

1. The empirical evaluation of regulatory policy reforms in network industries: some methodological issues

Massimo Florio*

1.1 INTRODUCTION

Economists are often confronted with the task of providing assessments of policy reforms. In fact, a very large share of the contemporary economic literature deals with the evaluation of either *ex post* or *ex ante* reforms just proposed or actually implemented by governments. These reforms can be represented by the specific change of the level of a given 'signal'. A signal (Drèze and Stern, 1990) is a microeconomic variable which 'directly' affects the behaviour and welfare of economic agents. Examples are the rate of an indirect tax, the level of a price-cap on tariffs of utilities, the number of years of compulsory primary education or the permitted level of carbon emissions. This is the context of the well-established literature originated by the theory of policy reform, particularly of taxation. For example, Ahmad and Stern (1984, p.259) summarize their study of indirect taxation reform in India in this way:

Given a set of value judgements, an initial state, and a model of the economy, one can ask whether some feasible tax change would increase welfare. We do this by defining the marginal cost in terms of welfare of raising an extra rupee from the *i*th good. The inverse optimum problem is the calculation of non-negative welfare weights on households which imply that the initial state is optimum. If no such welfare weights exist, then a Pareto improvement is possible. We illustrate the concepts and results using data from the Indian economy for 1979–1980. Directions of tax reform for a number of specific social welfare functions and for Pareto improvements are presented.

To understand the theory of reform this way has several founding fathers in public economics, starting with Ramsey (1927), the unpublished Samuelson (1951) memorandum to the US Treasury, Diamond and Mirrlees (1971), Guesnerie (1977, 1998), Drèze and Stern (1990), and for a

review see Boadway (2012). In spite of its taxation origin, this theoretical stream had the ambition to cover any specific reform in an applied welfare economics perspective (see Drèze and Stern, 1990; Florio, 2014 for marginal reforms in cost-benefit analysis (CBA), and Johansson and Kriström, 2015 for the evaluation of large-scale projects and policy changes). In this context, 'marginal' means that the reform is small, in the technical sense that while it may have general equilibrium effects, it does not change the trajectory of the economy (this is equivalent to the 'small project' assumption in cost-benefit analysis).

However, reforms can often take a rather different shape from the change of a microeconomic signal or macroeconomic instrument. For example, product market reforms are a broad change of public policy arrangements hindering competition in some industries. This 'mesoeconomic' change will only have an indirect impact on agents' utility. For an early discussion of mesoeconomics see Ng (1986). Here I shall refer more loosely to a context where it is the change of an institutional setting that should be evaluated, neither captured by a microeconomic nor macroeconomic variable. Examples are the privatization of public enterprises, unbundling in network industries or the promotion of competition in a formerly monopolistic market. These changes will be embodied in legislation, through several bills or regulatory decisions which are more or less closely linked together by common objectives of a general nature. For policy makers a reform is usually a more comprehensive package than the change of a specific signal (such as a statutory tax rate, a price-cap, the amount of a good provided by a government agency). The reform, in this wider meaning, is in fact a mechanism aimed at promoting a cascade of several punctual changes in a certain desired direction.

In this chapter, I discuss some methodological issues of the latter evaluation, the assessment of changing a 'policy framework', focusing on network industries, drawing from a very selective and purely illustrative review of some literature and from my own experience as an applied economist. My main concern here is to suggest an analytical framework and point out possible methodological errors and pitfalls that may bias the research design and the results when something more comprehensive than the change of a specific signal is at stake. I shall use the concept of 'Policy Framework Reform' (PFR) when the change concerns a set of rules more than the change of a specific signal in the Ramsey-Samuelson-Guesnerie tradition.

The topic is discussed as follows by Parker and Kirkpatrick (2012, p. 7), for example, in their review of the literature:

OECD member countries have been engaged with regulatory reform and improving regulatory processes for a decade or more, in the expectation that there will be significant improvements in economic welfare outcomes. But in the absence of clarity about how and why the changes should lead to improvements, policy failures are likely. The critical public policy challenge is to ensure that the expected economic benefits from regulatory changes are both achieved and outweigh any economic costs imposed. This requires firm evidence on how different policies perform. Evidence on the outcomes of regulatory policies should help policymakers design regulatory measures that work better.

The perspective adopted in this chapter is that of a concise discussion of some frequent pitfalls in the interpretation of results, when the traditional theory of reform is not applicable, or not appropriate, for the empirical analysis of a given context and one needs to rely on reform indicator variables.

The structure of the chapter as follows: Section 1.2 discusses the problem of how to properly define the object of analysis, clarifying the distinction between two different ways to consider policy reforms; Section 1.3 discusses issues in measurement of the change of variables representing the PFR; Section 1.4 presents a taxonomy of possible errors throughout this procedure; Section 1.5 turns to the selection of proxies for welfare changes, which is a necessary step for any evaluation of policy reforms; a generic empirical modelling approach is presented in Section 1.6; Section 1.7 discusses possible misinterpretations of what the empirics of policy reform evaluation can achieve; Section 1.8 concludes.

1.2 DEFINING THE OBJECT OF ANALYSIS

Properly defining the object of analysis in the policy reform context is crucial, and a less trivial step than often understood. In theoretical models, such as those reviewed in Guesnerie (1998) or developed for CBA of projects and policies by Drèze and Stern (1990), things may be mathematically and informationally demanding, but conceptually unambiguous. There is a given constellation of signals, that is, variables that potentially influence the behaviour of agents, households or firms. Signals may include prices, quantity rations, taxes, ownership shares and so on, and one wants to study the welfare effects of marginal changes of such signals. In some models, one may want to study non-marginal effects, which implies different techniques.

In this form, the problem is simple: there are functional relations linking signals, constraints, agents' reactions and government objectives to be maximized, and one studies how changes in signals ultimately have an

impact on objectives, possibly through a series of chain effects. Basically, if $W(s)$ is a social welfare function and s is a vector of signals, the theory revolves around the estimation of marginal social values of the form $dW(s)/ds$. This can be done by looking for solutions around the optimum of a program (a second-best way to understand the problem) or starting from the existing set of signals and gradually moving on (see Johansson and Kriström, 2015 for a discussion of the two different applied welfare economics strategies).

To give an example, when the vector s is a set of indirect tax rates, this is how a revenue neutral marginal tax change, in the context of discussing environmental policies, is introduced by Mayeres and Proost (2001, pp. 346–7), where t_m, t_k are indirect taxes, W is the social welfare function, B is exogenous government budget and MCF is the marginal cost of public funds:

We want to evaluate whether a revenue neutral marginal policy reform is welfare improving or not when starting from an arbitrary tax system and from an arbitrary level of public abatement. In a first instance the analysis concentrates on marginal reforms of the tax system . . . The effect on welfare of a revenue neutral tax change which consists of increasing t_m and reducing t_k is given by:

$$dW = \frac{\partial W}{\partial t_k} dt_k + \frac{\partial W}{\partial t_m} dt_m \quad \text{with} \quad dB = 1 = \frac{\partial B}{\partial t_m} dt_m = -\frac{\partial B}{\partial t_k} dt_k$$

Defining the marginal cost in terms of social welfare of raising one additional unit of government revenue via the tax on commodity m as:

$$MCF_m = -(\partial W / \partial t_m) / (\partial B / \partial t_m)$$

we find

$$dW \geq 0 \Leftrightarrow MCF_m \leq MCF_k$$

So welfare is increased (reduced) when the tax with the highest MCF is reduced (increased) and when simultaneously the tax with the lowest MCF is raised (reduced).

I shall come back later on this way of looking at the evaluation problem, but here I just claim that in the perspective of applied economists the main limitation in implementing this scheme is that, in the real world, policy reforms are often not easily tracked as a (marginal) change in a signal.

Frequently, economic policy reforms do not take the simple aspect of changing a tax rate or a regulated price, but are embodied in legal acts providing possible mechanisms that may lead to such changes under certain circumstances. Clearly, in the case of adding one year to compulsory education,

or lowering by $x\%$ an emission standard, marginal reforms can be analysed without major definitional issues, in the framework of applied welfare economics, such as in the Drèze and Stern (1990) theory discussed in Florio (2014). The general equilibrium frame can be a second-best economy or departures from the existing state of affairs (Johansson and Kriström, 2015), or partial equilibrium (Boadway and Bruce, 1984; Boardman et al., 2016).

Even so, one has to be careful when going from the theory of reform to empirics. For example, suppose a government adds one year to compulsory education, but in fact does not recruit more teachers. This implies that the existing stock of teachers has to work more and if their effort has diminishing returns in terms of students' achievements, the marginal net welfare effect of the reform would be misread without taking into account a quality measure. The quality of the service before-after the reform is assumed to be homogeneous, but it is not. Thus, we would need a 'quality adjusted additional year' as the proper unit of analysis. In some cases, the operative definition of the object of analysis may be problematic in practice – for example, see the notion of Quality Adjusted Life Years (QALY) in health economics as discussed in Johansson and Kriström (2015, p. 208) or the precise definition of an emission standard, but it is not conceptually difficult per se.

However, such policy reforms as the privatization of state-owned enterprises (SOEs), or the unbundling of vertically integrated network services, pose a definition problem: these are generic labels for a range of potential changes in the current circumstances, and in a deep sense such labels are not well defined in the perspective of applied welfare economics. Unbundling gas networks is a set of legislative and regulatory changes, and does not enter explicitly in the utility functions of agents. It can influence their behaviour in many ways. How to evaluate such a reform?

Applied welfare economics is about a quantitative assessment of the social effects of a change of circumstances. If the economic policy reform is described in a form that does not lead to empirical analysis in the specific form of a quantitative evaluation of effects, it may be an interesting object of study for economic historians, or for other social scientists, but not *sensu stricto* for applied economists. I am in favour of interdisciplinary studies in this area. For example, I think that the best approach to the study of privatization in principle should be a comparative political economy, which would blend together the perspectives of history, political sciences and applied economics (a review may be found in Obinger et al., 2016). However, without a definition of privatization or liberalization in the form of measurable variables, the contribution of an applied economist to the possible interdisciplinary policy evaluation would be severely constrained.

Thus, in the rest of this section, I discuss the situation whereby a policy reform cannot be treated in the Ramsey-Samuelson-Guesnerie tradition (RSG from here on) as the (marginal) change of the supply of a good by the public sector or of a signal. It is important to be precise in making the distinction between the RSG 'theory of reform' framework and what is discussed below. A simple example is privatization. In the general theory of CBA one may consider ownership of firms as a signal, taking the possible measurable dimension of a share of equity capital owned by government or households. Thus, a marginal change of such government ownership can be interpreted as a reform, and its welfare effect can be analysed in the same way as one would analyse the change of a price. Drèze and Stern (1990) show how this analysis of privatization may work in a general equilibrium, and there are also some loosely related ideas in Bös (1991) and Jones et al. (1990). This analysis is far from being easy, for example, a marginal reform of selling just one share of a SOE by a government to a private shareholder may or may not be significant according to the fact that such share insures a change of control or not: there are often these discontinuities in the value of signals which make the empirics less straightforward than one would expect (see, for example, Clò et al., 2015).

I turn now, however, to the concept I am more interested in here: a policy reform that is seen as a set of legislative or regulatory packages which are not immediately translated into a change of signal. To simplify, suppose that a government is able to pass a legislation that provides for privatization as a broad framework requiring that in certain industries SOEs' control must be relinquished to investors other than government itself and instructs managers of SOEs to implement such directives. This was the case of British privatization policy (Florio, 2004) – in fact, a set of different acts. When observed at this level, the object of analysis is not the marginal change of ownership of shares in one specific firm, a micro-economic change as in the above discussion, but rather a 'mesoeconomic' change: not as wide as would be described in terms of macroeconomic variables, but not so small to refer to a punctual transaction in the market for corporate control in relation to one firm.

Here the government announces and implements, possibly over some years, an entire set of changes. These can include, for example, the corporatization of SOEs, then listing them in stock markets, then seasoned selling of tranches of equity either by floatation or by deals with some domestic or international investors. Such deals may or may not be pre-determined in the legislation, when the control of a number of firms is actually passed from government to third parties.

A second example, particularly important for network industries, is an act providing for unbundling of some facilities which are the natural

monopoly core of the industry. It is very uncommon that such legislation or regulation takes the precise form that is suitable for the RSG theory of reform style of analysis. For instance, the legislation may say that a railway firm should divest tracks from the train operating company; however, such divestiture can take, as a preliminary step, the form of accounting separation, then of functional separation and, finally, of full divestiture of ownership and control (which goes beyond legal separation of ownership). There are many nuances around the implementation of such reforms, as they are not easily described as changes of signals, but are changes of frameworks which are then expected to eventually lead to changes in signals.

The two examples above, selling one share of a SOE by the government, or passing an unbundling legislation, are just illustrations of a more general issue in defining and hence in tracking policy reforms. To distinguish them from changes of signals, I would use the term PFRs. In some interesting cases, such changes are related to an even more comprehensive shift, a change of the policy paradigm, that is, of the theory (sometimes of the ideology) supporting in general the way a government thinks about its policy.

To summarize this discussion, an act passing a quality adjusted increase in compulsory primary education is a (marginal) reform affecting a signal; one or more acts changing the incentives for households to take advantage of primary education is a reform affecting a framework which in turn will change some signals; and a shift away from the view that government is ultimately responsible for the provision of primary education is a change of paradigm in the role of the welfare state.

I will not discuss changes of paradigms below because it is uncommon that an applied welfare economist is able to empirically analyse them in quantitative terms, and because this is the level where the interdisciplinary approach would be most useful. Instead, I claim that the 'mesoeconomic' changes of policy framework lend themselves to empirical analysis, albeit with due caution.

1.3 MEASURING CHANGES OF POLICY FRAMEWORKS

While a change of a signal requires only some ingenuity to appropriately define the unit of measurement of the appropriate variable (a tax rate, a QALY, a price, number of passengers per kilometre provided and so on) the empirics of PFR need the creation of artificial indicator variables, for which usually no natural units of account exist.

An example of this is the Energy, Transport and Communications

Regulation (ETCR) database created and maintained by the Organisation for Economic Co-operation and Development (OECD),¹ which for several industries in the energy, transport and communication sectors, tracks reforms in the European Union Member States, in the form of yearly scores given by experts (Figure 1.1).

I dwell here on this specific example because it has been used by several researchers for the empirical analysis of market reforms (there are, however, other examples of policy indicators released by international organizations – including the European Commission (EC), the European Bank for Reconstruction and Development (EBRD), the World Bank, or built by think-tanks and individual researchers.²

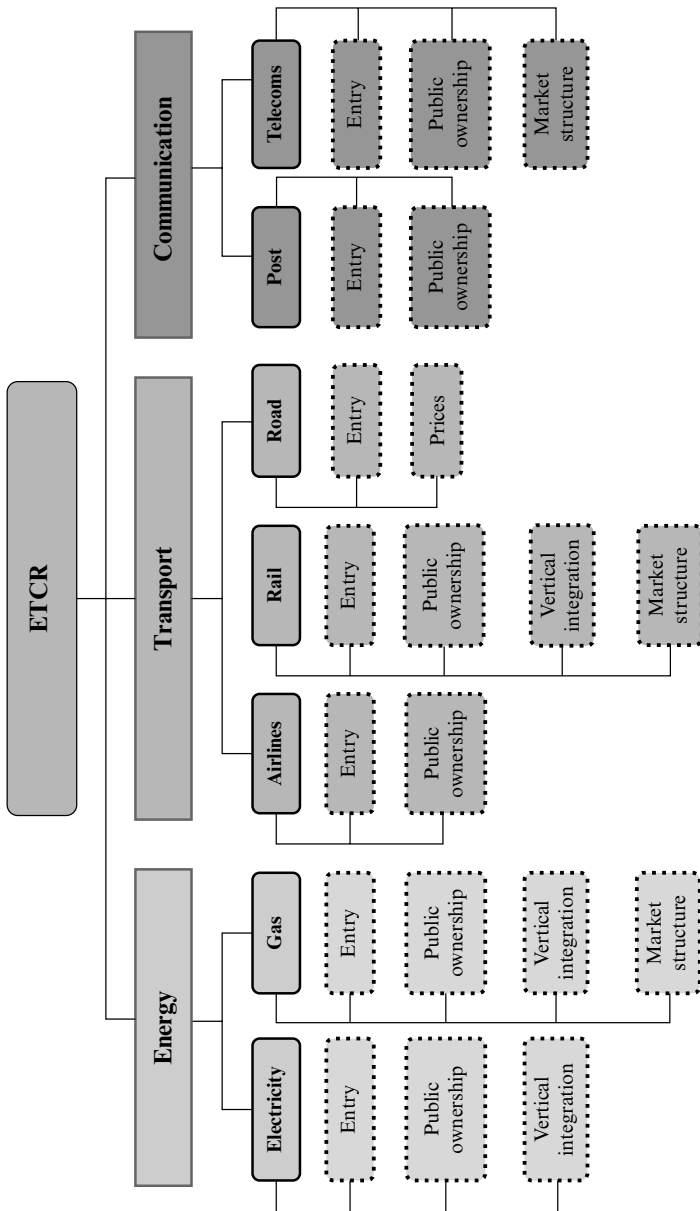
At a more elementary level one can think of ETCR as an exercise of assigning normalized scores by experts along certain dimensions of change of a PFR, sometimes based on surveys of policy makers, but also on reviews of legislation and even news in the media.

When scores are given consistently by an expert team over years and across countries, one could track reforms in several ways, including by figures such as Figure 1.2, where different PFR trends are visually displayed for the natural gas industry.

In the ETCR case there may be different layers of analysis of PFR. The progress of reforms in a specific country and industry implies that one may need to aggregate different levels of the scoring system. Table 1.1 shows an example of such layers of analysis from the ETCR.

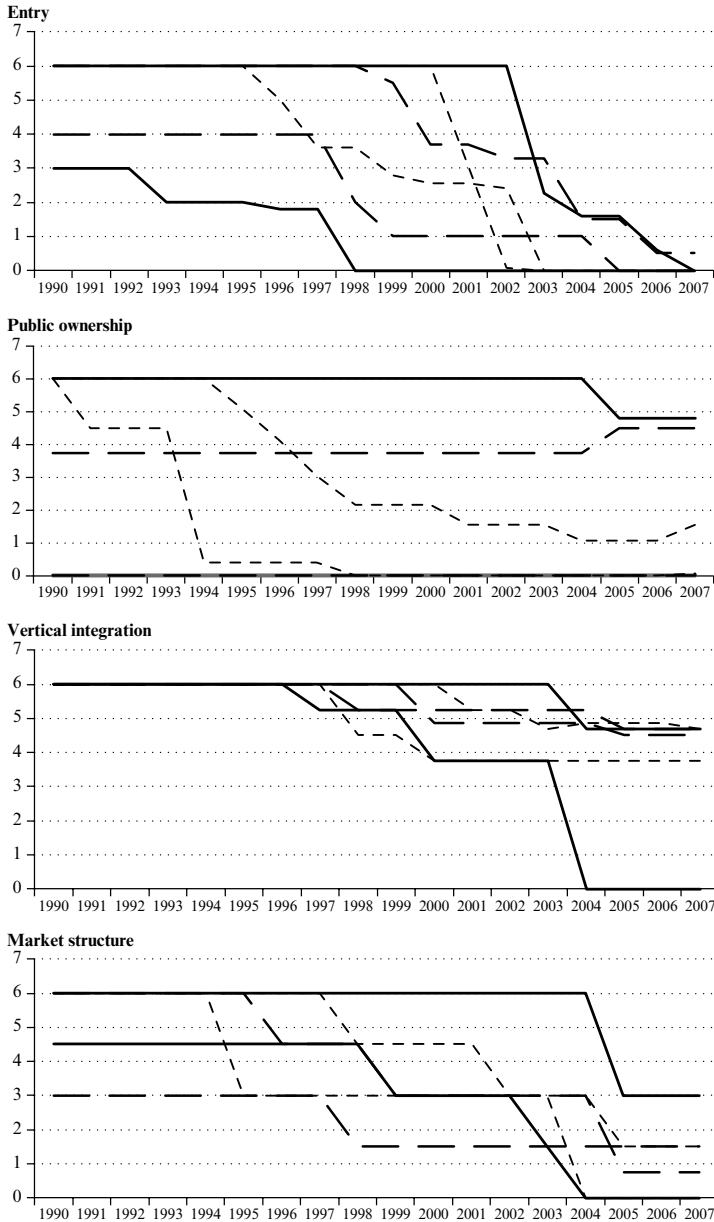
The methodological issues involved in building such scoring systems are non-trivial. The analyst should, first of all, collapse the legislation or regulatory decisions supporting the reforms along some simplified dimensions, which should be sufficiently recurrent but also variable across countries and time to allow for a consistent measure in different circumstances. Moreover, as the above ETCR example shows, there may be a mixture of qualitative and quantitative variables involved in the attempted measurement of the reforms. However, one should want to normalize the scoring system and this typically leads to transforming continuous variables (for example, market share of the incumbent electricity firm in the generation segment of the industry) into discontinuous scores, defining brackets (such as 0 score meaning that the market share is below $x\%$). For example, in the words of Parker and Kikpatrick (2012, p. 15):

Multiple regression analysis is central . . . for modelling and measuring the relationship between one dependent variable (e.g. GDP growth) and other, so-called independent or explanatory, variables that can be expected to cause changes in the dependent variable (e.g. labour input, investment and technology). An additional explanatory variable reflecting the nature of regulation or a regulatory change can be added to the independent variables, either as a separate variable



Source: OECD Economics Department Working Paper (2006), No. 550, 'Product Market Regulation of Non-manufacturing Sectors in OECD Countries: Measurement and Highlights'.

Figure 1.1 Structure of the ETCR indicators



Source: Florio (2013), *Network Industries and Social Welfare*, Oxford University Press.

Figure 1.2 PFR trends for the natural gas industry

Table 1.1 Sectoral indicator of regulatory reform: gas

	Weights by theme (b_j)	Question weights (c_k)	Coding of data		
<i>Entry regulation:</i>	1/4				
			Regulated TPA	Negotiated TPA	No TPA
How are the terms and conditions of third party access (TPA) to the gas transmission grid determined?		1/3	0	3	6
What percentage of the rental market is open to consumer choice?		1/3	(1-% of market open to choice/100)*6		
			No, free entry in all markets	Yes, in some market	Yes, in all markets
Do national, state or provincial laws or other regulations restrict the number of competitors allowed to operate a business in at least some market in the sector: gas production/import?		1/3	0	3	6
<i>Public ownership:</i>	1/4				
			None	Between 0 and 100%	100%
What percentage of shares in the largest firm in the gas production/import sector is owned by the government?		1/3	0	3	6
What percentage of shares in the largest firm in the gas transmission sector is owned by the government?		1/3	0	3	6
What percentage of shares in the largest firm in the gas distribution sector is owned by the government?		1/3	0	3	6

Table 1.1 (continued)

	Weights by theme (b_j)	Question weights (c_k)	Coding of data		
<i>Vertical integration:</i>	1/4		Ownership separation	Legal/ Accounting	Integrated
What is the degree of vertical separation between gas production/ import and the other sector of the industry?		1/2	0	3	6
What is the degree of vertical separation between gas supply and the other sector of the industry?		3/10	0	3	6
Is gas distribution vertically separate from gas supply?		1/5	0	3	6
<i>Market structure:</i>	1/4		< 50%	Between 50% and 90%	> 90%
What is the market share of the largest company in the gas production/import industry?		1/3	0	3	6
What is the market share of the largest company in the gas transmission industry?		1/3	0	3	6
What is the market share of the largest company in the gas supply industry?		1/3	0	3	6
Country scores (0–6)			$\sum_j b_j \sum_k c_k \text{answer}_{jk}$		

Source: Adapted from OECD Economics Department Working Paper No. 530 (2006), 'Product Market Regulation of Non-manufacturing Sectors in OECD Countries: Measurement and Highlights'.

or as an interactive variable (e.g. interacting with the level of investment). This may take the form of a – dummy or – binary form, with a value of 1 when there is good regulation or a regulatory improvement and zero in other cases or there could be a scaling between 0 and 1.

The ETCR scoring has a range 0–6, with 0 indicating the 'completed' reform and 6 the 'no reform'. The binary form is the extreme form of

apparently a loss of information, but the advantage is the opportunity to aggregate scores deriving from originally quantitative variables with scores deriving from qualitative variables, where the latter are typically discrete. Such discretization defines an ordering, and it is a delicate issue that is discussed further below.

Finally, the aggregation from the lowest to the highest level of scoring implies using weights and there are no natural ways to define such weights in most cases. For example, there is no theoretical justification for assigning equal 1/4 weights to each of the entry-privatization-unbundling-liberalization scores in ETCR (see Table 1.1). This leads to a discussion of potential errors in this area.

1.4 ERRORS IN MEASURING POLICY FRAMEWORK REFORMS

The above discussion suggests that several types of errors may occur in measuring PFR. We can classify these errors into five types:

1. Conceptual errors in defining the relevant reform dimensions.
2. Errors in discretization of continuous variables.
3. Errors in defining metrics and orderings.
4. Measurement errors deriving from misinterpretation of the information.
5. Scoring aggregation errors.

I briefly elaborate on this classification, with some examples, and how to manage the concerns arising from uncertainty surrounding the PFR data.

‘Conceptual errors’ are encountered when the team or individual researcher(s) providing the scoring system of the PFR defines a variable, either quantitative or qualitative, that per se does not work in the same way as a signal in the context of marginal reform theory. In the RSG context, there are natural units and natural orderings, such as a given increase in the provision of a public service by the government production plan. In network industries, ‘unbundling’ can be considered as a categorical variable with an ordering going from full vertical integration to accounting, functional, legal, ownership unbundling. This may be a relatively crude way to define a sequence of steps, but it is not illogical per se if one considers that a new paradigm promotes a PFR in a direction that goes from full vertical integration to full vertical disintegration of an industry. Clearly, each step should be sufficiently well described and unambiguous, but this is a different issue, which I discuss below. However, if one defines as a PFR

indicator what in fact is possibly an outcome variable, this may lead to some interpretation problems.

For example, the market share of the largest electricity generator in a country may or may not be legitimate policy reform information: it can be interpreted as PFR when the incumbent has been forced by legislation or regulation to divest generation capacity, thus the observed share is part of the policy, but this is not the case if the observed change is a response to different circumstances, for example, a technological shock decreasing the optimal size of the production for any firm, including obviously the incumbent one. If the latter shock is an omitted variable in the subsequent modelling of the welfare effects of the PFR, there is the risk of misunderstanding some effects as if they were caused by the reform, while they have a different origin. In some cases, some control variables, if available, can cure the problem, but if not available, then the conceptual error of attributing a possibly non-policy related change to a PFR may cause a serious interpretation problem. As a consequence, a prudent applied economist using secondary PFR information, that is, data created by somebody else, should double check if there is the above-mentioned risk in the data he is going to use.

Turning to type-2 errors, any scoring system is intrinsically based on some assumptions about how to build indicators from initial information, and this poses specific problems when the original data are in continuous form. 'Discretization' may or may not have alternatives, but in most cases PFR data are expressed on a multidimensional scale because of the complexity of programs, hence they should be normalized for aggregation, as discussed below. Discretization means that all the variables, whatever the nature of the initial information, are expressed in the form of discontinuous scores, as in the ETCR example. Mistakes in discretization may arise when the distribution of the original data are not such that it is obvious how to define the brackets that will be translated into scores. This problem is not specific to policy reforms and is widely discussed in econometrics. If the original data are not available, concerns arise about discretization; in some cases, it may be better to collapse the scores in a dichotomic variable, or aggregate them in wider brackets: this procedure indeed shrinks statistical variability, but also may reduce the discretization error.

The above-mentioned problem may or may not lead or coincide with 'mistakes in defining orderings and metrics' (errors of type-3). A possible problem arises when there are different reform policy options and it is not entirely clear whether each of them can be considered as 'superior' or 'inferior' to another along an ideal path leading from the less to the most ambitious or advanced. The second problem arises when the empirical analysis uses a categorical variable in such a way that de facto ordinal

information is treated as cardinal. While the ordering per se may be right, the metric involved in the implicit cardinalization may be wrong, for example, in linear regressions in which an indicator taking different values is used as the variable of interest. This happens because obviously ‘third’ is not the same as ‘three’, but in Ordinary Least Squares (OLS) or other regressions the latter may be taken as the former. Thus, inadvertently, the scoring process may introduce a bias in subsequent analysis. In these cases, again, it may be better sometimes to shrink the information to a dichotomous variable, thus dramatically reducing the error (but unfortunately also variability).

‘Measurement errors’ (type-4) affect any empirical analysis, but in the context of the study of PFR they are potentially particularly frequent because the source of information may include non-standard sources, for example, reports in the media, parliamentary acts and regulatory documents, in some cases interviews to policy makers or other stakeholders. Occasionally, it is possible to double check the scoring attributed to a PFR in a particular country and sector with alternative sources, and if this is possible a double check on data samples is advisable. As usual, irregular patterns may point to these types of errors and should be removed.

Finally, and most important, the ‘aggregation error’ (type-5) is pervasive in this context because there is usually no natural way to decide how to ‘sum’, for example, liberalization scores with unbundling scores. In many cases the existing PFR datasets simply count the items and aggregate them linearly, with equal weights. This error is potentially serious as it makes the aggregate reform score dependent on arbitrary coefficients, which are neither supported by theory nor estimation, but simply plugged into the system for computational ease. In some cases, the problem can be circumvented either by a Principal Component Analysis (PCA) of the individual items or entirely avoiding aggregate indicators and introducing the individual variables in the model, where one of them will be interpreted as the variable of interest and the remaining as additional controls

It is important to acknowledge that in spite of all these problems, in most cases there are limited alternatives to the use of such coded information to perform a quantitative analysis of the effects of reforms. This discussion should be seen more as a set of preliminary tests on the available PFR data, rather than generic arguments not to use them and to revert to a more traditional marginal reform approach, which in turn may be unfeasible or inappropriate.

To sum up, a careful interpretation of the data is always needed in this context, and applied economists working in this area should not take for granted that the information they use is meaningful just because it is provided by authoritative sources, such as international institutions. With this

caveat, empirical modelling may be more informative than a simple narrative, or of individual case histories of reforms, even if these may have their place in a comprehensive policy evaluation.

1.5 DEFINING AND MEASURING THE WELFARE EFFECTS

Having defined and measured the policy variable of interest, the other crucial ingredient in the evaluation of reforms is the identification of the dependent variable(s). The choice, which actually specifies the research question, should be firmly based on the analyst's perception of the relevant Social Welfare Function (SWF).

Exactly because the adoption of a SWF is somewhat controversial, one needs to be clear about the perspective of the research question. First, a policy evaluation always takes a normative perspective. A positive analysis of policy reforms is instead often embodied in wide 'political economy' questions such as: 'Are social-democratic governments against privatization?', 'To what extent are regulators independent in less developed economies?', 'Is government debt causing or hindering liberalization policies?' and so on. The core of positive empirical research on policy reforms is the understanding of the historical, political, social, financial and economic drivers of policy adoption, of its success or failure. The core of the normative perspective, however, is the objective analysis of the social welfare effects of such reforms. Examples of research questions are: 'Is privatization beneficial to the poor?', 'Does market opening to foreign investment destabilize the income of small farmers?'

While in the same paper one can try to combine the two perspectives, in this section I focus on the issue that any welfare analysis explicitly or implicitly assumes a SWF (for example, gross domestic product or GDP, which, however, is based on some special assumptions) and aims to explore the impact of a reform on it, at least partially. Any country, or portion of it, is ultimately a society, that is, a set of individuals that can be lumped in convenient aggregates: social classes (shareholders and workers), income groups (the rich and the poor), age groups (the elderly and the youth), regional sub-groups (North and South). Such partitions may overlap, as when welfare effects of a reform on consumers, taxpayers, shareholders, employees are simultaneously considered. In a general equilibrium context, simultaneity of effects is embedded in the structure of the analysis, as a 'local' reform can have 'distant' effects across social groups, regions and so on.

Having in mind a SWF does not imply that all the welfare effects should

be tracked, in many cases this could be a daunting task. Nevertheless, even if the study focuses on a specific segment of the society (for example, pensioners and rebalancing of their electricity tariffs), it is often implicit that in the SWF assumed by the researcher a particular group enters in the overall computation of effects with a certain weight. Sometimes such welfare weights are implicit, for example, when in a partial equilibrium frame, changes of consumer surplus and producer surplus are simply summed after a reform that increases the tariff of public service. There is nothing wrong in doing so, but only provided that this is compatible with either a very specific SWF or the welfare level of all the consumers and shareholders should be the same. A Bergson-Samuelson SWF compatible with simply summing consumer and producer surplus must have identical welfare weights (Florio, 2014) or for a larger class of SWF, one may assume that the income or other welfare related variables are more or less similar for the owner of the utility and his client. Both assumptions do not command a wide consensus, however, as just using the textbook sum of surpluses in policy evaluation reveals that the analyst is putting issues of equity under the carpet.

The issue is relatively less important when only a very narrowly defined social group is considered, in a partial equilibrium frame; however, even an analysis of effects on one relatively broad group, such as pensioners, may raise the question of differences in welfare level within the group, and it should be made explicit what welfare weights are assigned by the analyst to high and low income pensioners and so on. One may experiment with different welfare weights in this context, in a sort of sensitivity analysis of the conclusions to the SWF assumptions. Or at least qualitatively discuss the robustness of the normative conclusions to such assumptions.

Having said this, the outcome variable in many empirical analyses of reform is not explicitly the welfare change, but an empirical proxy of it. Thus, in the literature on economic policy reforms it is quite rare that one is able to read the results in the form of a change in a well-defined welfare measure, which necessarily requires at least the estimation of (compensated) demand curves or other appropriate functions, supply curves, quantity constraints, local equilibriums. Such results, if available, would then be expressed typically in a numeraire ('welfare' euro) or in the form of percentage change (before-after or without-with the reform, controlling for confounding factors). More frequently, the results are expressed in empirical variables, which are directly observable by the researcher, such as, for example, consumer or producer prices, profits, household income, quantity consumed and quality of service. This procedure is economical and sensible if there is at least some discussion of the assumptions linking these observables to the unobserved welfare change of the target social group(s).

In many cases, consumer prices are the most important signal relevant for a welfare analysis, as it can be easily demonstrated (Price and Hancock, 1998; Florio, 2004). Though important to use micro data on prices paid, these are rarely available, and one should be aware of the possible aggregation error across types of consumers, as this can be large.

While it may seem a less orthodox approach to welfare analysis, the use of consumers' satisfaction survey data adds an interesting dimension to potential proxies, that is, the perceived welfare change. For example, some Eurobarometer³ waves allow the individual responses to questions on the subjective appreciation of the fairness of prices, of quality of service, of accessibility to be recorded. Other surveys such as SILC⁴ and ECHP⁵ (Poggi and Florio, 2010) were particularly useful in the study of utilities poverty, and occasionally regulators commission specific surveys which can feed in research on perception of the welfare impact of PFR.

Interestingly, one can combine micro data originating from a survey with country data, for example, to double check if the average 'objective' information on outcome variables, such as unplanned interruption of electricity services, match with subjective data.

The relationship between subjective and objective information can be framed in terms of behavioural welfare economics, a broader topic that cannot be discussed here, but amounts to making appropriate hypotheses about the observable satisfaction (or happiness) of agents and their unobserved utility. This is a controversial topic as according to some behavioural welfare economists any divergence between happiness and utility can be traced as deviations from rationality, while others would question the notion of rationality itself in this context. This is a subtle controversy with wide epistemological implications for economics, but an applied economist could take a pragmatic approach, and simply focus on convergence or divergence of objective and subjective data, and try to understand the drivers of such divergence. In many cases, there are simple explanations, such as aggregation errors when average objective data on an outcome variable are compared with subjective micro data. Clearly, the question, if such divergence arises, is to rule out these simple explanations, before resorting to the most complex ones, pointing to cognitive issues. Of course, if one has the actual information on prices paid by a sample of individual users of a public service before and after a reform, or with and without a reform, and for exactly the same individuals satisfaction data are available, the usual hypothesis is that price decrease would be correlated to higher satisfaction, after controlling for any confounding factors. If the hypothesis is rejected, there are two possibilities: either there are omitted variables in the controls, or in some sense there are systematic measurement errors either in the subjective or the objective data or in both. Some of these issues are

discussed below. However, unfortunately, it is rare that one is able to take advantage of subjective and objective data simultaneously on the same samples of users, and the two sources of information can be seen as heterogeneous but complementary, and combined case by case with due caution.

1.6 MODELLING THE RELATION BETWEEN REFORMS AND OUTCOMES

The basic model for policy reform (of the PFR type) empirical evaluation often takes the generic form:

$$Y = f(R, D, S, Z) \quad (1.1)$$

where Y is the outcome dependent variable, R is a set of indicator variables which are proxies for the policy reform framework, D are demand-side controls, S are supply-side controls, and Z are additional controls, including, for example, time and fixed effects, and other controls such as individual characteristics (firm level or consumer level).

There is often no fundamental difference at this level with an empirical model where R is replaced by a specific signal, for example, a statutory tax rate, but the discussion below focuses on the PFR context, while in the next section we go back to issues of interpretation of the results.

The justification of the generic form (1.1) is as follows. We want to understand the correlation (if not the impact) between the reform indicators and the outcome variable, which is seen in turn as a proxy of the welfare effects, but we need to control for any possible confounding factors. It is convenient to think of three broad sets of such controls.

First, drivers of *demand* usually have an influence on most outcome variables. Obvious examples are prices of the service, which respond positively, that is, increase with demand and decrease with the price of substitutes. For example, the price of electricity for residential users is correlated to income per capita of households because – at least in the short or medium term – if prices are allowed to vary (that is, there is an upstream electricity market following liberalization) – they will respond to increased consumption of electricity: households will use more electric appliances during the day, or will increase their stock of appliances. Electricity used for cooking or heating, however, has natural gas as a substitute, and a decrease of the price of gas compared to electricity may lower the demand of electricity. Hence, in general at least income per capita and the price of a main substitute will enter into the determinants of price, but other demand-side effects may reflect shifts in preferences (for example, energy saving attitudes may

respond to cultural changes, usage of fixed telephony may be sensitive to higher mobility of users, which in turn increases the preference for mobile phones). When micro data are available some individual characteristics may be directly linked to preferences, and considered as drivers of demand (for example, younger people may have a preference for smartphones which goes beyond the price of the service).

On the *supply* side, the core drivers of outcome variables are changes in unit costs, which in turn are related to two sets of data: input prices and technology. Some network services may resort to different inputs, for example, electricity generation to gas, coal, nuclear, hydro, solar, wind, and changes in prices of such inputs have a direct impact on production costs. In relatively long run terms, technological shifts may dramatically change the cost of such services as telecommunications, but less than in relatively stagnant industries, such as water distribution to households. This calls for a good understanding of the essentials of the technological environment of service provision by the analyst.

Finally, there are characteristics which are relatively 'fixed', but are nevertheless important. In the context of international panel data analysis, which is particularly helpful in policy evaluation, country fixed effects may be able to capture demand and supply aspects that are not captured by D and S variables, but also other features that are not easily accommodated in standard economic variables. One example is the average 'optimism' of agents about their perspectives in the country, which may be a complex mixture of memory (for example, of past unemployment or inflation) and expectations (for example, about government effectiveness). Time fixed effects may capture shocks not otherwise identified. Most importantly, according to the type of data available, there are some individual characteristics (of households or firms) which while not being directly correlated to D or S may have an influence on outcomes.

Both objective and subjective outcome data can be fitted with the above-mentioned generic model, with opportune variations. For example, it is obvious that in the perception of price or quality of service, age, gender, ideology may enter as a confounding factor about the evaluation of the reform, and not because they may be directly correlated to D or S , but because social attitudes may vary between the old and the young, male and female, left-wing and right-wing respondents to a survey. However, even with objective micro data available, it would still be helpful in some cases to consider the individual characteristics. Suppose, for example, that we want to understand how profits of firms providing a service respond to regulatory changes. While D and S variables, and year and time fixed effects may be included in the model as controls, there may be firm-level characteristics which are important too: just consider the difference

between SOEs and those controlled by private investors, co-operatives, firms listed in the stock exchange and those controlled by private equity funds and unlisted. Ownership often implies different unobservable firm-level objectives, corporate governance mechanisms, managerial responses to shocks. Hence, omitting relatively fixed firm-level characteristics in a price model would imply that coefficients estimates of other covariates would be biased.

To simplify the argument, suppose there are two countries where the reform is unbundling of a network industry in the same year, and we want to understand to what extent this is correlated with a change of price for households after the reform, *ceteris paribus* in terms of cost of inputs, technology, demand drivers. However, in one country the industry after unbundling is split between a private owner of the network and several private competitors, while in the other country the owner of the network is a national government body, and the competitors are firms under the control of local governments. The latter wants to maximize the welfare of citizens under budget constraints (they are not allowed to incur losses), while the former wants to maximize the value to shareholders. In both countries there is a regulator that imposes a price-cap on access to the network, but there is no price regulation for service to the households. There is no reason to expect that if the markets in the two countries are not interconnected, the equilibrium price will be the same because the firms' managers will react to unbundling in different ways, given the difference of their objectives. Omitting the information about ownership of firms may lead to biasing the estimates of other coefficients, including the variable of interest.

Other issues are more technical, and shared with any kind of econometric approaches in similar contexts, for example, whether or not to include a lagged dependent variable among the covariates to account for dynamic effects, to what extent to introduce interaction terms, non-linearity, or how to pick up the appropriate estimation techniques. Some of these aspects are indeed not specific to the empirical analysis of policy reforms, but others take a particular form and are discussed in the next sections.

1.7 INTERPRETATION OF THE EMPIRICAL RESULTS

The main question that arises in our context is how to exactly interpret the estimated coefficients of the variables of interest.

In the more standard theory of reform model, the interpretation is relatively straightforward: for example, one may conclude that the public

provision of one additional place in a hospital has an x effect of QALY for the target population, and this result would lend itself to relatively simple cost-benefit tests, including the social cost of increasing distortionary taxes to support public spending in healthcare.

But this simple interpretation does not usually hold when the variable of interest is a PFR indicator. It is important to restate here that in some cases the RSG ‘theory of reform’ approach is either unfeasible or not appropriate. For example, one would be tempted to represent privatization policy with a decreasing change of a continuous variable such as the first difference of privatization proceeds for the Treasury or the change of the percentage share of firms’ assets in one country (or region) from year to year. If data are available, this is perfectly feasible and may be interesting, but not necessarily a really good proxy of the actual scope of reforms. If the intention of the government, through selling public enterprises, is to collect cash proceeds, perhaps to decrease the level of indebtedness of the public sector, it would be entirely appropriate to study the impact of one additional pound of privatization proceeds on the outcome variable, for example, GDP per capita, after controlling for confounding factors. Here the interpretation is simple: perhaps through changes in savers’ expectations or those of foreign investors the null hypothesis could be that the privatization proceeds are correlated to GDP because of diminishing debt, hence less burden of interest and repayment for public finance, and ultimately, according to some macroeconomic models, more growth. If one assumes that GDP is the relevant social welfare function (no equity and sustainability concerns) then a ‘positive’ estimated coefficient of the privatization variable would point to the expected effect. However, if the government objective is to get rid of the control of firms for the sake of assumed greater efficiency of private ownership, it may even be interested in under-pricing SOEs and the continuous variable of cash proceeds would not be the right signal. In such cases, discontinuous thresholds of equity shares held by government may be more appropriate, because what is at stake is a policy framework and not the marginal reform of a specific signal in the traditional meaning. Hence, an indicator variable may be a better description of the change, for example, setting a multinomial ordering such as full ownership at more than 50 per cent of equity, less than 50 per cent but more than 20 per cent, less than 20 per cent but greater than zero.

While this may be seen as reducing the variability of the information, it actually points to a different type of information, which would not be captured by a coefficient expressing the contribution of a small change in ownership to the social welfare proxy variable. Thus, one should read the coefficient as expressing the contribution of a policy reform in a different

meaning, as the shift from a regime to another one, each represented by the value taken by the reform indicator.

It is helpful to observe that this issue is not related per se to the well-known problem of causality in empirical analysis: it would arise even if one would loosely interpret the estimated coefficients simply as the expression of correlations. The issue is deeper, and it has to do with the appropriate interpretation of the 'meaning' of the reform and to what extent the empirical proxy captures it. Thus, one should read the coefficient as correlated to the welfare effect of a 'one step further' transition of policy regimes, when the variable of interest is expressed in an appropriate way. Clearly, no statistical test can clarify the issue of such appropriateness, only a careful qualitative interpretation of the 'history' behind the PFRs. Here the economist needs, perhaps more than in other contexts, to be a true social scientist.

1.8 CONCLUDING REMARKS

This introductory chapter contributes to the literature on the empirics of policy evaluation by making the distinction between, on one side, the approach of the theory of reform in the Ramsey-Samuelson-Guesnerie tradition and – on the other side – the approach here described as the evaluation of changes of policy reform frameworks.

Some methodological issues of the two approaches are similar, but the latter departs from the former in certain aspects, the most important being the way the reform is treated as an object of analysis. In the theory of reform approach a precise signal can be identified and its marginal change evaluated, in the PFR context what changes is a broader orientation of the policy makers on some economic governance mechanisms or structures.

This distinction has consequences in the empirical analysis, particularly of the reform of network industries. While some potential errors are common to the two approaches, others are more likely when a PFR is considered. Moreover, the interpretation of estimated coefficients differs because usually it is not meaningful to express it as quantitative changes of the variable of interest, but rather as a transition from a policy regime to another one along an ordering.

In both contexts it is apparent that the dependent variable should be clearly linked, even if indirectly, to a social welfare function assumed as the evaluation criterion. In principle, both objective and subjective outcomes may be considered, with some implications for the interpretation of the results. In the empirics of PFR, apart from being particularly careful in dealing with potential errors given the more fuzzy nature of the variables

of interest compared with changes in the RSG theory of reform, an applied economist needs to spend some time and effort in understanding the historical, political and social circumstances that supported an actual or proposed policy shift in the first place. The example of privatization or liberalization shows that indicator variables related to such changes of policy framework can have different meanings in different contexts, and this crucial characterization should be embodied in the empirical proxy of the policy variable of interest.

NOTES

Websites were last accessed April 2017.

- * The author is grateful for helpful comments on a previous version to Andrea Bastianin, Paolo Castelnovo and Valentina Morretta.
1. Koske et al. (2015) and more recent materials and papers are available at <http://www.oecd.org/eco/growth/indicatorsofproductmarketregulationhomepage.htm>.
 2. See, for example, the wide range of international data available at the Quality of Government Institute website <http://qog.pol.gu.se/data>.
 3. http://ec.europa.eu/public_opinion/index_en.htm.
 4. Statistics-on-Income-and-Living-Conditions (SILC) is available at <http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions>.
 5. European-Community-Household-Panel (ECHP) is available at <http://ec.europa.eu/eurostat/web/microdata/european-community-household-panel>.

REFERENCES

- Ahmad, E. and Stern, N. (1984). 'The Theory of Reform and Indian Indirect Taxes', *Journal of Public Economics*, **25**, 5, 259–98.
- Boadway, R. (2012). *From Optimal Tax Theory to Tax Policy: Retrospective and Prospective*, Cambridge, MA: MIT Press.
- Boadway, R.W. and Bruce, N. (1984). *Welfare Economics*, Oxford: Basil Blackwell.
- Boardman, A., Greenberg, D., Vining, A. and Weimer, D. (2016). *Cost-benefit Analysis*, 4th edn, London: Pearson.
- Bös, D. (1991). *Privatization: A Theoretical Treatment*, Oxford: Clarendon Press.
- Clò, S., Del Bo, C.F., Ferraris Fiorio, C.V., Florio, M. and Vandone, D. (2015). 'Public Enterprises in the Market for Corporate Control: Recent Worldwide Evidence', *Annals of Public and Cooperative Economics*, **86**, 4, 559–83.
- Diamond, P.A. and Mirrlees, J.A. (1971). 'Optimal Taxation and Public Production II: Tax Rules', *American Economic Review*, **61**, 3, 261–78.
- Drèze, J. and Stern, N. (1990). 'Policy Reform, Shadow Prices, and Market Prices', *Journal of Public Economics*, **42**, 1, 1–45.
- Florio, M. (2004). *The Great Divestiture: Evaluating the Welfare Impact of the British Privatizations, 1979–1997*, Cambridge, MA: MIT Press.
- Florio, M. (2014). *Applied Welfare Economics. Cost-benefit Analysis of Projects and Policies*, Abingdon and New York, NY: Routledge.

- Guesnerie, R. (1977). 'On the Direction of Tax Reform', *Journal of Public Economics*, **7**, 179–202.
- Guesnerie, R. (1998). *A Contribution to the Pure Theory of Taxation*, Cambridge: Cambridge University Press.
- Johansson, P.O. and Kriström, B. (2015). *Cost-benefit Analysis for Project Appraisal*, Cambridge: Cambridge University Press.
- Jones, L.P., Tandon, P. and Vogelsang, I. (1990). *Selling Public Enterprises: A Cost/Benefit Methodology*, Vol. 1, Cambridge, MA: MIT Press.
- Koske, I., Wanner, I., Bitetti, R. and Barbiero, O. (2015). 'The 2013 Update of the OECD's Database on Product Market Regulation', OECD Economics Department Working Papers.
- Mayeres, I. and Proost, S. (2001). 'Marginal Tax Reform, Externalities and Income Distribution', *Journal of Public Economics*, **79**, 2, 343–63.
- Ng, Y. (1986). *Mesoeconomics: A Micro-Macro Analysis*, London: Harvester Wheatsheaf.
- Obinger, H., Schmitt, C. and Traub, S. (2016). *The Political Economy of Privatization in Rich Democracies*, Oxford: Oxford University Press.
- Parker, D. and Kirkpatrick, C. (2012). 'Measuring Regulatory Performance. The Economic Impact of Regulatory Policy: A Literature Review of Quantitative Evidence', OECD Expert Paper No. 3.
- Poggi, A. and Florio, M. (2010). 'Energy Deprivation Dynamics and Regulatory Reforms in Europe: Evidence from Household Panel Data', *Energy Policy*, **38**, 1, 253–64.
- Price, C.W. and Hancock, R. (1998). 'Distributional Effects of Liberalising UK Residential Utility Markets', *Fiscal Studies*, **19**, 3, 295–319.
- Ramsey, F.P. (1927). 'A Contribution to the Theory of Taxation', *The Economic Journal*, **37**, 145, 47–61.
- Samuelson, P.A. (1951). 'Theory of Optimal Taxation', Unpublished Memorandum, US Treasury.