A non-parametric method to identify nonlinearities in global productivity catch-up performance

Bart Los

INTRODUCTION

Economic performances of countries cannot be solely explained by tendencies or policy measures within the countries themselves. Individual countries are parts of an intricate system of ever-changing interdependencies. Some of these interdependencies have a purely economic character, whereas others are of another nature but have strong implications for economic performance. A very important aspect of economic performance is productivity growth. Increases in productivity provide opportunities to attain higher levels of well-being, both in developed and underdeveloped countries. Consequently, economists have directed much effort at descriptions and explanations of international differences in productivity levels and their growth rates. Especially within the popular field of convergence analysis, interdependencies between countries play an important role, although this role is often implicit (at best).

The notion of productivity convergence refers to the idea that productivity levels of countries would become more similar over time. Although this ‘definition’ could (and maybe should) be qualified in many respects, it can serve well to summarize the types of interdependencies that could yield productivity convergence as emphasized by the two main strands of theory in this field.

Traditional mainstream economists focus on the international mobility of capital. Due to the law of diminishing returns to capital accumulation, new investment will be directed towards countries with still higher returns. By assumption, these have low capital intensities and productivity levels. Because technological progress is generally assumed to be exogenous and uniform across capital intensities, such international investment flows should equalize productivity growth rates and levels in the long run. Scholars who adhere to the technology gap approach adopt a different viewpoint.
They stress the non-immediate international diffusion of technology pertaining to productivity-enhancing innovations. Backward countries would need time to imitate or assimilate such technology, both in embodied and in disembodied form, but would eventually be able to do so. Consequently, more backward countries would ‘catch up’ faster than less backward countries. This mechanism would ultimately equalize productivity growth rates, leaving a constant proportional productivity gap between the leader country and the follower countries.¹

In this chapter, we will focus on the technology gap approach, although the core methodology we propose would be easily applicable to mainstream analyses as well. As said, the basic idea of technology gap models is that the dynamics of productivity differences is governed by two opposing forces. The differences tend to be enlarged by innovation in the leader country, while they tend to be reduced by catch-up through technology spillovers to lagging countries. The joint effect of these tendencies can be captured in a very simple regression equation, originally suggested by Abramovitz (1979):²

\[ \dot{g}_i = \alpha + \beta g^0_i + \varepsilon_i \]  

In this equation, \( g_i \) represents the productivity gap, defined as the logarithm of the ratio of the productivity levels of country \( i \) and the leader country. The superscript 0 refers to the value in the initial year. In terms of the simplest technology gap model that one could imagine, the intercept \( \alpha \) refers to the rate of innovation by the leader, and the slope coefficient \( \beta \) to the ability of countries to benefit from the pool of technology spillovers, which is larger for countries with a large gap (a strongly negative value of \( g^0 \)). The simple theory thus predicts a positive value for \( \alpha \) and a negative value for \( \beta \), which would imply that all lagging countries would converge towards an equilibrium productivity gap in the long run.

A more advanced model would allow backward countries to innovate independently from the leading country. In this case, estimates for \( \alpha \) should be interpreted as the average difference between rates of innovation attained by the backward countries and the leader country. Of course, the autonomous rate of innovation by backward countries can be endogenized to some extent by including right-hand side variables like ‘number of patent applications’ and ‘educational attainment’ by country \( i \) (see, for example, Fagerberg, 1988; Baumol et al., 1989). The rate of catch-up \( \beta \), though, is generally assumed to be identical across countries, whereas economic historians like Gerschenkron (1962), Ohkawa and Rosovsky (1973)
and Abramovitz (1979, 1986) emphasized the fundamental difference between potential spillovers (indicated by $g^0$ in Equation (1)) and actual spillovers. According to this literature, the degree to which lagging countries are able to turn potential spillovers into actual spillovers is dependent on their social and technological capabilities, which can be proxied by schooling, R&D, patent and infrastructural indicators, among others. Differences in capabilities could yield complex patterns of convergence and divergence among countries, if only because equilibrium gaps become country-specific.

This chapter adds to the mostly recent literature that seeks to develop methods to identify nonlinearities in catching-up performance and link these to the presence or absence of sufficient capabilities. In the next section we will briefly evaluate some important, earlier, related methodologies. The third section will be devoted to the introduction of a simple nonparametric test that alleviates some of the problems pertaining to the methods in use. In the fourth section, we will describe the data used to illustrate the methodologies. In the fifth and sixth sections the results will be discussed. The final section concludes.

ANALYSIS OF GROWTH NONLINEARITIES: STATE OF THE ART

The usual, linear specification of the catch-up Equation (1) suggests that countries will converge to a situation in which the technology gap between the follower countries takes on a negative, constant equilibrium value, unless the exogenous rates of innovation are equal ($\alpha = 0$). This equilibrium value equals $\alpha/\beta$, as can be derived in a straightforward way by setting the left hand side of Equation (1) to zero and solving the right hand side for the gap. The widely found empirical fact that there is no tendency towards such a common gap is not in line with this result. Country-specific values for $\alpha$ could cause this, but the evidence cited by economic historians suggests that country-specific values for the catch-up parameter $\beta$ are at least equally or probably even more important. A simple linear relation like Equation (1) cannot capture this idea, even if additional regressors would be added.

At least two approaches have been pursued to deal with such nonlinearities. First, Verspagen (1991) derived a nonlinear regression equation. Second, Durlauf and Johnson (1995), Fagerberg and Verspagen (1996) and Cappelen et al. (1999) proposed piecewise linear estimation techniques, as did Hansen (2000). Below, we will discuss these contributions briefly.
The Nonlinear Regression Approach

The point of departure of Verspagen (1991) is the observation that Equation (1) is a special case of a more general specification of the dynamics of the technology gap:

\[
\dot{g} = \alpha + \beta g e^{-g/\delta} \quad (\delta > 0)
\]  

(2)

Verspagen (1991) argues that the parameter \(\delta\) can be considered as an indicator of the ‘intrinsic learning capability’ of a country. Equations (1) and (2) are identical if and only if \(\delta = \infty\). In this specific case, actual spillovers equal a fixed fraction of the potential spillovers indicated by the gap itself. As mentioned, this situation will lead to an equilibrium gap of which the size depends on the relative strength of the innovation parameter and the catch-up parameter. For finite values of \(\delta\), however, Verspagen (1991) shows that convergence to such an equilibrium gap is not warranted. If \(\delta\) is lower than the threshold level \(\delta^* = \alpha e/\beta\), the productivity gap will tend to infinity, irrespective of the size of the initial gap. In other words, falling behind is the certain outcome of the model. For finite values of the intrinsic learning capability that exceed \(\delta^*\), two equilibria exist. If the initial gap is too big (relative to \(\delta\)), falling behind will result. If the initial gap is relatively small, however, the gap will tend towards a stable value. This value is always larger than \(\alpha/\beta\).

In his empirical analysis, Verspagen (1991) allowed \(\delta\) to vary across countries, assuming a common \(\alpha\) and \(\beta\). By means of nonlinear least squares techniques, he estimated a model derived from Equation (2) with \(\delta\) modelled as a simple function of either educational attainment or electricity generating capacity. The results for a comprehensive sample of 114 countries in the period 1960–85 by and large indicates that the nonlinear specification outperforms linear equations that involve the same variables, in particular if the intrinsic learning capability was measured by education variables. In an analysis of a somewhat more complex model along similar lines, Verspagen (1992) came to the conclusion that more than one-third of the countries in his sample had intrinsic learning capabilities that did not exceed the implicitly computed threshold value \(\delta^*\). Only three countries had higher intrinsic learning capabilities, but faced an initial gap so large that it prevented them from catching up with the world leader, the United States. The remaining countries belonged to the category of countries characterized by convergence in productivity growth rates.

The Piecewise Linear Regression Approach

Multiple productivity growth regimes could also prevail in mainstream models based on production factor accumulation. Azariades and Drazen
(1990), for example, suggested that returns to scale may be increasing locally, due to surpassing a ‘threshold value’ for an accumulable input. Durlauf and Johnson (1995) analysed this issue empirically. The scope of their ‘regression tree’ analysis is not confined to mainstream theory-inspired investigations, though. For ease of exposition, we will discuss the methodology with reference to the simple technology gap Equation (1).

The main idea is that the parameters $\alpha$ and $\beta$ may not be homogeneous across an entire sample of countries. Instead, threshold values of some variables might identify different regimes of innovativeness by laggards and/or their ability to catch up. These variables could for example relate to social and technological capabilities, such as schooling or initial productivity. The Durlauf and Johnson (1995) approach basically amounts to simultaneously identifying the apparently most important threshold variable and estimating the value of this variable. The relevant criterion is the maximum reduction of the residual sum of squares obtained by ordinary least squares estimation, added over the two subsamples implied by the ‘split’. The estimates $(\alpha_1, \beta_1)$ and $(\alpha_2, \beta_2)$, for the first and second subsample respectively, together constitute a piecewise linear function. This procedure is repeated for each of the two subsamples and, if any, the smaller subsamples that result from subsequent splits, until the residual sum of squares cannot be reduced any further or the number of observations within a subsample would become smaller than twice the number of parameters to be estimated. It should be noted that the first split does not preclude other variables to be the threshold variable in subsequent steps.

In usual regression analysis, adding explanatory variables will reduce the residual sum of squares. Without putting a ‘penalty’ on adding regressors, specifications with many regressors would always be preferred over simpler specifications. One way to avoid such ‘overparametrization’ is to base the choice of variables on a minimization of the sum of the residual sum of squares and a penalty roughly proportional to the number of regressors (Akaike’s information criterion). With respect to reducing the residual sum of squares, adding splits has effects similar to adding regressors. Durlauf and Johnson (1995) propose a variant of Akaike’s criterion, in which the penalty is roughly proportional to the number of splits. Application of this criterion (‘pruning the regression tree’) generally leads to a limited number of splits, the actual number depending on the weight of the penalty function relative to the residual sum of squares. This eventual weight is determined using a rather complicated cross-validation procedure.

In their study inspired by the mainstream augmented production function, Durlauf and Johnson (1995) identify four productivity growth regimes for the period 1960–85. The first split divides the sample of 96 countries into a low-initial productivity subsample (14 countries) and a higher-initial productivity
subsample (82 countries). The second subsample was further subdivided, into a low-literacy subsample (34 countries) and a higher-literacy subsample (48 countries), which could be split into an intermediate-initial productivity subsample (27 countries) and a high-initial productivity subsample (21 countries). The estimated coefficients vary wildly across subsamples.

The piecewise linear regression approach also got some attention from authors who have been investigating cross-country productivity growth differences from a catching-up perspective. Fagerberg and Verspagen (1996) included unemployment rates to split their sample of European regions into three subsamples. They found that the effects of EU support through R&D subsidies and investment loans for productivity growth were quite different across the high-, intermediate- and low-unemployment subsamples, which led the authors to consider European regions as ‘at different speeds’. In a follow-up study, Cappelen et al. (1999) restricted the analysis to a single split for each threshold variable, but considered many more variables. Besides R&D activity levels and unemployment rates, structural characteristics as the shares of agricultural employment (viewed as a retarding factor for growth) and the shares of industrial employment (viewed as an engine of growth) were considered as splitting variables and found to be useful. Other variables, such as population density, population growth and physical infrastructure led to less convincing results.

Assessing the Significance of Nonlinearities in Growth Regressions

Many growth theories suggests that nonlinearities play an important role in the productivity growth process. The empirical work reviewed above yielded interesting results that are often in line with intuition. Nevertheless, at least one pressing question remains. Do nonlinear estimation methodologies perform significantly better (in a statistical sense) than ordinary linear techniques?

Verspagen (1991) subjected his nonlinear regression results to two specification tests, a ‘nested’ and a ‘non-nested’ test. Discussion of these rather complicated tests goes beyond the scope of this paper, because the method we will propose in the next section belongs to the class of piecewise linear regression techniques. It should be mentioned, though, that Verspagen (1991) presents strong evidence for the claim that his nonlinear equation outperforms its linear counterpart.

Until the recent contribution by Hansen (2000), almost nothing could be said about the significance of splits found in the piecewise linear regressions. As discussed already, Durlauf and Johnson (1995) relied on a rather ad-hoc ‘pruning’ procedure to arrive at an apparently reasonable number of splits. In the absence of a known distributional theory for regression trees, it seems
impossible to draw strong conclusions on significance levels. Fagerberg and Verspagen (1996) do not report what criterion they used to find exactly three unemployment rate-determined productivity growth regimes. Cappelen et al. (1999) are much more explicit on this issue. They base their decision on the significance of a single split on the $F$-statistic used in well-known Chow-tests for structural change. As is explained by Hansen’s (2001) very accessible article on dating structural breaks in time series, this procedure leads to over-rejection of the null hypothesis of a single productivity growth regime. The Chow-test assumes that the location of the possible split (for example, an unemployment rate of 0.12) is given beforehand, and not derived from the data. In the procedure by Cappelen et al. (1999), however, the location of the split is first estimated by the rate of unemployment for which the reduction of the sum of squared residuals is largest. In general, this will be a split between an observation for the unemployment rate that yields a very negative residual in the linear regression based on the entire sample and an observation that yields a very positive residual, or the other way round.

Hansen (2000) suggested a technique to determine the significance of splits on the basis of asymptotic theory. First, the threshold value of the split variable is estimated in the way first put forward by Durlauf and Johnson (1995). Next, a Lagrange Multiplier test (based on a kind of bootstrapping procedure) is used to get an indication of the significance of the split. This is a parametric test, since it is assumed that the errors are normally distributed around zero. Hansen (2000) finds point estimates of the location of splits that are quite close to the estimates obtained by Durlauf and Johnson (1995). Interestingly, he is also able to present confidence intervals for the location, based on a parametric Likelihood Ratio principle. With regard to the first split (initial income is the threshold variable), the asymptotic 95 per cent-confidence interval appears to be very wide. Actually, for as many as 40 of the 96 countries in the sample, no definite answer can be given to the question to which of the two regimes they belong. Hansen’s (2000) approach can formally not be extended to testing for multiple splits, but repeating the same procedure for both subsamples can at least give some good indications of the presence and location of more splits.

This review of existing methods that deal with potential nonlinearities leads us to two interim conclusions, which should offer a justification for the content of the next sections.

First, Verspagen’s (1991) nonlinear regression framework is attractive in many respects:

1. it is derived from a simple but elegant theory of productivity growth;
2. it yields sensible results; and
3. it survives exposition to specification tests against linear specifications.
A potentially serious drawback, however, is that the model supposes linearity at a lower level. It assumes that capabilities to assimilate spillovers are proportional to variables like schooling and infrastructure. It could well be that this relation has a different shape in reality, which would lead to a different nonlinear relation between productivity growth and the initial gap. Such an alternative nonlinear specification might do better than Verspagen's nonlinear equation in specification tests. This does not seem to be a viable way to proceed, however, because theory does not offer any clue about the ‘right’ specification of the relation between capabilities and the inputs that shape capabilities. Empirical clues can neither be obtained, since capabilities are not observable.

Second, the piecewise-linear regression approach is based on an almost opposite modelling perspective. The procedure is entirely data-driven. No functional specification is required a priori, apart from the choice of variables that are included in the ‘grand equation’. Theory starts to play a role only at the stage in which the results have to be judged on their plausibility. Significance results rely on asymptotic theory and normality assumptions, which may well be untenable if samples are small. An advantage is that confidence intervals for the location of thresholds can be constructed.

In the next section, we will propose a non-parametric, exact way to test for the presence of two catching-up regimes. We will also propose a non-parametric, though asymptotic, way to construct a confidence interval for the location of the threshold.

BASIC METHODOLOGY AND EXTENSIONS

This section consists of two subsections. In the first subsection, we will explain how the location of a threshold is estimated and its significance will be assessed. The second subsection deals with the construction of confidence intervals for this location.

**Point Estimation of a Threshold Value**

In statistical terms the problem of detecting two catch-up regimes and assessing the significance of the splitting variable reads as follows. Consider the model

\[
\begin{align*}
\dot{g}_i &= \alpha_H + \beta_H g_i^0 + \epsilon_i, \quad \forall x_i > x_0 \\
\dot{g}_i &= \alpha_L + \beta_L g_i^0 + \epsilon_i, \quad \forall x_i \leq x_0, \quad i = 1, \ldots, n \quad x_0 \in X
\end{align*}
\]
x denotes the splitting variable, and \( x_0 \) the threshold value of this variable. It is assumed that the errors \( \varepsilon_i \) are independent and identically distributed. In order to rule out threshold values very close to the minimum and maximum values of \( x \) in the sample (which would yield regimes which would be relevant for just one or two countries) the range \( X \) will be chosen in such a way that a reasonable number of countries will be included in both regimes.  

\[
H_0: (\alpha_H, \beta_H) = (\alpha_L, \beta_L) \quad \forall x_0 \in X
\]

\[
H_1: \exists x_0 \in X \text{ for which } (\alpha_H, \beta_H) \neq (\alpha_L, \beta_L)
\]

It should be noted that the alternative hypothesis does not make a distinction between situations in which the two regimes differ with respect to one parameter and situations in which both parameters differ across regimes. The procedure we propose to test \( H_0 \) against \( H_1 \) can best be described as a sequence of steps.

**Step 0:** Estimate

\[
\hat{\dot{g}}_i = \alpha + \beta g^0_i + \varepsilon_i, \quad i = 1, \ldots, n
\]

Store the sum of squared residuals (SSR0). This involves estimating the ordinary linear catch-up Equation (1), for all observations.

**Step 1:** Order the \( n \) countries according to a splitting variable \( x \). For analysis of catching-up performance \( x \) should be some variable that affects the social capabilities of countries. The regressor \( g^0 \) may act as such, but \( x \) is not necessarily included in the piecewise linear function itself. Country 1 has the lowest absorptive capacity, country \( n \) the highest. Now, estimate the linear convergence equation for pairs of ordered subsamples. In the first subsample, the first \( n_1 \) countries are included, the second subsample contains the \( n - n_1 = n_2 \) remaining countries. This is done for all values of \( n_1 \), as long as both subsamples contain at least a pre-specified number of countries. For each estimation, store the sums of squared residuals SSR1abs (corresponding to the first subsample) and SSR2abs (second subsample). To complete Step 1, determine the value of \( n_1 \) for which SSR1abs + SSR2abs is found to be lowest and denote \( \hat{x}_{n_1} \) as the 'potential split' or 'potential threshold value'. Note that Step 1 is exactly identical to the first part of the procedures adopted to arrive at the earlier estimators of piecewise linear functions, as discussed in the previous section.

**Step 2:** Repeat Step 1 many times (for instance 10 000 times), with \( x^a \) as an artificial splitting variable. Each time, the values of \( x^a_i \) are obtained by means of a random number generator. Consequently, random orderings of countries are used to obtain artificial 'potential threshold values'. Store SSR1rand + SSR2rand for each of these. Finally, calculate the fraction of
random orderings for which $SSR_{1\text{rand}} + SSR_{2\text{rand}}$ is lower than $SSR_{1\text{abs}} + SSR_{2\text{abs}}$. This fraction equals the level of significance of $\hat{x}_{n1}$. If this significance level is lower than a pre-specified level, two catching-up regimes are identified. The ‘point’ estimate for the critical level of the absorptive capacity variable is in between its value for the countries $n_1$ and $n_1 + 1$ in the ordering.

In brief, the test procedure outlined above boils down to investigating whether a theoretically sensible splitting variable yields a larger reduction of the sum of squared residual than the large majority of random, theoretically nonsensical, splitting variables.

**Confidence Intervals for the Threshold Value**

The test procedure indicates whether the splitting variable $x$ defines two productivity growth regimes or not. If so, it also yields a point estimate of the value of $x$ that constitutes the boundary between the two regimes. In statistical terms, this point estimate is dependent on the distribution of the errors. Hence, drawing another hypothetical sample by assigning other realizations of the error generating process and applying the test procedure could yield a different point estimate for the threshold value. Repeating this redrawing many times (for instance 10 000 times) would give a distribution of estimated threshold values. One could cut-off the lowest $0.5\times\theta$ per cent and the highest $0.5\times\theta$ per cent of the estimated threshold values to obtain a $(1 - \theta)$ per cent confidence interval for $x_0$.

The main problem associated with this approach is that the true distribution of the errors $\varepsilon$ is not known. Often, see for example, Hansen (2000), a normal distribution with a fixed variance is assumed. In the case of growth regressions, this assumption seems hard to defend, if only because a glance at the performances of formerly low-productivity countries indicates that these vary across a much wider range than the performances of initially high-productivity country. Heteroscedasticity seems omnipresent and the usefulness of the normal distributions is rather doubtful, as is the justification of the use of test statistics based on this assumption.

Instead of relying on questionable assumptions regarding the distribution of unobserved errors, we propose to use the observed distribution of residuals as an estimate of the actual distribution of errors. We apply a bootstrapping methodology (Efron, 1979, is the classic reference). That is, we start from expected values for $(y_i^0, x_{ip}, i = 1 \ldots n)$ obtained from Model (3) and its least squares estimates (including the threshold value $x_0$). To each of these expected values we add a randomly drawn residual $\hat{u}$, from the distribution obtained for the least squares estimators of Model (3). These drawings are done with replacement, as the distribution of residuals
is considered as a discretized version of the actual continuous distributions. This solution to the problem of unobservability of errors allows for the construction of confidence intervals by means of the procedure outlined above.\textsuperscript{11}

**EMPIRICAL ILLUSTRATION**

**Data**

In this section, we will apply the test for multiple catch-up regimes and the procedure to construct confidence intervals as proposed in the previous section on country-level data. The majority of these are taken from the GGDC Total Economy Database (University of Groningen and The Conference Board, 2003). To include as many observations as possible, the labour productivity variable considered is GDP per worker, instead of, for example, GDP per hour worked. The average annual growth rates are computed over the period 1960–2000. The initial levels are evaluated for 1960. Productivity growth and the initial gap are defined as relative to the growth and initial level of the USA.

Following Verspagen (1991) and Durlauf and Johnson (1995), we consider educational attainment as a variable that might delineate multiple regimes. Many alternative measures of educational attainment are available, and a choice between them is mainly a matter of taste. We choose the average number of schooling years for the entire population (males and females) over 25 years old in 1960 as the variable of interest. These data were taken from Barro and Lee (2001). It should be mentioned that we deliberately choose to included educational attainment in 1960 as a threshold variable, instead of educational attainment in a specific year in between 1960 and 2000, or an average value of this variable for this time period. We agree with Durlauf and Johnson (1995) that one of the latter approaches would imply some endogeneity of the threshold variable, because a good productivity growth performance in early years could allow a country to spend more funds on education. Admittedly, a drawback of our choice is that levels of variables as far back as 1960 are supposed to explain patterns that were partly shaped less than ten years ago.\textsuperscript{12}

These choices with regard to variables and time interval left us with 44 observations. Roughly speaking, the sample seems to cover the world's economies rather well. Nevertheless, the rich OECD countries seem to be over-represented, while both (formerly) centrally-led economies and African economies are under-represented.\textsuperscript{13} The countries in the sample and the values of the variables of interest are included in Appendix A.
Test for Thresholds

Estimation of Equation (1) for the entire sample yields a result that does not point towards convergence:

\[ \hat{g} = 0.00797 + 0.00125g^0 \]

According to standard heteroscedasticity-consistent $t$-tests, the slope is far from significantly different from zero. The intercept is significantly positive (p-value: 0.034). This would point towards a general tendency for countries to converge to the US productivity levels. This would be caused by better abilities to innovate than the USA, which is of course highly counter-intuitive. Within the sample of lagging countries, however, no convergence in labour productivity levels seems to occur. This could lead to a conclusion that catching-up is no important factor in the process of productivity growth. It should be noted that the (admittedly simple) model has an extremely poor fit. Actually, only 0.5 per cent of the total variation of growth rates can be explained. SSR0 used to compute this value of $R^2$ equals 0.01018. Figure 10.1 presents the datapoints and the estimated technology-gap equation. The conclusions drawn above are quite clearly reflected in this diagram, in which the dashed line indicates the estimated equation.

Figure 10.1 Observations and basic regression results
Next, we re-estimated the model, but now allowed for a threshold effect. Following Durlauf and Johnson’s (1995) lead, we experimented with two potential threshold variable candidates. First, we looked at the initial labour productivity gap $g^0$. Second, we repeated the analysis for the educational attainment variable described in the previous section. In both cases, we decided that at least five countries should be included in a subsample. This minimum size falls within the trimming bounds of 5 to 15 per cent of the number of observations, as suggested by Hansen (2001).

For the initial gap as the splitting variable, the sum of squared residuals appears to be minimal for $n_1 = 9$ and $n_2 = 35$. The sum of squared residuals ($SSR_{1abs} + SSR_{2abs}$) is reduced to 0.00718. Application of the same procedure for educational attainment yields subsample sizes of $n_1 = 15$ and $n_2 = 24$. In this case, the sum of squared residuals turns out to be somewhat lower, 0.00652. Irrespective of the choice of threshold variable, the estimates for both parameters appear insignificant ($p$-values above 0.2) for the first subsample. For these, low-initial productivity and low-schooling subsamples, tendencies towards convergence seem absent. For the second subsamples a much clearer picture emerges. The estimated intercepts are not significant. This points to a situation in which the ‘intrinsic’ abilities to innovate are in general not different from the USA. For the catch-up parameter $\beta$, significant negative estimates are found. For the initial gap threshold variable, this estimate equals $-0.01141$, with a small associated $p$-value of 0.002. Schooling as a threshold variable yields a point estimate of $-0.01264$ and a $p$-value of 0.0004.

As mentioned, the sum of squared residuals is smaller for the educational attainment threshold variable than for the initial labour productivity gap threshold variable. It remains to be seen, though, whether the reduction is sufficiently large to conclude that a threshold is present indeed. We performed Step 2 of the testing procedure outlined in Section 3 10,000 times. In only 318 cases, the sum of squared residuals appeared to be lower than found for the educational attainment threshold variable and $n_1 = 15$. Thus, in view of the $p$-value of 0.032, educational attainment is found to define two catching-up regimes. Further, it should be noted that we could not reject the null hypothesis of a single regime against the alternative of two regimes identified on the basis of initial productivity gaps. The $p$-value for this test was 0.107.

A natural question to ask is whether the labour productivity dynamics is best characterized by two regimes. It might well be that three or more regimes are present. This is hard to test in a statistically sound way, as was also stated by Hansen (2000) for his parametric analysis. A first, and arguably not too bad indication can be obtained by repeating the testing procedure for the two subsamples. Thus, we first look at the ‘low-schooling’
subsample of 15 countries and tested for a single regime for this sample against the alternative of two regimes based on either initial gap or educational attainment. This subsample does not show any symptoms of multiple regimes. The \( p \)-values (again based on 10,000 random orderings) are 0.715 and 0.609 for the initial productivity gap variable and the schooling variable, respectively.

Next, we run the same procedure for the ‘high-schooling’ subsample of 29 countries. This yields a very significant threshold value, with a \( p \)-value of only 0.0022.\(^{16}\) The ‘medium-schooling’ sample consists of only six countries, mostly located in East Asia: Korea, Singapore, Spain, Sri Lanka, Taiwan and Thailand. This subsample is characterized by a rather high point estimate for \( \alpha \) (although just insignificant at 5 per cent, according to a \( t \)-test) relative to the high-schooling subsample. For both subsamples, the catching-up parameter \( \beta \) estimate is insignificant. This leaves us with a rather uncomfortable result: medium-schooling countries might well be the only countries with a positive innovation differential to the USA, whereas theory would suggest that this should only be the case for countries with a high educational attainment. Further, and maybe even more striking, catching-up to the world productivity leader does not seem to play a role within any of the three subsamples. An alternative, reasonable explanation can be offered, however. The model could suffer from omitted variables. The ‘medium-schooling’ subsample might therefore be a group of outliers under a single productivity regime, instead of being governed by a separate regime. An indication of the potential power of this explanation is given by the results found if we require each subsample to contain at least seven (instead of five) observations. In that case, a still very significant split (\( p = 0.014 \)) is identified for \( n_1 = 13 \) and \( n_2 = 16 \). The estimated equations are

- medium-schooling subsample: \( \dot{g} = -0.00319 - 0.01245g^0 \)
- high-schooling subsample: \( \dot{g} = -0.00734 - 0.02543g^0 \)

The coefficients in these equations all have \( p \)-values well below 5 per cent, except for the intercept in the medium-schooling subsample. More importantly, these results are much more in line with theory. That is, in general, follower countries do not have higher innovative abilities than the world productivity leader. Further, countries with higher educational attainments were catching-up at a faster rate. For the methodology, however, these results indicate a weakness: if the number of observations is small, clustered outliers can well affect the results if trimming is relatively weak. Therefore, more research is required to prescribe rules-of-thumb for the minimum subsample size.
Confidence Intervals for Thresholds

As discussed in more general terms on pp. 238–40, the analysis on catching-up regimes presented in the previous section yields a point estimate for threshold value between different regimes. This estimate of the threshold value depends on a single realization of a stochastic process, however. In this section, we will apply the procedure outlined on pp. 240–41 to produce a confidence interval for the threshold value of the schooling variable. We will restrict our analysis to the first split, which breaks our sample into just two subsamples.

In fact, our point estimate is not a ‘real’ point estimate. The actual point estimate tells us that the threshold value would be located in the range of average years of education of the population of 25 years old and over between 3.02 and 3.14. These educational attainment values relate to Peru and Singapore, the most ‘schooled’ country of the low-schooling subsample and the least schooled country of the high-schooling subsample, respectively. More precise point estimates are impossible to obtain, without strong assumptions on error distributions. Although this issue might have been more important if our estimated threshold had been located in between two countries that differed strongly with respect to their educational attainments in 1960, this identification problem is not the topic of this subsection.

Our main concern is illustrated by Figure 10.2. On the horizontal axis, the values of the educational attainment variable are depicted. The values on the vertical axis indicate the sum of squared residuals that are obtained by

![Figure 10.2 Sensitivity of sum of squared residuals to location of break](image)
splitting the sample for a specific value on the horizontal axis. These values are obtained in the process of testing for the significance of a threshold variable. The leftmost point thus refers to the sum of squared residuals obtained for the minimal size of the first subsample, that is $n_1 = 5$. Analogously, the rightmost point is the observation for the opposite extreme, $n_2 = 5$.

The minimum value is found for $n_1 = 15$. For smaller values of the threshold variable, however, the sum of squared residuals is almost as low. For higher values, the reduction in comparison to the single regime case is much smaller. Thus, we expect that other realizations of the stochastic error-generating process will often yield point estimates of the average number of years of schooling that indicates the border between the two regimes well below 3.0, and relatively seldom above that value.

The results for 10,000 replications of the procedure are summarized by Figure 10.3. The diagram presents the frequency distribution of the estimated $n_1$s, that is, the size of the low-schooling subsample. This frequency distribution is in line with our expectations. Most observations are for $n_1 = 15$, a value found for approximately 20 per cent of the observations. Further, much of the distribution’s mass is found for smaller values of $n_1$, while relatively few replications yield high values. Actually, only about 17 per cent of the replications yield a low-schooling subsample that contains 16 or more countries.

Notes:
Horizontal axis: number of countries in low-schooling subsample.
Vertical axis: frequency (10,000 observations in total).

Figure 10.3  Frequency distribution of estimated low-schooling subsample size
As argued on pp. 240–41, the empirical distribution function allows us to construct confidence intervals for the threshold value. A two-sided 90 per cent confidence interval is found by eliminating the 500 lowest values of \( n_1 \) and its 500 highest values. For \( n_1 \), this yields the interval \([8, 17]\). The corresponding values of the average number of years of schooling are \([1.45, 3.23]\). This can be regarded as a rather wide interval. Of course, similarly constructed 95 per cent confidence intervals are even wider, \([7, 18]\) for \( n_1 \) and \([1.20, 3.32]\) for the educational attainment variable, respectively.

This analysis provides a first indication of the range within which the threshold value is located. In the literature on bootstrapping, it is stressed that the reliance on point estimators such as our least squares estimator of the threshold value and the coefficients of the two productivity dynamics equations in Model (1) is only warranted if these are unbiased. If not, bias correction procedures should be applied. At present, we do not have reasons to believe that our estimators are unbiased, but we should prove this conjecture by means of simulation analysis.

CONCLUSIONS

In this chapter, we offered a non-parametric method to test whether a very simple catching-up model of labour productivity growth should account for the possibility that multiple catching-up regimes co-exist. Further, we proposed a way to construct confidence intervals for the threshold value between two such regimes, by means of bootstrapping techniques. The methods were implemented for a small dataset on labour productivity dynamics for aggregate economies in the period 1960–2000. We did not find evidence for two or more regimes if the labour productivity gap to the world leader (the USA) was used as an indicator of social capabilities. For an educational attainment variable, however, the test pointed to at least two regimes. The confidence intervals for the threshold value of this schooling variable appeared to be rather wide, in particular towards the lower end of the schooling spectrum.

An attempt to find evidence for a second threshold showed that the method is not flawless if the number of observations is very limited. In that case, the results appeared to be very sensitive to the minimum number of observations contained in a subsample. The method does not endogenously decide on this argument, thus introducing some arbitrariness. More investigations into this issue might reveal that related literature on parametric suggests solutions to this problem.

Another caveat refers to the construction of confidence intervals. Implicitly, we assume that our point estimators are unbiased. Although we do not see a reason why this would not be the case, we should study the
validity of this assumption thoroughly. If the assumption would turn out to be untenable, we should leave our very simple and intuitive methodology and opt for a more complex alternative. The basic idea, however, could remain unchanged. This also holds for a test that could cope with heteroscedasticity.

In this chapter, we emphasized methodological issues, instead of the economic interpretation of results. For instance, we did not discuss the plausibility of the result that multiple regimes are found for schooling as a potential splitting or threshold variable, while the initial productivity gap did not give reason to multiple regimes. This is an interesting finding in itself, though, because more neoclassically-oriented studies into the existence of multiple productivity growth regimes found contrary results. A study which would apply our methodology to related datasets could tell more about the robustness of this result, as could application of our methodology to a different (probably more advanced) model.

A final question to be discussed is of a more fundamental nature. Is it natural to suppose that worldwide labour productivity growth could be regarded as being governed by a finite number of linear regimes, as we implicitly assume? Or should we adopt the viewpoint that there is a smooth continuum of ‘regimes’, that relate the productivity performance of countries to each other in a nonlinear way? Some very preliminary analyses on the same database, using varying coefficient models (for example, Hastie and Tibshirani, 1993) seem to suggest that this might prove a promising avenue of research. Nevertheless, we feel that refinement and application of the techniques proposed in this paper deserves attention in our future work.

ACKNOWLEDGEMENTS

Earlier versions of this paper were presented at EMAEE 2003 (Augsburg, 10–12 April) and a Technology and Economic Growth workshop (Maastricht, 11 June 2003). I would like to thank Bart van Ark, Paul Bekker, Marcel Timmer and Bart Verspagen for very useful suggestions. My research has been made possible by a grant from the Dutch Organization for Scientific Research NWO.

NOTES

1. This description is purposefully simplified. Since its inception by Solow (1956) the mainstream view has been modified in many respects, for instance by including human capital as a separate production factor that captures some aspects of innovation (for example, Mankiw et al., 1992). Moreover, many adopters of the technology gap perspective (Gerschenkron (1962) is generally seen as the classic study) feel that sufficient capital
good accumulation is a precondition for innovation and catch-up (for example, Fagerberg, 1988). Bernard and Jones (1996), Basu and Weil (1998) and Los and Timmer (2005) offer approaches that combine elements from both perspectives on convergence in different ways.

2. Abramovitz (1979) actually computed rank correlation coefficients between the left hand side and right hand side variables (although not standardized by the performance of the leader), and found an inverse relationship.

3. To our knowledge, Nelson and Phelps (1966) were the first to relate the availability of human capital to the speed of technology diffusion in the framework of a formal model. The importance of technological capabilities at the firm level was assessed empirically by Cohen and Levinthal (1989), who coined this concept ‘absorptive capacity’. Note that governmental policies can affect the availability of capabilities, as opposed to variables like geographical location found in many studies that try to explain productivity growth differentials.

4. If $a < 0$, follower countries would eventually overtake the productivity level of the initial leader. Although productivity leadership has historically changed a couple of times, it has done so only among a very limited number of countries. Hence, it may be appropriate to confine our theoretical analysis to non-negative values of the intercept.

5. Durlauf (2001) views careful treatment of nonlinearities (as suggested by modern mainstream growth theories) as one of the most promising avenues for research in applied growth econometrics.

6. Fagerberg and Verspagen (1996) used a simplified version of the Durlauf and Johnson (1995) methodology, in which only one splitting variable was considered at a time. Initial labour productivity and EU-funded R&D activities did not produce as large reductions in the sum of squared residuals as unemployment rates did.


8. See Hansen (2001), who also suggests this ‘trimming’ to save desirable distributional properties of his test statistic.

9. The test outlined below supposes that the errors are homoscedastic, that is, they have an identical variance. In growth regressions including developing countries, this assumption is hardly defensible. The $F$-test and the tests by Durlauf and Johnson (1995) and Hansen (2001) share the same weakness.

10. This corresponds to $X$ in Model (3).

11. This procedure could be refined, for instance by adoption of methodologies to account for differences between the estimate of the threshold value and its actual value. In this chapter, we do not consider these. Results should therefore be taken with some caution.

12. Verspagen (1991) opted for the alternative choice, by defining one of his social capability indicators as a weighted average of electricity generating capacities over years for which productivity growth rates were analysed.

13. Although this assertion is rather speculative, the under-representation of low-productivity countries might well lead to underestimation of the number of productivity growth regimes (in view of the trimming procedure, which implies that each subsample should at least contain a predetermined number of observations).

14. Figure 10.1 suggests that Zaïre (the observation in the extreme southwest of the diagram) might drive the result of multiple regimes. A robustness check in which this observation was excluded did not yield qualitatively different results. The threshold value remained identical and the $p$-value of the test decreased to below 0.01.

15. Interestingly, application of the procedure adopted by Cappelen et al. (1999) would have yielded strong evidence in favour of multiple initial gap-based regimes. Their Chow-test (2, 40 degrees of freedom) would yield a value of 8.355, which corresponds to a $p$-value of only 0.0009. Careful inspection of Figure 10.1 shows that $n_1 = 9$ implies that the last observation included in the ‘low-initial productivity’ subsample has the second most negative residual in the one-regime regression, while the first observation included in the ‘high-initial productivity’ subsample appears to yield the third most positive residual. The Chow-test is only valid if the threshold value is chosen independently from the data.
16. Again, the initial productivity gap threshold variable did not produce a significant split at reasonable levels of significance.

17. If one or more estimators would be biased, this would invalidate the choice of the distribution of observed residuals as an approximation for the distribution of unobserved errors.

REFERENCES


## APPENDIX A: SAMPLE CHARACTERISTICS

Countries are ordered according to their initial labour productivity gaps. Column headers refer to the following variables:


**SCHOOLING:** average number of schooling years of entire population aged 25 and over (source: Barro and Lee, 2001).

<table>
<thead>
<tr>
<th></th>
<th>INITGAP</th>
<th>GROWTH</th>
<th>SCHOOLING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kenya</td>
<td>−3.090541</td>
<td>−0.007369</td>
</tr>
<tr>
<td>2</td>
<td>Burma</td>
<td>−3.034502</td>
<td>0.002940</td>
</tr>
<tr>
<td>3</td>
<td>Zaïre</td>
<td>−3.032776</td>
<td>−0.051021</td>
</tr>
<tr>
<td>4</td>
<td>Bangladesh</td>
<td>−2.834400</td>
<td>−0.007212</td>
</tr>
<tr>
<td>5</td>
<td>India</td>
<td>−2.725251</td>
<td>0.006144</td>
</tr>
<tr>
<td>6</td>
<td>Pakistan</td>
<td>−2.591427</td>
<td>0.013305</td>
</tr>
<tr>
<td>7</td>
<td>Thailand</td>
<td>−2.569596</td>
<td>0.024413</td>
</tr>
<tr>
<td>8</td>
<td>Indonesia</td>
<td>−2.368618</td>
<td>0.007990</td>
</tr>
<tr>
<td>9</td>
<td>Ghana</td>
<td>−2.337197</td>
<td>−0.018392</td>
</tr>
<tr>
<td>10</td>
<td>Sri Lanka</td>
<td>−2.306328</td>
<td>0.028008</td>
</tr>
<tr>
<td>11</td>
<td>Korea</td>
<td>−2.022904</td>
<td>0.035356</td>
</tr>
<tr>
<td>12</td>
<td>Philippines</td>
<td>−1.863815</td>
<td>−0.007191</td>
</tr>
<tr>
<td>13</td>
<td>Taiwan</td>
<td>−1.812441</td>
<td>0.035236</td>
</tr>
<tr>
<td>14</td>
<td>Malaysia</td>
<td>−1.783009</td>
<td>0.016348</td>
</tr>
<tr>
<td>15</td>
<td>Turkey</td>
<td>−1.653433</td>
<td>0.015677</td>
</tr>
<tr>
<td>16</td>
<td>Brazil</td>
<td>−1.438171</td>
<td>0.000661</td>
</tr>
<tr>
<td>17</td>
<td>Singapore</td>
<td>−1.381946</td>
<td>0.027299</td>
</tr>
<tr>
<td>18</td>
<td>Portugal</td>
<td>−1.331659</td>
<td>0.015496</td>
</tr>
<tr>
<td>19</td>
<td>South Africa</td>
<td>−1.322980</td>
<td>−0.010448</td>
</tr>
<tr>
<td>20</td>
<td>Spain</td>
<td>−1.266535</td>
<td>0.022186</td>
</tr>
<tr>
<td>21</td>
<td>Japan</td>
<td>−1.259673</td>
<td>0.022714</td>
</tr>
<tr>
<td>22</td>
<td>Peru</td>
<td>−1.240228</td>
<td>−0.013421</td>
</tr>
<tr>
<td>23</td>
<td>Colombia</td>
<td>−1.239998</td>
<td>−0.001749</td>
</tr>
<tr>
<td></td>
<td>INITGAP</td>
<td>GROWTH</td>
<td>SCHOOLING</td>
</tr>
<tr>
<td>----</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>24</td>
<td>-1.188852</td>
<td>0.014757</td>
<td>4.64</td>
</tr>
<tr>
<td>25</td>
<td>-1.062227</td>
<td>0.022590</td>
<td>6.45</td>
</tr>
<tr>
<td>26</td>
<td>-0.976179</td>
<td>0.000298</td>
<td>2.41</td>
</tr>
<tr>
<td>27</td>
<td>-0.851216</td>
<td>0.014406</td>
<td>5.37</td>
</tr>
<tr>
<td>28</td>
<td>-0.840021</td>
<td>0.008762</td>
<td>6.99</td>
</tr>
<tr>
<td>29</td>
<td>-0.783847</td>
<td>0.001528</td>
<td>4.99</td>
</tr>
<tr>
<td>30</td>
<td>-0.779124</td>
<td>0.012089</td>
<td>6.71</td>
</tr>
<tr>
<td>31</td>
<td>-0.765787</td>
<td>0.013393</td>
<td>4.56</td>
</tr>
<tr>
<td>32</td>
<td>-0.735929</td>
<td>-0.004476</td>
<td>4.99</td>
</tr>
<tr>
<td>33</td>
<td>-0.652546</td>
<td>0.011030</td>
<td>4.74</td>
</tr>
<tr>
<td>34</td>
<td>-0.525499</td>
<td>0.005093</td>
<td>7.65</td>
</tr>
<tr>
<td>35</td>
<td>-0.518960</td>
<td>0.008504</td>
<td>6.11</td>
</tr>
<tr>
<td>36</td>
<td>-0.508081</td>
<td>0.010663</td>
<td>7.46</td>
</tr>
<tr>
<td>37</td>
<td>-0.485456</td>
<td>0.005024</td>
<td>8.95</td>
</tr>
<tr>
<td>38</td>
<td>-0.478826</td>
<td>0.008937</td>
<td>5.78</td>
</tr>
<tr>
<td>39</td>
<td>-0.462938</td>
<td>0.003472</td>
<td>7.67</td>
</tr>
<tr>
<td>40</td>
<td>-0.295109</td>
<td>0.001064</td>
<td>9.43</td>
</tr>
<tr>
<td>41</td>
<td>-0.210945</td>
<td>-0.003858</td>
<td>7.30</td>
</tr>
<tr>
<td>42</td>
<td>-0.189182</td>
<td>-0.002390</td>
<td>5.27</td>
</tr>
<tr>
<td>43</td>
<td>-0.178020</td>
<td>-0.008866</td>
<td>9.56</td>
</tr>
<tr>
<td>44</td>
<td>-0.141383</td>
<td>-0.002200</td>
<td>8.37</td>
</tr>
</tbody>
</table>