance and for a clear emphasis upon industrial expansion.10

Two decades of reconstruction
In the 1980s, in the midst of coping with the debt crisis, came a new beginning for the region, both politically and economically. The ubiquity of military government that had begun in the 1960s and reached its height with the Pinochet intervention in Chile in 1973, faded extraordinarily rapidly. Interestingly, the problem of managing the debt – whose accumulation had occurred under military regimes – was a decisive element in speeding their demise. New constitutions and new civilian governments appeared

---

Table 65.3 Latin American per capita income (1990 Geary–Khamis dollars)

<table>
<thead>
<tr>
<th>Country</th>
<th>1950</th>
<th>1981</th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>4987</td>
<td>7655</td>
</tr>
<tr>
<td>Brazil</td>
<td>1673</td>
<td>4984</td>
</tr>
<tr>
<td>Chile</td>
<td>3827</td>
<td>5933</td>
</tr>
<tr>
<td>Colombia</td>
<td>2089</td>
<td>4272</td>
</tr>
<tr>
<td>Mexico</td>
<td>2085</td>
<td>5582</td>
</tr>
<tr>
<td>Peru</td>
<td>2263</td>
<td>4292</td>
</tr>
<tr>
<td>Venezuela</td>
<td>7424</td>
<td>9637</td>
</tr>
<tr>
<td>Regional average</td>
<td>2614</td>
<td>5528</td>
</tr>
<tr>
<td>Relative to USA</td>
<td>0.27</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Source: Angus Maddison, Monitoring the World Economy, OECD Development Centre, 1995
throughout the region. Leaders sought support in fairer elections in which larger proportions of the population figured than had previously. Democracy, and inevitably a negative response to continuing inflation, whose inevitable consequence was redistribution of income away from the poor, emerged and strengthened over this period.

The ongoing debt problems of the region in the 1980s cried out for solution. All that initially emerged were low growth, negative capital flows, real devaluations and postponement of external payments. Finally, by the decade’s end, the principle of debtors paying less was accepted. Private banks in the USA and elsewhere had finally emerged from the danger of their own failure, and now, in conjunction with the IMF, were amenable to the substitution of new Brady bonds – so named after the US Secretary of the Treasury – that saw reductions of some 40 percent of initial value. Mexico was the first country to proceed along these lines; Brazil was the last. Commercial banks got out of the business of extensive lending for capital investment, and back into providing shorter-term finance for international trade and other needs.

As this solution of the external problem evolved, there emerged a parallel recognition that fiscal deficits were the prime causal force in explaining Latin American inflation. Only if that governmental excess were credibly, and visibly, curtailed would prices stop their continuing acceleration. Some stabilization plan was necessary, and one that was more immediate in impact than past exercises. National outcomes varied in their particular style, their timing and the number of attempts required. Sometimes, as in Argentina and Brazil, multiple efforts were necessary. Sometimes, as in Nicaragua and Peru, populist measures were vainly instituted as alternative strategies. And sometimes, as in Bolivia in 1985, after achieving rates as high as an annualized 50,000 percent, results were virtually immediate after curtailing government deficits.

This was the essential key. In almost every instance, a firm price anchor was initially provided by a fixed exchange rate, and abetted by freer access to imports that helped to restrain future price increases. Ultimately, however, the key variable to brake inflation expectations remained the fiscal surplus, and when that was not realized, very high real interest rates and reliance on external capital flows were forced temporarily to substitute. These could work, but only for a short period. That is what we learned from the ‘Tequila’ crisis in Mexico at the very end of 1994, the Brazil crisis in 1999, and the Argentine collapse at the end of 2001.

What impresses is the extent to which past inflationary experience had to be overcome. Cutting back dramatically on inflation in the short term was not too difficult. The problem was ensuring continuity. Modest governmental deficits – by comparison with those of Europe, Japan and the
United States – gave rise to very high rates of price increase in Latin America. Internal credibility was slow to be achieved. Today, despite the variety of political leadership found in the region, the battle against inflation seems to have been achieved. That success had a notable and positive effect on the distribution of income, even when other factors intervened to worsen the aggregate.

Concern about fiscal deficits had a direct spillover upon two related subjects. The first of these was privatization. The impulse to sell off substantial state assets in energy, telephones, railways, steel and other intermediate sectors, petroleum exploration and production, airlines, banks, and so on came less from a philosophical commitment than from financial necessity. Initial levels of tax revenue, post-reform efforts to stem inflation, were generally insufficient to cover continuing current expenses. Virtually the only feasible solution was massive disposition of state property; the revenues thereby gained immediately eased the problem. When the need for revenue was very large, the terms of privatization were necessarily more favorable to the private buyers. Employment was almost always directly affected. There was excess labor, at higher wages than paid in the private sector, and almost every instance of privatization saw reductions in jobs, reflected in increased productivity, but also increases in formal sector unemployment.

Additionally, there was renewed attention to collecting larger revenues on a regular basis. Reform of the public sector to reduce current expenses was always announced, but rarely implemented. As a result, revenue collection generally increased much more than outlays were reduced. New taxes were frequently imposed: Brazil is at the regional extreme, with a total inflow of something like 37 percent of product, compared to an initial imposition of 25 percent. In other countries, beginning with Chile in the 1980s, attention turned to privatization of the social security system as a means of reducing future expenses. But these conversions required a surplus of current revenues as an enabling mechanism, thereby providing further incentives to ensuring adequacy of tax receipts.

Still a third area of great change during this period was a movement away from domestic protection and toward greater engagement with enhanced trade flows. Import flows, actual and potential, were an important competitive force capable of checking domestic price increases. Liberalization was therefore frequently utilized as part of the anti-inflationary effort. As countries privatized, external inflows to purchase, partially or entirely, the former nationalized assets became substantial. Currencies appreciated. The familiar problem of balance-of-payments limitations to growth soon reasserted itself. Numerous countries experienced serious crises as a consequence. There are the cases of Mexico in 1994, Brazil in 1999 and Argentina in 2002. But the commitment to openness, although occasionally frayed,
has remained. Tariff levels declined from an average in excess of 60 percent at the end of the 1980s to something between 15 and 20 percent in the mid-1990s.

These dramatic changes in the region inspired much criticism from the Left. As they progressively took place over the decade from 1985 to 1995, there was vocal opposition that took form in the electoral process. But newly elected presidents, frequently seemingly critical of such neoliberal policies, soon adopted the same measures. There was an initial positive consequence upon economic growth almost everywhere. But the new circumstances of the Tequila crisis, declines in Asia, the strong US dollar, failure in Russia and finally, Brazilian devaluation in 1999 took their toll: the promise that such reforms – termed neoliberal by their critics – might permit a resumption of sustained expansion was frustrated.

This is no way better seen than by viewing the data in Table 65.4. For virtually every country, with the prominent exception of Chile, and to some degree Mexico, the lost decade of the 1980s has transformed into a virtual repetition in the 1990s and even beyond. Note especially the dramatic decline in the ratio of income relative to that of the United States. Over the course of the entire twentieth century, there has been a dramatic decline; instead of convergence, there has been a disappointing divergence.

The present

Political opposition to these ‘Washington Consensus’ reforms has continued to increase in recent years. Much of that opposition – within, as well as outside, Latin America – emanates from intellectuals. This opposition

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Argentina</td>
<td>8206</td>
<td>6834</td>
<td>6436</td>
<td>8005</td>
<td>8544</td>
</tr>
<tr>
<td>Brazil</td>
<td>5198</td>
<td>4917</td>
<td>4923</td>
<td>5296</td>
<td>5556</td>
</tr>
<tr>
<td>Chile</td>
<td>5738</td>
<td>5168</td>
<td>6402</td>
<td>8612</td>
<td>9841</td>
</tr>
<tr>
<td>Colombia</td>
<td>4265</td>
<td>4282</td>
<td>4840</td>
<td>5418</td>
<td>5096</td>
</tr>
<tr>
<td>Mexico</td>
<td>6289</td>
<td>6218</td>
<td>6119</td>
<td>6027</td>
<td>7218</td>
</tr>
<tr>
<td>Peru</td>
<td>4205</td>
<td>3631</td>
<td>2955</td>
<td>3505</td>
<td>3686</td>
</tr>
<tr>
<td>Venezuela</td>
<td>10139</td>
<td>8521</td>
<td>8313</td>
<td>8950</td>
<td>8415</td>
</tr>
<tr>
<td>LA average</td>
<td>5412</td>
<td>5052</td>
<td>5053</td>
<td>5460</td>
<td>5838</td>
</tr>
<tr>
<td>Relative to USA</td>
<td>0.29</td>
<td>0.24</td>
<td>0.22</td>
<td>0.22</td>
<td>0.21</td>
</tr>
</tbody>
</table>

starts from a legitimate concern about highly unequal income distributions and excessively high rates of unemployment that have been the counterpart of low growth. They then place blame upon a macroeconomic policy that has seemingly been too much in search of price stability, and too little in pursuit of economic expansion. The IMF has become an easy target for its insistence upon stabilization: reduced government outlays, higher rates of taxes, but above all, high real rates of interest. Privatization has been lamented, and its reversal sought, largely ineffectively. Now the argument has shifted to demands for greater regulatory control over private operation and decision. Greater protection against imports has been sought by industrial producers, eager to re-establish their former degree of dominance in internal supply. Finally, the initial aim of the 34 countries of the hemisphere – excluding Cuba – for a comprehensive free trade area, opening trade in goods and services, has not met its goal. Instead, there have been only an increasing number of bilateral agreements with the United States, with overt opposition from Venezuela, Mercosur and Bolivia.

But outright reversal of the great changes since the 1990s hardly guarantees the future development of Latin America and the Caribbean. Much that has occurred has been necessary, and much overdue. Market forces and wider trade are now a prominent feature almost everywhere in the world. Ignoring them has a real cost. Better to build upon what has been accomplished, and to recognize that more reforms, not fewer, are needed. Recent high rates of growth in the region since 2003 provide a singular opportunity to seize the moment.11

As essential first step is a substantial and ongoing commitment to educational reform and expansion.12 Latin America, for a variety of reasons, has been very late to recognize the importance of universal education to economic growth as well as to a more equal income distribution. In recent years, catch-up has begun to occur: the number of years of education available to the young has expanded, but still lags behind the impressive efforts in Asia. Repetition of beginning years of schooling continues, poor quality of teaching remains, and free public universities absorb significant budgetary resources. Past mistakes inevitably linger as individuals with limited literacy persist in the labor force. Reform is further complicated by its necessary continuity: one requires consistent policies over a generation, not over a single presidential term. So despite expenditures as a percentage of national product that approach 5 percent, few countries stand out positively. Comparable international test results confirm this regional backwardness.

Advance in coping with inequality requires continuing advance in the educational system. That is clear even from the extensive comparative report put together by the World Bank that correctly stresses other
Education is hardly the only element influencing the distribution of income, but it frequently comes out as a principal one. Most relevantly, the quality of education is likely dramatically different for individuals whose parents are poor versus those who are wealthier. The appropriate conclusion is that 'educational disparities account for an important share of Latin America’s high income-related inequality, but are not the only explanatory factor'.

A second area requiring attention is continued macroeconomic reform. In particular, the extraordinary regional difference in public finance seems to suggest a fundamental area of needed advance. Brazil, at one extreme, has revenues that exceed 38 percent of product, while Mexico, excluding oil receipts, barely enters into double digits. Neither level is likely to be efficient or effective. Even when large expenditures are made, governmental investment is small, and an inadequate infrastructure (physical, as well as individual, where health, nutrition, and so on weigh heavily) results. Countries, for example, have invested little in research and technology. Consequently, they have less capability to follow on closely the advances achieved elsewhere. Solving the inflation problem, as most countries have, does not mean the end of concern; instead, it is just the beginning.

Regional countries equally must expand their savings rates if they are again to achieve in the twenty-first century the per capita expansion accomplished in the 1960s and 1970s. Something of the order of 25 percent of national product is required, of which the foreign component should not exceed a small proportion. An inversion is needed. Historically, the private sector saved, allowing the public sector to invest. Now primary surpluses should become the rule, financing not only public capital formation, but permitting private firms to borrow at lower interest rates and for longer terms.

Trade advance should continue, on a more geographically as well as product-diversified basis. In the midst of the present rise in the terms of trade, and strong trade surpluses, there is greater enthusiasm about international trade within the region. Latin America continues to be a region where trade has been less significant than objective indicators suggest it should. Chile and Mexico are now the exceptions rather than the rule. The real test will come if, and when, the commodity boom begins to slacken. Latin America has had that experience before, and the inevitable result was to emphasize the internal market and strengthen protection against import competition, rather than to upgrade the quality of its exports and compete more effectively.

Whether an increasing commitment to the international market will persist, is the question. Globalization is seen in many countries as a hindrance rather than an opportunity. Last time, in the expansion of income
through 1980, that option was largely ignored. Now one hears increasing doubts about whether an open market should persist. Latin America needs the chance to enter, however belatedly, into the international marketplace. For it to choose an alternative path will again deter rather than enhance the region’s emergence.

A final, but important, subject is the need for institutional reform. This refers not merely to reform of the judiciary and the need for persistent and independent standards, but also to the area of regulation of the newly privatized activities. As natural monopolies, compounded by the reality of extensive foreign participation, these activities require supervision and reasonable certainty. Otherwise, investment required in infrastructure will continue to lag behind. Not only economic institutions are involved. Political reform is likewise needed in so many countries, where electoral rules are inadequate, and fundamental changes in structure are required.  

**Conclusion**

The countries of Latin America and the Caribbean find themselves challenged once again, just as they have been over the entire post-World War II period. This time, after having lost out to South Korea and Taiwan in the 1980s and 1990s, the task is much greater. China, India, South Africa and still other Asian countries are now the new competitors in world markets for trade and foreign investment. These are much larger challengers to cope with. And, over the last several years, they have been impressive performers.

In the face of this new challenge, some have advocated greater South–South exchange, as was the mantra in the 1960s and 1970s. Some see the present World Trade Organization (WTO) Doha Round as the place for large reductions in agricultural protection that will assure fairer trade and better opportunities for Latin American advance. Some have called for a resurgent state to replicate the advances achieved in the earlier era, a return to subsidies and import substitution, but this time fully committed to reduction of inequality in the distribution of income. Second-stage reform, as advocated here, is rejected in favor of a different approach.

That route of denial, attractive as it sometimes seems, runs a substantial risk of casting away the advances of almost a generation of reforms. International competition is inevitable, and increasing, in the present age of globalization, as even the developed nations have been discovering. Rather than try to deny that reality, the countries of the region must seek to confront it. Only through a process of internal productivity advance, fueled by advances in technology and greater domestic savings, will Latin America be able to compete, and thereby achieve greater continuing growth.

There is not much time left to choose.
Statistical appendix

There are now four comprehensive series of estimates of national income dating back to the nineteenth century, and encompassing the countries of Argentina, Brazil, Chile, Colombia, Mexico and Venezuela. Peru is the additional country incorporated in the Maddison estimates. Uruguay is also sometimes found. Although these series do differ with the particular base year selected, as well as sources utilized, which does explain some difference in absolute value relative to the United States or OECD countries, they are quite similar in their reported longer-term rates of growth.17 Table 65A.1 provides these results. Shorter-term movements in individual countries do differ, however, and sometimes importantly.

I will not enter here into the index number problems presented by such long-term comparisons, and the familiar Laspeyres bias found as growth incorporates new sectors into production over time. These, and other methodological issues, are discussed extensively by Prados de la Escosura.18 But less time seems to have been spent in searching out and considering critically the national historical estimates that underlie the aggregates. For Brazil, for example, three different series seem to have been chosen by the authors. Ironically, the most recent estimates by Prados de la Escosura choose a series of Raymond Goldsmith, put together in the 1960s, well before recent research on historical Brazilian development flourished.

Table 65A.1 Comparative per capita rates of growth

<table>
<thead>
<tr>
<th></th>
<th>(1) LA 6</th>
<th>(2) LA 6</th>
<th>(3) LA 6</th>
<th>(4) LA 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1900–13</td>
<td>2.2</td>
<td>2.5</td>
<td>2.2 (2.6)</td>
<td>2.4</td>
</tr>
<tr>
<td>1913–29</td>
<td>1.0</td>
<td>1.5</td>
<td>2.4 (1.5)</td>
<td>1.6</td>
</tr>
<tr>
<td>1929–50</td>
<td>1.4</td>
<td>1.6</td>
<td>1.8 (1.8)</td>
<td>1.5</td>
</tr>
<tr>
<td>1950–80</td>
<td>3.1</td>
<td>2.7</td>
<td>2.5 (3.0)</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Notes:
LA 6 is Argentina, Brazil, Chile, Colombia, Mexico and Venezuela
LA 7 includes Peru
a 1929–45
b 1945–81

Sources:
Col (1) Prados de la Escosura (2004)
Col (2) Thorp (1998, Statistical Appendix), calculated from Total GDP minus population growth
Col (3) Hofman, (2000, p. 169)
Col (4) Maddison (2001)
I have chosen to use the Maddison estimates for Latin America – despite his use of the Goldsmith estimates for Brazil – principally because of their greater comparability with those for other regions. That database extends back to 1820 and incorporates information not only on the present developed countries but also on many developing nations in Africa, Asia and Southern and Eastern Europe. In all, there are 56 countries utilized, far larger than in any other source.

Notes
1. There has been a proliferation of new research covering Latin American economic history over the course of the last two decades. Much of this work is quantitative and a large part has been the product of Rosemary Thorp, who merits special credit. She authored Progress, Poverty and Exclusion: an Economic History of Latin America in the Twentieth Century for the Inter-American Development Bank in 1998, and co-edited three volumes of background papers in the series An Economic History of Twentieth Century Latin America, published by Palgrave in 2000. I have made much use of these volumes, without explicit citation. But additionally, there are the recent contributions of Leandro Prados de la Escosura, as well as that of Andre Hofman. And, of course, there has been much independent research done on a national level. In the brief ‘Statistical appendix’, I explain my choice of Angus Maddison’s national income estimates as the basis for long-term comparisons.
2. Lewis (1978).
3. Bulmer-Thomas (2003, p. 144), finds an R-squared of 0.82.
7. The classic exposition is found in CEPAL (1951).
8. This can be easily seen from the formula for revenue from the inflation tax: \( p(M/P) \), the rate of inflation times the public holdings of money. Of course, as inflation occurs, desired holdings of money decline, which is why the process tends to become cumulative.
9. The Economic Bulletins issued by CEPAL at the beginning of the 1960s give evidence of this shift, as do the Annual Economic Surveys.
10. Ironically, many sometimes forget the degree to which the gains were associated with military governments, as well as substantial debt accumulation.
11. Nancy Birdsall has put together another 11 measures beyond the Washington Consensus, ten that are seen as required for the future. They stress the need for great social equity in the region. There is some overlap with my modest four; Birdsall et al. (2001).
12. PREAL, Quantity without Quality, 2006 Report Card, provides extensive details on the reforms required, and the poor marks countries continue to receive for only partial commitment.
13. de Ferranti et al. (2003).
15. This is true of various gravity models using physical characteristics such as distance from markets as well as income levels. See Carillo-Tudela and Li (2004).
17. For some reason, Andre Hofman (2000) has chosen to present his aggregate results for per capita income, as well as for other measures, as a simple average of the individual country results, rather than weighting by the relevant magnitudes. This is appropriate were one interested exclusively in differences among countries, rather than also concerned with group, that is, Latin American, performance. In Table 65A.1, I have provided his original averages in parentheses, and recalculated the LA 6 totals.
References


Inter-American Development Bank (various years), *Economic and Social Progress Report*, Washington, DC: IDB.


Introduction

In 1960, economic development was the mandate and keenly sought-after province of a founding generation of African political leaders. Visions of economic progress were central to liberation rhetoric (Ake, 1996) and were widely embraced within the broader development community (for example, Karmarck, 1971). Yet development failed, and it failed uniquely. Over the next 40 years, as extreme poverty fell dramatically on a worldwide basis, poverty in sub-Saharan Africa (henceforth SSA or ‘Africa’) increased. At the turn of the millennium, nearly one in two Africans or 300 million in total consumed less than $1 a day, a proportion twice as high as the global rate and double the number prevalent in Africa in 1970. Home to 10 percent of the world’s population, the region now accommodates 30 percent of the world’s poor – who spend a quarter less than the Asian poor on their livelihood (World Bank, 2005). The African development challenge has become the global development challenge. How and why did it become so, and what can we expect, looking ahead?

We approach these questions through the lens of economic growth. The second section describes the African record from 1820 to the present, focusing on the period since 1960 and emphasizing demographic and other features that differentiate African growth patterns from those of other developing regions. In the third section we outline two main structural explanations for the African experience, based in turn on governance and resource endowments. We develop the former theme in a detailed exploration of the political economy of African growth. The fourth section documents the revival of growth that got under way in the mid-1990s; we interpret this revival as a lagged response to the economic and political reforms of the late 1980s and early 1990s. We argue that Africa now faces a window of opportunity, with politically stable countries facing the prospect of mutually reinforcing declines in fertility rates and increases in capital formation and growth. We conclude with a summary of lessons from the African experience, both for the region itself and for development economics.

The growth record

Any adequate account of Africa’s development experience must acknowledge the overall weakness of Africa’s growth record, its juxtaposition with
population explosion, and its variability over time and across countries. We begin with Africa’s long-run growth experience, drawing first on the painstaking historical research of Maddison (2001) and then on World Bank data for 100 developing countries since 1960.² Our central concern here is the relative stagnation of economic growth in SSA in the period since 1950. This performance is not an outlier in historical terms: in Table 66.1, Africa’s per capita growth rate of 0.7 percent after 1950 matches that of the rest of the developing world over the previous century. But African populations missed out on the economic transformation that took place in the developing world – particularly in Asia – in the second half of the twentieth century. The result was that by the 1950s, African incomes, which had gained considerable ground in relative terms since 1913, had begun to diverge powerfully from incomes elsewhere in the developing world. By comparison with East Asia and Pacific, a shortfall of less than 50 percent in purchasing power parity (PPP)-adjusted terms around 1960 rose to well over 300 percent by the end of the century. The consequences of this growth failure are apparent in Table 66.2, which provides a snapshot comparison of human development, first at the outset of the 1960–2000 period and then at the end. With the exception of the primary enrollment rate, which was already high outside of Africa in the early 1960s, Africa fell further behind the rest of the developing world. Regress was not as severe for non-income measures as it was on income or poverty: by 2000 Africa exceeded the levels of primary enrollment, adult literacy and life expectancy that had prevailed elsewhere in 1960. But the failure to raise per capita incomes significantly had critically undermined these achievements.

At the sectoral level, slow growth has gone hand-in-hand with limited structural diversification. Traditional agriculture continues to absorb the majority of the labor force in many African countries, a feature no longer observed in any other region of the world (O’Connell and Ndulu, 2000). Irrigation is expensive and extremely sparse, with the result that African agriculture remains largely rain-fed and subject to periodic drought. Exports have tended to remain concentrated in a narrow band of primary commodities (Berthélemy and Soderling, 2001, 2002), including exploitation of mineral resources. Collier and O’Connell (2007) use global data to identify ‘resource-rich’ economies as those that exceed threshold values for the ratios of primary commodity rents (from energy, mineral and forest resources) to gross domestic product (GDP) and primary commodity exports to total exports on a sustained basis. Comparing SSA with other developing regions (excluding the Middle East and North Africa), they find that a stark difference already existed in 1960, with 12.5 percent of the SSA sample classified as resource-rich and only 7 percent of the non-SSA sample. This difference expanded over time, with another 16.7 percent of
Table 66.1  Long-run growth rates by developing region.

<table>
<thead>
<tr>
<th></th>
<th>SSA</th>
<th>Total</th>
<th>LAC</th>
<th>ASIA</th>
<th>MENAT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1820–70</td>
<td>0.3</td>
<td>0.2</td>
<td>1.3</td>
<td>0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>1870–1913</td>
<td>0.7</td>
<td>0.6</td>
<td>1.6</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>1913–50</td>
<td>1.7</td>
<td>1.0</td>
<td>2.0</td>
<td>0.9</td>
<td>1.3</td>
</tr>
<tr>
<td>1950–2001</td>
<td>2.6</td>
<td>2.0</td>
<td>2.3</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>1950–60</td>
<td>2.2</td>
<td>2.1</td>
<td>2.8</td>
<td>2.0</td>
<td>2.6</td>
</tr>
<tr>
<td>1960–2001</td>
<td>2.7</td>
<td>2.0</td>
<td>2.2</td>
<td>1.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Peak rate*</td>
<td>3.00</td>
<td>2.52</td>
<td>2.87</td>
<td>2.52</td>
<td>3.34</td>
</tr>
<tr>
<td><strong>Real GDP (PPP-adjusted 1990 dollars)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1820–70</td>
<td>0.6</td>
<td>0.1</td>
<td>1.2</td>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>1870–1913</td>
<td>1.1</td>
<td>1.3</td>
<td>3.5</td>
<td>1.1</td>
<td>1.6</td>
</tr>
<tr>
<td>1913–50</td>
<td>2.7</td>
<td>1.5</td>
<td>3.4</td>
<td>0.9</td>
<td>2.5</td>
</tr>
<tr>
<td>1950–2001</td>
<td>5.3</td>
<td>5.0</td>
<td>4.0</td>
<td>5.4</td>
<td>4.9</td>
</tr>
<tr>
<td>1950–60</td>
<td>4.1</td>
<td>5.6</td>
<td>5.1</td>
<td>5.8</td>
<td>5.6</td>
</tr>
<tr>
<td>1960–2001</td>
<td>3.1</td>
<td>4.9</td>
<td>3.7</td>
<td>5.3</td>
<td>4.7</td>
</tr>
<tr>
<td><strong>Real GDP per capita (PPP-adjusted 1990 dollars)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1820–70</td>
<td>0.2</td>
<td>-0.1</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.5</td>
</tr>
<tr>
<td>1870–1913</td>
<td>0.4</td>
<td>0.7</td>
<td>1.8</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>1913–50</td>
<td>1.1</td>
<td>0.5</td>
<td>1.4</td>
<td>-0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>1950–2001</td>
<td>0.7</td>
<td>3.0</td>
<td>1.7</td>
<td>3.4</td>
<td>2.2</td>
</tr>
<tr>
<td>1950–60</td>
<td>1.9</td>
<td>3.4</td>
<td>2.3</td>
<td>3.7</td>
<td>2.9</td>
</tr>
<tr>
<td>1960–2001</td>
<td>0.4</td>
<td>2.9</td>
<td>1.5</td>
<td>3.4</td>
<td>2.0</td>
</tr>
<tr>
<td><strong>No. of countries</strong></td>
<td>53</td>
<td>90</td>
<td>27</td>
<td>42</td>
<td>21</td>
</tr>
</tbody>
</table>

Notes:
SSA is Maddison’s ‘Africa’, excluding Algeria, Egypt, Morocco, and Tunisia, but including Libya (see below). LAC comprises ‘Latin America’ and includes the Caribbean. ASIA is Maddison’s ‘East Asia’. MENAT (Middle East, North Africa and Turkey) corresponds to ‘West Asia’ plus North Africa. Libya could not be separated out and we therefore include it here in SSA rather than in MENAT. Note that these growth rates are based on regional totals and are therefore dominated by the large countries in each region. More specifically, the population, real GDP and real GDP per capita growth rates correspond to population-, GDP- and (approximately) GDP-weighted individual-country growth rates. Note also that data before 1950 unavoidably contain major ‘guesstimates’. For example, the African growth rate for 1870–1950 is based on case study work on Algeria, Egypt, Ghana, Morocco, South Africa and Tunisia. Maddison applies the average growth rate of GDP per capita for this group of six (= 0.90) to all countries in SSA other than Ghana and South Africa. Our SSA aggregate for 1870–1913 combines this indirect estimate for 51 countries with Maddison’s direct estimates of 2.07 and 1.25 for Ghana and South Africa.
* Peak year and peak rate correspond to the earliest year after which all subsequent population growth rates are lower.

Source:  Ndulu and O’Connell (2007), Table 1.4. Calculations are based on country and regional data from Maddison (2001).
Table 66.2  Regional growth comparisons

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>Real GDP per capita (1996 PPP$)</th>
<th>Gross primary enrollment rate, 1970</th>
<th>Adult illiteracy rate, 1970</th>
<th>Life expectancy at birth</th>
<th>Real GDP Growth (1960 or earliest year before 1965, or as indicated)</th>
<th>Real GDP per capita</th>
<th>Real GDP per worker</th>
<th>Real GDP per capita</th>
<th>Real GDP per worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSA</td>
<td>35</td>
<td>1278.1</td>
<td>53.8</td>
<td>55.8</td>
<td>41.1</td>
<td>3.20</td>
<td>2.63</td>
<td>0.56</td>
<td>-0.15</td>
<td>0.72</td>
</tr>
<tr>
<td>OTHER</td>
<td>43</td>
<td>2591.5</td>
<td>90.5</td>
<td>26.5</td>
<td>53.2</td>
<td>4.28</td>
<td>2.16</td>
<td>2.12</td>
<td>0.23</td>
<td>1.90</td>
</tr>
<tr>
<td>DEM'ING</td>
<td>22</td>
<td>3338.4</td>
<td>99.1</td>
<td>17.4</td>
<td>56.4</td>
<td>3.52</td>
<td>2.08</td>
<td>1.44</td>
<td>0.42</td>
<td>1.03</td>
</tr>
<tr>
<td>LAC</td>
<td>22</td>
<td>3338.4</td>
<td>99.1</td>
<td>17.4</td>
<td>56.4</td>
<td>3.52</td>
<td>2.08</td>
<td>1.44</td>
<td>0.42</td>
<td>1.03</td>
</tr>
<tr>
<td>SASIA</td>
<td>5</td>
<td>934.4</td>
<td>58.6</td>
<td>55.5</td>
<td>45.3</td>
<td>4.34</td>
<td>2.23</td>
<td>2.10</td>
<td>-0.32</td>
<td>2.42</td>
</tr>
<tr>
<td>EAP</td>
<td>9</td>
<td>1833.1</td>
<td>94</td>
<td>20.4</td>
<td>50.6</td>
<td>5.48</td>
<td>2.07</td>
<td>3.41</td>
<td>0.16</td>
<td>3.29</td>
</tr>
<tr>
<td>MENAT</td>
<td>7</td>
<td>2402.5</td>
<td>81.9</td>
<td>42.3</td>
<td>51.7</td>
<td>5.09</td>
<td>2.48</td>
<td>2.61</td>
<td>0.13</td>
<td>2.48</td>
</tr>
</tbody>
</table>

Notes: Full values are provided in Table B.4.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>INDUST</td>
<td>22</td>
<td>8507.6</td>
<td>101.8</td>
<td>70.2</td>
<td>3.45</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>3433.3</td>
<td>80.4</td>
<td>38.1</td>
<td>3.72</td>
</tr>
<tr>
<td>SSA v SASIA</td>
<td></td>
<td>(136.8)</td>
<td>(91.8)</td>
<td>(100.5)</td>
<td>(90.7)</td>
</tr>
<tr>
<td>SSA v Other</td>
<td></td>
<td>(49.3)</td>
<td>(59.4)</td>
<td>(210.6)</td>
<td>(77.3)</td>
</tr>
</tbody>
</table>

**Notes:**
Except in the final 2 rows, the numbers in parentheses give the relevant developing-country mean as a percentage of the industrial-country mean. The final 2 rows show the SSA mean relative to the SASIA mean and the mean for all non-SSA developing regions.

Regions: SSA = Sub-Saharan Africa; OTHER DEV’ING = Other Developing (LAC = Latin America and Caribbean, SASIA = South Asia, EAP = East Asia and Pacific, MENAT = Middle East, North Africa and Turkey); INDUST = Industrial countries.

**Source:** Ndulu and O’Connell (2007), using PWT6.1 and *World Development Indicators.*
the African sample acquiring resource-rich status by 1990, as compared with only 10.5 percent of the non-African. A consequence of Africa’s delayed structural transformation has been the continued vulnerability of its population to shocks to rainfall and world commodity markets.

Tables 66.1 and 66.2 document the demographic explosion that is a striking correlate of Africa’s economic stagnation over the post-1960 period (O’Connell and Ndulu, 2000; Lucas, 2003). With the exception of life expectancy rates, standard indicators of demographic pressure differed only modestly between Africa and the rest of the developing world in the early 1960s (Figure 66.1). But the demographic transition was already well underway in Latin America, where population growth rates had peaked in 1960 (Table 66.1). For the next 25 years, total fertility rates fell sharply outside of Africa while remaining virtually unchanged within Africa. Population growth rates therefore diverged sharply, and from the early 1970s through the remainder of the century, the population of SSA grew more rapidly than the non-African developing-country population had grown at its peak. The ratio of (overwhelmingly young) dependents to working-age population grew steadily, exceeding historical developing-country norms by 1970 and remaining above these through 2000. The fertility rate began to fall in Africa in the mid-1980s, suggesting entry into the final phase of the demographic transition. We will return to this observation, which is complicated by the huge impact of HIV/AIDS on life expectancies starting in the late 1980s.

The aggregate growth record conceals considerable variability, both over time and across countries. Table 66.1 indicates that a sharp deceleration of growth took place in Africa between the end of the colonial period (1950–60) and the remainder of the century. Within the latter period, a further distinction can be made (Figure 66.2), between the moderate growth rates of the 1960s and late 1990s and the deep contraction of 1974–94. African economies continued to grow in the 1960s. This growth was already weaker than that of other developing regions, however, and the global deceleration of the 1970s therefore took substantial portions of the continent into outright contraction. Between 1960 and 1994, nearly half of African countries with comparable data suffered per capita income losses exceeding 20 percent in constant domestic currency (Rodrik, 1998). The bulk of these losses occurred between 1974 and 1994, a period that began with a set of shocks to energy and tropical commodity markets (1974–79) and ended with a concentrated wave of African democratic reforms (1989–94). As indicated in Figure 66.2, the growth shortfall between 1974 and 1994 is much larger on a population-weighted basis than in the simple averages. This is partly a large-country phenomenon within Africa, but after 1980 it is driven primarily by the dramatic growth performance of
Source: Ndulu and O’Connell (2007), using World Development Indicators 2005. The figures show simple averages of country observations, for all countries with continuously available data.

Figure 66.1 Demographic pressure: SSA versus Other developing regions, 1960–2000

The averages discussed here also mask wide variation in the growth performance of individual countries. Many African countries have experienced limited episodes of robust growth. Nearly half of the African economies studied by Pritchett (1998), for example, exceeded a per capita income growth threshold of 1.5 percent through the mid-1970s. Ghana and more strongly Uganda have consistently exceeded a 2.5 percent threshold (roughly the long-run median for developing countries) since the mid-1980s. Botswana and Mauritius grew spectacularly; their long-run records compare favorably with those of the East Asian miracle economies. A diversity of outcomes also characterizes the period after 1994, during which 15 African countries have seen growth rates of total GDP in excess of 5 percent. Focusing on the most recent five-year period (1999–2004) and excluding the oil countries, median per capita growth rates in the fastest-growing, middle, and slowest-growing thirds of the African sample – each comprising 12 or 13 countries – were 2.8, 1.0 and −1.8 percent respectively.4

Source: Ndulu and O’Connell (2007), using World Bank data. The figure shows averages of country observations, for all countries with continuously available data. The suffix ‘popw’ refers to Population-Weighted averages. The data are smoothed using a Epanechnikov kernel-weighted polynomial smoother of degree 1; alternative approaches yield similar results.

Figure 66.2 Smoothed average growth in real GDP per capita (countries with full set of growth observations)
Outside of the mineral-exporting group, rapid growth after 1990 has also been associated with substantial diversification of production and exports. Finally, the weak contribution of increases in measured physical and human capital per worker to African growth is yet another significant feature. Weak investment effort is part of this story (see below): standard growth accounting exercises assign about half of the post-1960 growth shortfall relative to other developing regions to a shortfall in measured capital deepening per worker. But the data also suggest profound problems in translating investment effort into effectively utilized capital (Pritchett, 2000). In order to reconcile observed growth outcomes with measured capital inputs, one must conclude that the average productivity of African inputs per worker fell considerably over time, not just relative to productivity elsewhere but in absolute terms. Ndulu and O’Connell (2007), for example, find that the country-level cumulative changes in total factor productivity over 1960–2000 were as likely to be negative within Africa as positive (see also Hall and Jones, 1999). The correlation of physical capital accumulation with growth, moreover, is considerably lower within Africa than in the rest of the developing world, even over periods of a decade or longer. The latest example of this is Africa’s recovery starting in the mid-1990s, which was not accompanied by a commensurate boom in aggregate investment (Berthélemy and Soderling, 2001).

Explaining poor growth performance in Africa
Why did the African environment prove hostile to economic growth after 1950? Two broad lines of argument dominate the literature, based in turn on economic mismanagement and structural impediments to growth. With some imprecision we refer to these below as the ‘governance’ and ‘geography’ views. Demographic trends pose distinct structural challenges, and we treat these separately. A synthesis of these strands has yet to be achieved, though we make a tentative approach after reviewing the main arguments.

The governance critique
A critique of economic management has dominated the literature on African economic performance since the early 1980s. The central themes of what we will call the ‘governance critique’ were laid out in the World Bank’s 1981 Accelerated Development in Sub-Saharan Africa (the ‘Berg Report’, after its chief author Elliot Berg), Robert Bates’s classic 1981 Markets and States in Tropical Africa and, with a ten-year lag, Paul Collier’s 1991 African Affairs article on agencies of restraint. In its initial and most influential form, the governance critique sought to explain what it characterized as excessive intervention by African governments in economic markets. Bates (1981) argued that African policy-makers
had sacrificed both the agricultural sector and industrial efficiency in order to divert resources to favored interests. The 1980 Berg Report had located African economic stagnation in overtaxation of export agriculture, over-protection of import-competing industries, and oppressive state control of finance, industry and agricultural marketing. Bates (1981) argued that while these policies were economically inefficient, the ‘urban bias’ they displayed was rational from the point of view of political elites. Farmers would lose out, and their losses would exceed the gains to the political elites and to the urban interests more generally (the civil service, the military, and labor and capital in the formal private sector or state enterprise sector). But farmers faced deep collective action problems; while numerous, they were too poorly organized to constitute the ‘selectorate’ on which the government depended to retain power (Bates and Devarajan, 2001).

Governments would therefore penalize agriculture and support urban-based industrialization far beyond what could be justified by the correction of market failures. Moreover, they would do so using inefficient quantity-based instruments rather than price-based interventions (that is, quotas, exchange controls and marketing monopolies rather than tariffs and explicit export taxes), because the distribution of policy-generated rents was central to their political security. Growth would fail, but the protected urban electorate would be well served. Exceptions to the urban bias pattern could be explained, in the Bates analysis, by appealing to the rural political roots and business interests of founding political leaders in countries like Kenya, Malawi and Côte d’Ivoire.

The Berg–Bates contribution provided a description of African policy biases, a link from these to growth outcomes, and a grounding of these choices in the interests of African political elites. Each element of the argument provoked important voices of dissent or moderation. But the empirical content of the critique gave it substantial appeal to Africa’s donors and to economists seeking to understand the continent’s lagging growth performance. For donors, the governance critique provided a rationale for using conditional lending to push market-based reforms – a tendency undoubtedly strengthened by the ascendancy of conservative governments in the United States and Europe in the early 1980s. Donors could act as the agents of disenfranchised African populations, imposing conditionality on easily monitored policy reforms like trade liberalization and exchange rate unification. Since the existing policies penalized the poorer rural sector, their removal would simultaneously improve growth and distribution. For economists, the view that policies represented the largely autonomous choices by a self-interested political elite provided a causal interpretation of regression evidence linking measures of policy distortion with economic growth (for example, Sachs and Warner, 1995).
A generalization of the critique to misaligned interests

Subsequent contributions developed the governance critique beyond its initial application by restating it in terms of a conflict of interest between African policy-makers and African populations. The basic argument is that under authoritarian rule, development policy in a large number of African countries has tended to be captured by a narrow political elite operating under relatively weak institutional constraints. Characterized by weak legitimacy and tenuous bureaucratic control, autocratic regimes in Africa until the 1990s did not function as agents of the public interest, tending instead to sacrifice growth in favor of patronage-based redistributive politics or outright predation (see also Adam and O’Connell, 1999; Ndulu and O’Connell, 1999; Humphreys and Bates, 2001).

An anti-growth bias may become severe if policy-makers undervalue future interests relative to present. Policy-makers may discount the future excessively if they have little expectation of remaining in power or being held accountable. Using African data from the 1980s, McMillan (2001) finds that high rates of presidential turnover predict inefficiently high rates of export taxation in Africa, particularly for crops with large fixed inputs. Fosu (2002) documents the frequency of attempted coups in Africa and finds a strongly negative causal link to overall growth. In an extreme case of misaligned interests, the expected tenure of an authoritarian leader or oligarchy may therefore be inversely related to growth performance. Diamond (1977) appeals to global evidence, for example, to argue that economic success tends to bring democratization. If autocratic leaders internalize such a link, then development may represent a threat rather than an investment in future rents (Robinson, 1997). Political elites may then actively oppose development – as in some cases the colonial powers themselves did, fearing the contestability development might create in the economic and political spheres.

A combination of autocracy and tight economic controls during 1960–85 presented a particularly potent environment for patronage and predation. Using data on the type of African leadership Ndulu (2007) shows that 80 percent of autocratic regimes since independence imposed soft or hard controls (see Collier and O’Connell 2007) for a substantial portion of their tenure.

Economic controls were not new to Africa at the time of independence. Colonial administrations had embraced an activist, developmental role by the late 1940s, reflecting the suddenly temporary nature of their remaining trusteeship and the allure of state intervention following depression, wartime mobilization and the emergence of the Soviet Union as a great power. A mentality of market regulation was therefore in place well before the formal transfer of sovereignty in Africa, as were some of its key institutional mechanisms including monopoly export marketing boards,
exchange restrictions and economic plans (Fieldhouse, 1986). But the period from 1960 to 1975 saw a dramatic expansion of the regulatory presence of the African state. Although this sharp expansion was part of a global phenomenon and grounded in the global development paradigm of the day, it was propelled further by rent-seeking behavior.10

Late in the colonial period, the colonial powers had begun to introduce the rudiments of democratic self-government, and independence constitutions reflected the institutional structures of Western democracies, with contested multi-party elections, checks and balances, and substantial civil liberties. But the degree of *de facto* democracy deteriorated steadily following independence. As in the case of economic controls, authoritarian government was not an African innovation. Its growth impact in SSA, however, appears to have been far from beneficent, in contrast to the broad Asian experience with authoritarian rule during the same period (Alesina and Perotti, 1994). The governance critique asserts that in the African context, causality runs from government institutions to growth.

**Limitations and extensions of the governance critique**

While the governance critique had instinctive appeal both to Africa’s donors and to economists seeking to understand the continent’s lagging growth performance, it also raised serious conundrums.

**Institutions** Van de Walle (2001) argues that the unifying feature of African political economy is not the power of urban interests but rather the absence of coherent domestic political interest groups of any kind. In this view, domestic interests are everywhere too weak to restrain the behavior of the tiny political elite that holds or shares power in its own interest. To explain the persistent ‘choice’ of stagnation, he appeals to the low capability of African public bureaucracies and the dysfunctional influence of strategically motivated external donors. These conspired to undermine the medium-run coherence of virtually any policy initiative, he argues, while simultaneously protecting high-level African elites from personal responsibility for economic decline.

In an influential extension of the governance critique, Collier (1991) similarly appealed to an institutional vacuum that left too much discretion and too little accountability in the hands of political elites. Collier focused on agencies of restraint – public institutions designed to protect national assets (including privately held ones) from predation. Before 1960, this role was provided by colonial administrations operating under tight mandates of internal security, fiscal solvency, and openness to metropolitan trade and finance. Immediately following independence, the new counterparts to colonial institutions – export marketing boards, national central banks,
multiparty parliamentary systems, independent judiciaries – struck a balance between flexibility and restraint that reflected the conservative mandates of their colonial predecessors. But political leaders sought to consolidate power and could appeal to the need for ambitious development programs. By the mid-1970s, political leaders had systematically relieved national ‘agencies of restraint’ of their powers to restrain executive action (see also Collier, 1982; Bratton and van de Walle, 1997). Some palpable gains in policy flexibility were observed – Botswana, for example, left the Rand Monetary Area (RMA) and subsequently achieved lower inflation than the RMA countries. But in most cases such gains appear to have been overwhelmed by increased macroeconomic instability and deterioration in the protections afforded to private investment (points developed in detail in Collier and Pattillo, 1999). 11

In the mid-1990s, the cross-country growth literature began to develop a broader institutional version of the governance critique, based on the tendency of African countries to cluster in the lower ranks of institutional performance measures that are correlated with growth on a global basis. The attentions of policy reformers meanwhile began to shift from conventional economic reforms to problems of improving public service delivery, reducing bureaucratic corruption and strengthening the rule of law. While economists still know relatively little about how durable improvements in public sector performance are achieved, three observations seem relevant to the African situation. First, when institutions are initially weak, the initiative of high-quality political leaders is critically important in determining how well or poorly existing institutions actually perform (Reinikka and Collier, 1999; Glaeser et al., 2004). Second, as suggested above (and as measured by shares in formal sector employment or total investment), African governments have tended to be large and overbearing rather than small and efficient. Third, institutions are known to display persistence: learning takes place and interests form around existing patterns of behavior (North, 1990). These observations suggest that much of the logic of the governance critique applies directly to the performance of public sector institutions over the 1960–2000 period. Wherever political leaders were unable to reconcile the benefits of a market-friendly institutional environment with their own priorities, institutional performance deteriorated and growth suffered.

Did Africa’s abrupt democratization between 1988 and 1994 improve the institutional ground of policy-making? Contested elections and a free press are among potentially important agencies of restraint cited by Collier (1991). Partly for this reason, our overall answer is a positive one.

Fractionalization, polarization and nation-building We next look at the impact on policy of the patterns of sub-national identity that existed at the
time of independence and that in many cases became the dominant mode of political mobilization and conflict. The countries of SSA came to political independence both later and more rapidly than those of other developing regions (Table 66.3). While only Ethiopia, Liberia and South Africa existed as independent states at the end of 1955, fully three-quarters of colonial Africa, representing the vast bulk of its population and GDP, had achieved political independence by 1966. In 1966 the average independent state in SSA had held sovereignty for fewer than ten years; its counterpart in the rest of the developing world had been independent for the better part of a century.

Colonial structures of political control were both arbitrary – with boundaries cutting across historical patterns of politics and trade – and effective. Their abrupt departure meant that the challenge of economic development was in many cases confounded from the outset with an acute problem of nation-building. Nigeria provides a telling example of the impact of *ex ante* regional polarization on political and economic development. But similar patterns of internal polarization, often created or reinforced in the encounter with conquering European powers, existed throughout the continent in 1960. Azam (2007) emphasizes the salience of coastal–interior cleavages in West Africa, operating as in Nigeria on a North–South axis and tending to separate a nomadic, pastoralist Muslim interior from a more sedentary, educated, Christianized coast. In the Horn of Africa, both Sudan (Arab North, Christian and Animist South) and Ethiopia (federated with richer Eritrea after World War II, to guarantee Ethiopia’s access to the coast) have engaged in ethno-regionally based civil wars since the early 1960s. In Central Africa, Belgian favoritism towards

---

**Table 66.3 Dates of political independence, developing countries**

<table>
<thead>
<tr>
<th>Region</th>
<th>n</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Developing</td>
<td>66</td>
<td>1830</td>
<td>1830</td>
<td>1946</td>
<td>1961</td>
<td>1981</td>
</tr>
<tr>
<td>of which:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAC</td>
<td>27</td>
<td>1818</td>
<td>1825</td>
<td>1840</td>
<td>1962</td>
<td>1981</td>
</tr>
<tr>
<td>ASIA</td>
<td>20</td>
<td>1816</td>
<td>1933</td>
<td>1948</td>
<td>1956</td>
<td>1975</td>
</tr>
<tr>
<td>MENAT</td>
<td>19</td>
<td>1816</td>
<td>1932</td>
<td>1948</td>
<td>1962</td>
<td>1971</td>
</tr>
<tr>
<td>Total</td>
<td>112</td>
<td>1822</td>
<td>1907</td>
<td>1960</td>
<td>1964</td>
<td>1993</td>
</tr>
</tbody>
</table>

*Source:* Gleditsch and Ward (1999) database, as compiled by Ndulu and O’Connell (2007), Table 1.2.
the Tutsi minority produced enduring cleavages that erupted in civil war in the 1990s. In South Africa, and in those portions of colonial Africa with large settler populations – including Kenya and Zimbabwe – race-based geographical discrimination was a matter of state policy; in these cases the pattern of *ex ante* polarization would eventually require determining the status of settler minorities.

While the salience of ethno-regional polarization was clear to political scientists in the early 1960s (for example, Carter, 1966), economists have only recently begun to come to grips with the implications of nation-building for African economic growth. Two approaches have been important. The first is due to Easterly and Levine (1997), who focused on fractionalization rather than on polarization *per se*. Easterly and Levine noted that the degree of ethno-linguistic fractionalization – measured by the probability that two randomly chosen individuals in a given country spoke a different first language – was extremely high in many African countries, by comparison with global norms. Moreover, on a global basis ethno-linguistically heterogeneous countries tended to grow more slowly, as a result of weaker public sector performance. Miguel (2004) reports a similar finding for Kenya and Tanzania, using data on local provision of public services. Collier (2000) finds, however, that the adverse impact of heterogeneity is strongly contingent on political institutions. In democracies, ethno-linguistic heterogeneity has no impact either on overall growth or on microeconomic efficiency (as measured by the economic return on World Bank projects), while in dictatorial regimes the adverse impact is strong.

Azam (1995, 2007) focuses directly on polarization, defining a polarized society as one like Nigeria’s, in which there are two or three large sub-national ethnic groups that dominate population and politics in separate regions. Azam argues that in a situation of *ex ante* ethno-regional polarization, regionally-based redistribution may be required to buy off the threat of armed conflict. The existence of such a risk is consistent with the global evidence of Collier and Hoeflinder (2004), who find the risk of civil war maximized under conditions of polarization: homogeneous societies have low exposure to civil war, but so do heterogeneous societies. In cases of *ex ante* polarization, then, the Azam analysis may force a reinterpretation of what is conventionally viewed, within the governance critique, as distortionary redistribution. If the absence of redistribution invites armed conflict and economic collapse, then a program that distorts efficiency relative to an irrelevant peaceful counterfactual may in fact be growth-promoting relative to the true counterfactual of civil war. This shifts the ground of the governance critique from redistribution *per se* to the instruments that are employed to achieve it. Political elites attempting to ‘buy the peace’ should be observed doing so transparently and credibly.
(perhaps via constitutional means), and with a minimum of distortion; and they should simultaneously employ instruments directly targeted at reducing polarization.

**Geography**

Starting in the mid-1990s, Jeffrey Sachs and Adrian Wood began to build an empirical case for the adverse influence of resource endowments and geography on African economic growth. Wood argued that in a world of capital mobility, comparative advantage was determined by endowments of immobile factors: primarily unskilled labor, human capital and natural resources. Africa’s rich endowment of natural resources relative to human capital implied a deep comparative advantage in the production and export of primary commodities (Wood and Berge, 1997). The failure of African countries to achieve competitiveness in manufacturing was therefore largely independent of trade policy or the quality of governance, although these factors may have affected the long-term evolution of factor endowments. Sachs argued that high transport costs and a hostile disease environment conspired to make capital accumulation and productivity growth much more expensive in Africa than elsewhere in the developing world (Sachs and Warner, 1995, 1997, 2001; Bloom and Sachs, 1998; Gallup and Sachs, 1999).

**Distance and landlockedness** African populations are internally fragmented and isolated from world trade by unusually large land distances, unhealthy lowland coastlines, a sparse network of ocean-navigable rivers, and multiple political borders (Gallup and Sachs, 1998; Faye et al., 2004). With its 48 economies, the region has by far the highest density of countries per land area of any developing region; on average, each country shares a border with four neighbors (Ndulu, 2004). Nearly 40 percent of the African population lives in countries that are landlocked or virtually so. The unusual distance of African population concentrations from coastlines and ocean-navigable rivers seems to follow in part from the inland locations of water resources critical to agriculture, including the Great Lakes, major non-ocean-navigable rivers, and fertile rain-fed uplands.

The remoteness of African population concentrations may have severely limited the scope for Asian-style growth patterns based on proximity to global markets, scale economies and agglomeration. Africa’s relatively more sparse distribution of population, significantly low population density and relatively lower rate of urbanization raises significantly the transport intensity of its economic activities (C. Kessides, 2005). Furthermore, the lower population density and urbanization tend to increase the amount of infrastructure investment required to produce similar levels of income (Esfahani and Ramirez, 2003).
Unusually high internal transport costs accentuate this remoteness. Limão and Venables (2001) estimate that it costs nearly twice as much for the median African country to move a 40-foot container from a coastal port to its in-country destination, as it does for countries in other developing regions. These costs are particularly damaging for manufacturing, where the share of traded intermediate inputs is relatively large. They also penalize physical capital accumulation by raising the relative price of investment. Investment in Africa is unusually expensive in terms of local income, so that a given national saving rate delivers a lower increment to real capital accumulation in Africa than in other regions. The average relative price of investment goods for sub-Saharan Africa was 70 percent higher than for Organisation for Economic Co-operation and Development (OECD) countries or East Asia. Artadi and Sala-i-Martin (2003) find that the high relative price of investment goods reduces Africa’s predicted growth rate by 0.44 percent on an annual basis, holding saving effort constant.

Notwithstanding variations across countries in the region, for most African countries distance from their primary markets and the high transport intensities of their products (low value, high weight and sparsely produced) are major impediments for production and trade (Esfahani and Ramirez, 2003). Using a gravity model Limão and Venables (2001) estimated the elasticity of trade with respect to transport costs, and found it typically to be quite high at \(-3\). Distance to key markets is an important impediment to trade as expected, but in their model poor infrastructure (measured by an index combining road, rail and telecom density) accounts for 40 percent of the predicted transport cost for coastal countries and up to 60 percent for landlocked countries. The median landlocked country has only 30 percent of the trade volume of a median coastal country. What is also striking from this study is that, holding activity levels and direct distances between trading partners constant, improving internal infrastructure within the landlocked country itself is as important as improving the infrastructure in the transit country.

Landlockedness adds a political dimension to remoteness. Transport costs now depend crucially on the infrastructure investments and pricing policies of coastal neighbors. These neighbors may also be important but unreliable hosts for the export of labor services. Sachs and Warner (1997) and others find that landlocked status reduces predicted growth by up to 1 percent per year on a global basis.

**Disease burden** Sachs and Warner (2001) and Masters and McMillan (2001) emphasize the high burden of human and animal disease in tropical climates and its impact on life expectancy, human capital formation, labor
force participation and economic growth. Ninety-two percent of SSA lies within the tropics, as compared to 60 percent for East Asia. Following Sachs and Warner (2001), Artadi and Sala-i-Martin estimate the foregone growth in Africa as a result of malaria prevalence at 1.25 percent per annum, a figure that surely reflects the influence of other highly correlated aspects of the health environment. Acemoglu et al. (2001) take a very different, institutions-based approach to linking disease burden with growth. They argue that the quality of contemporary institutions reflects the nature of the institutions introduced by European powers during the colonial period. Where the local disease environment was inhospitable, Europeans introduced extractive institutions, leaving a legacy of predation and violence that continues to undermine the rule of law and the security of property. Where the disease environment was favorable to European settlement, colonial regimes set up institutions conducive to long-term growth (see also Easterly and Levine, 2003).

**Demography**

In contrast to the experience of other regions, a fertility transition has not happened in Africa despite sharp reductions in infant mortality since the late colonial period and (until the HIV/AIDS epidemic starting in the 1990s) gradual improvements in life expectancy across the age distribution. The distinctive demographic features of African countries weigh unusually heavily on national saving and undermine the building up of the human capital needed for growth (Bloom and Sachs, 1998). Indeed, although we saw earlier that human development indicators have not diverged as strongly as income levels when comparing SSA averages with averages for other developing regions, these indicators have nonetheless deteriorated in relative terms. Enhancing human capacity by increasing the longevity of working life and improving skills and organizational effectiveness are important components of a strategy to close the growth differential with other regions.

Two distinct consequences of continued high fertility stand out. Firstly, the average population growth rate is at least one full percentage point above that for other developing regions. This increases the amount of national saving required to achieve any given increase in human and physical capital stocks per capita. It also increases the age dependency ratio, which reduces the per capita purchasing power associated with any given level of output per worker. High dependency ratios may also undermine the quality of human capital accumulation by spreading educational resources more thinly.

Secondly, until the early 1990s, rapid population growth produced not just a high but also a rising age dependency ratio. As we saw earlier, as Africa’s age dependency ratio gradually increased, the rest of the developing world
experienced a fertility transition that lowered population growth rates sharply and gradually reduced the ratio of dependents to working population over time.

In a regression-based counterfactual exercise, O’Connell and Ndulu (2000) estimate that Africa’s average growth is reduced by 0.85 percentage points relative to the sample mean, and by nearly 1.5 percentage points relative to East Asia as a consequence of its distinctive demographic patterns. This situation is made worse by the fact that HIV/AIDS has become epidemic and added to the burden particularly for survivors. HIV/AIDS patients in Africa account for 60 percent of the world’s people living with HIV/AIDS. This has a profound social and economic impact due to the large number of premature deaths of people in their prime age of employment and parenting.

A window of opportunity and sustained growth since the mid-1990s

Over the decade since 1995, 16 countries have had annual GDP growth of 5 percent or higher, by comparison with only five during the previous decade (Table 66.4). These countries account for 35 percent of the sub-Saharan Africa population. Some of the fastest-growing countries have also done relatively well in terms of poverty reduction, as demonstrated by a group of eight low-income African countries that grew at an average rate of 2.9 per capita per year and reduced poverty headcounts at an annual rate of 1.5 percentage points (World Bank, 2005). Excluding the oil-producing countries, income per capita in the fastest-growing one-third of African countries grew at a median rate of 2.8 percent over the five years 1999–2004. During the same period, the slowest-growing countries – predominantly those affected by conflict – saw their economies contract at a median rate of 1.8 percent. These host 21 percent of the region’s population. There has nonetheless been a striking decline in the number of countries posting negative growth rates of total GDP: four during the most recent period, down from 13 during the first half of the 1990s. In the middle of the growth distribution (accounting for some 13 percent of Africa’s population) the median growth rate was 1.0 percent per capita over 1999–2004.

These trends reflect important changes that are taking place across the continent. Policies and institutions are improving, peace and security is returning to the region, and African governments are increasingly taking control of their own economic destiny. Increased political participation and competition are giving Africans a greater stake in their own future. Demographic trends appear to have taken the first turn towards a transition that will reduce pressures on fiscal resources, encourage savings and support productivity growth. We elaborate below on each of these areas of progress.
Table 66.4 Average GDP growth rates: sub-Saharan Africa

<table>
<thead>
<tr>
<th>Country</th>
<th>85/94</th>
<th>95/04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angola</td>
<td>−0.9</td>
<td>7.7</td>
</tr>
<tr>
<td>Benin</td>
<td>2.9</td>
<td>5.0</td>
</tr>
<tr>
<td>Botswana</td>
<td>8.2</td>
<td>5.5</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>3.6</td>
<td>4.7</td>
</tr>
<tr>
<td>Burundi</td>
<td>2.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Cameroon</td>
<td>−1.6</td>
<td>4.6</td>
</tr>
<tr>
<td>Cape Verde</td>
<td>4.4</td>
<td>5.8</td>
</tr>
<tr>
<td>Central African Republic</td>
<td>0.2</td>
<td>1.8</td>
</tr>
<tr>
<td>Chad</td>
<td>4.2</td>
<td>7.4</td>
</tr>
<tr>
<td>Comoros</td>
<td>1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Congo, Dem. Rep.</td>
<td>−3.6</td>
<td>−0.6</td>
</tr>
<tr>
<td>Congo, Rep.</td>
<td>−0.4</td>
<td>3.1</td>
</tr>
<tr>
<td>Côte d’Ivoire</td>
<td>1.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Equatorial Guinea</td>
<td>2.9</td>
<td>22.3</td>
</tr>
<tr>
<td>Eritrea</td>
<td>−</td>
<td>2.3</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>1.6</td>
<td>5.2</td>
</tr>
<tr>
<td>Gabon</td>
<td>1.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Gambia, The</td>
<td>2.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Ghana</td>
<td>4.6</td>
<td>4.5</td>
</tr>
<tr>
<td>Guinea</td>
<td>4.0</td>
<td>3.9</td>
</tr>
<tr>
<td>Guinea-Bissau</td>
<td>3.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Kenya</td>
<td>3.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Lesotho</td>
<td>5.0</td>
<td>3.3</td>
</tr>
<tr>
<td>Liberia</td>
<td>−18.7</td>
<td>16.6</td>
</tr>
<tr>
<td>Madagascar</td>
<td>1.2</td>
<td>2.9</td>
</tr>
<tr>
<td>Malawi</td>
<td>1.7</td>
<td>4.1</td>
</tr>
<tr>
<td>Mali</td>
<td>1.7</td>
<td>5.9</td>
</tr>
<tr>
<td>Mauritania</td>
<td>3.3</td>
<td>4.3</td>
</tr>
<tr>
<td>Mauritius</td>
<td>6.4</td>
<td>5.0</td>
</tr>
<tr>
<td>Mozambique</td>
<td>4.2</td>
<td>8.0</td>
</tr>
<tr>
<td>Namibia</td>
<td>3.5</td>
<td>3.4</td>
</tr>
<tr>
<td>Niger</td>
<td>2.2</td>
<td>3.7</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4.7</td>
<td>3.6</td>
</tr>
<tr>
<td>Rwanda</td>
<td>−4.3</td>
<td>10.9</td>
</tr>
<tr>
<td>Sao Tome and Principe</td>
<td>1.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Senegal</td>
<td>2.2</td>
<td>5.1</td>
</tr>
<tr>
<td>Seychelles</td>
<td>5.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>−1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Somalia</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.6</td>
<td>2.7</td>
</tr>
<tr>
<td>Sudan</td>
<td>3.6</td>
<td>6.2</td>
</tr>
<tr>
<td>Swaziland</td>
<td>6.6</td>
<td>2.9</td>
</tr>
</tbody>
</table>
A significant and durable improvement in the policy and institutional environment

During the last decade a large number of reforming African countries have re-established sustained macroeconomic stability, committed credibly to more open trade regimes, and consolidated market-based economic reforms that have improved the conditions for private sector involvement in the economy.

Sustained macroeconomic stability has returned to a large number of countries in the region, as evidenced by significantly lower inflation, narrower fiscal and external trade deficits, and a widespread move to Article VIII status in the International Monetary Fund (IMF), implying a commitment to currency convertibility for current account transactions – a status hardly conceivable only a decade earlier.

Consumer price inflation has persistently and sharply fallen within a decade, from 27 percent in 1995 to about 6 percent by 2004. This has resulted from a combination of significantly stronger fiscal controls across a wide range of countries and a shift by central banks towards a focus on price stability as the primary goal (Ndulu, 2004). In the median African country, government spending as a proportion of GDP fell sharply in the past decade, as it has in other developing countries in the world. During the 1990s, fiscal deficits among the 31 low-income African countries for which comparable data exist dropped from double digits to 5.2 percent before grants and only 2.5 percent after grants (World Bank, 2000). The residual deficit has been financed largely through non-inflationary sources. Where financing through external grants is assured, fiscal deficits (before grants) have tended to be higher, but without destabilizing the macroeconomic environment. At the turn of the century black market premia were extremely low across the continent, averaging just 4 percent outside of a few countries like Zimbabwe in acute political turmoil. Through unilateral trade reforms, African countries have also compressed both the tariff rates and categories with average tariff rates of 15 percent. In these respects the

<table>
<thead>
<tr>
<th>Country</th>
<th>85/94</th>
<th>95/04</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tanzania</td>
<td>3.1</td>
<td>5.1</td>
</tr>
<tr>
<td>Togo</td>
<td>1.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Uganda</td>
<td>4.6</td>
<td>6.7</td>
</tr>
<tr>
<td>Zambia</td>
<td>0.6</td>
<td>2.8</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>3.7</td>
<td>−1.5</td>
</tr>
</tbody>
</table>

Source: World Bank data.

Table 66.4 (continued)
continent now more resembles other developing regions, where reforms have been pursued in earnest for prolonged periods.

We also use here the Country Policy and Institutional Assessment (CPIA) data from the World Bank to show the extent of improvement in the policy and institutional environment since the mid-1990s. Annually the World Bank assesses the quality of borrowers’ policy and institutional performance in areas relevant to economic growth and poverty reduction. These assessments began in the late 1970s but the criteria of good performance have evolved over time. While in earlier years assessments focused mainly on macroeconomic policies, they now include other aspects such as institutional quality and governance. CPIA scores for individual countries have been found to be highly correlated with private-sector country ratings, such as the Institutional Investor, International Country Risk Guide (ICRG) and Euromoney ratings (Easterly, 1993). The CPIA scores range between 0 and 6 and are benchmarked globally. A score of 3.5 is broadly considered as a threshold of good performance. The average CPIA score for African countries rose sharply starting in the mid-1990s, and between 1997 and 2004 the number of countries scoring above the 3.5 threshold tripled, from 5 to 15 – the latter number accounting for nearly a third of all countries in the region.

Peace and security is spreading in the region
After protracted periods of conflict, peace and security has returned to many parts of the region. Southern Africa entered the millennium as a region of stability following the resolution of long-standing conflicts in Mozambique, Namibia, Zimbabwe and Angola (2002) and the transition to majority rule in South Africa. West Africa likewise has seen the end of conflicts in Chad, Sierra Leone and Liberia. In Eastern Africa and the Horn of Africa, the resolutions of conflicts in Burundi and more tentatively in Sudan and Somalia hold out the prospect of durable change. Africa’s progress is significant relative to other regions of the world, as reflected in global data on the incidence of civil war. The incidence of civil war shows a sharp increase in both absolute and relative terms in SSA in the early 1990s, a period that coincided with the wave of first democratization documented above. Since the mid-1990s, however, there has been a sharp decline in the proportion of countries under civil conflict in the region; and the proportion of African population in such countries has fallen even more sharply, dipping below the average of other developing regions at the turn of the century. This positive change is corroborated by a new global database on the incidence of violent conflicts (Gleditsch, 2004). Collier and Hoefler (2004), using this data, show that immediately after the end of the Cold War the incidence of wars declined. The number
of wars peaked in Africa in 1992 and since then the number of civil wars seems to be declining.

*Increased political participation gives Africans a greater stake in their own future, laying a stronger foundation for domestic accountability and policy restraint*

During 1990–94, SSA experienced a sharp wave of democratization (Bates, 2007). By the mid-1990s, this episode had fully made up for nearly three decades of absolute and relative deterioration, placing the region well above the norms prevalent in the rest of the developing world during 1960–85.

Political competition and participatory processes improved by more in Africa during the 1990s than in other regions. In 1982, only one-tenth of African countries and two-tenths of other developing countries had competitively elected executives. As late as 1991, Africa showed virtually no improvement, while other developing countries had doubled their figure to 40 percent. By 1995, however, the gap was nearly closed, despite continuing increases in other regions. In 2002, Africa was ahead of the other regions by about eight percentage points.

The political democratization drive in Africa has created space for peaceful regime changes, deeper debates about societal development visions, and greater respect for human rights. Although we observed earlier that the initial wave of democratization was associated with a sharp rise in civil conflict in the region, this situation appears to have reversed itself as democratic practices have taken hold. The democratization process has also raised the expectations from citizens throughout the region. Analysis of recent Afrobarometer surveys and the World Values Survey show that the majority of Africans believe democracy is good for the economy. They also prefer democratic political systems to authoritarian alternatives, as indicated in Figure 66.3. The African public expects democracy to deliver access to the basic necessities of life, including food, water, shelter, and also education. These ‘values’ surveys also show that Africans care about equity and public action to reduce poverty. They report discomfort with wide wealth differentials and a strong commitment to political equality (‘voting not a privilege of the better-educated’).

*African governments are pursuing collective action to improve the region's attractiveness and taking increasing control of their economic destiny*

There is now a strong revival in regional integration initiatives in Africa, with a change in focus from preoccupation with preferential trade arrangements to an approach that emphasizes market integration and promoting the region as an attractive investment destination for foreign and African
capital. The African Union and the New Partnership for Africa’s Development (NEPAD) have embraced the latter two objectives. These objectives can be met if Africa as a region can: (1) achieve a critical mass of countries with a policy environment friendly to capital accumulation and private business; (2) improve cross-country infrastructure links; (3) moderate the risks faced by domestic and foreign capital, and employ risk-mitigating instruments more effectively; and (4) strengthen as well as retain its pool of human skills.

As part of the effort to improve the region’s collective reputation and attractiveness, African governments are taking action to improve governance and connectivity under the African Union (AU) and the NEPAD initiatives. These initiatives are designed to: (1) push African countries to be assertive about ownership and to assume leadership and accountability for their development programs; (2) improve the reputation of the region through certification of good practices in governance for a critical mass of African countries under the African Peer Review mechanism; (3) increase regional connectivity to improve capacity to trade within the region and with the outside world through regional initiatives to scale up collaborative effort in improving infrastructure; and (4) enhance the capacity of a rationalized system of regional bodies to provide regional public goods – such as cross-country transportation and power-sharing networks, coordination in managing pandemics such as HIV/AIDS and malaria, and protection of regional commons such as the Nile river basin and the Great Lakes.


Figure 66.3 African Political Values
Implications for future development strategies

We have anchored the analysis of Africa’s development experience in the overall weakness of the region’s growth record, the juxtaposition of this record with a population explosion, and the variability of experience across countries and over time. If recent improvements are to be sustained and deepened, a combination of addressing the fundamental bottlenecks to scaling-up growth, and accelerating the demographic transition now slowly under way, is fundamental. The wide diversity of opportunities and constraints in the region rules out generalizations about the country-level growth strategies appropriate to these goals. Nevertheless there are a few broad lessons worth highlighting by way of concluding remarks.

The development constraints African countries face are neither static nor decisive. Geographical disadvantages and natural-resource dependence are not destiny, as their effects can be offset or ameliorated. Botswana, the fastest-growing economy in Africa (and among the fastest globally) since its independence in 1966, presents a striking example. It is landlocked and natural-resource dependent. Arguably, the strength of its state capacity, together with its being part of Southern Africa’s relatively effective infrastructure system, customs union and monetary area (for a long period), helped offset the negative effects of remoteness and served as a commitment instrument against rent-seeking.

The historical analysis showed clearly that Africa virtually missed the rapid development that has taken place in other developing regions in the closing four decades of the twentieth century and indeed the region can be considered the last frontier of the global development challenge. Being a late starter has its advantages and disadvantages. Access to knowledge from development experience and technological progress presents an opportunity to fast-track the development process and leapfrog. At the same time, Asia’s cumulative success presents a challenge to the competitiveness of late-starting Africa, as trade preferences are eroded and opportunities to learn before facing intense competition shrink. Among the resource-poor coastal economies, the incumbency of the successful Asian coastal economies has probably created a more challenging playing field for African export diversification starting in the 1990s than existed in earlier decades.16

We can categorize the constraints to growth discussed above into three groups: risks, transactions costs and capacity. Risk is largely associated with macroeconomic instability and absence of credibility to commit due to weaknesses in governance and institutions. The higher transactions costs in the region are largely associated with its unfriendly geography, climate and bureaucratic processes. Capacity constraints relate primarily to low human capital, partly associated with the late demographic transition and weak institutions.
There is no doubt that raising the level and efficiency of investment is critical if Africa is to close the gap in growth with other regions of the world. Getting the right policy environment in place and sustaining such an environment is a key signaling instrument for credibility to investors (both local and foreign). Indeed, given the central role of the modern state in defining the incentive environment for private economic activity, the failure to engender sustained growth must ultimately be traced to unsuccessful policy choices. Collier and O’Connell (2007) estimate the contribution of anti-growth policies to the growth differential between Africa and other regions (p. 24); taking 40 years of African growth experience as a whole and controlling for differences in the composition of opportunities, they conclude that policies inimicable to growth account for more than half of the overall growth differential with the rest of the developing world, or as much as 1.8 percentage points out of an overall (population-weighted) differential of 3.5 percent. This result is confirmed by a regression analysis in which they estimate the effect of anti-growth policies, controlling for shocks and differences in growth opportunities across countries. This magnitude of relative importance of policy mistakes in explaining the growth performance differential with other regions is corroborated by evidence from cross-country growth studies reviewed extensively in O’Connell and Ndulu (2000).

Notwithstanding the importance of policy choices in explaining Africa’s growth differential with other developing regions, however, a substantial portion of the growth differential is accounted for by geographically based proxies for differential growth opportunities (Collier and O’Connell, 2007, pp. 81–8). Separating landlocked and resource-poor, coastal and resource-poor, and resource-rich countries, Collier and O’Connell find that while African countries tended to underperform in each category, nearly a third of Africa’s overall growth shortfall is associated with the unusually high share of African countries that are either landlocked and resource-scarce or resource-rich – both relatively low opportunity categories as compared to the coastal and resource-scarce group, based on the global evidence. The underperformance was most severe for Africa’s coastal resource-scarce economies (suggesting that missed opportunities dominate the story) and least severe for Africa’s land-locked resource-scarce economies (suggesting poor potential). Two sources of this differential are important – proneness to policy errors and the higher cost of doing development business, discussed above.

Proneness to policy mistakes is likely to be more important in resource-rich and coastal countries where the rent-seeking stakes are higher (from resource and trade rents), while higher transactions costs of economic activities from geographical disadvantages are likely to be more important
in landlocked countries. The convergence of resource wealth and ethno-regional fragmentation in many of the countries of SSA presents a particularly high risk of adoption of anti-growth policy regimes. The management of resource rents under conditions of ethnic diversity is one of the most important and distinct policy challenges for SSA.

A third challenge is dealing with the consequences of a dramatically delayed demographic transition in Africa compared to other regions. The fertility rate began to fall in Africa in the mid-1980s, suggesting entry into the final phase of the demographic transition. But unlike the experience in other regions, the pace of this transition is very slow; and the HIV/AIDS pandemic complicates the situation further. Although population changes are longer-term phenomena, as Srinivasan (1988) urges, it may be worthwhile to look at potential policy responses in the shorter term to help accelerate the demographic transition. These include education policy, population policy and gender equality to induce greater attention to the quality of children.

Much of the above discussion has employed *ceteris paribus* counterfactuals in order to isolate the individual effects on growth of policy mistakes, underprovision of public goods or weaknesses in human capital. If there are critical synergies across these factors, or if there are sharply increasing returns to individual state variables over some interval (as suggested by Azariadis and Drazen, 1990 and Berthélemy, 2005 in the case of human capital and Sachs et al., 2004 for both human capital and public infrastructures), then the achievement of rapid growth may require complementing policy improvement with a ‘big push’ to deal with poor governance, conflict and insecurity, and inadequate infrastructure. In terms of their ambition, contemporary versions of the big push are not novel; the record of Africa’s early decades is replete with the ambitions of visionary leaders who sought to engineer a sharp break from patterns of specialization and distribution inherited from the colonial era (Ndulu, 2007). What is different this time around, and provides a critically important window of opportunity for stakeholders, is that the core functions of market-friendly governance are widely understood to be among the binding constraints.

**Acknowledgment**

We thank Lopa Chakraborti for research assistance.

**Notes**

2. Here and throughout the chapter, we confine group totals to countries with continuously available data. This eliminates potentially misleading compositional effects within
groups, but it also means that group composition can differ by table or variable (for example, World Bank data are available for only 35 countries in SSA, while Maddison provides estimates for 52). For each variable (or table, as indicated), we report group totals for the full set of countries with continuously available data.  

3. Nigeria, Ethiopia, Democratic Republic of the Congo and South Africa together account for 45 percent of African population among countries with available data; these economies contracted at an average rate of 1.8 percent over 1974–94.  

4. We exclude oil countries, which grew at a median per capita rate of 3.4 percent. The upper, middle and lower thirds of the growth distribution comprise 35, 13 and 21 percent of Africa’s population, respectively.  

5. A recent contribution in this line is van de Walle’s *The Politics of Permanent Crisis* (2001), discussed further below.  

6. On within-Africa variation in agricultural policy and export taxation, see Mkandawire and Soludo (2000). On the weakness of within-country links from policy to long-term growth, see Easterly et al. (1993). On the ideological (as opposed to self-interested) motivations of African policy-makers, see Ndulu (2007), who notes the profound influence of Fabian socialism and dependency theory on founding African leaders. On the relative weakness of initial urban interests see Ndulu and O’Connell (1999), who observe that urban interests were nonetheless created by policy in some cases, and thereby came to undermine subsequent prospects for reform (see also the discussion of van de Walle, 2001 in the text).  

7. Sahn (1996) argues cautiously, on the basis of calibrated general equilibrium models, that if market-based reforms had been fully implemented in the 1980s, their impact would indeed have been mildly progressive.  

8. In McMillan’s analysis, policy-makers seek to maximize the revenue from taxing agricultural exports. They face a time-consistency problem, however: once farmers have sunk planting costs in the hope of receiving high producer prices, there is an incentive for policy-makers to pay very low producer prices that cover only the costs of harvesting. Evidence on crop- and country-specific export tax rates confirms that this incentive is strongest when policy-makers have urgent revenue needs and short planning horizons, and when the ratio of harvest to planting costs is small.  

9. In a celebrated article, Acemoglu et al. (2001) argued that colonial regimes brought development-oriented institutions only where local health conditions supported the establishment of a large settler presence (see p. 507).  

10. Migdal (1988) describes the development paradigm spanning this period as one in which the state is the ‘primum mobile’ of socio-economic progress. The idea of ‘developmentalism’ and the idea of state intervention were seen as inseparable, and policies and planning were seen as offering boundless possibilities for social engineering. It was taken for granted by multilateral and bilateral development agencies that the state had a pivotal role to play in transforming societies from backwardness to modernity (Ljunggren, 1993, pp. 7–8).  

11. Collier argued that in the resulting situation of executive dominance, effective restraints in the areas of trade and monetary policy would have to mimic the reciprocal and supranational structure of international trade agreements, where countries reciprocally commit to growth-promoting policies and to penalty structures capable of enforcing them. Donor conditionality, he argued, was ill-suited to fill the institutional vacuum; donors had their own constituencies and could not credibly threaten to terminate aid based on poor policy performance. As examples of partially successful supranational arrangements he cited the maintenance of low inflation in the CFA countries (the 13 members of 2 monetary zones issuing respectively the West African CFA – Communauté financière d’Afrique – franc and the Central African CFA – Coopération financière en Afrique centrale – franc) and Rand Monetary Area, and the avoidance of highly distorted trade regimes by members of the Southern African Customs Union.  

12. The Democratic Republic of Congo belongs in the latter category; perhaps also the Sudan, with its vast internal territory and limited access to its Red Sea coastline, and Ethiopia before the independence of Eritrea in 1994.
13. Limão and Venables (2001) estimate the median transport cost for a 40-foot container, from coastal port to destination (including transshipment), at $7600 for African countries. The comparable figures for Latin America and the Caribbean, East and South Asia, and the Middle East and North Africa are $4600, $3900, and $2100.

14. The 16 countries are Angola, Benin, Botswana, Cape Verde, Ethiopia, Mali, Mauritius, Mozambique, Rwanda, Senegal, Tanzania, Uganda, Chad, Equatorial Guinea, Liberia and Sudan (see Table 66.4).

15. These countries are Senegal, Mozambique, Burkina Faso, Cameroon, Uganda, Ghana and Cape Verde.


17. There is a wide range of other studies that corroborate this same conclusion. Ndulu (1998) reported the results of counterfactual simulations of Africa’s growth performance if conditions obtaining in East Asia were present in the region. Using results earlier obtained by Elbadawi and Ndulu (1995), Easterly and Levine (1997) and Elbadawi et al. (1997), a better policy environment as obtaining in East Asia would have raised growth by an additional 1.5–2.6 percent.

References


Berthélemy, Jean-Claude (2005), ’Conférence du G8: Pour chaque dollar de dette annulée, autant sera supprimé dans l’aide’, Marché tropicaux et méditerranéens, 60 (3112), July: 6–8 [interview].


Robinson, James A. (1997), ‘When is a State Predatory?’ mimeo, Department of Economics, University of Southern California, Los Angeles, October.
Introduction

The growth and development performance of the Middle East and North Africa (MENA) region presents one of the major anomalies that current economics literature seeks to resolve, which is how to reconcile the existence of massive natural resources with the high unemployment, low growth and general underdevelopment of the region. In this debate, much attention is focused on the problems arising from: (1) state-oriented inward-looking economic policies; (2) lack of ‘integration’ with the world economy; (3) underdeveloped financial sectors and chilling investment climate; and (4) low levels of human capital development. In this chapter, we attempt to present a summarized yet more balanced and hopefully more insightful analysis of the growth and development experience of the countries in the region, with special attention given to the existing bottlenecks hindering future development prospects.

While discussing the MENA region as a whole we will divide the countries into five subgroups: (1) oil-rich labor-importing states (Bahrain, Kuwait, Oman, Libya, the United Arab Emirates, Qatar and Saudi Arabia); (2) oil-rich labor-abundant states (Algeria, the Islamic Republic of Iran, Iraq, Syria); (3) oil-poor labor-abundant NICs (Egypt, Morocco, Turkey); (4) oil-poor limited natural resource states (Israel, Tunisia, the West Bank and Gaza, Jordan, Lebanon); and (5) natural resource-poor states (Sudan, Yemen) (Richards and Waterbury, 1996). Although the inclusion of Turkey, Israel and Iran is controversial as the trajectory of the Arab and other Middle Eastern countries constitute a more appropriate whole, they share many commonalities as well. However, unless stated otherwise, the general statements will exclude Turkey and Israel.

The economic history of the MENA region is characterized by several cycles of growth and accumulation. In retrospect, the region formerly enjoyed higher levels of economic development and prosperity compared to its counterparts in Europe. While Istanbul with its 700,000 inhabitants in the sixteenth century was the largest city in the world, North Africa overall was much more urbanized than Europe (Paris with 125,000 inhabitants versus Cairo with 450,000 in around 1500) (Bairoch, 1997, pp. 517–37). However, in the last of these cycles, the region experienced a decline in its growth and development indicators starting from the early eighteenth century, with the...
factors that precipitated this decline remaining a source of continuing debate. The current chapter will focus on the region’s most recent cycle, namely that after the 1950s following the gaining of political independence and control over natural resources by the countries in the region.

**Growth and development in the MENA**

From the late 1950s to the late 1970s (and in some cases, till the late 1980s and early 1990s) the economic structure of the region was characterized by an import-substituting industrialization (ISI) regime, the main features of which (as elsewhere) included strict quantitative controls on international trade, overvalued exchange rates and severe rationing in foreign exchange and credit markets.

Following the hikes in petroleum prices in the early 1970s, growth and development indicators in the MENA region improved rapidly. The sudden increase in investment and growth rates in the oil-exporting countries spread to the rest of the region through increases in worker remittances, and capital flows. In addition, gross capital formation jumped to exceptionally high rates, generating a locomotive effect on growth rates and overall standards of living. On the financial front, considerable amounts of financial savings were accumulated abroad, resulting in the famous expansion of the Eurodollar market through the recycling of petrodollars.

In contrast, the downside of the above picture has been the high level of volatility of gross domestic product (GDP) growth since the 1970s: the average volatility of GDP growth in the region as a whole has been twice that of the developing-country average, and twice more volatile in the oil-rich economies than the rest of the region (Abed and Davoodi, 2003; Hirata et al., 2004, pp. 62–3).

The single most important determinant of growth in the MENA (where fuel products account for about half of the region’s GDP and around 90 percent of total exports in the oil-rich countries) has been the fluctuations in international oil prices. In addition to growth volatility, as a result of high dependence on oil revenues, fiscal policy in the oil-rich countries is also volatile and procyclical. Likewise, the oil-poor labor-abundant countries are also oil price-sensitive because a large part of their economies is dependent on worker remittances as well as on development aid and tourism revenues from the oil-rich labor-poor countries. The non-oil-producing sectors, on the other hand, suffer from the ‘Dutch disease’ where the continuous flow of large oil revenues result in an appreciation of the real exchange rate, making it less competitive.

In large part due to the collapse in oil prices during the 1980s and 1990s the growth rates in the region experienced marked declines. Low growth rates failed to provide the rapidly expanding labor force with sufficient
employment opportunities and led to a deterioration of living standards and a rise in poverty rates. As a result, growth performance and GDP per worker and total factor productivity growth rates in the MENA region since the early 1980s has been near zero and negative, closer to Latin America and lagging far behind East Asia. Furthermore, despite substantial improvements since the gaining of political independence, the region lags behind both East Asia and Latin America in the UN Human Development Index (that is, in adult illiteracy rate, life expectancy at birth, and so on) (Bosworth and Collins, 2003; UNDP, 2002).

Nevertheless, there is considerable heterogeneity in performance across countries within the region. While the per capita incomes of the oil-producing countries declined at an average annual rate of –0.79 percent between 1980 and 2000, those in the non-oil-producing countries increased by around 2 percent over the same period. Looking at the oil boom years we get a similar picture where non-oil-producing countries grew almost twice as much as oil-producing ones. What are the reasons behind this diverse yet overall poor performance? We will turn to this question in the coming sections.

State, institutions and development

Despite the diversity in state structures, resource endowments and economic performance a characteristic shared virtually by all countries in the region (including Turkey and Israel) is the dominant role played by the public sector in the development process (Richards and Waterbury, 1996).

The majority of the states also experimented with the usual sequencing in terms of development models, going through an ISI period (accompanied by land reform) which would be disrupted (usually following an economic and/or political crisis, such as Egypt in 1967, 1974; Turkey in 1980) and be replaced by an outward-oriented development model where the role of the state is attempted to be downsized through domestic and external liberalization programs and public sector restructuring.

The first ISI attempts in the region were launched by Turkey in its first five-year plan of 1934, emphasizing the establishment of state enterprises in textiles, primary commodities and minerals, ceramic and glass, paper, chemicals and cement, and iron and steel as well as state banks for financing these enterprises. The Turkish model would provide a guide for the other MENA countries and be replicated throughout the region (Richards and Waterbury, 1996). In the oil-rich labor-abundant countries, the availability of massive oil rents, or what are termed ‘soft-budget constraints’, allowed the large countries to launch ambitious ISI projects. This led to a proliferation of capital-intensive (often turnkey) industries producing protected intermediate and final products for the domestic market. However, tariff
protection and credit access was often granted wholesale and no technological upgrading or other performance measures were required as was the case in East Asian countries.

During this period, despite the presence of a generally hostile attitude by the state bureaucracy, the private sector in many countries benefited largely from intermediate products supplied by the state enterprises at discounted prices or from other subsidies in the form of cheap credits or foreign exchange. One of the characteristics of the ISI era was that during this period the accumulation process was highly dependent on politics rather than markets. The political and economic environment thus created opportunities for wide-ranging rent-seeking behavior within the business community, as businesses competed for the special set of incentives (subsidized credit and foreign exchange, import licenses, and so on) provided by the state. The pre-liberalization ISI era thus gave rise to a narrow distributional coalition between the state bureaucracy and the business community.

Regarding financial and banking sector development, the region continues to suffer from the lack of an efficient banking system with long-term credit availability for private investment projects (excluding Israel). Furthermore, the use of state banks for political rent distribution in the form of distributing cheap credits on non-economic grounds manifests itself in the accounts of large ‘duty losses’ of these banks (OECD, 2001; Mitchell, 1999, pp. 29–30). In the case of capital market deepening, the money markets are mostly dominated by short-term government securities, while capital markets in private securities remain underdeveloped throughout the region.

Moreover, the tax system of the countries in the region has been characterized by the inability or unwillingness of policy-makers to implement an efficient and fair tax scheme which, in addition to resulting in a narrow tax base and high tax evasion, further contributes to the unequal distribution of the tax burden on low-income groups. As a result in virtually all countries in the region the business environment for private firms with no political ties with the state bureaucracy is not encouraging. Informal tax traps are common and lack of clear-cut and consistent tax laws result in incorrect assessments of tax obligations (for example EIU, 2004).

On the other hand, following independence the survival of these mostly autocratic regimes required distribution of economic rents to a wider group of supporters. Indeed, what is common in the region is that the continuous flow of revenues (mostly from oil rents) has helped postpone economic and political reforms, since the region as a whole (excluding Turkey and to some extent Israel) could manage to avoid the balance-of-payments crises that other developing countries faced at the last stage of their ISI experience.
Natural resources and development
After gaining full independence and national sovereignty in the post-colonial era, oil-producing MENA countries (which account for about three-quarters of the world’s proven crude oil reserves and 35 percent of global oil production) reaped the benefits of increasing oil revenues. Once national governments secured control over their oil production and pricing, oil revenues started to flow in cascades. For example, in the case of Saudi Arabia, crude oil revenues increased from around $10.4 million in 1946 to $104.2 billion in 1980 (Owen and Pamuk, 1998, p. 210).

However, this development has led to a Dutch disease with destructive effects on non-oil industrial sectors while retarding economic diversification and growth (Sachs and Warner, 2001). Appreciating domestic currency resulting from large oil-related foreign exchange inflows created an unsuitable environment for the development of domestic industries by making non-oil exports less competitive. Furthermore spending on massive construction projects further turned the terms of trade against manufacturing. Another major reason for the misalignment is the pegged or fixed exchange rate regimes adopted in the region as a whole (excluding Turkey) (World Bank, 2003, p. 110). Nabli and Veganzones-Varoudakis (2002) argued that MENA countries experienced overvaluation of more than 20 percent a year in their real exchange rates from the mid-1970s to 1999. They also suggest that the exchange rate policy explains losses in competitiveness and in manufactured exports in the region as a whole where real exchange rate overvaluation has decreased the ratio of manufactured goods to GDP by 18 percent a year.

On the other hand, in contrast to the abundance of oil resources, with 5 percent of the world’s population, the MENA countries have only 1 percent of the world’s renewable fresh water. According to the World Bank the region’s per capita supply ‘stands at only one-third of its 1960 level, and water availability is expected to halve over the next 25 years if the present pattern of use continues’ (World Bank, 2004, p. 4). Water shortage means that in addition to the strain of providing clean water to a rapidly increasing population, the countries are also increasingly dependent on food imports. Moreover, conflicts over water distribution and sharing have been exacerbated due to a lack of adequate regional conflict resolution mechanisms.

Trade and development
Historically, the MENA region was a thriving center of trade both originating within the region and as a crossroads for trade routes between Europe, East Asia and southern Africa. However, the shift in the balance of power between the MENA and Europe over the seventeenth and eighteenth centuries and Europe’s subsequent industrialization instituted a new
pattern of trade, that of manufactures exports from Europe in return for primary products and raw materials, and led to the subsequent decline and decimation of existing manufactures and crafts production that the region had enjoyed. During this period, any attempt by the region to industrialize was forcefully prevented (most notably by Britain and France), such as the industrialization efforts by Muhammed Ali in Egypt (Issawi, 1966, p. 363). This not only significantly shifted the pattern of production and trade, but also served to disrupt intra-regional trade in agriculture and manufactured goods, which had expanded under the consolidation of the region under the Ottoman rule (Owen, 1993).

More recently, the fortunes of the region (excluding Turkey and Israel) in the post-World War II period have been dependent on two types of trade. The first is the inter-regional export of fuels and other primary products (for example natural gas, iron phosphates), which during oil price booms reached almost 50 percent of GDP in the oil-exporting countries, ‘with 35 to 40 percent of GDP “spillover” effects for the region as a whole’ (Shafik, 1998). The spillover was mainly due to the second, intra-regional trade in labor, which has been a vehicle of transmitting the rents throughout the region, reaching as high as 20 percent of GDP for some countries such as Jordan and around 5–10 percent of GDP or higher for several countries such as Egypt, Syria, Morocco, Jordan and Tunisia (Galal, 2000).

Both the oil-rich labor-poor and labor-abundant countries have fuel exports that made up around 85 percent of their total exports as of 2000. The oil-poor small states, on the other hand, have successfully diversified their exports whereby manufactures made up around 75 percent of total exports in 2000. Furthermore, Israel and Turkey have highly diversified exports compared to the rest of the region, with Israel emerging as a world leader in high-technology exports.

Although regional integration or ‘Pan-Arab unity’ has been a prominent topic in the region, intra-regional trade in the MENA has never exceeded 8 percent of exports and is the lowest of any region in the world (Galal, 2000). The lack of diversified production structures has undoubtedly been a hindrance for intra-regional trade: gravity model estimations reveal that MENA countries trade about a third less than otherwise identical countries (Rose, 2002).5

In order to reverse this trend, the Arab Free Trade Agreement has been established, with 18 countries signing the agreement in 1997. Furthermore, as a sign of expanding regionalism, in addition to intra-Arab treaties, several MENA countries have signed bilateral association agreements with the EU, with others to follow (Fawzy, 2003).

The uncertainty in gains from regional integration is a risk for regimes that are wary of engaging in potentially destabilizing reforms. Moreover,
the availability of windfall rents has allowed the regimes to appease domestic constituencies, and prevented the formation of coalitions pressuring integration or other deep structural transformations (Carkoglu et al., 1998).

**Labor markets and human capital in the MENA**

Regarding demographical challenges, the MENA region has the second-highest population growth rate in the world after sub-Saharan Africa, exacerbating labor market problems. Although the rate has been decreasing in recent years, average annual growth in the labor force is still expected to be 3.4 percent a year in 2000–2010, which is twice that of other developing countries, with adverse effects on per capita incomes (World Bank, 2003, p. 19).

The presence of a disproportionately high share of the young in the population (under-30-year olds constitute almost two-thirds of the population on average), low growth rates and lack of skill development has resulted in high unemployment throughout the region.

It is estimated that 16 MENA countries that represent 60 percent of the regional population need to provide 47 million new jobs between 2002 and 2012 just to keep up with the increasing labor supply (Keller and Nabli, 2002). As a result, the unemployment (and underemployment) rate is quite high in the region, and despite underestimated official figures stands at around 15 percent in the Arab countries (UNDP, 2002). Furthermore, since 1981 the labor force has grown faster than population growth and can be expected to increase further with increasing female participation rates.

In the case of oil-poor countries, another problem lies in their dependence for job growth on the oil-producing countries. As of 1997, for example, foreigners in Kuwait held 99 percent of private sector and 42 percent of public sector jobs. The ratio for the foreign to domestic workforce is 90 percent in the UAE, 83 percent in Qatar and 69 percent in Saudi Arabia (McMurray, 1999, p. 19).

Regarding human capital, following political independence the MENA countries faced a daunting task to educate their population, with adult illiteracy of 70 percent in Syria and 85 percent in Algeria, Iraq and Libya around independence (El-Ghonemy, 1998). The colonial powers had established parallel systems of education and the systemic discrimination in education left the majority of the population, especially in rural areas, with dilapidated and low-quality public schools while the expatriates, the urban elites and sectors friendly to colonial powers enjoyed high-quality educational establishments (El-Ghonemy, 1998).

Since the post-independence period MENA countries have invested a high proportion of their GDP towards education and health, and have
made remarkable gains on both counts. Average illiteracy rate dropped from 60 percent in 1980 to about 43 percent in the mid-1990s, while enrollment at all levels went up from 31 million to 56 million during the same period (UNDP, 2003). However there is still widespread illiteracy among youth and adults and even higher rates among women and the rural poor.

A side-effect of the industrialization attempts by the MENA countries was an allocation of resources towards secondary and higher education, which typically have lower social rates of return than primary education. The result has been the oddity of unemployed highly educated workers, while having large numbers of illiterate adults and youth (Richards and Waterbury, 1996).

**International conflicts and socio-political instability**

The region has been plagued with ongoing conflicts since the eighteenth century, starting with the decline and the following collapse of the Ottoman Empire and the erection of colonial regimes. Having borders drawn by the colonial powers based on politics rather than historical, cultural or ethnic backgrounds or social consensus led to subsequent ethnic and religious civil conflicts (for a detailed list of these conflicts, see for example Elbadawi, 2005, pp. 306–7).

In addition, since 1948 the Middle East has witnessed: four wars between Israel and several of its Arab neighbors; three wars with Western countries; the full occupation of Iraq and Palestine and the partial occupation of Egypt, Lebanon and Syria; extended periods of economic sanctions on Syria, Iraq, Sudan and Libya; and several *coup d’état* instigated from within and outside the region. The Iran–Iraq war alone left around 1 million dead and 2.5 million refugees, with an estimated cost of $200 billion. These conflicts have had a direct impact on state structure and overall trajectory of development.

The artificial mapping of the region with sovereign borders overlapping with different ethnic and religious groups further fed into the authoritarian state structure, thanks to the excuse that the survival of the unity of the country is dependent on the suppression of popular demands by different groups.

The majority of publications on socio-political risk and the investment and growth relationship find a negative correlation between these variables. Venieris and Gupta (1986), Alesina and Perotti (1996) and others find an inverse relationship between political instability and growth or investment, or savings rate. In addition, Asteriou and Price (2001) found that socio-political instability not only negatively affects the growth rate but also increases its volatility. Similarly, Rodrik finds a significant negative relationship between external shocks and growth in countries where there
are latent domestic social conflicts and poor conflict management institutions, as in the case of the MENA (Rodrik, 1998). The idea behind the above research is that socio-political unrest and instability disrupts market activities and investment decisions by increasing uncertainty and risk while directing limited resources to non-productive security-related expenditure.

As a result (or on the pretext) of non-stop civil or military conflicts the existing regimes have devoted a sizable portion of their budgets to military spending. The average military expenditure to GDP ratio in the region is 6.6 between 1990 and 2004 with a maximum of 21.8 in Kuwait and minimum of 1.8 in Tunisia. Comparatively, the averages were 1.4, 0.5 and 2.5 in Argentina, Mexico, and Malaysia for the same period (SIPRI, 2005). Such military spending creates a substantial potential for peace dividend in the region. However, for the peace dividend to materialize, the peace must be ‘real and durable, and perceived as such’ (Fischer et al., 1993).

Economic reform in the MENA

Despite the presence of a general consensus among policy-makers and economists on the need for reform, the question regarding which path to follow remains unanswered. Several countries in the region have embarked on structural adjustment programs (SAPs) under the guidance of the International Monetary Fund (IMF) and the World Bank. The reforms included standard policy packages by the twin institutions such as fiscal reform (introducing value-added taxes – VATs; eliminating state subsidies; increasing transparency in public expenditures), liberalizing trade and capital accounts, and shifting to more flexible foreign exchange regimes. Despite differences, the countries that have enjoyed higher rates of growth since the early 1990s have been those that implemented reform programs (that is, Egypt, Jordan, Morocco, Tunisia) (Hirata et al., 2004).

Nevertheless, it is difficult to make a generalized statement on the success of the SAPs in the MENA due to credits and debt reliefs extended to certain reforming countries based on political considerations during the adjustment period (for example Egypt, for its support to the first Gulf War) (Gray, 1998).

Despite the implementation of comprehensive trade and financial liberalization programs including tariff reductions, privatization, tax breaks and eased restrictions on foreign ownership, as well as establishment of free trade zones and other incentives to encourage foreign direct investment (FDI), capital flows to the region remain minimal. The region’s share of FDI fell to 0.7 percent in 2000 from 2.5 percent in 1980 (Hirata et al., 2004). In addition, most of the capital flowing into the region appears to be in short-term funds. In the case of Turkey, capital account liberalization has
exposed it to the uncertainties and instabilities associated with short-term capital flows, which have demonstrated themselves in three major crises in 1994, 2000 and 2001. Furthermore, as shown in the case of banking sector crises and subsequent cost of bank defaults resulting from endemic corruption as well as rent-seeking promoting IMF engineered policies (such as 100 percent state insurance on private bank deposits in Turkey), the countries in the region need a major restructuring in their financial systems.

On the other hand, income inequality and poverty rates have increased since the implementation of reform policies (Ali and Elbadawi, 2002; Fergany, 1998). The region had previously enjoyed the lowest incidence of poverty and income inequality of any region in the developing world (Adams and Page, 2003). As a result, only 5.6 percent of the population in the region lived on a less than the $1 a day benchmark compared with 14.7 percent in East Asia and 28.8 percent in Latin America (Shafik, 1995). Adams and Page (2003) pointed out two statistically and economically significant reasons for this: international remittances, and public sector employment and the welfare state. The policy of public sector employment and subsidized public services and pricing to generate popular support for the survival of the political regimes appears to be the common element in the region. However, the slow-down in growth as well as neoliberal reforms, which have scaled back the role of the state, have reversed the trend of lowered inequality (Ali and Elbadawi, 2002; Fergany, 1998).

**Conclusion and policy suggestions**

Contrary to the view that finds ‘little reason for gradualism’, as in World Bank (2003, p. 7), there may arise significant socio-political costs to a ‘big bang’ approach in the region. In a majority of MENA countries, certain sectors and groups of people (that is, peasantry, civil servants, organized labor) will stand as absolute losers from the reform programs, at least in the short run. Furthermore, the economic and political failures of the past have created an unstable environment pregnant with socio-political fault-lines, which are further exposed by slow growth rates, increasing unemployment, and increasing income inequality and poverty among different income groups and different regions.

The worsening economic performance has radicalized the divide between urban and rural, secular and Islamist, and ethnic identity groups, and these politicized fault-lines have, in turn, been accompanied by increasing authoritarian governance in the region (Lubeck, 1998, p. 299).

As a result, increasing hegemony of neoliberal policies along with economic liberalization and deregulation of markets may have the opposite effect on political liberalization and consolidation of democracy in the
region, by further deepening such divisions through increasing economic insecurity and social dislocation among the public.

The experiences of MENA countries suggest that historically determined institutional characteristics and the political environment of a country are of crucial importance in determining both the nature of the adjustment process and subsequent economic performance. Given that the market-led, outward-oriented reform programs have not produced the anticipated results so far, there is a continuing debate among economists about the underlying reasons. This chapter follows the line that developing countries share common structural problems in their institutional settings, and that policies that are designed to liberalize their economy (and political and civil life) may also generate serious instabilities without necessarily eliminating the previously existing ones. The existence of strong state hegemony in the form of military, legislative and economic institutions with a lack of clear-cut lines between private and public spheres resulted in a lack of democratic accountability and transparency during the design and implementation of reform packages. In addition, previously formed rent-seeking coalitions have prevented the implementation of a comprehensive reform program designed according to the needs of the countries in the region.

As a result, instead of removing the state, the neoliberal reform programs helped the state become instrumental in distributing rents to a new group of rentiers that make their living from financial rents (Mitchell, 1999, p. 30; Yeldan, 2001; Demir, 2004, 2005). Hence the state(s) ‘now subsidizes financiers instead of factories, speculators instead of schools’ (Mitchell, 1999, p. 31).

The recent experience of MENA countries suggests the state and the institutional infrastructure need to be reformed before embarking on reform programs that may undermine the legitimacy of the state structure and lead to socio-political instabilities. In other words, sequencing of reforms is a must both for the sustainability of the reforms and for their further deepening. The future of the region in terms of economic and political outcomes depends on the following:

1. Institutional reform in the form of judicial, legal, administrative and prudential regulation including the rule of law should be established. In addition, rent-seeking groups need to be controlled if any economic reform is to be successful.
2. Providing social safety nets for the disadvantaged and the losers during transition.
3. Political liberalization including reforming the state and making it democratically accountable rather than populist.
4. Privatization of the ownership or the management of State Economic Enterprises (SEEs) and public banks are needed to avoid corruption, rent-seeking and subsidized credit distribution to a few wealth groups based on political considerations.

5. Boom–bust cycles need to be stabilized especially in the case of resource-rich countries, which are dependent on the changes in the oil markets. As also argued by the World Bank (2003, p. 10), the countries need to: establish rules that shield fiscal spending from fluctuations in oil revenues; create deposit accounts for oil revenues to be set aside for future generations; and avoid misalignment in exchange rates. This may have solved the ‘resource curse problem’ in the oil-rich countries by offering an alternative to investing revenues in non-profitable and non-competitive domestic investment projects. This may also pave the way to avoid currency appreciation and support competitive domestic sectors. Also, this may provide an outlet for intergenerational resource distribution for future generations, given the limited supply of oil reserves.

Notes

1. Israel is considered an industrialized country and in that sense is in a separate category to other MENA countries.

2. Our classification here is slightly different from Richards and Waterbury and is meant to highlight export structure, intra-regional labor migration, patterns of industrialization and dependence on oil revenues. For example, the proven oil reserves of Syria are negligible; however it was highly dependent on oil exports for revenues during the ISI period and until today as fuel exports made up 76 percent of total exports in year 2000.

3. Some recent scholars attempted to explain the lagged performance in the region with the legacy of Islam. Kuran (2004) for example, blames the Islamic *waqf* or trusts that locked capital into a dysfunctional institution, Islamic inheritance law which dispersed inheritance among multiple heirs, and the individualism of Islamic law as preventing capital accumulation à la Europe. However, the lagging performance of the region *vis-a-vis* Europe came far too recently on a historical scale to be pinned on the influence of religious (or cultural) institutions. Moreover, as Inalcik (1969) emphasized, Islamic society and law ‘shaped themselves from the very first in accordance with the ideas and aims of a rising merchant class’ (Inalcik, 1969, p. 101). Finally, any attempt to explain the decline in economic performance of the region after the eighteenth century with the religious and cultural factors or institutions should also be able to explain how the same institutions could create the opposite results prior to that date.

4. In Turkey, unlike others, the state assumed a direct role in creating and supporting the development of a national business class; Bugra (1994).

5. The colonial legacy on regionalism warrants greater attention. As Ventura-Dias (1989) argues, colonial powers promoted intra-regional trade in Asian countries which allowed ‘permanent marketing channels to be established’ in contrast with both MENA and Latin America, where colonial intervention disrupted intra-regional trade.

6. The only country in the region with a definite plan to limit the harmful effects of the oil curse is Kuwait, which preferred to utilize its oil revenues on investments abroad (Owen and Pamuk, 1998, p. 216).
References


Economic Intelligence Unit (EIU) (2004), Country Finance Egypt, London: Economic Intelligence Unit.


Many economists believe that China today is largely a market economy. Yingyi Qian (2003), a leading economist on China, observes: ‘In the last 22 years of the 20th century, China transformed itself from a poor, centrally planned economy to a lower-middle-income, emerging market economy.’ Barry Naughton, another leading economist on China, echoes this view in his recent textbook (Naughton, 2007). This chapter counters this claim with evidence that shows that China today is far from a market economy – defined as one predominantly based on private ownership.

This chapter first reviews China’s growth experience since 1978. I will note briefly the enormous gains China has made, but the emphasis here is to highlight aspects of Chinese performance that are less well known, such as the increasingly investment-driven growth, the slowdown of productivity growth, and some notable microeconomic inefficiencies (as compared with, say, India). Social performance also deteriorated in the 1990s.

I then turn to ask the question, ‘Just how capitalist is China?’ The answer, surprisingly, after nearly 30 years of reforms, is not at all clear. There is evidence that China today is a commanding-heights economy similar to some of the most statist economies of the 1970s. I use data on fixed-asset investments by what I call the registered domestic private sector to show that in the 1980s the domestic private sector developed vigorously but in the 1990s the pace of liberalization and denationalization slowed down. The final section concludes with some broad implications of this analysis.

China’s growth experience
China’s economic success is indisputable. Its gross domestic product (GDP) growth has led the world in the growth table. Between 1978 and 2004, according to the World Bank, real GDP growth per annum averaged 9.73 percent, the fastest in the world. Poverty has fallen dramatically since the onset of the reforms in the late 1970s. By one estimate, the overall poverty level in China – measured as the percentage share of the population living under the poverty line – declined from 53 percent in 1981 to only 7.97 percent in 2001 (Ravallion and Chen, 2004). In the 1970s, the Chinese economy was plagued by shortages; today, it is often blamed as the source of the worldwide deflation as the economy has become a powerful export engine.
I do not repeat these well-known success indicators. Rather, this is a focused treatment of those topics that have received less attention in the academic literature. These include the non-economic foundation of China’s rapid growth, the heterogeneous growth record during the reform era, and the divergence between the GDP indicators and other development indicators. The purpose of the discussion is to highlight the complexities involved in trying to understand China’s growth experience.

The social and political foundations of growth
Economists assign a heavy weight to the reforms when explaining China’s economic success. This is not wrong, but it is important to point out that cross-country evidence on the linkage between policy reforms and economic growth is not nearly as clear-cut. Yet, it is uncontroversial that China’s impressive growth clearly followed the policy changes introduced in the late 1970s. Even if we acknowledge the importance of the reforms, the appropriate way to frame the discussion is to ask whether the policy changes introduced in the late 1970s interacted with some of the favorable pre-existing conditions in the country and whether it was an interaction effect, rather than just the policy changes alone, that spurred China’s growth.

In this context, it is useful to compare China with India. When examining the details of India’s growth record, Hausmann et al. (2004) correctly note that India’s growth rate began to pick up relative to the historic benchmark in the 1980s and that this acceleration of growth coincided with a number of modest policy changes. In addition, they note that China’s growth also followed seemingly modest policy changes in the 1980s. If we accept this characterization of the respective policy changes in China and India, we are still left to explain why China’s growth was so much stronger than India’s growth, as indicated in Table 68.1. For the period on which Hausmann et al. focus, that is, from 1978 to 1990, the average annual GDP growth in China was 9.28 percent, almost twice India’s growth rate of 5.01 percent.

This differential suggests the importance of identifying some initial differences between China and India that might have contributed to their divergent growth performances in the 1980s. The most important initial difference in favor of China has to do with the social conditions. For complex reasons, during the orthodox socialist period (1949–78) China invested heavily in health and education, especially in the rural areas. For example, as early as 1965 the life expectancy of Chinese women at birth was 55 years, compared with only 44 years in the case of Indian women. In the 1980s, infant mortality in China was substantially lower than infant mortality in India. (In the mid-1980s, China’s infant mortality rate was 54 per
Furthermore, China’s primary education enrollment ratio was far higher than that of India as early as 1975. China also had a more equal initial distribution of income compared with that in India.4

Panel (A) of Table 68.1 gives the average annual growth percentages of real GDP for China and India. Panel (B) gives the average annual growth percentages of real GDP per capita. Comparing the data in Panel (A) with the data in Panel (B) reveals an interesting pattern: Although China’s aggregate GDP performance was better than that of India, its per capita GDP performance was even better. For the period from 1978 to 2004 as a whole, China’s GDP growth rate was 1.81 times that of India (9.37 percent compared with 5.37 percent). However, the ratio of Chinese GDP per capita growth to that of India was 2.51 for this period (8.44 percent compared with 3.37 percent), substantially exceeding the GDP growth ratio of 1.81 of the two countries. It is important to understand the source of this differential.

One hypothesis centers on population control. China has a draconian population control program, but for political reasons India cannot replicate this aspect of China’s ‘development strategy’. Because both countries have a chronic surplus of labor, it is plausible to argue that the higher GDP growth per capita in China – far in excess of its aggregate GDP growth – is

### Table 68.1 GDP growth and GDP per capita growth, China and India, 1978–2004 (%)

<table>
<thead>
<tr>
<th></th>
<th>Panel A: Average annual % growth of real GDP</th>
<th>Panel B: Average annual % growth of real GDP per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>9.73</td>
<td>9.28</td>
</tr>
<tr>
<td>India</td>
<td>5.37</td>
<td>5.01</td>
</tr>
<tr>
<td>Ratio of China to India</td>
<td>1.81</td>
<td>1.85</td>
</tr>
</tbody>
</table>

**Notes:** GDP data are calculated on an exchange rate basis.

**Source:** Data are downloaded from *World Development Indicators* at [http://devdata.worldbank.org/dataonline/](http://devdata.worldbank.org/dataonline/).
partially due to the coercive capacity of the state. Many economists neglect this fact when they tout China’s supposedly superior record of economic reforms; in reality it is the political, not the economic, management that matters.

**Heterogeneous growth experience**

It is interesting to explore the substantial heterogeneity in China’s growth experience during the long time-span since reforms (1978). I deal with three respects: the importance of investments for China’s recent growth performance, productivity development and social performance.

Let us revisit Table 68.1, which breaks down the reform era into three sub-periods. GDP growth was fastest pace during the 1991–97 period, averaging 11.5 percent per year. The period from 1978 to 1990 came second, at an annual 9.28 percent on average. The most recent period from 1998 to 2004 turned out to be the least impressive, averaging 8.76 percent per year.

These fluctuations in the GDP growth coincided with huge changes in the investment levels. In the 1980s, the gross fixed capital formation as a percentage ratio to GDP averaged around 30 percent. During the recessionary years of 1989 and 1990, this ratio declined to 25 percent and then it surged to 35 percent in 1993. Starting in 1997, the ratio increased sharply, reaching 40 percent in 2004 and then 48 percent in 2005, another new high.5

So there was a change in the drivers of growth over time. Of the three sub-periods presented in Table 68.1, China’s GDP performance is the least impressive during the 1998–2004 period, but this is also a period when China was investing at its highest level. Thus, at 8.76 percent per year, although China was still leading the world in GDP growth, it was achieving this performance at a substantially higher level of investments than when it was growing faster in the 1980s and the early 1990s.

Is China’s latest growth spurt as sustainable as the one it experienced in the 1980s and the early 1990s? The answer depends on the productivity of the latest investment surge. If the high levels of investments are accompanied by or directly lead to technological progress, then an investment-driven growth pattern can be sustainable. Research on Western economies has shown a sustained, long-term positive correlation between productivity growth and capital deepening (Wolf, 1991).

The evidence, however, suggests that the latest investment growth did not lead to improved productivity. An exhaustive survey of the various studies on China’s total factor productivity (TFP) reveals one consistent pattern: TFP performance declined beginning in the late 1990s relative to earlier periods. The key findings are summarized in Table 68.2. These studies differ on the TFP estimates but they converge on trend developments: TFP growth during the last period, that is, in the late 1990s or early 2000s, was
Table 68.2  Estimates of the annual TFP growth in the Chinese economy (%)  

<table>
<thead>
<tr>
<th>Sources of estimates</th>
<th>Level of data</th>
<th>1980s</th>
<th>1990s</th>
<th>2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>First half</td>
<td>Second half</td>
<td>First half</td>
</tr>
</tbody>
</table>

Note:  a  Kuijs and Wang (2005) estimate TFP growth to be 3 percent during the 1978–2004 period and 2.7 percent during the 1993–2004 period. I have calculated TFP growth to be 3.74 percent during the 1978–93 period on the basis of their estimates.

Sources:  See the table for citation information.
considerably more moderate than TFP growth in the 1980s and the early 1990s. For example, Zheng and Hu (2004) report that TFP grew annually by 3.26 percent between 1978 and 1995, but during the 1995–2001 period TFP growth virtually disappeared (0.32 percent). Focusing only on Chinese industry, Ren and Sun (2006) report a reduction in TFP growth of a similar magnitude.6

Although during the reform era as a whole China made substantial progress in eradicating poverty, progress was uneven across both space and time. The largest gains in poverty reduction occurred in the first five years of the 1980s. According to Ravallion and Chen (2004), between 1980 and 1985 poverty declined drastically. In the rural areas, the incidence of poverty – measured by the headcount of those living below the poverty line – declined from 75.7 percent to only 22.7 percent. Income distribution improved in the early 1980s, as indicated by a reduction of the Gini coefficient.

Since then, the pace of poverty reduction has been considerably more measured and there have been episodic setbacks. In 1998 rural poverty was at 11.6 percent, and in 2000 it rose to 13 percent and in 2001 to 12.5 percent. Income distribution also deteriorated. In 2001 China had a Gini coefficient of 39.45, compared with 27.98 in 1980. This remarkable speed of reversal, according to two experts, is ‘almost unheard of in the developing world’ (Khan and Riskin, 2001).

The retrogressions in the late 1990s warrant special scrutiny. First, productivity indicators for this period deteriorated as well, so there was no economic and social trade-off. Second, according to the data provided by Ravallion and Chen (2004), the poverty level increased in the late 1980s as well. But the circumstances of the late 1980s differed substantially from those of the late 1990s. The surge in poverty in the late 1980s can be easily explained by the macroeconomic shocks – the Chinese economy went into a severe contraction in the late 1980s. It is more difficult to explain the increase in poverty during the economic boom period of the late 1990s.

The adverse social developments in the late 1990s are a sign that Chinese growth may have acquired an inherent anti-poor bias. The magnitude of the effect is substantial. Consider the increase in poverty incidence from 11.4 percent in 1999 to 12.96 percent in 2000. In percentage terms this increase may not be striking, but because China has a huge rural population, such a seemingly small rise in poverty incidence in fact corresponds to 11.24 million rural residents newly thrust into poverty, equivalent to the entire population of Greece. Other – far less known – indicators are also telling. For example, the World Bank has documented that during the reform era China has underperformed – both against other countries such as India and against its own economic potentials – in terms of reducing
infant mortality. The World Bank has also reported that China is one of only seven countries in the world to have a higher infant mortality rate among girls than among boys.\(^7\)

The most recent evidence is even more alarming. The World Bank has just reported that the income of China’s poorest 10 percent of the population declined by 2.4 percent between 2001 and 2003 (McGregor, 2006). This is the first documented evidence that a large number of Chinese people – about 130 million people – has actually experienced an absolute reduction in their living standards. The issue is no longer one of ‘relative deprivation’, about which economists tend not to be as concerned, but one of ‘absolute deprivation’. If this trend continues, it will have serious implications for the prospects for China’s growth and political stability.

Creating output vis-à-vis creating value

Bai et al. (1997) draw the distinction between the technical capabilities on the part of the SOEs to produce outputs and their economic capabilities to create value. State-owned enterprises (SOEs) can produce a massive quantity of a product, but the product itself may lack demand. In this case, there is a divergence between technical and economic measures of efficiency.

We can apply the same reasoning to the Chinese economy as a whole. It is well known that India’s GDP growth has lagged that of China for much of the last two decades of the 1980s and 1990s. But a little-known fact is that the Indian economy is able to create more value and wealth for a given unit of GDP than is the Chinese economy. It is intriguing to note that India has a higher manufacturing value added per worker than China. The value added per worker in manufacturing was 2885 dollars per year during the 1995–99 period for China, but 3118 dollars per year during the same period for India. In fact, the value added in manufacturing declined between the mid-1980s and the mid-1990s in the case of China but increased in the case of India.\(^8\)

An International Monetary Fund (IMF) paper shows that India ‘overly’ specialized in highly skilled industries at the expense of low-skilled industries.\(^9\) But even if the two countries differed in their initial starting points, it still begs the question why the manufacturing value added declined over time in China. The aforementioned IMF paper shows that between 1981 and 1996 China’s share of output in skill-intensive industries was not only lower than that of India, but it was also declining over time. The declining value added in Chinese production suggests that Chinese firms were not climbing up the value chain during a period of massive boom.

Just how capitalist is China?
The conceptual framework explaining China’s transition to a market economy is gradualism – the idea that the reforms are an endogenous
process whereby ‘unhooking a single key connection can cause the entire fabric to unravel’ (Naughton, 1996, p. 311). A critical empirical benchmark is the size of the private sector. China started out with a very small private sector, but due to the increasingly supportive policy environment, the private sector grew and overtook the previously dominant state sector.

I do not question the logic of a gradualist approach nor the empirical basis for applying a gradualist approach to China in the 1980s. The findings of a deterioration of TFP performance, worsening social performance and the increasingly investment-driven nature of the high growth since the mid-1990s raise a question whether gradualism is still an accurate empirical characterization of China today. Apart from the fact that it sheds light on the sustainability of China’s growth, the TFP performance can be a proxy measure of the reforms. In this regard, the across-the-board finding that TFP growth began to deteriorate in the late 1990s is significant. Did the reforms stall in the 1990s?

I examine a key benchmark in the gradualist interpretation of China’s reforms – private sector development. I ask, ‘After nearly 30 years of transition, just how capitalist is the Chinese economy?’ Surprisingly, the answer is not straightforward. I distinguish between an output-based measure and an input-based measure of the size of the private sector and I show that there is a huge discrepancy between these two measures. First, while the output-based measure shows the size of China’s private sector to be quite large, the input-based measure gives rise to a far smaller estimate. Second, while the output-based measure shows a continuously rising private sector, the input-based measure indicates severe setbacks – and even retrogressions – in private sector development in the 1990s. The latter finding directly contradicts the predictions of the gradualist framework.

In the following paragraphs, I will first describe the conceptual and methodological problems underlying many of the output-based measures of the Chinese private sector. I will then present detailed statistics based on a critical input – fixed-asset investment – on the evolution of the registered private sector in China. This measure shows a robust development of the private sector in the 1980s and a sharp contraction in the 1990s.

The registered private sector firms, either individual businesses (getihu) or privately operated enterprises (siying qiye), refer to newly established private businesses registered as such with the government. For sure, this is a narrow and conservative measure of the private sector but, as I will explain, an examination of the development of this sector provides an important insight into the course of the economic reforms in the 1980s and 1990s.
Output-based measures

The most frequent measure of private sector development used by economists is the share of the private sector in production. By this measure, China’s private sector has made huge strides. For example, Naughton (2007, p. 300) shows that the domestic private sector accounted for 19 percent of industrial output value in 1996, up from zero in 1978. This type of data is often used to support the argument that the policy environment improved steadily for the domestic private sector.

An output-based measure incorporates two very different effects. One is the ‘policy effect’: the increase in the private sector share that resulted from a more favorable policy environment. But this measure also incorporates what might be called an ‘efficiency effect’. The private firms are more efficient than the SOEs and, therefore, even given a very narrow business space, they can outcompete the SOEs. This suggests that the ratio of the private to the state sector can rise without any improvement in the policy environment for private sector firms and with rising inefficiencies of the SOEs. Thus, this measure tells us as much about China’s policy environment as about the huge inefficiencies of the state sector, and we cannot distinguish which of the two dynamics is driving this ratio. As an illustration, in 1985 the industrial output of the private sector was about 2.9 percent that of the state sector; by 1997, this ratio had risen to 70.2 percent. Even if the argument is correct that the policy environment improved between 1985 and 1997, it would be highly misleading to conclude that the policy environment facing the private and state sectors converged at about 70.2 percent in 1997.

China economists use a broader measure than the registered private sector to showcase China’s transition success. The Chinese style of reforms has spawned a variety of hybrid and highly ambiguous ownership forms, such as SOEs with some private revenue rights, collective firms controlled at the local levels, and private–state joint-ownership firms. It is difficult to sort out who actually controls these myriad firms. The most careful analysis to decompose the ownership of Chinese firms has been carried out by two economists at the Organisation for Economic Co-operation and Development (OECD) (Dougherty and Herd, 2005). For their estimation, they use a detailed industrial firm data set maintained by the National Statistical Bureau (NSB).

One feature of the NSB dataset makes this estimation possible: the data set identifies the controlling shareholder of the firm. The OECD economists then use the shareholding structure information to generate estimates of the size of the Chinese private economy. In the NSB industrial data sets, the shareholders are classified among the following categories: (1) state (direct or indirect); (2) collective (that is, local governments); (3) individuals;
domestic legal persons; and (5) foreign companies. The OECD economists make the crucial assumption that individuals, domestic legal-persons and foreign companies comprise the private ownership. They conclude that the private economy accounted for 52.3 percent of industrial value added in 2003, compared with 27.9 percent in 1998.

The most serious problem with the OECD study is the assumption that domestic legal-person shareholders are private. Legal-person shareholding refers to cross-shareholding by firms. The term implies a type of *keiretsu* arrangement whereby firms own one another’s stocks. The difference with Japan, however, is that in China much of the legal-person share capital originates in the state sector, that is, SOEs establishing or holding significant equity stakes in other firms. The subsidiaries of the SOEs, on account of their final ownership, should be classified as SOEs. However, the OECD calculation classifies the entire output of these firms as ‘private’. As an example, the OECD methodology would classify Shanghai Automotive Industry Corporation (SAIC) as a private firm. SAIC is a quintessential SOE but its largest shareholder is a legal-person shareholder, an investment company of the Shanghai government.

But let us take at face value the claim that the Chinese private sector – inclusive of the foreign firms – is producing 52.3 percent of industrial production and let us place that claim in perspective. Here a comparison with India in the 1970s is revealing. India at that time was at the apex of its commanding heights after Indira Gandhi had nationalized all major banks, significantly expanded the scope of the ‘License Raj’, and created numerous barriers for the private sector. But even at the height of the ‘License Raj’, the importance of the Indian private sector far exceeded the level of the Chinese private sector in 2003. One estimate puts the share of private sector firms in total manufacturing GDP at 93 percent in the early 1960s and at 69 percent in 1983–84. The share of fixed-asset investments of the private sector was around 58 percent, a ratio that is several multiples of the Chinese level today. Thus, even a generous accounting of the current size of the Chinese private sector puts China roughly in the same league as some of the world’s most statist economies of the 1970s.

An input-based measure of the domestic private sector

I focus only on the registered domestic private sector firms. This definition covers newly established private businesses (exclusive of foreign firms). This is a narrow definition of the private sector and, by construction, it understates its true size. The usefulness of this measure is that it is a test of a dynamic claim in the gradualist framework – that China chose the politically and economically prudent path to encourage the entry and the organic growth of new private businesses over time. Comparing this measure across
different time periods provides a way to assess the claim that the size of newly established private businesses has grown over time.

The input we focus on is capital allocated for fixed-asset investments. Fixed-asset investments, compared with similar activities in a market economy and with other economic activities in the Chinese economy, are heavily controlled by the government. All investment projects above a fairly low threshold require government scrutiny and approval. For this reason, fixed-asset investments are a superior indicator of the ownership policies of the state as compared with the output measure, because they are not subject to the confounding influences of efficiency differentials between the state and the private sectors.

Table 68.3 provides the fixed-asset investment data on four ownership types: SOEs, collective firms, the ‘individual economy’, and firms of other ownership. (The Chinese statistical system records fixed-asset investments by the registered private sector under the ‘individual economy’.) These four categories are exhaustive and mutually exclusive and thus their totals add up to 100 under Column (6).

One striking pattern emerges from this table. The investment share of the individual economy in the 1990s was actually smaller than it was in the 1980s. In the first six years of the 1980s, between 1980 and 1985, the ‘individual economy’ accounted for 20.7 percent of the total fixed-asset investments. This share climbed slightly in the second half of the 1980s, to an average of 21.9 percent between 1986 and 1990. In contrast, during the 1991–95 period, the ‘individual economy’ share declined to 13.2 percent and during the 1996–2000 period it was 13.9 percent. Panel (B) of Table 68.3 provides annual data for selected years. In 1993, the ‘individual economy’ only accounted for 11.9 percent of total fixed-asset investments, a full 10 percent drop from that prevailing in the second half of the 1980s (at 21.9 percent). After 1993, this ratio climbed slowly to 15 percent in 2002 and then fell back to 14.2 percent in 2004, just one percentage point higher than that at the very onset of the reforms in 1980.

Because our measure only covers fixed-asset investment activities in the registered private sector, and because of the possibility that the individual economy category may not fully record activities by the established private firms, we must return to the question of whether this measure is too narrow. In particular, the ‘other’ ownership category exploded from effectively zero in the second half of the 1980s to 11 percent in the 1991–95 period and then to 18.7 percent during the 1996–2000 period. If the ‘other’ ownership category encompasses mainly private sector firms, then the ‘true’ investment share of the private sector should be the sum of the individual economy and the ‘other’ ownership. That would put the investment share of the private sector during the 2001–2003 period at 43.2
percent (14.4 + 28.8), roughly double the level of the early 1980s at 20.7 percent.

But assigning all the firms in the ‘other’ ownership category to the private sector is a massive overstatement. The ‘other’ ownership category encompasses four types of firms: (1) joint-ownership firms; (2) shareholding firms; (3) foreign-invested enterprises (FIEs); and (4) unclassified firms. Shareholding firms and FIEs dominate this category of firms. These two types of firms accounted for 95.6 percent of the fixed-asset investments in this category during the 1996–2000 period and 97.2 percent during the 2001–03 period.

Table 68.3 Ownership composition of fixed-asset investment (%)

<table>
<thead>
<tr>
<th>Year</th>
<th>(1) SOEs</th>
<th>(2) Collective firms</th>
<th>(3) Individual economy</th>
<th>Of individual economy:</th>
<th>(5) ‘Other’ ownership</th>
<th>(6) Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel (A): Period data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980–85</td>
<td>66.7</td>
<td>12.7</td>
<td>20.7</td>
<td>1.6</td>
<td>0.0*</td>
<td>100.0</td>
</tr>
<tr>
<td>1986–90</td>
<td>64.8</td>
<td>13.4</td>
<td>21.9</td>
<td>2.9</td>
<td>0.0*</td>
<td>100.0</td>
</tr>
<tr>
<td>1991–95</td>
<td>59.0</td>
<td>16.3</td>
<td>13.2</td>
<td>2.7</td>
<td>11.0</td>
<td>100.0</td>
</tr>
<tr>
<td>1996–2000**</td>
<td>52.5</td>
<td>15.0</td>
<td>13.9</td>
<td>4.1</td>
<td>18.7</td>
<td>100.0</td>
</tr>
<tr>
<td>2001–03</td>
<td>42.7</td>
<td>14.1</td>
<td>14.4</td>
<td>7.6</td>
<td>28.8</td>
<td>100.0</td>
</tr>
<tr>
<td>Panel (B): Annual data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>61.5</td>
<td>17.9</td>
<td>11.9</td>
<td>2.7</td>
<td>8.8</td>
<td>100.0</td>
</tr>
<tr>
<td>1997</td>
<td>52.5</td>
<td>15.4</td>
<td>13.8</td>
<td>3.0</td>
<td>18.3</td>
<td>100.0</td>
</tr>
<tr>
<td>2000</td>
<td>50.1</td>
<td>14.6</td>
<td>14.3</td>
<td>5.5</td>
<td>21.0</td>
<td>100.0</td>
</tr>
<tr>
<td>2001</td>
<td>47.3</td>
<td>14.2</td>
<td>14.6</td>
<td>6.6</td>
<td>23.9</td>
<td>100.0</td>
</tr>
<tr>
<td>2002</td>
<td>43.4</td>
<td>13.8</td>
<td>15.0</td>
<td>7.8</td>
<td>27.9</td>
<td>100.0</td>
</tr>
<tr>
<td>2003</td>
<td>39.0</td>
<td>14.4</td>
<td>13.9</td>
<td>8.1</td>
<td>32.7</td>
<td>100.0</td>
</tr>
<tr>
<td>2004</td>
<td>14.1</td>
<td>14.2</td>
<td>100.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
*: Constructed as zero since this category did not exist prior to the 1991–95 period.
**: In 1997 the government changed the investment reporting and approval procedure. The investment reporting threshold was revised from 50 000 yuan to 500 000 yuan, but this change only applied to SOEs and urban collective firms. The effect of this change is that the published amount of fixed asset investments in the state and urban collective sectors is smaller than the actual amount. For 1996, the government published both the revised and unrevised data. In the unrevised data, the SOEs invested 1205.6 billion yuan in fixed assets and the collective firms invested 366 billion yuan. In the revised data, the SOEs invested 1200.6 billion yuan and the collective sector invested 365.2 billion yuan. This is about a 0.4 percent and 0.2 percent difference, respectively.

Sources: Based on various sources on fixed asset investments compiled by the NSB. See the text for a detailed explanation.
A critical empirical issue is whether shareholding firms are private. During the 2001–03 period, they accounted for 68.2 percent of the fixed-asset investments in the ‘other’ ownership category. The largest shareholding firms are SOEs that have issued shares on the stock market; and only 6.97 percent of the shareholding firms were private Initial Public Offerings (IPOs) between 1990 and 2003. These shareholding firms are firmly in the hands of the state although they have some private revenue rights. According to a detailed study of over 600 firms on the Shanghai Stock Exchange and the Shenzhen Stock Exchange, the three main groups of shareholders – the government, legal-persons and private individual investors – each controlled about 30 percent of the outstanding shares (Xu and Wang, 1997). But the control rights are far less dispersed. According to the same study cited above, on average individual shareholders controlled only 0.3 percent of the board seats of those firms, whereas the government retained 50 percent of the board seats and state-owned institutions controlled the remainder.

Another test of the gradualist claim is to compare firms that have clear, straightforward ownership rights at the extreme ends of the ownership spectrum. One useful indicator is the ratio of fixed-asset investments in the registered private sector relative to the state sector. If, as commonly alleged, ownership biases against the private sector declined over time, it must be the case that the ownership biases against registered private sector firms relative to the treatment of the explicit state sector firms must have declined. A decrease in the ownership biases should be associated with a rising ratio; an increase in the ownership biases should be associated with a declining ratio.

The ratio in fact declined over time. Table 68.4 presents data on fixed-asset investment in the registered private sector as a ratio of fixed-asset investment in the state sector, collective sector and firms of ‘other’ ownership under Columns (1a), (2a) and (3a). In the 1980s, the fixed-asset investments undertaken by the registered private sector in both urban and rural areas already amounted to about one-third of the fixed-asset investments in the state sector. The ratio of the individual economy to SOEs, under Column (1a), was 0.31 during the 1980–85 period and 0.34 during the 1986–90 period. But this ratio declined sharply between 1991 and 1995, to only 0.22. Between 1996 and 2000, the ratio rose moderately, to 0.27. Between 2001 and 2003, despite a period of rapid growth and economic reforms widely perceived as bold, the ratio of fixed-asset investment by purely private to state firms only managed to recover to the level prevailing at the very onset of the reform era. For much of the 1990s there is no evidence that the ownership biases abated by this measure. In fact, the evidence suggests the opposite. This pattern holds when comparing the
individual economy with the collective sector and with firms in the ‘other’ ownership category.

How do we reconcile the above findings with the widespread view that China’s private sector developed rapidly in the 1990s? Part of the answer is provided in Tables 68.3 and 68.4 where it is clear that the investment share of the urban private sector rose dramatically in the 1990s. Thus, the boom in the private sector occurred in the easily observable urban sector. But it is important to stress that private sector development in China was overwhelmingly rural in origin. In the 1990s, the private sector in the rural areas faced increasing difficulties and because of their sheer weight the problems in the rural areas weighed down the total size of the private sector.\textsuperscript{21}

### Table 68.4 Fixed asset investment ratios of the individual economy to ‘other’ firms

<table>
<thead>
<tr>
<th>Year</th>
<th>(1a) Urban &amp; rural</th>
<th>(1b) Urban only</th>
<th>(2a) Individual economy/Collective SOE ratios</th>
<th>(2b) Urban &amp; rural</th>
<th>(3a) Urban &amp; only</th>
<th>(3b) Urban only</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980–85</td>
<td>0.31</td>
<td>0.024</td>
<td>1.64</td>
<td>0.13 (0.41)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1986–90</td>
<td>0.34</td>
<td>0.045</td>
<td>1.64</td>
<td>0.22 (0.51)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1991–95</td>
<td>0.22</td>
<td>0.045</td>
<td>0.80</td>
<td>0.16 (0.58)</td>
<td>1.15</td>
<td>0.23</td>
</tr>
<tr>
<td>1996–2000</td>
<td>0.27</td>
<td>0.078</td>
<td>0.93</td>
<td>0.27 (1.24)</td>
<td>0.74</td>
<td>0.22</td>
</tr>
<tr>
<td>2001–03</td>
<td>0.34</td>
<td>0.18</td>
<td>1.02</td>
<td>0.54 (2.9)</td>
<td>0.50</td>
<td>0.27</td>
</tr>
</tbody>
</table>

### Panel (B): Annual data

<table>
<thead>
<tr>
<th>Year</th>
<th>(1a) Urban &amp; rural</th>
<th>(1b) Urban only</th>
<th>(2a) Individual economy/Collective SOE ratios</th>
<th>(2b) Urban &amp; rural</th>
<th>(3a) Urban &amp; only</th>
<th>(3b) Urban only</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>0.19</td>
<td>0.044</td>
<td>0.66</td>
<td>0.15 (0.56)</td>
<td>1.35</td>
<td>0.31</td>
</tr>
<tr>
<td>1997</td>
<td>0.26</td>
<td>0.056</td>
<td>0.89</td>
<td>0.19 (0.93)</td>
<td>0.75</td>
<td>0.16</td>
</tr>
<tr>
<td>2000</td>
<td>0.29</td>
<td>0.11</td>
<td>0.98</td>
<td>0.38 (1.8)</td>
<td>0.68</td>
<td>0.26</td>
</tr>
<tr>
<td>2001</td>
<td>0.31</td>
<td>0.14</td>
<td>1.03</td>
<td>0.47 (2.4)</td>
<td>0.61</td>
<td>0.28</td>
</tr>
<tr>
<td>2002</td>
<td>0.35</td>
<td>0.18</td>
<td>1.09</td>
<td>0.57 (3.1)</td>
<td>0.54</td>
<td>0.28</td>
</tr>
<tr>
<td>2003</td>
<td>0.36</td>
<td>0.21</td>
<td>0.96</td>
<td>0.56 (3.1)</td>
<td>0.43</td>
<td>0.25</td>
</tr>
</tbody>
</table>

**Sources:** Based on various sources on fixed asset investments compiled by the NSB. See the text for a detailed explanation.
Conclusion
According to a famous formulation, the Chinese reform has been ‘Pareto-optimal’ in that it has created winners without creating losers (Lau et al., 2000). It is time to reassess this claim about the Chinese reforms. The decline in income of 130 million Chinese in recent years and the massive forcible land seizures are at direct odds with the view that the Chinese reforms have been ‘Pareto-optimal’. A substantial portion of the Chinese population may have lost absolutely.

It is also time to reassess a central claim in the gradualist framework – that the Chinese reforms have continuously deepened over time and have succeeded in establishing a market economy. Many economists marvel at China’s speed of transition but, as I have shown, by an investment measure China may have retreated in establishing a market economy in the 1990s, and by an output measure the Chinese economy today is less private than the Indian economy in the early 1980s. Let us also keep in mind that in 2008 China will celebrate the thirtieth anniversary of its reform program (1978–2008). This is one year longer than the duration of the orthodox phase of central planning in China (1949–78). While some argue that China has chosen a different reform strategy as compared with other transitional economies (Stiglitz, 1999), a more plausible hypothesis is that the Chinese may have a different goal: they have reformed to preserve socialism, not to institute capitalism.

Notes
1. I thank Professors Jaime Ros and Amitava Dutt for comments on an earlier draft and Nancy Hearst for editorial assistance. The usual caveats apply. The empirical component of this chapter is based mainly on my book (see Huang, 2008).
2. For a comprehensive account of the Chinese economy, see Naughton (2007).
3. The GDP data are calculated on a foreign exchange basis and the source of the data is the World Development Indicators, compiled by the World Bank (available at http://devdata.worldbank.org/dataonline).
4. The data on social development in China and India can be accessed in the World Development Indicators (available at http://devdata.worldbank.org/dataonline).
5. The data on fixed-asset investments are reported in the China Statistical Yearbook, various years. For the latest data, see National Statistical Bureau (2006).
6. One exception to the findings reported here is Wang and Meng (2001), who report that TFP growth averaged 7.3 percent during the 1992–97 period but only 2.5 percent during the 1978–91 period. However, the authors themselves dismiss this finding as ‘a statistical error’ since they cannot locate the sources of this dramatic acceleration of TFP growth.
7. It is not known whether this is because of female infanticide. If it is, it is useful to determine whether female infanticide increased or decreased in the 1990s. Demographers believe that economic hard times tend to be associated with a higher rate of female infanticide. See the findings by the World Bank in research notes on China’s health sector in World Bank (2005a) and World Bank (2005b).
9. The output measures here refer to the ratios of output in labor-intensive (skill-intensive) industries to the output in less labor-intensive (less skill-intensive) industries. High (low) labor-intensive industries are those industries above (below) the median value of labor
intensity. The skill-intensity measure is similarly derived. For details, see Kochhar et al. (2006).


11. On TFP in China, see Chen et al. (1988), Borensztein and Ostry (1996) and Hu and Khan (1997). Groves et al. (1994) link specific reform measures to some aspects of the operating improvements of the SOEs. The disagreements over TFP center around the relative importance of productivity improvement vis-à-vis factor accumulation to explain China’s growth, and evidence on whether the state sector also experienced productivity growth. See the debates between Jefferson et al. (1992) and Woo and Fan (1994).

12. The other problem is that the study treats domestic private sector firms and foreign-invested enterprises (FIEs) as a single homogenous category. This treatment does not recognize that China has favored foreign firms at the expense of domestic private sector firms. Thus, the estimate implicitly incorporates a substitution effect between FIEs and domestic private sector firms.

13. An analogy would be those firms owned and controlled by Temasek, the holding and investment arm of the Singaporean government. Whether Temasek behaves as if it is a private firm is a separate question, but, from an accounting point of view, because Temasek itself is state-owned, the firms controlled by Temasek ought to be classified as state-owned as well.


15. One huge difference separates China today from India of the late 1970s – the role of foreign direct investment. The government of Indira Gandhi severely restricted FDI whereas China today welcomes FDI. It should be stressed that the OECD’s estimate of China’s private sector at 52.3 percent is inclusive of foreign firms and a comparison of only domestic private firms would be even more unfavorable to China.

16. The data on fixed-asset investments used in this section come mainly from a series of NSB publications specifically devoted to covering fixed-asset investment activities. We have checked the data in these specialized publications with those published in the annual China Statistical Yearbooks. In comparison with the Chinese data on output, the Chinese data on fixed-asset investments are remarkably consistent across a number of publications. The data used in the text come from NSB (1987), NSB (1989b; National Statistical Bureau 1989); NSB (1991), NSB (1992), NSB (1997) and NSB (1999). The data for some of the later years are from NSB (2003).

17. On the extent of state controls of fixed-asset investments, see Rawski (2001b).

18. The individual economy includes households, that is, self-employment proprietorships.


20. Another type of shareholding firm – known as a shareholding cooperative – is genuinely private. These were converted from township and village enterprises (TVEs) or small SOEs and are typically majority-owned by their employees. But they are very small. As of 2002, the shareholding cooperatives accounted for only 2.89 percent of China’s industrial output by value, as compared with 11.7 percent for the privately-operated enterprises (siying gijie). Therefore, this is not a serious downward bias. It should be noted that the NSB no longer uses the ‘individual economy’ in its data series on industrial output, although it still uses the ‘individual economy’ category for its fixed-asset investment reporting. The 11.7 percent quoted in the text refers only to siying gijie and presumably does not include industrial getihu. See NSB (2003).

21. I explore this topic in detail (Huang, 2008).

22. One issue that clouds an assessment of Chinese performance is whether the Chinese growth rate is as truly impressive as the official figures suggest. Youg (2000) and Rawski (2001a) raise questions about the veracity of the Chinese data. This is a complicated issue that I will not deal with here, except to note that better research is needed to reconcile the well-documented microeconomic inefficiencies in the Chinese economy with the apparent macroeconomic success.
References


Huang, Yasheng (2008), Capitalism with Chinese Characteristics, New York: Cambridge University Press.


World Bank (2005a), ‘China’s Progress Towards the Health MDGs’, Briefing Note No. 2, March.

World Bank (2005b), ‘Rural health insurance: Rising to the challenge’, Briefing Note No. 6, May.


Similarities among South Asian economies

The countries of the South Asian region differ in size, resource endowment, the specificities of class configurations and the nature of the ruling regimes. Nonetheless they share certain common structural characteristics: high degrees of inequality of asset ownership, especially in Pakistan and India; the presence of substantial underemployment; a strong dualism between organized and unorganized sectors, especially in manufacturing, which sometimes (but not always) translates into the dualism between large-scale and small-scale economic activities; the continuing significance of agriculture as a major employer; the recent emergence of service activities as the largest incremental employers; and the involvement of the dominant share of the workforce in what are essentially low-productivity activities, often in the form of self-employment.

There is also an apparent synchronicity of policies and processes across the region, despite very differing social and political pressures. All the economies of the region had import-substituting industrialization strategies and substantial state regulation over economic activity for the first few decades after independence. From the 1980s onwards, all of them moved, in varying degrees, to a strategy of development based on export-orientation, internal deregulation, trade liberalization and privatization. The process started in Sri Lanka, as the Jayawardene government in 1977 moved towards liberalization and dismantling of the earlier universal food security system. Thereafter, especially from the early 1990s, all the governments in the South Asian region introduced policies of internal and external liberalization and privatization (Mahmud, 2000).

There was reduction in state control in terms of administered prices, regulation of economic activity and direct responsibility for a range of goods and services. Along with internal deregulation there was trade liberalization, entailing shifts from quantitative restrictions to tariffs and sharp reductions in the average rate of tariff protection. Financial liberalization involved reductions in directed credit (especially to agriculture and small industries), freeing of interest rate ceilings and other measures which raised the cost of borrowing for governments, peasants and petty producers. There was privatization of state assets, often in controversial circumstances.
All the economies moved towards ‘market-determined’ exchange rates, liberalization of current account transactions, and some degree of capital account liberalization, such as easing rules for foreign direct investment, allowing non-residents to hold domestic financial assets and making it easier for domestic firms to access foreign commercial borrowing. The implications of such external liberalization were very different in tradition-ally aid-dependent economies such as Pakistan and Bangladesh compared to India or Sri Lanka: in the latter, net capital inflows kept real exchange rates at levels that generated current account deficits, while in the former capital inflows were a substitute for aid.

In fiscal policy, most countries experienced some degree of ‘rationalization’ (a euphemism for reduction) of direct and indirect tax rates. This was associated with declining tax–GDP (gross domestic product) ratios in several cases, as tax buoyancy failed to meet the optimistic expectations that had justified tax rate cuts. In particular, the cuts in import tariffs (and the associated cuts in domestic duties required to establish ‘level playing fields’) involved lower aggregate collections relative to GDP in most of these countries, except Bangladesh. Attempts to reduce fiscal deficits typically involved cutting back public productive investment and social expenditure, reducing subsidies to farmers and increasing user charges for public services and utilities. Ironically, fiscal deficits in most cases did not fall relative to GDP, as the largest increases in expenditure came about in interest payments, partly due to the burden of past debt and partly because of the increased costs of public sector borrowing. In addition to falls in public investment as a proportion of GDP, there was also a resource crunch at regional and lower levels of government. This tended to reduce per capita spending in important areas such as basic infrastructure development, health and education (as in India), and affected the viability and legitimacy of local government institutions (as in Nepal and Pakistan).

**Results of the transition from dirigisme to neoliberalism**

The results of this process had similar outcomes in most of these economies, despite their very different initial conditions. On the positive side, growth rates on average have increased and this has generally been associated with greater macroeconomic stability in terms of lower inflation and avoidance of balance-of-payments difficulties. There has been some increase in private investment in all of these countries due to the immediate effects of liberalization and increased export orientation. However, aggregate investment rates have not increased much except in Bangladesh and Sri Lanka, and very recently in India.

However, income inequalities have increased in all the economies of the region. Growing economic inequalities are evident between rural and
urban residents; between households in various size-classes of expenditure; between sub-regions and provinces within countries. This has been associated with increased social and political tensions in the region, which have often been expressed not so much in direct demands for redressal of income imbalances, but in terms of other ethnic, social, cultural or regional demands.

There has generally been a deceleration of employment growth, compared to the dirigiste period. This has occurred despite an improvement, or at least the same trend level, of growth in aggregate economic activity. In general, employment has not kept pace with the increase in population, resulting in higher rates of unemployment and underemployment, and also in declining labour force participation (which is not fully explained by increased involvement in education). The quality of employment also appears to have deteriorated, with declines in regular work and increases in either casual contracts or self-employment in adverse conditions. Wage shares of income have typically declined; and real wage rates have stagnated or declined.

The relative decline of manufacturing, especially in the small-scale sector, and the stagnation or decline of manufacturing employment, is marked across the region, with the exception of the garments industries in Bangladesh and Sri Lanka. Agriculture and/or services appear to have become residual refuge sectors for workers who cannot find productive employment in industry. Across the region, there appears to have been relatively little link between rates of aggregate economic growth and employment generation.

The decline of institutional credit has been a major factor affecting the viability of agriculture and small-scale industrial development in most countries of the region. This has become a particularly severe problem in the recent past, when trade liberalization in the context of stagnant or declining world trade prices for South Asian crops has put additional competitive pressure on farmers, and contributed to an agrarian crisis across all South Asian countries.

Standard indicators of human development have improved on the whole, but the rate of improvement is much lower than desirable, and some indicators have even worsened in some cases. Literacy rates and primary enrolment have improved across the region, but the progress is uneven, with some pockets of backwardness according to region and social group; and average levels remain low everywhere except Sri Lanka. Human development indicators have improved faster only in Bangladesh, where public expenditure in social sectors has been high. Nutritional improvements have been slow or minimal, and per capita calorie consumption has fallen even in supposedly fast-growing countries like India. Infant mortality rates
have actually increased in recent years in Pakistan and in some states of India.

Notwithstanding some of these outcomes, the transition from *dirigisme* to neoliberalism has been seen by many as a vindication of the latter, especially in the context of India where the retreat from *dirigisme* has been followed by a fairly creditable export performance in some spheres and a higher rate of economic growth than most of the developing world since the mid-1990s (Joshi and Little, 1996). However, such sanguineness about neoliberalism is unwarranted. This is evident from the analytical reconstruction, within a political economy context, of India’s development story, which we attempt below, and from our brief sketches of trends in other South Asian economies that follow.

**India**

The economic policy regime erected in the 1950s in India had its roots in the freedom struggle. The economy had been dominated by metropolitan capital and metropolitan commodities before independence. Freedom meant freedom from this domination; and this could not be ensured without giving the state in independent India a major role in building infrastructure, expanding and strengthening the productive base of the economy, setting up new financial institutions and regulating and coordinating economic activity. This was necessary for building capitalism itself, although some saw it as a means of transition to socialism. State capitalism and state intervention were essential instruments for the development of a relatively autonomous Indian capitalism, displacing metropolitan capital from the pre-eminent position it had occupied in the colonial economy.

This intervention however occurred within a certain context. Despite talk of land reform, of providing ‘land-to-the-tiller’, and curbing the concentration of economic power, little was done to attack or redress asset and income inequality. The worst forms of absentee landlordism were done away with, but concentration of landownership remained largely intact. And while some monopolistic practices were curbed, asset concentration in the industrial sector was never really challenged. India’s monopolists were thus able to use state intervention as a device to consolidate and expand their positions.

The persistence of asset and income inequality imposed a constraint on the expansion of the market for mass consumption goods in the country. The absence of any radical land redistribution meant that the domestic market, especially for manufactured goods, remained socially narrowly based. It also meant that the growth of agricultural output, though far greater than in the colonial period (where the last half-century had witnessed virtual stagnation), remained well below potential. For the country
as a whole, the benefits of such agricultural growth as did occur were largely confined to a relatively narrow stratum of landlords-turned-capitalists and sections of rich peasants who had improved their economic status. Meanwhile, industrial growth was not sufficiently employment-generating to create large increases in demand from this source.

In this context, continuous growth in government spending became crucial for imparting a dynamic to the system. Import-substitution policies provided domestic capitalists with a large initial market for manufactures, which the government sought to expand through its current and capital expenditures. At the same time the state supported the domestic capitalist class through its infrastructure investment and through the creation of a number of specialized institutions for financing private investment. While this strategy succeeded in overcoming the long stagnation in agriculture, established a range of basic and capital goods industries, and achieved on average a rate of industrial growth in excess of 7 per cent during the 1950s and the early 1960s, by the mid-1960s it was clearly running out of steam. Not only was the initial stimulus offered by import substitution exhausted, but the capacity of the state to continue to provide the stimulus to growth was undermined by its inability to raise adequate resources.

This reflected the contradictory nature of the roles the State was required to fulfil (Patnaik, 1994). While increasing government expenditure was necessary to keep the domestic market expanding, the state also became the most important instrument for what Marx had called ‘primary accumulation’ by the domestic bourgeoisie, which was reflected in the fiscal crisis of the state. This implied that the government had either to cut back the tempo of its investment, or to maintain this tempo through increased borrowing with inflationary consequences, since the dirigiste economy was not a demand-constrained one.

The state’s inability to impose a measure of discipline (essential for viable capitalist functioning) on the capitalists, which made tax evasion rampant and contributed to the fiscal crisis, also made the attempts at regulation through licensing and other instruments quite ineffective. The imbalance between the pattern of domestic production envisaged in the plans, and the pattern of demand emanating from the relatively affluent sections who were the main beneficiaries of growth, gave rise to unutilized capacity through the illicit diversion of resources towards unplanned end-uses. The consequent ‘irrationality’ reflected inter alia in an increase in the capital–output ratio, compounded the problem arising from the increasing fiscal constraints upon the state.

This basic contradiction of the dirigiste regime, namely that it could sustain the tempo of development only at the cost of accentuating inflation...
which undermined its own social stability, became apparent after the mid-1960s. After 15 years of rapid industrial expansion until the mid-1960s, industrial growth slowed over the next decade. This was mainly because of a cutback in public investment in the face of strong inflationary pressures, accentuated by the oil price hike of 1973. Public investment recovered somewhat after the mid-1970s, by putting a check on inflation, first by turning the terms of trade against agriculture and then, after the mid-1980s, by going in for external borrowings. But such borrowings made the economy vulnerable to capital flight. One such episode, triggered by the payments difficulties arising from the Gulf War in 1991, led to the explicit introduction of a neoliberal economic regime.

Growth had already picked up from the mid-1980s. The annual rate of growth of real GDP rose from an average of over 3.5 per cent per annum in the three decades after independence, to between 5 and 6 per cent in the 1980s and 1990s. Rather than import liberalization per se, the growth of the 1980s was related to a strong revival of agriculture through the spread of the ‘Green Revolution’ to Eastern India. It also owed much to higher government spending, financed increasingly by external commercial borrowings in the late 1980s. While domestic demand was kept up by higher government spending, import liberalization (especially of capital goods and components for manufacturing) reduced inflationary pressures and imparted an impetus to final good production. Inflation was also kept under control by lower relative prices in agriculture.

The inter-sectoral terms of trade for agriculture in the 1980s and early 1990s remained low compared to the early 1970s; but, for almost a decade prior to the mid-1990s, Indian agriculturists were protected from the far more adverse international movement of terms of trade against primary products. Relative prices for agriculture worsened again only in the late 1990s, when trade liberalization exposed farmers to declining world prices. This marked the onset of the period of agrarian crisis and increase in peasant suicides.

While the internal contradictions of the dirigiste regime were bringing it to a dead-end, major international changes also undermined dirigisme. The easier access to private capital flows negated the basic assumption of a binding foreign exchange constraint. More fundamentally, the process of globalization of finance sought to break down all barriers to the free flow of finance capital. It has been argued that the economic reforms after 1991 reflected an acceptance of the validity of the neoliberal economic paradigm (Bhagwati, 1993). It is more correct to locate the policy shift in the totality of circumstances produced by the interplay between the changing external context and the accentuating domestic contradictions within the earlier regime.
These contradictions encompassed other elements too (Chandrasekhar and Ghosh, 2004). A comparatively narrow social segment had provided the main source of growth in domestic demand under dirigisme. This social segment, eager to emulate the lifestyles and consumption patterns of the metropolitan centre, was not satisfied with having more domestically produced goods; its demand was increasingly for the new goods produced in the metropolitan centres, which could not be locally produced using only indigenous resources and technology. This imbalance increased over time because of further innovations in the metropolitan economies, creating a powerful and growing pressure from the more affluent groups for a dismantling of controls.

The emergence of newer strata, through a process of proliferation and diversification of the Indian capitalist class during the years of dirigisme, was another contributory factor. New capitalists operating outside the traditional bases of existing monopolistic groups, such as trade, finance, services of various kinds, had come into being, and operations abroad by non-resident Indian groups had become significant. These new entrants sought to diversify into manufacturing, and therefore welcomed deregulation. Because of access to newer technology, they were less averse to import competition. The established big businesses too which were, to start with, beneficiaries of state controls of various kinds, began to chafe against these controls at a certain stage in their search for new avenues for investment and production.

Support for economic liberalization also came from other quarters: new businessmen involved in what were essentially ‘parallel market’ transactions; a section of the top bureaucracy; the large and politically powerful urban middle classes; the more prosperous farmers whose real incomes increased in the 1980s. The technological and media revolutions imparted a significant impetus to the international demonstration effect, further fuelling neoliberal and consumerist demands. The ‘Non-Resident Indian (NRI) phenomenon’, whereby a significant number of people from the Indian elites and middle classes actually became resident abroad, also contributed to demands for opening up the economy.

The neoliberal reform strategy did not have much impact on growth to start with. The growth rate during the decade of the 1990s was scarcely any higher than during the 1980s; in the material commodity-producing sectors, agriculture and industry taken together, it was lower. Agricultural growth in particular decelerated dramatically during the 1990s, which became the first decade since independence to witness a decline in foodgrain output per capita. So great however was the compression in domestic demand, especially in rural India where government expenditure was sharply reduced, that per capita foodgrain absorption fell even more
dramatically, resulting in the accumulation of 63 million tonnes of food-grain stocks by mid-2002, of which 41 million tonnes were excess stocks. Per capita absorption of foodgrains for the country as a whole fell from the post-war triennial peak of 178 kg per annum during 1988–91 to 157 kg by 2001–04, which was the level in ‘British India’ on the eve of World War II. The excess stocks were largely dumped on the international market, where they were bought up to be used as animal feed for the rich countries. This was because neoliberal orthodoxy was hostile to the use of foodstocks on employment schemes, lest it enlarge the fiscal deficit (even though such deficit entailed no inflationary consequences in a demand-constrained system, and would not even raise the government’s net indebtedness much, as the foodstocks were with the government-owned Food Corporation of India).

GDP growth accelerated only after 2001, reaching more than 8 per cent between 2003–04 and 2005–06, mainly because of certain ‘newer’ services and in some export-oriented manufacturing of garments and chemicals. This period however also saw an even sharper divergence between agricultural and non-agricultural growth. Foodgrain production showed a zero trend, so that notwithstanding fiscally caused demand compression in the countryside, demand-pull inflation resurfaced by 2006. This was worsened by the government’s policy of stock reduction and winding down of public procurement operations in foodgrains.

Does liberalization per se account for India’s more successful recent export performance? If we leave aside ‘gems and jewellery’ where India was a successful exporter even before ‘liberalization’, and garments where India and her low-wage neighbours have been traditionally powerful, the one new area of strength that has emerged recently is information technology (IT)-related services and business process outsourcing (BPO). Here, the existence of a substantial educated manpower, whether with high skills as in the case of software exports, or with low skills as in the case of BPO, has been an important contributory factor. But the credit for this must go to the earlier dirigiste regime which defied conventional wisdom in setting up institutions of higher education, including several of excellent quality. Likewise, even the recent surge in exports of cement, steel and construction material are forays into the world market of industries that were set up and strengthened during the dirigiste period. All these avenues of export success would have eluded the Indian economy had it adopted neoliberal policies from the beginning, and not broken decisively out of the inherited colonial pattern of international division of labour.

Indeed, this was precisely the idea of P.C. Mahalanobis, the architect, along with Jawaharlal Nehru, of the dirigiste strategy in India. Mahalanobis assumed a closed economy in his 1950s plan model (inspired
by Soviet experience) because of his export pessimism on the basis of the then-existing production structure. However, he strongly believed that by breaking out of this production structure through protectionism and state intervention, and imparting divergence and depth to it, the economy would be able to emerge as a more successful exporter at a later date. Recent Indian experience appears to vindicate him more than his neoliberal critics.

The recent acceleration in growth however cannot be explained by export performance. True, the export performance, especially in the service sector, has boosted domestic incomes and consumption to an extent at the upper end of the spectrum. But much of this boost is self-propelling, giving a lift to ‘animal spirits’ à la Keynes, which explains the recent increase in investment ratio. Underlying this boom are: a consumption splurge by the upper income groups, driven by the international demonstration effect of metropolitan living standards; a construction-cum-real estate boom, led by an enormous appreciation in land values; and an explosion in corporate profits. Two of the most palpable features of the contemporary Indian economy fall into place here: the increase in the unemployment rate even in the midst of this unprecedented boom, since structural-cum-technological change is in the direction of high-productivity sectors; and the rampant drive to displace peasants from land in the name of ‘infrastructure projects’, ‘Special Economic Zones’ and such like, all of which camouflage land speculation.

This growth has been accompanied by significant increases in inequality – both across regions of India and within regions across different economic and social categories. There is a widening gap between incomes in agriculture and non-agriculture: the ratio of per-worker domestic product in non-agriculture to that in agriculture which was about 2 in the 1950s has increased to well over 4 in the early 2000s. The wage share of national income has also fallen sharply and the wage share of value added in organized manufacturing declined to only around 10 per cent in 2004. Consumption surveys show absolute declines in the income and consumption of a substantial share of the population. During the last two decades of the twentieth century, the urban top 20 per cent of households experienced historically unprecedented increases in per capita consumption in real terms, while the per capita consumption of the bottom 40 per cent of the rural population actually declined.

Agriculture, which continues to employ the bulk of the workforce, has been afflicted by an intense and prolonged crisis. Its problems are closely related to more open trade combined with domestic deflationary policies. From the mid-1990s, the financial viability of cultivation has been eroded, owing not only to longer-term problems such as declining soil fertility, changing weather conditions and excessive dependence on depleted
groundwater, but also to a combination of sharply rising input prices and volatile and declining output prices. Indian farmers were encouraged to shift to cash crops, and exposed to international competition from the highly subsidized crop production in the North, even as they were facing reduced support from central and state governments in the form of inadequate availability of institutional credit, decline in public extension services, insufficient regulation of some inputs and reduced subsidies for other inputs such as fertilizers, reduction in timely crop price support, and poor storage and marketing facilities.

Poor employment generation remains a critical issue. Agricultural employment has fallen, due to both agricultural stagnation and technological and cropping pattern changes that reduced labour demand in agriculture. Such changes were also hastened by the growth of landlessness (as peasant cultivation became less viable because of rising input costs and falling or stagnant crop prices) accompanied by a shift to commercial cultivation with hired labour. In urban India, manufacturing is increasingly characterized by more capital-intensive techniques, and therefore declining employment elasticity of production. Even the ‘newer’ and more dynamic services such as IT-enabled activities that have increased their share of output still remain minuscule in employment terms. All IT-related activities currently employ less than 0.2 per cent of the total workforce. Recent employment increases have been mainly in the form of subsistence self-employment in low-value services, despite the economic boom.

Until the late 1980s, there was a secular trend towards declining poverty. Subsequently, while official estimates show a decline in poverty, nutritional indicators suggest the contrary. The coverage and quality of public services has worsened, which has had particular impact upon the condition of women and girl children (Dreze and Sen, 1994). The persistence of illiteracy, especially among females; the inability to ensure even primary education to all children and high drop-out rates over successive years of schooling; the poor indicators of health and the recent stagnation of infant and maternal mortality rates; the absence of proper sanitation for a large proportion of the population – all these provide an indication of the current state of the development project in India.

The external sector, however, has displayed a degree of overall stability in the balance of payments and a relative absence of the boom-and-bust cycles that marked some other emerging markets. To some extent this reflects the relatively limited extent of capital account liberalization over much of the period, and the fact that the Indian economy was not really ‘chosen’ to be a favourite of international financial markets until 2002. The greatest stability to the balance of payments was imparted by the substantial inflows of workers’ remittances from temporary migrant workers in the
Gulf and other regions, which has amounted to more than all forms of capital inflow put together. Since 2004 there has been a sharp increase in portfolio capital flows and external borrowing, but FDI remains relatively small.

The alternative to neoliberalism in India consists of policies to ensure the viability and sustainability of agriculture, and greater emphasis on public expenditure with high direct and indirect effects on employment generation, especially in infrastructure, health and education. This requires higher resource mobilization from the rich. It is also necessary to counter some of the adverse effects of trade liberalization on employment, apart from more directly addressing the basic structural issues of asset and income inequality and the persistence of low-productivity employment. Of course, such a policy shift requires political will and therefore a change in political configurations.

South Asian scenario
The political economy of the transition from *dirigisme* to neoliberalism in other South Asian countries, while differing in detail from that of India, was roughly similar: it represented everywhere the abandonment of the domestic bourgeoisie’s quest for a relatively autonomous trajectory of capitalist development, and the pursuit of an alternative trajectory with much closer integration between domestic and metropolitan capital, between finance and industry, and between the local and the global. It therefore meant a growing hiatus between the bourgeoisie and the urban upper income groups on the one side and the bulk of workers (facing insecure employment), peasants (facing agrarian crisis), petty producers and small businessmen (facing closures), and agricultural labourers (facing shrinking employment) on the other.

Pakistan
Despite a respectable per capita growth of around 5 per cent per annum over the second half of the twentieth century, Pakistan systematically underperformed on most social and political indicators, including education, health, sanitation, fertility, gender equality and political instability. In general, output growth has been associated with very low employment growth, at the trend rate of only 2 per cent per annum for the long period 1960–99. Employment growth worsened after the imposition of an International Monetary Fund (IMF) structural adjustment programme in 1987–88. In the 1990s, economic growth reduced on average and also became much more volatile. This was associated with historically low rates of investment, as private investment failed to revive or compensate for the decline in public spending. The investment–GDP ratio declined from 17.3
per cent in 1998–89 to 14.7 per cent in 2000–01, largely due to the collapse in public investment. Industrial growth rates almost halved from 8.2 per cent to 4.8 per cent per annum. Further, the earlier success at reducing poverty was reversed in the 1990s, as the percentage of households living in absolute poverty increased from 21.4 per cent in 1990–91 to 40.1 per cent in 2000–01 (Hussain, 2004).

The initial years after Pakistan's third military coup witnessed a worsening of the macroeconomic situation, with increasing poverty and unemployment, falling real wages and worsening income distribution. However, recent geopolitics has impacted in some positive ways upon Pakistan's economy, mainly because of the willingness of the Musharraf regime to be a key ally of the USA. This caused the waiver or rescheduling of more than one-third of Pakistan's external debt, an increase in foreign aid flowing back to Pakistan, and the reinstating of export quotas in textiles and garments. Pakistani workers abroad have contributed to the massive recent increase in remittances, to as much as 14 per cent of GDP. However, since the domestic investment rate is still below the savings rate, the inflow of aid and remittances has not really contributed to economic activity, but is simply stored as foreign exchange reserves.

_Bangladesh_

Bangladesh shows a different and somewhat more optimistic economic trajectory than other South Asian countries. Since independence in 1971, there has been a slow but accelerating improvement in living standards. GDP grew at an average rate of 3.7 per annum in the 1980s, 4.8 per cent in the 1990s, and 5 per cent in the 2000s. Per capita income has grown even faster as population growth has slowed down in the recent years. With the fertility rate reduced from 6.3 children in 1975 to 3.3 in the mid-1990s, the rate of population growth has fallen from over 3 per cent to less than 1.5 per cent in a space of three decades. Per capita income growth, consequently, has doubled from about 1.6 per cent in the 1980s to over 3 per cent in the following decade and a half (Muqtada et al., 2002). This has been associated with a reasonable degree of macroeconomic stability, with inflation rates falling from over 10 per cent in the 1980s to just over 5 per cent in the subsequent 15 years, and moderate falls in the fiscal deficit and trade deficit ratios to GDP. Strong export performance in the garments sector and a steep increase in workers' remittances from abroad have played an important role in improving both external and internal balances of the economy, as the inflow of foreign aid has experienced a secular decline.

The rate of investment in Bangladesh increased slightly from 17 per cent of GDP in the 1980s to close to 20 per cent in the subsequent 15 years, with
both public and private sectors contributing to this increase. A major factor was the increase in public expenditure over this period, despite declining foreign aid. This was partly because of the rising tax–GDP ratio, reflecting higher indirect tax revenue after the imposition of a uniform value-added tax. The share of social sectors, such as education and health, increased from 15 per cent of the growing total budgetary expenditure to over 20 per cent. The share of physical infrastructure also increased. In consequence, the headcount poverty ratio declined from 71 per cent in 1973–74 to 40 per cent in 2000, with the 1990s witnessing a more rapid decrease. Human development indicators also improved: the decline in infant mortality in Bangladesh, from 95 (per 100 live births) in 1972 to 30 (per 100 live births) in 2000, was among the fastest in the developing world. In spite of these achievements, the basic development challenges remain, requiring continued government intervention to ensure income diversification and improvements in living standards in the future.

Sri Lanka
Sri Lanka has been often hailed for its high achievements in human development, despite limited growth. Since independence in 1948, Sri Lanka has registered an average annual growth rate of 3–4 per cent while achieving adult literacy rates of 92 per cent and life expectancy of more than 72 years by 2001. Nevertheless, Sri Lanka remains a low-income, food-deficit country with more than 34 per cent of households below the poverty line in 2001.

In the 1970s, poverty alleviation strategies included free or highly subsidized social and economic services, such as education and health care. There were food subsidies as well as producer subsidies intended to help agricultural producers. These created the highest levels of human development in the region. However, since 1977, Sri Lanka pursued an economic liberalization policy. This was initially accompanied by increased public investments financed by external borrowing, but this could not be sustained (Dunham and Jayasuriya, 2000). From the mid-1980s, the government pursued a macroeconomic stabilization course along with mounting military expenditures in the civil war. From the early 1990s, deregulation and liberalization policies were supposed to bring about economic growth, but success here has been confined to the garments industry. Employment has stagnated, real wages have fallen throughout the 1977–2001 period, and most of the population continues to be employed in commodity production activities as farmers or production workers, and possesses limited education. Remittances (largely from women workers abroad) have been crucial in stabilizing the balance of payments and providing buffer incomes for the poor.
Nepal
A landlocked least-developed country, Nepal has major infrastructure gaps and much of it is physically inaccessible. There are marked regional differences between the Terai, hill and mountain areas, with decreasing infrastructure and increasing poverty as one moves up. Agriculture dominates the production structure, and rural people dominate the poor. The problems of backwardness are compounded by a long and relatively open border with India, which creates a high degree of dependence upon the Indian economy and also a vulnerability to India’s macroeconomic policies. The extremely adverse material conditions have generated extreme political movements, which in turn have generated much political instability.

There was a systematic decline in public investment from around 25 per cent of GDP in the late 1970s, to 18 per cent in the late 1980s, to only 11 per cent in the 1990s. This was associated with low and falling rates of private investment. Neoliberal reforms from the early 1990s have been associated with stagnation in agricultural incomes. The manufacturing sector has focused on exports, of garments, carpets and pashminas; there has been little development of manufacturing for the home market. Import penetration has prevented the development of infant industries and led to closure of small units and low employment generation in manufacturing. Financial sector reforms have been associated with reduced flow of institutional credit to agriculture and small enterprises, and fewer possibilities of subsidies for providing micro-credit to the poor through government channels. Open unemployment rates are low, but rates of underemployment are very high, estimated to be around 43 per cent of the labour force. However, recent political changes in Nepal give rise to guarded optimism about the future, and the possibility of economic strategies that will focus more on livelihoods of the poor and productive employment generation in a diversified economy.

Conclusion
Dirigisme outside India was always more vulnerable, since the countries concerned were small in size. Paradoxically however, unlike in Latin America, no effort was made during the entire dirigiste period for closer economic integration between the South Asian countries. On the contrary, competition between the South Asian countries, such as between Sri Lanka and India in tea, and between Pakistan and India in jute, was intense and yielded suboptimal outcomes, compared to what cooperation even after an initial phase of competition could have achieved.

South Asia’s quest for economic development has been accompanied by a parallel quest for building modern nation-states encompassing multiple religious, linguistic, ethnic and regional groups. The difficulties of this
process, exemplified by Tamil secessionism in Sri Lanka, by the break-up of erstwhile Pakistan, by the periodic eruptions of secessionist movements in India, and by a host of conflicts dotting the entire region at any time, have been compounded by the consequences of the development trajectories pursued. In a dirigiste regime the distribution of the ‘gains’ from development across the proto-bourgeoisies belonging to different ethnic, linguistic and such other groups is determined by the nature of the state, by who has greater control over the state apparatus. In a neoliberal regime there exists a spontaneous tendency towards increasing divergence starting from certain initial differences. The resulting discontent under neoliberalism allows scope for self-serving intervention by powerful elements of metropolitan capital backed by metropolitan states, which creates a tendency towards a fracturing of the nation. The growing inequalities become particularly significant in this context: since any growing hiatus necessarily has some regional or linguistic or religious dimension, it can become a means of fragmentation of the nation. Dirigisme in South Asia did not always succeed in keeping these contradictions in check. Neoliberalism is further accentuating these contradictions. The development of humane societies in this region still has a long way to go.

Note
1. The ‘poor’, officially, are those with consumption below the official poverty line. This is determined by applying price-index adjustment to the observed cost of that 1973–74 consumption basket at which a certain number of calories were just accessed.

References
The image of rapid growth throughout East Asia following Japan has implied a more coherent region than actually exists, and a corresponding tendency to see economic progress in the region as similar in origin and nature, leading to regional terms such as the ‘Far East’, ‘Asia-Pacific’ and ‘Pacific Asia’ preceding others such as ‘East Asia’, ‘yen bloc’, ‘flying geese’, ‘tigers’, ‘mini-dragons’ and so on. On the other hand, rather amazingly, the World Bank’s (1993) influential *East Asian Miracle* volume hardly considered the significance of geography or location. For the Bank, it did not seem to matter that the eight high-performing Asian economies (HPAEs) – incredibly, with the benefit of hindsight, excluding China – were all from the same part of the world and were not only contiguous but often also linked by trade, finance and investment relations.

The next section starts with a critical review of the World Bank’s (1993) influential study of the East Asian economic miracle. The following section emphasizes the diversity of East Asian experiences and the significance of recognizing this diversity for drawing appropriate lessons. The chapter then considers implications of the East Asian financial crises of 1997–98, before the final section addresses some issues in drawing lessons from the East Asian experience.

**Interpretations of the ‘East Asian miracle’**

There are at least three major competing explanations for the rapid growth and industrialization of several East Asian economies, often referred to as the ‘East Asian miracle’. The dominant version in the 1970s and 1980s presented sustained rapid growth in the East Asian region as essentially due to unfettered market forces (Little et al., 1970). The obvious policy implication was to liberalize, as well as to open up or globalize. In the late 1980s, this was challenged by an almost opposite, *dirigiste* position, which emphasized the leading role of the ‘developmental state’ (Johnson, 1982; White, 1988; Amsden, 1989; Wade, 1990). The World Bank’s 1993 study seemed to offer an intermediate view by acknowledging the role of the state, while emphasizing the importance of the market.
The World Bank’s (1993) *East Asian Miracle* identified at least seven types of state interventions. It approved of the first four, deemed functional interventions, namely: (1) ensuring macroeconomic discipline and macroeconomic balances; (2) providing physical and social infrastructure; (3) providing good governance more generally; and (4) raising savings and thus investment rates. Functional interventions were said to compensate for market failures, and hence, were deemed acceptable, if not desirable, and less distortive of markets. Three types of market-distortive strategic interventions were considered in the areas of finance: directed (that is, subsidized) credit, trade policy and industrial policy.

The interventions were not just market-conforming, but also played important proactive roles which have been more than simply market-augmenting, as suggested by the World Bank’s analysis. Of the more controversial strategic interventions, the Bank conceded that financial interventions had been important and successful in East Asia, particularly in North-East Asia, that is, in Japan, Korea and Taiwan, but considered other strategic interventions to be failures. However, the Bank maintained that other developing-country governments were not capable of successfully pursuing the types of policies North-East Asians had successfully implemented because their state capabilities were almost unique and virtually non-replicable.

The Bank volume’s evaluation of the role and record of Japan’s Ministry of International Trade and Industry (MITI) and its counterparts elsewhere in the region is more predictable, arguing that government interventions have been trade-distortive and generally unsuccessful in East Asia, although export promotion in particular comes in for much praise. The Bank disapproves import protection and fails to see the connection with export promotion. The Japanese, South Korean and Taiwanese governments implemented import substituting industrialization (ISI) policies from the 1950s, but also pursued export-promotion soon after to ensure that their industries quickly become internationally competitive by requiring a rapid transition from import substitution to export orientation, that is, ‘effective protection conditional on export promotion’ (EPconEP). EPconEP is quite different from just import substitution or export promotion for example in export-processing zones (EPZs). Both import substitution and export promotion (for example subsidies) are trade distortions. It is also misleading to suggest that EPconEP is tantamount to ‘simulated free trade’ because the ‘bad’ import-protection distortion is negated by the ‘good’ export-promotion distortion. EPconEP allows industries and firms to enjoy rents from import protection to develop new industrial and technological capabilities in order to become internationally competitive in terms of both cost and quality.
Infant industries have often been provided with effective protection conditional on export promotion, which has had the effect of forcing firms and industries quickly to become internationally competitive. By giving firms protection for certain periods, depending on the product, and by also requiring that they begin exporting certain shares of output within specified periods, discipline was imposed on the firms in return for the temporary trade protection they enjoyed. Such policies forced firms to reduce their unit production costs as quickly as possible, for example by trying to achieve greater economies of scale and accelerating progress up learning curves. Requiring exports has also meant that producers have had to achieve international quality standards quickly, which imposed pressures to progress technologically in terms of product design and quality as well as technological processes. With strict discipline imposed, but also some flexibility in enforcement, many firms managed rapidly to achieve international competitiveness.

**Regional diversity**
Some important differences among the East Asian miracle economies suggest that not all East Asian economies have been proceeding inexorably in the same direction in a similar manner. The East Asian experiences are far from constituting a single model. Some of the major differences in East Asia are themselves very instructive. In the case of the role of foreign direct investment (FDI), tremendous contrasts are found. In the case of Singapore, FDI has constituted about a quarter of gross domestic capital formation and about 15 per cent in Malaysia. At the other end of the spectrum, in the case of Japan and Korea, the percentage has long been below 2 per cent. Some of the other countries fall between these two extremes, with few near the mean for developing countries of around 5 to 6 per cent. Those most successful in developing industrial capacities and capabilities in East Asia – namely Japan, South Korea and Taiwan – have hardly depended on FDI.

FDI’s far greater importance in South-East Asia has been due to a variety of reasons, which have not been entirely economic. One reason for the major role of FDI in Singapore and Malaysia is political. After Singapore seceded from Malaysia in 1965, the regime decided to attract FDI in massive quantities to ensure its own survival, so that the major foreign powers would quickly develop a stake in the survival of the Singapore regime. Subsequently, this FDI preference has been justified in terms of improving access to foreign markets and technology.

Malaysia has long had ethnic rivalries and an ethnic affirmative action policy. Some policy-makers tried to limit ethnic Chinese control of the economy by encouraging foreign direct investment. Again, one finds a
political motivation for the important role of FDI in Malaysia. In other words, political considerations have been a very important reason for attracting, even privileging foreign investment in Singapore and Malaysia.

There has also been considerable diversity in the role and performance of public investments, including state-owned enterprises (SOEs), in East Asia. In South Korea, Japan and, of course, Hong Kong, SOEs are hardly important today, but historically, state-owned enterprises were important in Japan before World War II, and in South Korea until more recently. Conversely, however, one finds that state-owned enterprises have been extremely important in Singapore and Taiwan. Again, this is partly explained by political factors, but there are also economic considerations. And very importantly, the performance of these SOEs has also been quite impressive.

Industrial and technology policies in East Asia have also been quite diverse. One extreme, of course, is Hong Kong, where there has been relatively little industrial policy, although more than most opponents of industrial policy care to admit, especially in recent years. It is far more detailed and sophisticated in Japan and Korea at the other end of the spectrum. In Korea, industrial policy is largely oriented towards large firms, whereas in Taiwan, much more emphasis is given to medium-sized and relatively smaller enterprises.

Industrial policies in the region have also had different orientations, emphases and instruments. For example, trade policy has been very important in almost all economies in the region except Hong Kong and Singapore, while financial policy has been important in all the countries, including Singapore, but again, with the exception of Hong Kong before the 1998 crisis. Since Hong Kong’s reversion to China in mid-1997, there have been many indications of greater government interventions in the territory, presumably in line with its new status and China’s envisaged role for the deindustrialized financial centre.

The World Bank recommended that the rest of the developing world emulate South-East Asia, not North-East Asia because of important differences between them. These misleading claims require us to recognize the far more impressive achievements and superior economic performance of the first-tier East Asian newly industrialized economies (NIEs) (including Singapore), compared to the second-tier South-East Asian NIEs. The World Bank (1993) argued that the South-East Asian high-performing economies were the preferable model for emulation by other countries seeking late development.

According to Yoshihara (1988), the South-East Asian economies have been characterized by ‘ersatz capitalism’ because of the compromised and inferior role of their states, their maltreatment of ethnic Chinese and their failure to develop better technological capabilities. Jomo et al. (1997)
criticized the World Bank’s claims, suggesting various problems associated with the growth experiences of the South-East Asian economies praised by the Bank. In any case, the currency and financial crises of 1997–98 radically transformed international opinion about the East Asian models, with earlier praise quickly transformed into condemnation (Jomo, 1998).

The Malaysia, Indonesia and Thailand (MIT) economies as second-tier or second-generation East Asian newly industrializing countries (NICs) share some common characteristics with Singapore, which is also in the region. However, they are not only far less advanced in developmental terms, but also quite different from the city-state's heavy reliance on trade and financial services besides manufacturing. Essentially, the MIT economies have had somewhat different, even ersatz developmental states and industrial policies, compared to the first-generation East Asian newly industrialized economies (NICs). Although Singapore too has pursued industrial policy, it has used fewer trade policy instruments and has been far more reliant on foreign direct investment compared to the other East Asian NIEs. Though Singapore, like Hong Kong, has eschewed trade policy instruments, it has used state-owned enterprises (SOEs) – usually referred to as government-linked corporations (GLCs) in the island republic – more than any other East Asian economy, and perhaps any other economy in the world in the early twenty-first century.

Most importantly, the South-East Asian high-growth economies (including Singapore) have relied much more heavily on foreign direct investment (FDI) to develop most of their internationally competitive industrial capabilities. Trade policy instruments in the region have been less well formulated and implemented, with rather mixed consequences, but have nonetheless been part of the region’s industrial policy story. Generally, government interventions in the region have been influenced by a variety of considerations besides economic development and late industrialization. Consequently, industrial policy has also varied in nature, quality and effectiveness. Yet, the economies in the region would not have achieved as much as they did without selective government interventions, including industrial policy.

Growth performance has been superior in North-East Asia over the long term despite the much greater resource wealth of South-East Asia. Over the period studied by the Bank, that is, from the 1960s until the early 1990s, growth in the former averaged about 8 per cent, compared to about 6 per cent for the latter. A 2 per cent difference, compounded over a period of a quarter-century or more, adds up to a lot. Very importantly, except in Hong Kong (due to immigration from China) and perhaps Singapore, population growth has been much lower in the former compared to the latter. Immigration into Hong Kong and Singapore involves a very high
proportion of the labour force, thus raising the average labour utilization rate. Political factors have also ensured far more equitable distribution of economic welfare than would otherwise have been the case in the first-tier NIEs, whereas such considerations have been less influential in the second-tier South–East Asian NICs despite Malaysia’s ethnic ‘social contract’ and Indonesia’s rural development efforts to achieve political stability.

Improvements in per capita income and economic welfare have been much more significant in North–East Asia, compared to South–East Asia (with the exception of Singapore), despite the greater resource wealth of the latter. Income inequalities have also been far less in North–East Asia, although there is some evidence of rapid recent increases in inequality. In other words, what South–East Asia has achieved has been less impressive in some critical ways. Drawing from this contrast, some people now argue that resource wealth is not a blessing, but a curse, in so far as it may have postponed the imperative to industrialize.

The North–East Asian NIEs have generally had much more sophisticated and effective industrial policy compared to South–East Asia’s NICs. This accounts, in no small way, for the very important differences in industrial and technological capabilities between North–East Asia and South–East Asia. Also, South–East Asian industrialization is still driven by FDI, whereas North–East Asian industrialization is primarily an indigenous phenomenon.

Japan and the first-generation NIEs began to industrialize in the very specific economic and political conditions of the post-World War II Golden Age and Cold War. North–East Asia grew rapidly in the immediate post-war period under a ‘security umbrella’ provided by the Americans, including aid and privileges no longer available to others. Besides providing generous aid, the Americans were anxious for them to ‘succeed’ economically in order to be showcased as attractive alternatives to their neighbours under communist rule or influence. Hence, the Americans were quite happy to tolerate trade, finance, investment, intellectual property and other policies violating neoliberal economic norms that they are now strongly opposed to. Such conditions are simply no longer available to others, and hence, their experiences are more difficult to emulate. To discourage other developing countries from trying to emulate the first-generation East Asian NIEs, it is now often argued that their state capabilities are culturally unique and impossible for others to emulate.

The Guomindang government of Taiwan was the same regime driven out of mainland China by the communists because of its incompetence and corruption. One could say the same of the Rhee regime in Korea in the 1950s. Japan has hardly been scandal-free in recent years, and most observers trace recent abuses to the nature of Japan’s post-World War II
political economy. The superior policy-making and implementation capabilities of the North–East Asian decision-makers was, at least until the 1997–98 financial crisis, widely acknowledged, but this in itself does not prove that policy-makers were thoroughly competent and incorruptible.

Some also claim that East Asia cannot be emulated owing to its very different initial conditions. Such differences are real, but often exaggerated. There is no doubt that Japan as well as the first-tier East Asian NIEs have also been distinguished by much higher levels of educational achievement. However, the level of literacy in Korea in 1950 was lower than the literacy rate in contemporary Ethiopia, which has one of the lowest literacy rates in Africa today. The educational achievements of contemporary South Koreans reflect the tremendous human resource investments in East Asia in the post-World War II period.

Some fortuitous circumstances must also be considered. Japan, South Korea and Taiwan all implemented relatively virtuous American-sponsored land reforms soon after the end of World War II. There was also significant redistribution of other non-land assets in Japan, most notably, of the pre-war and war-time zaibatsu industrial conglomerates. Much of the motivation for such redistributive reforms was, of course, anti-communist, that is, to undermine and minimize support for the communists by those desiring asset redistribution.

In contrast to South–East Asia, more egalitarian asset redistribution in Japan, South Korea, Taiwan and China have also been important. The Americans were not uninfluenced by the left, partly because of the nature of the wartime anti-Axis alliance and the nature of the most influential scholarship available. During the post-World War II American occupation of Japan, it was widely presumed that the zaibatsu ‘military industrial complex’ had been responsible for the militarization of pre-war Japan. The American occupation forcibly broke up family control of the zaibatsu to create a unique, corporatist ‘stakeholder’ economy. Assets were sold to employees, and to local communities, thus developing worker and community stakes in the companies.

Hence, the peculiarly Japanese economy was created by deliberately redistributive policies with unique outcomes. The ‘human relations’ school of industrial relations influenced the post-World War II development of guaranteed life-long employment and the seniority wage system, which have effectively strengthened a strong employee commitment to the fate of their firm.

From miracle to debacle
Although East Asian economic performance before 1998 was debated, the East Asian debacle of 1997–98 was not anticipated, partly because it was
not principally due to a failure of the real economy despite various recog-
nized economic weaknesses (Jomo et al., 1997). The financial crisis from
mid-1997 was precipitated by an eventually successful currency attack on
the Thai baht, overvalued after the strengthening of the US dollar after
mid-1995. The crisis was greatly exacerbated by herd-like panicky with-
drawals from the entire East Asian region, inducing currency and stock
market collapses (Jomo, 1998). Those who control financial assets usually
enjoy disproportionate policy influence in most contemporary economies,
especially in ‘emerging markets’. The greater role of foreign capital in
South–East Asia subordinated domestic industrial capital in the region,
allowing finance capital, both domestic and foreign, to become more
influential in the region, thus rendering it more economically vulnerable
(Jomo, 1998).

Finance capital in the region had developed complex symbiotic relations
with politically influential rentiers, dubbed ‘cronies’ in the aftermath of the
1997–98 crisis. Although threatened by the full implications of inter-
national financial liberalization, East Asian financial interests were quick
to identify and secure new possibilities for capturing rents from arbitrage
as well as other opportunities offered by gradual international financial
integration. Thus, foreign dominance of South–East Asian industrializa-
tion facilitated the ascendance of financial interests and politically
influential rentiers.

Such increasingly influential alliances were primarily responsible for pro-
moting financial liberalization in the region, both externally and internally.
However, insofar as the interests of domestic financial capital did not
tirely coincide with international finance capital, international financial
liberalization was necessarily partial. The processes were necessarily
uneven, reflecting the variety of interests involved and their varying
strengths in different parts of the region. Such flows were also desired to
finance current account deficits in both countries, principally due to service
account deficits (mainly for imported financial services as well as invest-
ment income payments abroad) and growing imports for consumption and
output of non-tradeables, mainly in the property (real estate) sector. There
is little evidence that such capital inflows contributed significantly to
accelerating economic growth, especially of the tradeable sectors. Instead,
they often contributed to asset price bubbles, consumption binges and
‘over-investments’.

Challenges
There are important lessons to be drawn from East Asia, but clearly, there
is no single model as such. For other reasons as well, it does not make
much sense to emulate any particular economy in East Asia. Most other
developing countries would find it impossible to do so even if they wanted to. Nevertheless, some important lessons can be drawn from the East Asian experiences. Such lessons are best drawn from careful analysis rather than broad-brushed generalizations about a rather diverse region.

Economic liberalization, including globalization, since the 1980s has fundamentally changed the environment and conditions for selective industrial policy and, hence, for aspiring developmental states. Most importantly, economic liberalization – at both national and international levels – has seriously constrained the scope for government policy interventions, especially selective industrial promotion efforts. This is true of both the international and domestic policy environments, where policy conditionalities and World Trade Organization (WTO) membership obligations have radically reduced the scope for national economic policy initiatives.

The mid-1980s onwards has seen widespread, sweeping and rapid opening up of trade, investment, finance and other flows. Very often, such liberalization has been externally imposed by the Bretton Woods institutions as part of conditions imposed to secure access to emergency credit during the debt crisis of the 1980s, and more recently, in the wake of the financial crises since the mid-1990s. Various policy packages for (price) stabilization in the short term or for structural adjustment in the medium term have involved such conditionalities. The new intellectual and policy environment from the 1980s – the so-called ‘Washington Consensus’ – has promoted such policy reforms.

Such policy changes as well as limited government capabilities have meant little preparation in terms of a proactive strategy to anticipate and cope with the new international competition. Few industrial policy instruments of the past are viable or feasible options today, including many tools used successfully in post-World War II East Asia. Many, if not most of, the main industrial policy tools still available are already intensively used by most advanced industrial economies.

These policies are probably necessary, but certainly not sufficient for stimulating and sustaining economic growth and structural change for developing countries ‘catch-up’. Special policies are urgently needed to prevent such economies – already at a historical disadvantage in various respects – from falling further behind, if not to begin to close the gap with the industrially more developed economies of the North as well as the industrial economies that have emerged in recent decades, that is, during the last third of the twentieth century.

As a region, East Asia has led other developing-country regions in terms of economic performance, growth rate, increased exports as well as technological progress. Lall (2003) notes the great divergence between East
Asian countries with and without selective industrial policy, and finds the latter (mainly in South–East Asia) far more economically vulnerable.

Industrial development in the new circumstances clearly requires international competitiveness, and such competitiveness is increasingly defined in manufacturing, related services and institutions, and not simply in terms of wage costs or exchange rate competitiveness, as important as these may be. Inability to compete effectively implies being bypassed, and likely, stagnation at the lower end of the technological and income ladder.

Appropriate industrial policy will require selective interventions as well as effective coordination among firms, clusters and factor markets, which should presumably be consistent with a clear and coherent ‘vision’ of the future as well as the ‘road-map’ towards policy goals (Lall, 2003). For this purpose, there are still many useful lessons to be drawn from the varied experiences of the more successful East Asian NIEs and China, as well as the more modest and flawed achievements of the South–East Asian NICs.

Notes
1. This chapter is based on considerable earlier work. I am grateful to all who have provided me with critical feedback on this work. Needless to say, however, no one else bears responsibility for this version.
2. After the East Asian crisis, even the IMF seemed to back off from its previous advocacy of financial market liberalization.

References
71 Transition economies: lessons for development

_Ruud Knaack and Henk Jager_

**Introduction**

‘Transition’ is a designation that is widely used for the transformation of state socialist countries, characterized by state-controlled economies and political dictatorship, into democratic market economies. Transition started to receive intensive attention when a number of countries freed themselves in 1989 from either particularly the Soviet control (think of Poland, Hungary, Czechoslovakia, East Germany and the Baltic states) or mainly domestic tyranny (such as Romania and Albania). Transition was extended in 1992 to the former Soviet Union (FSU) states, also dubbed the Commonwealth of Independent States (CIS), after the disintegration of the Soviet Union in December 1991.

We also will consider countries such as China and Vietnam as transition economies, though there is a reason to exclude them. In China and Vietnam political power remained in the hands of the leaders of the communist parties. In contrast, the member countries of the FSU and the Central and Eastern European (CEE) countries have undergone a dramatically rapid democratization process – with some exceptions, such as Belarus, Turkmenistan and Uzbekistan. If a high speed of political turmoil were a determining feature of transition, countries such as China and Vietnam would stay out. The differences between the outcomes in, on the one hand, the countries in CEE and of the FSU and, on the other hand, China are striking. From the countries of the first group, only three countries were able to recover from the ‘transition depression’ of gross domestic product (GDP) after ten years. However, China did not experience a decline in GDP in any year from 1978 onwards – China’s transition period – while it had a high and stable annual GDP growth figure in the order of 8 percent on average. It is, of course, quite interesting and informative to analyze the reasons for such economic divergences. For that goal, we will not employ a rapid political transformation as a distinctive feature of transition, so that we will focus on economic transition.

At the start of transition in Europe, around 1990, knowledge about the preferred materialization of the transition process was partly borrowed from the experience obtained from the process of economic growth in
developing countries. Now, more than 15 years later, there is ample information about the transition process to consider the reverse question: What are the lessons of this transition from practice for developing countries that are eager to give additional spurs to growth? The search for these lessons is the aim of this chapter. The chapter consists of three parts. It starts with a short history of political developments, mainly in Russia, the motor of state socialism after World War II. The next part describes the stylized facts of the transition process in the countries in CEE and of the FSU, and China as well. The following part discusses the lessons for developing countries that can be drawn from the transition process.

A short political history
After the October Revolution of 1917, the Bolsheviks found all political power in an internationally isolated and underdeveloped country rested in their hands. They considered it their main task to industrialize the country at full speed, using internally raised investment funds. It was for this purpose that the Soviet planning system was created (Knaack, 1996).

The Soviet Union experienced a ‘golden age’ in the 1950s. The country grew rapidly, propelled by increases of capital, labor and raw materials. Economic growth directly benefited consumers as their diet and housing improved apace (Schroeder, 1992). Space flights and Nobel prizes symbolized the achievements of Soviet science.

However, from the 1960s onwards, the Soviet economy settled on a slower growth path. According to Ellman and Kontorovich (1992, pp. 10–12), there are three explanations for this slowdown of economic growth. First, a loss of control of the economy. If an economy becomes more complex, coordination from above becomes increasingly difficult. Second, a reduction in the growth rates of both the volume and the productivity of production factors. One could not, for example, increase the participation rate of women any more, and also stocks of natural resources were running out. Third, a weakening of the ‘entrepreneurial spirit’. In a command economy, pressures from above provide the main source of dynamics in an economy. Brezhnev’s policy of ‘stability of the cadres’ represented a codification of the process of slackening the pressures from above.

Declining growth rates alone cannot explain the collapse of the Soviet Union at the end of the 1980s. According to Kornai (1992), the command system was still able to guarantee the population a decent way of life. What caused the crisis was its weak economic performance relative to that of the USA and other Organisation for Economic Co-operation and Development (OECD) countries. The dynamic efficiency argument was the raison d’être of the Soviet Union. Consequently, the relatively poor growth figures of the 1980s threatened the political legitimacy of the whole system.
Economic reform under Gorbachev must be understood as part of his effort to revive and modernize the Soviet economy. The results of the reform process were disappointing, mainly because Gorbachev’s policies were not feasible (Hewitt et al., 1987). But his policies also had unintended consequences. This happened especially for his glasnost policy intended to unmask bureaucrats sabotaging the reform process.1

On 1 November 1989, the Berlin Wall fell. From that moment on in CEE one country after another unlinked itself from the Soviet dominance and started a process of conversion from a centrally governed economy to a market economy. On 30 June 1990, the two parts of Germany were officially reunited and the German economic, monetary and social union was created. In Czechoslovakia, the Velvet Revolution ended the ruling of the Communist Party in November 1989. The new government of Prime Minister Klaus introduced a series of measures as from 1 January 1991, aimed at the integration of the Czechoslovakian economy into the world economy. The Soviet Union was also contaminated with this spirit of the time. After the breakdown of its economic system the new rulers strove for a rapid transition towards a new system characterized by market relations, private ownership, and a liberal democracy. After the abortive coup d’état of August 1991, the Soviet Union disintegrated and Yeltsin became President of the new Republic of Russia. On 2 January 1992, the Gaidar administration introduced a number of market reforms, which inflicted an enormous shock to the Russian economy.

Stylized facts of the transition process

Central and Eastern Europe and the former Soviet Union

The processes of change in the countries of CEE were not based on a blueprint showing how a formerly communist country ought to be restructured as a capitalist country. According to the prevailing opinion, that was not necessary. Due to the German unification, East Germany took the West German legislation and rules over at one stroke and Eastern German enterprises were privatized at a quick pace. Other countries had the possibility to copy a large part of the existing and tried and tested recipe as applied earlier by, for example, Spain at its accession to the European Union and developing countries at their integration into the world economy. It was the recipe developed by the International Monetary Fund (IMF) and the World Bank (Taylor, 1993) and is often dubbed the Washington Consensus. The Washington Consensus comprises the ‘Holy Trinity’ of stabilization, liberalization and privatization (Rutland, 1999). This Consensus initially was advice for a reform agenda for the Latin American countries at the end of the 1980s to adapt their policies and institutions.2 Applied to the former
communist countries at the time, this trinity would read as follows. ‘Stabilization’ refers to the need to reduce inflation, both open and suppressed, to create a currency with a stable external value, and establish balance-of-payments equilibrium. To achieve these goals, a standard package of fiscal and monetary measures was recommended, together with price liberalization so as to eliminate suppressed inflation. ‘Liberalization’ means the freeing of enterprises and individuals of the old planning institutions. The initial expectation was that the disappearance of these planning institutions would create space for the new market institutions to arise spontaneously. ‘Privatization’ was considered crucial for both political and economic reasons. Politically it would create powerful interest groups with a stake in transition and, for that reason, strongly opposed to the old system. In addition, it would secure active support from the West. It was thought, moreover, that only privately owned enterprises could operate successfully in a market economy. The collapse of the communist regimes in CEE after 1989 and the dissolution of the Soviet Union in 1991, spread optimism about the chances of rapid economic growth.

The transition process appears to have several characteristics, both intended and unintended (Campos and Coricelli, 2002; Ellman, 2005; Knaack and Jager, 2007). First, all countries of the former Eastern bloc faced a dramatic fall in output. The individual country patterns of recession and recovery cover the years 1989–97. They have largely been of the L-curve shape (sharp fall, followed by slow recovery), rather than the initially hoped-for J-curve type (small fall, followed by fast growth). Figure 71.1 displays the growth rates of real GDP in the years of transition for the distinct country groups. The strongly negative GDP growth rates for each group in the first years of transition indicates that real GDP levels should exhibit L-curves. The fall of investment expenditures, especially for inventories and housing, was even larger than the fall in real GDP. Defense expenditures on equipment and materials have declined sharply as a proportion of GDP, especially in the countries of the FSU. Consequently, during the period of falling output, private consumption has declined only moderately relative to production in most transition economies.

Second, a large inter-country variation in the adjustment patterns is observable, with large differences in both depth and length of the L-curve. According to Figure 71.1, for the whole group of CEE countries the average growth rate already became positive again in 1993, after a depth of −9 percent for the sub-group EU8 (the first eight CEE countries that joined the European Union, that is, Poland, Hungary, the Czech Republic, Slovakia, Slovenia and the three Baltic States) and −14 percent for the countries in the sub-group SEE (or South-Eastern Europe), both in 1991. The CIS, however, only experienced for the first time since 1989 positive economic growth in
1997. The low-income CIS countries had their depth in economic growth in 1992, with a decline in GDP of over 20 percent, and the middle-income CIS countries in 1994 with a negative growth rate of 15 percent. The variation in the cumulative fall of real GDP per country in the beginning of the transition process has been large. It ranges from moderate for Poland (6 percent, in two consecutive years of output decline) to high for Russia (40 percent, in seven years of consecutive years of output decline) and extremely high (over 60 percent fall) for Armenia (63 percent, in four consecutive years), Moldova (63 percent, seven years) and Georgia (78 percent, five years).

Third, like output, the level of foreign trade in transition economies has followed a pattern of decline and recovery. In the countries of Eastern Europe, foreign trade declined by 62 percent in the period 1990–93 and then rose to 71 percent of 1990’s export level in 1998. As for trade redirection, the share of the Western countries in the export of the CEE countries has increased tremendously. It nearly doubled from 35.8 percent in 1992–93 to 67.5 percent in 1998–99 (World Bank, 2002, p. 7). For the CIS countries this share did not change at all: it remained at 28 percent. On the other hand, the
former CMEA trade collapsed. Appropriate exchange rate changes to promote competitiveness proved to be a useful help in redirecting and promoting foreign trade. For example, the Czech Republic was successful in redirecting its trade to the West by devaluing its currency by 50 percent. In contrast, East Germany could not benefit from devaluation at all. Instead, it suffered from a managed real revaluation. Given the one-to-one substitution of the West mark for the East mark, in practice the competitiveness of East Germany badly worsened, creating mass unemployment.

Fourth, the sectoral composition of GDP in current prices changed during transition in favor of services and away from manufacturing. Largely this was due to the relatively larger contraction of manufacturing during recessions. However, productivity gains in the manufacturing sector and, therefore, declining relative prices of manufacturing products during economic recoveries also contributed to this sectoral change.

Fifth, during the transition the participation rates in the labor markets changed rapidly. Under communism, the participation rates were high, in particular for women. During the transition there was a significant increase of flows out of the labor force, especially for women and older persons near to retirement age. At the same time, officially registered unemployment rates increased much faster in Eastern Europe relative to the FSU countries. Labor also moved in great numbers from the state sector to the private sector. In contrast, geographical labor mobility remained very low.

Sixth, the production fall in the formal sector led to sharp increases of the official unemployment rates. Across the transition countries, the employment rates differed widely. The outcome for a specific country strongly depended on whether or not the loss of employment in the old enterprises was compensated by the creation of jobs in new (de novo) enterprises. In transition countries where restoring sustained growth has proved relatively illusive, new enterprises account for a low share of employment and value added – both between 10 and 20 percent (World Bank, 2002, p. xxv). According to the World Bank, in practice the transition process proves only to get momentum when the share of medium- and small-scale enterprises in the national employment is more than 40 percent. The CEE countries reached this percentage in 1996. In this respect, the countries of the FSU stayed behind. In these countries the growth of small-scale enterprises stagnated completely; here the share of small enterprises did not rise above 20 percent. Both the government and the Mafia seem to be responsible for that, as they viewed the new enterprises as cash cows instead of centers of new economic initiatives that have to be stimulated.

Seventh, a major result of the transition has been the emergence of a large informal sector. Partly, this concerns new enterprises, which wish to escape the attention of bureaucrats and tax authorities. Partly, it concerns
enterprises that already existed under the old system. These enterprises undertake substantial volumes of activity ‘off the books’ and make extensive use of barter trade. In addition, there has been a widespread informalization of the labor market. This takes the form of, for instance, unilateral determination of wages and conditions by the management of the firms, regardless of laws about these matters. Other examples are employers that disregard paying wages on time, and employees that do not receive maternity leave, are not protected from dangerous working conditions and are not represented by trade unions.

Eighth, a dramatic result of the transition has been a huge increase in relative and absolute poverty, as well as income inequality. According to the World Bank (2002, p. 8), between 1990 and 1998 the population living on less than US$1 a day in the CEE and CIS regions increased from 1.5 to 5.1 percent. It was the only area in the world for which this share increased in that period (in the poorest area worldwide, that is, sub-Saharan Africa, the share remained stable at 47 percent). The norm of US$1 is in effect not appropriate for the transition region as the costs of living are higher there than in other regions. Think, for example, of the cost of heating. Adapting the poverty line for region-specific costs, the World Bank estimates that between 1987–88 and 1993–95, the share of the poor in the total population in the FSU and CEE rose from 3 percent to about 25 percent, and in number of persons from 7 million to 89 million. This means that the transition process pushed more than one-fifth of the population below the poverty line. A group of the population that was particularly badly affected was children. Pensioners also suffered heavily. As a survival strategy, the population in all countries involved put much time and energy into cultivating food in the many allotments. For example, in 1996 this line of food production was estimated to yield 43 percent of Russia’s total food production. The deterioration of the living conditions led also to demographic changes. There has been a striking increase in mortality, concentrated among adult men in Russia and Ukraine. This increased the gender gap in life expectancy (in Russia women have a life expectancy that is 13 years longer than men). There was also a dramatic decline in birth rate and an increase in emigration. The combined effects of the current trends in mortality, birth rate and migration are that the population in many countries in CEE and the FSU declines.

Poverty increased not just because of the fall in output, but also due to greater income inequality in all European and FSU transition countries. Table 71.1 contains these developments over time in the years 1987–98, expressed through the Gini coefficients. Without any exception, the countries included in the table had a higher Gini coefficient, and thus a more unequal income distribution, at the end of that period. The CIS exhibited,
on average, a much larger increase than the CSB (Central and South-East European countries and the Baltic States). The Gini coefficient of Hungary hardly increased, whereas Armenia showed the largest increase. The picture of changes in the Gini coefficient in the years after 1998 (until 2003) is divergent: for example, Poland and Romania experienced further increases in the coefficient, for Hungary there is no change, while Russia and Armenia exhibit substantial declines in the coefficient (see, World Bank, 2005, p. 15).

Ninth, growth of crime and the widespread criminalization of society has been a significant feature of transition. This has been particularly

### Table 71.1 Income inequality during transition

<table>
<thead>
<tr>
<th>Countries</th>
<th>Gini coefficient of income per capita</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CSB</strong></td>
<td>0.23</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>0.23</td>
</tr>
<tr>
<td>Croatia</td>
<td>0.36</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>0.19</td>
</tr>
<tr>
<td>Estonia</td>
<td>0.24</td>
</tr>
<tr>
<td>Hungary</td>
<td>0.21</td>
</tr>
<tr>
<td>Latvia</td>
<td>0.24</td>
</tr>
<tr>
<td>Lithuania</td>
<td>0.23</td>
</tr>
<tr>
<td>Poland</td>
<td>0.28</td>
</tr>
<tr>
<td>Romania</td>
<td>0.23</td>
</tr>
<tr>
<td>Slovenia</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>CIS</strong></td>
<td>0.28</td>
</tr>
<tr>
<td>Armenia</td>
<td>0.27</td>
</tr>
<tr>
<td>Belarus</td>
<td>0.23</td>
</tr>
<tr>
<td>Georgia</td>
<td>0.29</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>0.30</td>
</tr>
<tr>
<td>Kyrgyz Republic</td>
<td>0.31</td>
</tr>
<tr>
<td>Moldova</td>
<td>0.27</td>
</tr>
<tr>
<td>Russian Federation</td>
<td>0.26</td>
</tr>
<tr>
<td>Tajikistan</td>
<td>0.28</td>
</tr>
<tr>
<td>Turkmenistan</td>
<td>0.28</td>
</tr>
<tr>
<td>Ukraine</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**Notes:**
- Not available.
- Median of countries with data.
- CSB is the acronym of Central and South-East European countries and the Baltic States.

**Source:** World Bank, 2002, p. 9.
marked in the FSU. In some countries, kleptocracy has been an important part of the political system. In these countries, there are also close links between the criminal, political and business worlds. This is connected with the inability of the state to perform even its night watchman functions. Hence, the supply of property protection and rule enforcement are privatized, that is, taken over by criminal organizations.

Tenth, during transition the old institutions collapsed, creating an institutional vacuum (Schmieding, 1993; Knaack, 1996). As the transition experience has advanced, it has become clearer that the disparity between the rapid progress in liberalization and privatization and the slow development of institutions that support markets and private enterprises directly and negatively affects overall economic performance.

China
The reform processes in the countries of Eastern Europe and the former Soviet Union can be better evaluated when we compare them with the reform process in China. It is interesting to compare, for example, Russia and China, because there are so many similarities between them as economies in transition and yet the outcomes of the transition process were so different. As said before, the fall in GDP in Russia in the period 1991–94 was more than 40 percent, while in the period 1978–2002 the annual growth rate in China was 9.4 percent. During this period in China, the per capita income of rural and urban population, measured at constant prices, increased 5.3 and 4.7 times, respectively (Lin, 2004). How can we explain these differences?

There are many similarities between Russia and China as economies in transition (Buck et al., 2000). These include the enormous geographical scale, abundance of cheap labor, and large potential markets to attract foreigners. In addition, both countries inherited similar economic and political ideologies from their Stalinist and Maoist periods, with a common emphasis on Party control, high defense budgets, large industrial enterprises, heavy bureaucratic and tariff protection against manufactured imports, and subsidized public services.

Of course, there were important differences too. The main difference was the economic structure. On the eve of the start of the transition process (in Russia in 1991 and in China in 1978) China was largely an underdeveloped and rural country with 71 percent of the workforce active in agriculture, despite heavy industrialization efforts in the 1950s and 1960s. Russia was an industrialized country with only 13 percent of the population working in agriculture. China’s financial system was also relatively underdeveloped.

To some extent, these differences in initial conditions and structural characteristics can explain the divergence of transition policies. Since
agriculture was the biggest sector, accounting for 37 percent of output and 71 percent of employment, agriculture was the biggest Chinese economic problem in 1978. It explains why the transition process started in agriculture and gradually spread to industry. When China started its transition in 1978, initially the government did not question the feasibility of the old system. Its attempt was simply to improve the system by giving agents in collective farms and state enterprises some autonomy so that a closer link between personal rewards and individual efforts could be established. Agricultural prices were raised substantially, mandatory quotas for sown area and output eliminated, and compulsory procurement quotas reduced, with the sale of above-quota output on free markets and increased possibilities for so-called sideline activities. Finally, the commune system gave place to the family responsibility system, accounting for 94 percent of the peasant households in 1984. In response to these reforms, per capita grain production rose by one-sixth between 1978 and 1985, the per capita output of cotton rose by two-thirds, and that of oilseeds, sugar and tobacco doubled (Balassa, 1987, p. 411).

With respect to industry, the Chinese government adopted a dual-track policy, liberalizing the entries to the labor-intensive sectors, but also creating conditions to address the viability issue of those firms in the capital-intensive high-priority sectors. The enterprises that were most able to take advantage of the entry option were collective enterprises sponsored by local governments, particularly in rural areas. These township and village enterprises had started already in the Maoist period, but expanded rapidly after 1978. They already produced 33 percent of total industrial output in 1991. Private firms then produced 11 percent, implying that the share of the state sector reduced to a mere 56 percent of total production (Naughton, 1994, Table 1). These developments in industry had two effects. The intensified competition resulted in the disappearance of monopoly rents and the occurrence of a division of labor between the state and non-state sectors. The state sector increasingly consists of large firms in heavy industry, while the non-state sector consists of medium-sized and small firms in light industry.

The success of the reforms in agriculture and industry had a positive effect on the macroeconomic stability of the economy. Formal rationing was in place for more than 20 consumer goods in 1978, accounting for more than 50 percent of consumer expenditures. Shortages of even the basic commodities were common. Shortages of consumer goods disappeared quickly, even though price controls remained in place in China. These policies had two effects: there was a significant increase in living standards and, therefore, an increase of popular support for the reform measures. Due to high saving rates of the Chinese population, there was no collapse of
investment in China, but a dramatic change in the way these investments were financed. Savings by government and by state enterprises has dropped sharply, while savings by private business and households has increased sharply. By the late 1980s, households were saving 15 percent of their income, compared with 2 percent before 1978.

The dynamics of the non-state enterprises exerted a heavy pressure on the state enterprises and triggered a state policy of managerial reforms. These reforms had some success. For example, during the 1980s the output of the state enterprises grew 7.7 percent annually – though other ownership forms of production exhibited even more rapid growth rates. Total factor productivity also grew in the state sector, although less than in the non-state sector. As a result, the state sector is still less competitive than the non-state sector. The relatively low export levels in the state sector compared with the export levels of comparable firms in the West are a clue to that (Buck et al., 2000, p. 393). The continued government subsidization and protectionism of the state enterprises through low-interest loans and monopolistic practices is both cause and consequence of this disappointing competitiveness. The government hopes to eliminate these distortions in the near future.

China also put into effect a policy of opening up to the outside world, but the scope of the opening to the outside world was only expanded gradually. The first step was to set up special economic zones in the south near Hong Kong and Macao, as laboratories for market reforms. Enterprises in these zones had the right to retain most of the foreign exchange they earned and were more flexible in firing workers. Furthermore, foreign investors in these zones could repatriate profits and own land. After the successful performance of these zones, a number of coastal cities were opened, followed by economic areas along some rivers, such as the Yangtze River, and along the borders, and finally capital cities of the various provinces were turned into open cities. In the first years, it was especially the ‘overseas Chinese’, including Taiwanese, who made use of the new possibilities. Later China could welcome a huge influx of foreign direct investment (FDI), also compared, for example, with Russia. In the period 1989–95, Russia attracted FDI to the amount of $3.9 billion, or $1.1 per capita, in 1995. In the same period, China received $121.7 billion of FDI, equivalent to $18.2 per capita (Buck et al., 2000, p. 384). Virtually all FDI was in the form of joint ventures. Joint ventures were made possible, although the government limited foreign ownership and control of businesses. In the late 1970s, foreign involvement was limited to 35 percent of any venture. This was raised to 49 percent in 1985, while from 1988 onwards majority foreign ownership is allowed, though still subject to state approval. The joint ventures were crucial for the success of the Chinese reform process, because the foreign
companies delivered the technological knowledge for the Chinese partners to produce goods that could compete on the world market.

An important difference between China and the CEE and the FSU is that in China the Communist Party kept control, while in the other transition countries the power of the Communist Party dissolved in favor of a democratic political system. The main argument of the Chinese leadership is that the continued rule of the Communist Party guaranteed social stability, which is conducive for the success of the economic reform. Leaders occupied with competing for power would create uncertainty about whether or not the economic reforms would be continued. This is not to say that there were no political reforms in China. Actually, economic reform often contains some elements of political reform. In the case of China, there was devolution of power from the center to the provinces. This empowerment of the regions has created what is now the major driving force behind economic reform (Woo, 1994, p. 289).

**Lessons to be learned**

For most economists and politicians the depth and duration of the depression which accompanied the transition process in CEE and the FSU came as a surprise. The initial idea was that the removal of the overwhelming apparatus of political control over economic activity could only imply additional prosperity in the medium term. The previous system was characterized by a myriad of distortions, and the removal of most of them would lead to a vigorous impulse to output. This optimism was not a monopoly of neoclassical economists. Well-known heterodox economists, like Janos Kornai, also held this view (Campos and Coricelli, 2002). This raises the question of how to explain this anomaly.

**Economic stabilization**

As mentioned before, in order to suppress open and hidden inflation, most transition countries adopted a standard package of restrictive fiscal and monetary policy (sometimes supported by exchange rate and income policy). Experience has shown that, despite widespread initial skepticism, such packages – if persisted in – are successful in reducing macroeconomic instability (Ellman, 2005). However, the time for the package to work may be far longer than anticipated. For example, Poland implemented the stabilization package at the beginning of 1990, but only nine years later inflation fell below 10 percent. Moreover, this restrictive macroeconomic policy may contribute to a sharp decline in output and welfare, as happened in Russia after the unsuccessful attempt at shock therapy by Prime Minister Gaidar. As a result, Kornai (1994) argued – writing on the Hungarian situation where inflation was about 20 percent – that growth must be the main
economic objective when the danger of accelerating inflation does not exist anymore. Stiglitz (1998) strengthens this view, summarizing the evidence that only high – and not moderate – inflation is costly: When countries cross the threshold of 40 percent annual inflation in an upward direction, they fall into a high-inflation, low-growth trap. Below that level, there is little evidence that inflation is costly. Recent research even suggests that low levels of inflation may improve economic performance relative to what it would have been with zero inflation.

**Optimal sequencing**

Economic stabilization is a prominent part of the discussion on the optimal order of reforms. The transition process of the 1990s has intensified and widened this discussion, which previously was concentrated on the reform process in developing countries. It concerns reform on three levels of aggregation. On the highest level of aggregation, it is about the optimal order over time of economic liberalization, economic stabilization, privatization and the creation of supporting institutions, necessary for a smooth working of the markets. On a lower level of aggregation, the optimal-sequencing discussion focuses the order over time of the different parts of economic liberalization, in combination with economic stabilization. Here one distinguishes the liberalization of domestic goods and labor markets, international trade, domestic financial markets, international capital flows, and the foreign exchange market. On the lowest level of aggregation, the focal point is the optimal order of domestic sectoral reform: agriculture prior to industrialization, or perhaps the other way around?

A very useful empirical analysis of the transition lessons of the optimal order of the highest aggregation level is Beyer (2001). His data set consists for each of 14 CEE countries of the months in which they switched to the new regime, the majority of their prices were liberalized, their most substantial attempt for stabilization was undertaken, and a new or adjusted constitution was adopted (as an indicator of a country’s new institutional system). By using groups of countries with a similar order of reform it appears that the sequence over time of stabilization, privatization and liberalization is significantly the best sequence in terms of GDP level six years after the system switch. Beyer dubs this sequence the graduality approach. Slovenia and Hungary have followed this adjustment path. These two countries started the reform process with constitution-building. If instead liberalization took place together with stabilization at the start of the reform process, Beyer considers the adjustment process would be a ‘big bang’ approach. The countries that belong to this group (Albania, Bulgaria, the Czech Republic, Poland and Slovakia) display a worse development of GDP relative to the gradualists, but a significantly better outcome than the
transition countries that liberalize and/or privatize before stabilization. Early privatization appears to work badly, as Belarus, Romania, Russia and the Ukraine witness. However, Estonia and Lithuania are counter-examples.

The previous paragraph gives the important clue for optimal sequencing at the lower level of aggregation: that stabilization should in any case not come after liberalization. This gives a strong support to the earlier literature on optimal sequencing for developing countries, which concluded that stabilization should be carried out first, followed by liberalization. The standard outcome of that literature with respect to the optimal sequence within liberalization is: domestic goods and labor markets, international trade simultaneously with unifying the exchange rate and realizing the equilibrium value of that rate, domestic financial markets, and finally, freeing international capital flows from border restrictions.

The transition gives some support to this optimal order, though only of an anecdotal character. The anecdotes that follow have to do with the space for the exchange rate to find its equilibrium value in time. In former East Germany, the unification with West Germany implied a ‘big bang’ liberalization and the introduction of a unified, but highly overvalued currency from former East Germany’s viewpoint. The politically motivated choice of a one-to-one conversion of the East mark into the West mark resulted in a huge deterioration of competitiveness of former East Germany and a concomitant disaster with respect to its GDP, creating mass unemployment. This outcome gives support to the earlier-mentioned optimal liberalization order, which requires that the exchange rate is able to reach its equilibrium value in the process of trade opening.

The collapse of the CMEA trade soon after the start of transition and the resulting loss of jobs in the big state enterprises had to be counterbalanced by an increase of exports to mainly the West and the creation of new jobs by small and medium-sized enterprises. The Czech Republic, for example, was successful in both respects. The large devaluation of its currency resulted in a strong swing of its foreign trade to the West and the process of ‘small’ privatization contributed to strong employment growth in the private sector. Obviously, the Czech Republic also profited from its geographical position and the possibilities of the tourist industry, especially in Prague.

Like East Germany, Russia has opted for, ultimately, a less successful road. As with the Czech Republic, it devaluated its currency in 1992, but made the mistake to choose a more or less fixed exchange rate of the rouble against the dollar under conditions of high internal inflation. The cause was that Russia had already liberalized international capital movements before the economy was sufficiently stabilized. The potential instability was the lax policy stance on fiscal deficits. Large capital inflows initially allowed
the government to finance a continuing fiscal deficit at relatively low interest rates. This can be considered the so-called good equilibrium (see Gros and Steinherr, 2004, p. 243). However, in the meantime, for Russia the debt-to-GDP ratio increased and so did the country’s debt service burden. This development gradually undermined the country’s credibility. Reinforcing simultaneous developments were a growing overvaluation of the rouble, in response to the large capital inflow, and inflationary pressure, due to capital inflow as long as the central bank stabilized the exchange rate. Both reinforcing developments usually worsen the country’s current account over time – in a gradual, but inevitable way. The real appreciation of the rouble did not lead so much to a deterioration of the trade balance, given the strong export potential of the gas and oil reserves. But this appreciation resulted in the crowding-out of the Russian industrial production in line with ‘Dutch disease’ features. Industry became more and more expensive and lost its possibilities to export. The industrial loss of sales became still more severe because the real rouble appreciation stimulated the Russian consumers to opt for cheaper foreign consumer goods. In this state the country was ripe for a shift in expectations leading to the so-called bad equilibrium (see Gros and Steinherr, 2004, p. 243). Given the openness of the capital account, Russia was exposed to sudden withdrawals of foreign capital. This fear became reality during the summer of 1998, after which the rouble collapsed and a severe economic crisis occurred.

With respect to the third level of aggregation and the concomitant optimal sequencing of sectoral reform, the different reform roads of China and Russia are informative. From the success of the economic reform process in China, some economists concluded that the Chinese road of agriculture first was also applicable to Russia. By starting the reform in industry, Russia was unable to gain the credibility that probably would have come if it had started the reform process in agriculture. This position is debatable. The situation in China and Russia was quite different. In China, the agricultural sector was the biggest sector, accounting for 37 percent of the output and 71 percent of employment. In Russia, the agricultural sector employed only 13 percent of the labor force and generated 18 percent of gross national product (GNP). Moreover, relative to China the Russian agriculture was much more mechanized. It is much easier to assign property rights to the individual plots that farmers have been working on than to assign property rights to the capital equipment that workers have been using jointly.

However, the argument gains weight when we include in agriculture the activities of the townships and village enterprises. In the 1990s they already produced 33 percent of industrial output. Including the private sector, the enormous dynamism of the non-state sector had a positive effect on the
supply of consumer goods and the living standards of the population, and therefore created popular support for the reform measures. A rapid growth of the non-state sector can also absorb the unemployed in the state sector. This happened also in the Czech Republic. In the period 1989–1995, about 25 percent of the workers in the state sector left that sector. They could easily find new jobs in the new private sector, especially in the new service sector. Especially in the Prague area, the new private sector showed amazing growth. A recent report of the World Bank (2002) confirms the importance of the stimulation of the non-state sector in an early phase of the reform process. According to this report, a key for economic growth in transition countries is the shift from capital-intensive to labour-intensive enterprises. The last group consists overwhelmingly in small enterprises (with a maximum of 50 workers). According to the World Bank, the transition gets momentum when the share of medium- and small-scale enterprises in the national employment is more than 40 percent. China fulfills this condition, as well as to a lesser extent some CEE countries.

Compared with China and the Czech Republic, the position of the small and medium-sized enterprises in Russia is delicate. In the years 1995–97 employment in Russian businesses with a maximum of 50 employees fell officially by 50 percent. Disappointing productivity growth cannot be an explanation. On the contrary, these firms often had good performance (Commander et al., 1996, Chapter 8). The true explanation is the ‘grasping hand’ of the Russian government and the Mafia. They compelled small enterprises to pay highly unpredictable taxes and regular payments to their ‘protectors’. This explains why the transparency of government behavior and a reduction of risks in the business environment are important determinants of the success of the reform process.

**Institutions**

After the fall of the Berlin Wall in 1989, in principle all CEE countries followed a liberalization process directed at the breaking down of the planning systems. It was expected that markets would arise spontaneously as soon as the old planning bureaucrats disappeared. In other words, the policy-makers expected that a process of ‘organic growth’ would create the political and economic institutions necessary for the smooth functioning of a market economy. Obviously they trusted that the fundamental propensities of human nature to ‘truck, barter and exchange one thing for another’, as postulated by Adam Smith, were not foregone during the decades of communist rule (Knaack, 1999, p. 357).

However, they did not take into account the fact that proper functioning markets require an institutional infrastructure and that it takes a lot of time before the new institutional system and the persons who have to work in
those markets are adapted to the new circumstances. A clearly delineated system of property rights; a regulatory apparatus curbing the worst forms of fraud, anti-competitive behavior and moral hazard; a moderately cohesive society exhibiting trust and social cooperation, the rule of law and clean government – these are the social arrangements that economists usually take for granted, but which were absent in the transition economies (Rodrik, 2000). Not only must new institutions be created, but they must also prove their value during a time-consuming process of trial and error. Each economic transition process is fundamentally an incremental process, during which the country constantly experiments with new forms and finally keeps that form which is satisfactory. In this way, the existing institutional structure actually improves.

The abolition of the old planning system in one stroke without the construction of new institutions of a market economy has irrevocably led to an institutional vacuum. That vacuum has many forms (Knaack, 1999, p. 363). The old rules lose their value, but the enterprises have not yet learned how to behave in the new situation. Further, the information structure of the old system disappeared, while the new market signals were not yet fully developed. For the enterprises it was difficult to find new customers, and when they finally succeeded it was difficult to assess their creditworthiness. As a result, the enterprises operate in an environment characterized by an extreme uncertainty.

It is obvious that the institutional vacuum must be filled. In the CEE countries and the countries of the FSU this happened in different ways. In the case of East Germany, the country took over in one stroke the institutions of West Germany. The Czech Republic profited heavily from the neighborhood of the large markets of Germany and Austria, and the country also learned quickly from the international trade relations. Moreover, from 1995 onwards the Czech Republic gradually adopted the acquis communautaire, the legal structure of the European Union. It must be stressed that imported blueprints do not do their work instantaneously. The main reason is that blueprints are highly incomplete. Much of the knowledge to operate with the blueprints has not been written down and has to be learned. However, blueprints give a direction for knowledge acquirement.

Russia did not have these advantages. For a big country, it has a surprisingly huge international trade. However, nearly all export is energy and raw materials. Actually, these are the features of a small and open developing country. With regard to the possibility of the import of institutions it only had to fulfil the requirements of the IMF when it borrowed some money. Consequently, much more than the other small CEE countries, Russia had to fill the institutional vacuum on its own terms. Given the fact that the
creation of new institutions is a time-consuming process, one can understand that it fell back on its old routines and that, given the weakness of the state, organizations such as the Mafia also filled the vacuum.

Compared with the CEE and FSU countries, the problem of the institutional vacuum in China was less acute. First, we have to remember that after the reforms of 1978 the overwhelming majority of the economic relations in China was still shaped and guided by the official planning system. Only in a very gradual way was there a shift from the planning system to more market relations. Consequently, firms and individuals had enough time to adapt to the new circumstances and to learn the rules of a market economy. Second, insofar as the enterprises in the economic zones had to obey immediately the rules of the market, they could profit from the experiences, knowledge and economic networks of the so-called ‘overseas Chinese’, businessmen especially from Hong Kong, who invested heavily in the zones. Third, sales in the big cities of agricultural surpluses and light industry products of the village enterprises need relatively little organization and structure. Permission from the local authorities to sell on a street corner or a square is sometimes enough. This also explains why, for example, in the Czech Republic it took so little time to start small enterprises in the service sector.

**Political reform**

There is an intensive debate about the relationship between political regime type and economic performance. Based on the experience in a handful of economies in East and South-East Asia, which (until recently at least) registered the world’s highest growth rates, under authoritarian regimes, one could conclude that economic development requires a strong hand from above. To embark on self-sustained growth, deep economic reforms are often needed, which cannot be undertaken in the messy pull and push of fragile democratic politics. The main argument is that economic reform necessarily imposes costs on some segments of society, and that political openness would provide the losers with the opportunity to form coalitions to stop the reform. An example of this occurred in 1992 in Russia when the apparatchik Chernomyrdin replaced the reformist Gaidar. The first new acts were to squeeze the thousands of small shops that had appeared since January 1992 and to extend cheap credits to the industries under the Ministry of Oil and Gas that Chernomyrdin had headed (Woo, 1994, p. 288). A strong and committed leadership can also push economic reforms against the interests of some interest groups. For example, Buck et al. (2000) describe that the Chinese Communist Party stimulated joint ventures with foreign partners against possible dissent of insiders of state enterprises excluded from the deals. The central authorities stimulated
foreign investors to select from the state enterprises the best physical and human assets to form joint enterprises, usually geographically separated from the unreformed parts of the state enterprise left behind. Normally, the incumbent manager and workers repel any outside investor, unless they are prepared to give employment guarantees for all branches of the enterprise, including those involved with social provisions.

This line of thought met a lot of criticism. Rodrik (2000), for example acknowledges that in effect the Asian countries have prospered under authoritarianism, but that many more have seen their economies deteriorate – think of Zaire, Uganda or Haiti. Moreover, some of the most successful economic reforms in South America were implemented under newly elected democratic governments – witness the stabilizations in Bolivia (1985), Argentina (1991) and Brazil (1994). Moreover, the transitions in the democratic European countries were more successful than the transitions in the authoritarian FSU countries, like Belarus, Turkmenistan and Uzbekistan. It is obvious that we cannot subtract from these examples a clear-cut answer about the relationship of political and economic reform. But it is evident that the reform process is helped when a strong government is able and willing to create the necessary market institutions and resist the interest, especially of the insiders in the state enterprises.

Conclusions

After the fall of the Berlin Wall in 1989, one country after another in Central and Eastern Europe freed itself from Soviet domination and started a transition process from a centrally planned economy into a market economy. In this transition process, they followed the recipe from the IMF and the World Bank, developed earlier for developing countries, mainly in Latin America. In all transition countries in Central and Eastern Europe, the results were rather disappointing. In 1999, only three of all these transition countries surpassed the 1989 national income levels. The new countries that belonged to the former Soviet empire underwent an even more severe income fall. The national income reductions were much more profound than initially expected. When we compare these figures with China’s experience, the difference is striking. China started its reform process in 1978, and for many years had double-digit positive growth figures without any intermediate fall.

The length and depth of the recession in most countries can be explained by the fact that the reform process was based on an incomplete theory about the functioning of a market economy. The policy-makers recognized too late the precondition for the functioning of a market economy, namely an institutional infrastructure, and the dynamics of the reform process, namely that it takes time before the new institutional infrastructure and the
persons who have to work in it are adapted to each other and the new situation. The collapse of the old planning institutions placed the enterprises in an extremely uncertain situation, in which it was difficult to find new customers and to decipher how trustworthy they were. From this perspective the length and depth of the depression depended on the time it cost to build new institutions, for example the new private property rights, and the time for the market players to adapt to them.

Our study also reveals that the negative aspects of transition can be compensated for, and in the Chinese case even more than fully compensated for if the authorities allow structural flexibility. This takes two forms. First, the speed with which the de novo enterprises can expand is important for the success of the transition process. This change is observable in the Czech Republic and especially in China. According to the World Bank the transition gets momentum if the share of medium-sized and small-scale enterprises in national employment is more than 40 percent. This condition is fulfilled in China. In Russia, to the contrary, the de novo enterprises were unable to expand. The government could not protect the new enterprises against the negative practices of the Mafia and the already existing big enterprises. Behaviour of the government itself, such as an erratic tax burden, also was counterproductive. Second, the loss of the export market that the COMECON was before the regime switch had to be offset by an increase of exports to mainly the West. The Czech Republic was particularly successful in this respect. The strong devaluation of its currency resulted in a strong swing of foreign trade to the West. China’s export possibilities to the West also profited strongly from an undervalued domestic currency, the yuan. In contrast, Russia did badly in the 1990s due to an overvalued rouble, leading to a crowding-out of the traditional industry.

Besides these institutional lessons, which are also applicable to developing countries, the transition process in Eastern European countries and the newly independent countries that arose from the former Soviet Union also produced some lessons about the order of reform measures. Institutional adjustments and economic stabilization, both early in the reform process, prove to have positive effects on a rapid restoration of the pre-transition national income level. Late stabilization, in contrast, is devastating in this respect. A late adjustment over time of the exchange rate system towards more flexible – and thus less rigid, disequilibrium – exchange rates appeared to be economically costly in the transition countries: witness the negative experience of former East Germany and Russia, and the positive experience of the Czech Republic. This is an implicit support for the optimal sequence of liberalization steps as developed before the transition experience of the 1990s.

Summing up, the success of a transition process, and thus a development policy in general, not only depends on the building of a viable market
sector. It also depends on the existence of a strong government that is able and willing to create the necessary market institutions, fight the vested interests, and formulate an economic policy that aims at an immediate and thorough economic stabilization. Privatization and full liberalization can come later. As soon as domestic markets function, a rapid opening of international trade and stimulus of the international trade relations, including the introduction of a unified exchange rate which closely approaches its equilibrium value, should be focal points of economic policy.

Notes
1. For example, latent nationalism was fuelled by new publications about the Chernobyl catastrophe and the contents of the Molotov–Ribbentrop pact.
3. CMEA is the group of countries that belonged to the Council for Mutual Economic Assistance. The latter was the body that was supposed to govern trade among Soviet-bloc nations.

References


Index

‘1-2-3’ model, see dependent economy model

Acemoglu, D. 266–70, 385–6
Afghanistan 101
Africa 24, 458–9, 463
poverty in 220
skilled worker migration from 122
Africa, sub-Saharan 103, 125, 152, 459, 463, 490–516
brain drain from 119
child labor in 254
children's health in 250–51
democratization in 512
demographic change in 495, 507–8
development strategies 514–16
economic structure 491–5
gender gaps in education in 237
demographic disadvantages 505–7
growth experience 490–98
interpretations of poor growth performance 498–508
poverty in 220, 490, 508
African Development Bank 157
aggregate demand 557–8, 560–61
agricultural productivity
and gender inequality 239
agriculture 37, 397–8, 400
in China 588, 593
effects of war on 430
foreign aid and 108
gender issues and 228, 229, 230, 231
in India 557–8, 562
in Russia 593
in South Asia 556
AIDS 154, 221, 224, 225, 251, 383, 435, 507–8
Aitken, B. 142
AK model of growth 4
Amsden, A. 168
Arab Free Trade Agreement 527
Asia 73, 75, 76, 78
gender inequality in 236
Asian Development Bank 157
Asian financial crisis 72, 74–6, 84–5, 463, see also financial crisis in East Asia.
asymmetric information 342
in international capital markets 88
in technology markets 138, 140
international migration and 124–5
balance of payments 151, 451–2
of India 563–4
Baland, J-M. 394–406
Balasubramanyam, V. 22–3, 48–58, 143–4
Bangladesh 556, 565–6
Bank for International Settlements 66
bank lending
international 71, 79–88
banks 79–88, 89–91, 385, 453, 525
Bardhan, P. K. 381–93
Barro, R. J. 22, 195, 413, 420
Bates, R. 270, 498–9
Bauer, P. 98
Bebbington, A. 414
behavioral economics 389
Benabou, R. 201
Berg, E. 498–9
Beyer, J. 591
Bhagwati, J. 5–6, 51, 170
Birdsall, N. 193–212, 461
Blecker, R. 170
Blecher, H. 24, 26
Bloom, D. 126
bonds
GDP-indexed 65
local currency-denominated 65–6
Borenszttein, E. 144
Borjas, G. 125
Boserup, E. 225–6
Brady plan 481
brain drain, 119–22, 132–3, 171
Branstetter, L. 145
Braunstein, E. 242–3
Brazil 72, 154, 156
economic growth 261
ethnic/racial inequality in 264
Canada
  ethnic/racial inequality in 264–5
  capabilities 312, 418, 420
  capital account liberalization 531
    in India 563
  capital accumulation 32, 35, 166–7,
    349–50, 458–9
  capital flows see international capital
    flows
  capital markets 382, 384, 525
  capital mobility 295
  capital requirements 68
  Caribbean
    skilled worker migration from 121
  CEPAL see ECLAC
  centrally planned economies,
    lessons from 282–4
  Chang, H-J. 367
  Chen, S. 219
  child labor 253–4, 255–6, 259–60
  children
    and development 250–61
  effect of female education on 238
  Chin, J. 173
  China 54, 78, 165, 457, 463, 536–50,
    579, 587–90
  comparison to India 537–8
  FDI in 51–3
  gender issues in 234–5
  growth experience 536–42
  income distribution in 189–90, 541
  investment levels 539
  legal system in 387–8
  population control in 538–9
  poverty in 220, 536, 541
  private sector participation 543–9
  state owned enterprises in 542,
    544–5, 548
  total factor productivity 539–41
  Chiswick, B. 124
  civil war 265, 423–41, 511 see also war
  class 390–91
  classical economists 165

  Coase theorem 342, 385
  coefficient of variation 182–3
  collective action 398–9
  Collier, P. 270, 426, 427, 498, 501–2
  Colombia 427
  colonies 31, 167, 267–9, 529
  commodity agreements 158
  Commonwealth of Independent States
    (CIS) 579 see also transition
    economies
  comparative advantage 16, 21–2, 31,
    153, 167
  Compensatory Financing Facility
    (CFF) 151
  Computable General Equilibrium
    (CGE) Models 347–8, 354
  conditionality
    for foreign aid 111–13
    IMF 151–2
  contagion, financial 60, 81, 88
  contract enforcement 203, 411
  contracts, informal 388
  convergence, of income across
    countries 169, 172
  corporate governance 387
  corruption 199–200, 369–77, 386
    costs of 371–2
    in East Asia 574–5
    foreign aid and 107, 113
    and growth 290, 372
    levels and types of 370–71
    measurement of 369–70
    policies towards 372–6
    and size of government 374
  cost-benefit analysis 32, 351–3
  Cote d’Ivoire 425
  credit markets 198–9, 384, 389
  crime 203
    in transition economies 586–7
  cross-country growth regressions 10–12
  culture 407–22
    definition of 409
  cycles, effect of international capital
    flows on 90–94
  Dahi, O. 522–35
  Damon, A. 250–61
  Darity, W.A. 262–77
  data problems 215–16, 217–18, 229–30
  international migration and 120
De Soto, H. 411
Deardorff, A. 173
debt crisis 88, 151, 454
debt relief 101
deflation 297
Deininger, K. 208
Demir, F. 522–35
democracy 408, 412, 531–2
and war 427–8
dependency theory 167, 478
dependent economy model 322–4, 331–2
Development Assistant Committee (DAC) 99
discrimination
against women 231
economic 197
ethnic/racial 264
distortions 5–6
market 32–3, 37–8
policy-induced 6
divergence, in growth 455–6
division of labor, international 287
Doha development round 151, 153, 154, 155
dollarization 298
Douglas, M. 417
dual economy 36
Dutch disease 43, 322, 333, 523, 526
foreign aid and 108
remittances and 132
Dutt, A. 22, 163–76
East Asia 48, 203–7, 459, 569–78
asset redistribution in 575
diversity of experiences 571–5
foreign direct investment in 571–2
industrial policy in 572–3
interpretations of growth experience 569
lessons from 576–7
trade policy in 572–3
Easterlin, R. 194
Easterly, W. 98, 201–2, 264, 269, 270, 411
Economic Commission for Latin America, (ECLAC) 158, 474–6, 478
education 199, 209, 217
of children 251–3, 256–8
effects of war on 435
inequality of 206
in Latin America 484–5
in MENA 528–9
policy 224, 438
policy, for children 259
primary 207
in South Asia 556
university 206
efficiency 55, 342–3
of private firms and state-owned enterprises compared 361–4, 366–7
Elsun, D. 227–49
enclaves 460–61
Engel’s law 20
Engerman, S. 201, 385
England 381
entitlements, individual 430–32, 434–5
entrepreneurs 359
environment 173–4
environment, and war 428, 429
environmental Kuznets curve 173
environmental planning models 349
equity 342–3, 360
Ethiopia 24
ethnic diversity 270–75
ethnic fractionalization 504
ethnic/racial conflict 264–6, 275
ethnicity 262–77, 423–4, 531
ethno-linguistic fractionalization (ELF) index 270–74
Europe, Central and Eastern 579, 581
see also transition economies
evolutionary economics 405–6
exchange rate 206, 243, 245, 526
devaluation of 335
as a nominal anchor 297–8
policy 40, 42, 293, 295–6, 526
stability 296
targets 301, 336
exchange rate regime
fixed 296–8
floating 296–8
intermediate 296–8
export processing zones
in East Asia 570
export promotion 10, 21, 50–51, 144 in East Asia 570
export-oriented industrialization women and 232
exports 33, 35, 39, 40, 54
effects of war on 430
and gender inequality 239
of India 561
manufacturing 449–50, 459
technology transfer and 137
externalities 130, 167, 168, 169, 342,
384, 395, 396, 399, 400–401, 447
from skilled emigration 132–3
factor price equalization 168
families,
bargaining power of women in 232,
235
migration decision by 126
famines 418
Fel’dman model 350
female education
effect on children 238
effect on fertility 237
effect on growth 236–7
effect on population growth 237
fertility 237
finance 388
financial crisis 71, 72, 74–5, 454, 531
in East Asia 62, 575–6
in Mexico 72, 74–6, 82–3
in Russia 62
financial liberalization 88, 89, 90, 91–4,
452–3, 458, 463
in East Asia 576
financial markets 452–3
women and 240–41
financial volatility 60, 71, 89–90
effects on growth 59–60
Findlay, R. 166
fiscal deficit 312–13
relation with growth 313
in Latin America 477, 481–2
in Sub-Saharan Africa 510
fiscal policy 91, 305–17
characteristics in developing
countries 305–11
and gender equality 241–3, 244
in India 558
and privatization 360–61
procyclical 311
in South Asia 555, 557
fisheries 402–3
Fishlow, A. 469–89
FitzGerald, V. 319–40
food security 438
foreign aid 98–118, 430, 436–8
amounts of 100
definition of 99
donor motives for 105
donors of 101, 104
foreign policy and 104
recipients 101–3
tying of 104–5, 110
types of 99, 110
foreign debt 170
foreign direct investment 48–58, 61, 64,
170, 172, 448, 451
in China 589
determinants of 49–53
in East Asia 571–2
effect of 53–6
effects of war on 433
export oriented 143
and gender equality 244
import substituting 143
in Latin America 476–7
tax incentives for 54
and technology transfers 138–46
foreign exchange 54, 106, 447–8
effects of war on 430
Foster-Greer-Thorbecke index 215
fragility, economic 77–8
France, legal system in 387
Friedman, B. 410
Friedman, M. 98, 390
Fukuyama, F. 413–14
gap models 106, 350
gender,
and development 227–49
distribution, effect of growth and
development on 227–36
distribution, effect on growth and
development 236–41
inequality, in India 563
General Agreement on Trade and
Tariffs (GATT) 150
genocide 265–6
Ghosh, J. 554–68
Gibson, B. 341–55
Gini coefficient 179, 182–3, 195–7, 208, 463
Glass, A. 137–49, 173
Glewwe, P. 250–61
global warming 174
globalization 71, 119, 234, 288, 445–6, 473, 485–6, 577
Goulet, D. 417–19
governance 134, 388
and foreign aid 113
and growth 290
government 39, 343, 436
failure 201, 286, 343
policy intervention 39
revenue 359
foreign aid and 107–8
in transition economies 596
government expenditure 209
effects of war on 433–4
reform 316
Gray, C.W. 369–80
greed and grievance, and war 426, 429
Griffith-Jones, S. 59–70
Grilli, E. 27
gross domestic product
indexed bonds 65
Grossman, G. 173
growth 23–4, 30, 33, 37, 88, 165, 203–7, 447, 450, 454–6,
effect of brain drain on 133–4
effect of ethnic diversity on 270
effect of foreign aid on 105–10
effect of income distribution on 194–201
effect of international capital flows on 90–94
effect on ethnic/racial inequality 262
effect on gender inequality 235–6
effect on well-being of children 254–8
effect of war on 432
in India 560–61
in MENA 524, 529–30
models 349–50
in South Asia 555
of transition economies 582
relation with life expectancy 223–4
relation with literacy 223–4
relation with poverty 223–4
Hagen, E. 412
Hamilton, A. 168
happiness 194, 207–8
Harrison, A. 142
Harrison, L. 411, 412, 416
Harrod-Domar model 32, 36, 106, 321
Hatton, T. 125–6
Heal, G. 342, 349
health 199, 217–18, 383
of children 250–51, 256–8
effects of war on 435
foreign aid and 106, 107
impact of brain drain 132
in MENA 528–9
policy 224, 257, 438
policy, for children 258
in South Asia 556–7
in transition economies 585
Heckscher, E. 31
Heckscher-Ohlin-Samuelson model 129–30, 167, 168
Heinz, J. 242–3
Helpman, E. 173
HIV 221, 224, 225, 435, 507–8
Hoksbergen, R. 407–22
Homer-Dixon, T. 428
Hong Kong 463
horizontal inequality, see inequality, horizontal
Huang, Y. 52–3, 536–53
human capital 55–6, 133, 144, 171, 270
effect of war on 432
in MENA 528
Human Development Index 218, 418
for Bangladesh 566
ethnic/racial inequality in 263–4
Hymer, S. 49–50
identity 409, 416
immigration policy 127
imperfect information 389
imperialism 167
import substitution 10, 21, 32, 33, 36, 42, 43, 50–51, 144, 288, 523, 524–5
in East Asia 570
in India 558
in Latin America 477–8
income, measurement of 215–16
income distribution 193–212, 364, 398, see also income inequality

Amitava Krishna Dutt and Jaime Ros - 9781848442818
Downloaded from Elgar Online at 07/19/2019 02:48:07AM
via free access
effect of migration on 129
international migration and 124
income inequality
aggregation problems in measuring 181–5
in China 541
comparisons across countries 184–5
comparisons over time 185
determinants of 185–8
in different regions 183–4
and economic growth 188–90
in Latin America 484–5
measurement of 179–91
policies towards 190
increasing returns to scale 130, 167, 168, 169, 171, 386, 403–4, 446–7
India 156, 203, 415, 557–64
FDI in 51–3
gender issues in 230, 234–5
information technology sector in 133, 561, 563
international migration from 120
legal system in 387–8
uneven pattern of poverty reduction in 221
Indonesia 139, 425
industrial policy 168, 460
and gender equality 244
industrialization 188, 475–6
industrial sector, gender issues and 228
industry, in India 558
inefficiency 33, 395, 396
inequality 194–201, 399, 557, see also equity, income distribution between women 236
in distribution of education 199–200
effect on growth 194–201
ethnic/racial 262–4
horizontal 208–9, 425, 426, 428
in India 562
international 164–5, 456–7
international migration and 124
in landownership 199
in MENA 531
North–South 30
in South Asia 555–6
in transition economies 585–6
vertical 427
inflation 295, 301, 306
effects of war on 433
in Latin America 477, 481
in Sub-Saharan Africa 510
inflation targeting 294, 298–9, 325–30
and central bank independence 300
experience in developing countries 299–301
prerequisites 299–300
informal employment 232, 240
informal sector 37, 433–4, 437
in transition economies 584–5
information technology 133
in India 561, 563
infrastructure, effects of war on 432
innovation 134, 172, 385
input–output models 344–5
institutional economics 381, 385
new 394
institutions 41–2, 194, 201, 202, 266–9, 385–6
dysfunctional 385–6
in transition economies 587, 594–6
insurance markets 198, 389
intellectual property rights 138, 145–6, 172, 172, 382–4
Inter-American Development Bank 157
International Bank for Reconstruction and Development (IBRD) 104, 150–52, 157, see also World Bank
international capital flows 55, 59–68, 71–118, 169–70, 328, 448–9, 452–3, 463, see also foreign direct investment
boom–bust cycles of 59–61, 64
effects of war on 430, 432–3
main characteristics of 60–64
policies towards 64–8
regulation of 66–7, 77, 88, 94, 156, 295
to India 564
international capital markets 88
International Development Agency (IDA) 151
International Development Association (IDA) 104
international institutions 150–62
international migration 119–36, 171
effects of 128–33
growth in 120
returns to 123–4
International Monetary Fund (IMF) 53, 66, 111–12, 114, 151, 153, 155–7, 159, 243, 288, 319, 530–31, 581

Basic Financial Programming Framework 320–21

international production chains 451, 460

international trade 3–13, 31–48, 167–9, 287, 448
effects on income distribution 7
effects on growth 4, 8–12
effects of war on 433

gains from 4–8, 16, 167
liberalization 11, 38–9, 172, 449–52, 482, 510–11
liberalization, and gender equality 243

mainstream view of 3–4

in MENA 526–8

policy 21–2, 30–47, 168, 225, 460

of transition economies 583–4

International Trade Organization 150

investment 35, 36, 197, 451, 458
effects of war on 432

foreign aid and 106, 108
public 458–9

Iraq 101

Islam 533

Jager, H. 579–600

Jha, R. 305–18

joint ventures 144

Jomo K. S. 569–78

Kaldor, N. 207

Kaminsky, G. 71–97

Keen, D. 426

Kerala 222, 224

Keynes, J. M. 27, 150

Kindleberger, C. P. 54

Knaack, R. 579–600

knowledge,
codified 36
tacit 41, 43, 44

Korea, South 21, 39, 40

Kornai, J. 580

Korten, D. 416

Kozul-Wright, R. 445–68

Krugman, P. 168, 171–2

Kuran, T. 533

Kuznets curve 188–90

Kuznets, S. 188, 193

labor markets 123–4
effects of war on 430

in MENA 528–9

labor, skilled 55

Lall, S. 56

Lamb, J. 420

land 394, 397, 402

markets, women and 231

reform 201, 386, 557

rights, women and 230–31
titles 382

Landes, D. 413

Latin America 38, 203–7, 459, 469–86
capital flows to 72, 73, 75, 76, 78
gender inequality in 236
gender issues in 230
growth experience 470–83

law

and development 381–93
civil 387

commercial 388

common 387

contract 381

rule of 381–2, 386, 391

learning by doing 167, 168, 172, 384

Leontief model 345

Leontief, W. 344

Levine, R. 264, 270

Lewis, W. A. 36, 166, 227–8, 275, 286

liberalization 445, 581

in India 560–61

in MENA 531–2

in South Asia 545–55

licensing, technology 137, 139, 144

life expectancy 219–24

linear programming models 346–7

linkages 188–9

liquidity 77

List, F. 160

literacy rate 219–20, 223–4

loan-pushing 170

lock-in 167

Lorenz curve 182, 184, 463

Lucas, R. 171

lump sum transfers 6–7

Lutz, M. 23–4
macroeconomic policy 225
and gender distribution 241–9
Mahalanobis model 351
Mahalanobis, P. C. 561–2
Malaysia 54, 78–9
Mali 24
Malloch, T. 414
Mansfield, E. 140
market 41, 342, 354, 394, 401, 408–9, 446–8
failures 198–9, 285–6, 342
imperfections, capital 126–7
role in development 281–91
structure 20
and technological progress 282
in transition economies 594
Marshall Plan 150, 157–8
Masson, P. 293–304
Mattoo, A. 153, 154, 155
McCleary, R. 413, 420
McClelland, D. 411–12
McNamara, R. 151
Mexico 72, 400
financial crisis 72, 74–6, 82–3
migration from 120, 130
microfinance 240
middle class 201–2
Middle East and North Africa (MENA) 522–35
migration, determinants of 122
Milanovic, B. 165
military expenditure 434, 530
Millennium Development Goals 157, 237
Mishra, P. 130
modernization theory 407, 410, 416
monetary policy 91, 293–303
and development 293–4
and domestic nominal anchors 298–9
and gender equality 241–3
independence 294–5
see also stabilization policy
Montalvo, J. 273–4
moral hazard, in technology markets 140
Moran, T. 143
Morgan, W. 27
Morley, S. 179–92
mortality, child 107
mortality, infant 250–51, 255–6, 425, 435
mortality rate
adult 435
child (under 5) 219–20, 250–52, 255–7
multinational firms, see transnational corporations
mutual funds 72–9, 95
Myrdal, G. 417
natural disasters 105
natural resources 394, 395, 396, 397, 400–402, 404, 450, 461, 526
war and 426, 428
Ndulu, B. 490–521
need-for-achievement 411
neoclassical approach 167, 168
Nepal 425, 567
new growth theory 22, 167, 168, 171, 350
New International Economic Order (NIEO) 159
newly-industrialized countries (NICs) 174, 460
Nigeria 101
ethnic/racial inequality in 264
non-government organizations (NGOs) 99, 416, 434, 437
Norback, P.-J. 139
norms, social 202, 389
North Africa 552–3
North America 201
North, D. 371
North–South inequality 164–5
models 21, 145, 165–7
relations 160, 163–76
trade 167–9
Nurkse, R. 286
nutrition
effects of war on 434–5
in South Asia 556
nutritional status, of children 257
O’Connell, S. 490–521
official development assistance 99
Ohlin, B. 31
oil 523, 526
Okediji, T. 274
Olson, M. 399, 411
open access 394, 395, 398
openness 38–9, 450, 460
growth and 462
Organization for Petroleum Exporting Countries (OPEC) 158–9
outsourcing 561
outward-oriented development 40, 524

Pack, H. 141
Pakistan 564–5
patents 383, 391
Patnaik, P. 554–68
pharmaceuticals 383
planning 341–55, 594–5, 598
economy-wide 342–4
in India 561–2
models 341, 344–50
see also centrally planned economies
Platteau, J-P. 394–406
polarization 273
policies
and effects of foreign aid 108
to promote child welfare 257–60
towards economies in conflict 437–8
policy 343, see also under education,
fiscal, health, industrial, monetary
and trade policy
planning models and 353–4
political constraints, on privatization 365–6
political economy
of India 560
of migration policy 127
of South Asia 564
political instability 200, 529
political reform, in transition economies 596–7
political voice 217–18
politically relevant ethnic groups index 271–2
pollution 173
population 237, 428
portfolio flows, international 71
Posner, D. 271–2
poverty 213–26, 235, 417–18
absolute 216
in China 536, 541
foreign aid for reducing 104
gap 214
in India 563
line 214–17, 568
measurement of 214–15
in MENA 531
one dollar a day 216–17, 221
rate 219
relative 216
in sub-Saharan Africa 490, 508
targeted programs 325
in transition economies 585
two dollars a day 217, 221–2
and violent conflict 427
war and 428
Prebisch, R. 18, 21, 27, 158, 475, 478
Prebisch-Singer thesis, 18, 20–22, 25–6,
27, 158
primary commodities,
exporters of 25
income inelasticity of 20
price inelasticity of 20
principal agent problem, foreign aid and 111, 113–14
Pritchett, L. 165
privatization 208, 356–9, 364–7, 400,
401, 403, 533, 581
in Latin America 482
reasons for 360–64
in transition economies 592
productivity 446
growth 34, 141, 459
project appraisal 351–2
property rights 38, 201, 202, 267, 381,
389, 395, 401, 404, 405, 411
property, common 389, 396, 400, 401
protection
conditional on export promotion 570–71
import 38, 153, 201
import in OECD countries 153
psychological factors 410–12
public goods 282, 399
global 174
public investment 206
public sector 524
purchasing power parity 216
Putnam, R. 202, 264, 269, 413–14
Puttermann, L. 281–92
Quah, D. 164
Quesnay, F. 344
race 262–77
Radelet, S. 98–118
Raffier, K. 150–62
Ramachandran, V. 140
Randomized experiments 258–9
Ravallion, M. 220
Rayment, P. 445
reforms,
in China 589–90
in MENA 530–32
sequencing of, in transition economies 591–4
regional planning models 350
Reinhart, C. 88
religion 412–14, 417, 420, 423–4
remittances,
migrant 119, 127, 130–32, 523
migrant, to South Asia 563–4, 565
rent-seeking 33, 201, 202, 264–5, 525, 532
Research and Development 55, 138, 382–3
reverse engineering, 55, 137
Reynal-Querol, M. 273–4
Ricardo, D. 15–17, 21, 31
risk
indicators 77–8
political 529
Rivera-Batiz, F. 119–36
Rodríguez, F. 3–15
Rodrik, D. 152, 200–201, 382
Romeo, A. 140
Rostow, W.W. 286
Roy model 124
Russia 594
financial crisis 62, 74–6
Rwanda 265
Saavedra-Rivano, N. 170
Sachs, J. 98
Saggi, K. 137–49, 173
Sala-i-Martin, X. 164
Samuelson, P. 168
sanctions, economic 438, 529
Sapsford, D. 16–29
Sarkar, P. 19
saving 197
foreign aid and 106, 107, 108
private 324
relation with growth 313
school attendance 217
school enrollment
gender gap in 229
primary 217, 219, 251–3, 435
secondary 253, 255
Schumpeter, J. 407–8
Seguin, S. 236, 238
self-employment 230, 240
services, in India 563
shadow price 346, 351–2
Sierra Leone 225
Singapore 54, 463
Singer, H. 18–21, 23–4, 27, 170
Sjaastad, L. 122
skilled workers 119, 120–22, 124, 130, 133, 168, 171
slavery 167, 267, 270
Smarzynska, B. 140
Smith, A. 16, 165, 207, 381, 394
Social Accounting Matrices 347
social capital 202–3, 413–15
effects of war on 430
social conflict 200
social diversity index 274
social inclusion 217–18
Sokoloff, K. 201, 385
Solow growth model 43
South Africa 154
South America 201
South Asia 459, 554–68
child labor in 254
children's health in 251
poverty in 200
South East Asia 460
Soviet Union 341, 580–81
speculative attack 297, 302
spillovers
technology 55–6, 141–3, 145, 452, 462
wage 142
spiritual capital 414
Spraois, J. 27
Squire, L. 207, 208, 213–26
Sri Lanka 222, 224, 566
Srinivasan, T. N. 5–6
stabilization 285, 581
effect on women's employment 233
new Keynesian approach to 330
Index 611

policy 319–37
in transition economies 590–91
Stark, O. 126
state 427
capture 386
as coordinator of development 287
developmental 208, 385, 462, 524
in India 557, 562
intervention 390
in MENA 531–3
role in development 281–91
state-owned enterprises 356, 359–60, 524–5
in China 542, 544–5, 548, 589
in East Asia 572–3
reasons for establishing 359–60
Stern, N. 98, 152
Stewart, F. 423–41
Stiglitz, J. 98
stock markets 77, 89, 90, 95, 365, 384
effect of financial liberalization on, 81
Stolper-Samuelson theorem 7
Streeten, P. 56–7
structural adjustment 112, 152, 288, 319–37, 356, 458
effect on women’s employment 233
and income distribution 335–6
in MENA 530–32
structuralist CGE models 347–8
subjective well-being 194, 207–9
Subramanian, A. 153, 154, 155
subsidies
on FDI 50, 54
in MENA 532
to Northern farmers 154
surplus labor 286, 330
Taiwan 39, 40
Tanzania 203
tariffs, import 32, 50
tax buoyancy 310
tax-GDP ratio 306
tax reform 313–16
Taylor, J. E. 126
Taylor, L. 166
teacher absenteeism 259
technological capability 172
technological change 20–21, 56, 134, 403, 447–8
technology 53, 385
diffusion 145
transfer 36, 53–5, 137–49, 171–2, 384, 449, 462
transfer, foreign aid and 106
terms of trade
agriculture-industry 559
between primary products and manufactured goods 17–20
international 16–29, 20–21, 451, 461
net barter 17, 26
North–South 18–19, 31, 170
volatility of 23–4
Thompson, E. P. 391
Throsby, D. 411, 418
total factor productivity 141, 142, 384
growth of 33, 206
township and village enterprises 588, 593–4
trade, see international trade
trade, free 39
Trade-related Intellectual Property Rights (TRIPs) 154, 383
Trade-related Investment Measures (TRIMs) 154
transaction costs 389, 394, 397, 401–3, 405
transition economies 579–600
capital flows to 73, 76–8
Triplett, R. E. 262–77
trust 202
Turkey 524–5, 530–31
underemployment 233, 242
unemployment 233, 242, 459
in MENA 528
in South Asia 556
in transition economies 584
uneven development 22, 166–8, 169, 171
United Nations 157
United Nations Conference on Trade and Development (UNCTAD) 158–9, 453
United Nations Development Programme (UNDP) 158, 218, 263
unpaid work, women and 230, 231, 232, 233, 234
urban bias 499
Uruguay round 154, 155, 172
USA 154
legal system in 387
migration to 130
Value at Risk models 60–61
values 409, 412–13, 418, 419
Venezuela, technology spillovers in 142
Vietnam, well-being of children in 257
violence 203
violent conflict 423–41
causes of 423–9
consequences of 429–37
in South Asia 567–8
Voluntary Export Restrictions 154
von Neumann model 345
wage inequality, gender 232
Walrasian CGE models 347
war 225, 423–41, 529–30, see also violent conflict
foreign aid and 107
Washington Consensus 37, 39–40, 42, 44, 483–4, 577, 581–2
water, in MENA 526
Weber, M. 386, 407
Weil, D. 413
Weingast, B. 381
Wilber, C. 407–22
Williams, E. 266–70
Williamson, J., 44
Williamson, J. G. 125–6
Williamson, O. 415–16
women, missing 234–5
Women-in-Development literature 228, 229, 231
Wood, A. 168
working time, for men and women 234
World Bank 27, 38, 104, 109, 111–14, 158, 216, 243, 319–21, 414, 457–8, 530, see also International Bank for Reconstruction and Development (IBRD)
interpretation of East Asian miracles 569–70, 572
Revised Minimum Standard Model 321–5
World Food Program 158
World Health Organization 383
World Trade Organization (WTO) 11, 48, 150, 153–6, 158, 159, 172–3, 383
Yang, M. 27