

On the Power of Absolute Convergence Tests¹

Rómulo A. Chumacero²

February 2002

¹I would like to thank Rodrigo Fuentes, William Easterly, and Klaus Schmidt-Hebbel for helpful comments and suggestions. Francisco Gallego provided able research assistance. The usual disclaimer applies.

²Department of Economics of the University of Chile and Research Department of the Central Bank of Chile. *E-mail address:* `rchumace@econ.uchile.cl`

Abstract

This paper analyzes whether or not the econometric methods usually applied to test for absolute convergence have provided a “fair” chance to this hypothesis. We show that traditional (absolute and conditional) convergence tests are not consistent with even the simplest model that indeed displays convergence. Furthermore, claims of divergence on the grounds of bimodalities in the distribution of per capita GDP can be made consistent with models in which neither divergence nor twin peaks are present in the long run.

Key Words: Economic Growth, Convergence, Bimodality.

JEL Classification: C15, C31, E13, O41, O50.

1 Introduction

With the possible exception of Mincerian regressions (Mincer, 1974), few other subjects in applied economic research have been as studied as the convergence hypothesis advanced by Solow (1956) and documented by Baumol (1986).¹

In simple terms, it states that poor countries or regions tend to grow faster than rich ones. In its strongest version (known as absolute convergence), an implication of this hypothesis is that, in the long run, countries or regions should not only grow at the same rate, but also have the same income levels.²

This hypothesis has been tested using different methodologies and data sets, and appears to be strongly rejected by the data. In view of these results, several modifications of the absolute convergence hypothesis have been advanced and tested. Nevertheless, they usually lack both theoretical foundations and econometric rigor and discipline.

This paper analyzes whether or not the econometric methods usually applied to test for absolute convergence have provided a “fair” chance to this hypothesis. The document is organized as follows: Section 2 presents a brief review of some of the tests for convergence advanced in the empirical literature and documents their shortcomings. Section 3 develops simple theoretical models that imply absolute convergence and discusses how likely would it be for time series generated from them to accom-

¹An admittedly incomplete list of representative studies of this line of research is Aghion and Howitt (1997), Barro (1991), Barro and Sala-i-Martin (1992), Mankiw et al (1992), Durlauf and Johnson (1995), Jones (1995), Kocherlakota and Yi (1996) and Kocherlakota and Yi (1997).

²This interpretation has been challenged by Bernard and Durlauf (1996).

moderate the results of tests such as the ones described on Section 2. Finally, Section 4 provides the conclusions.

2 Results from the Empirical Literature

This section presents a brief review of the main results found in empirical growth analysis in order to test the convergence hypothesis.

2.1 Absolute Convergence is Strongly Rejected

The first stylized fact that appears uncontroversial is that independently of the type of data set used (cross-section of countries or panel data), the data strongly rejects absolute convergence (Barro and Sala-i-Martin, 1995).

The simplest test that can be devised to verify this claim using cross-section observations takes the form:

$$g_i = \zeta + \vartheta \ln y_{i,0} + \varepsilon_i \tag{1}$$

where $y_{i,t}$ is the per capita GDP in period t for country i , and g_i is the average growth rate of per capita GDP of country i ; that is:

$$g_i = \frac{1}{T} \sum_{t=1}^T \Delta \ln y_{i,t} = \frac{1}{T} (\ln y_{i,T} - \ln y_{i,0})$$

If pooled data were used, tests for absolute convergence usually take the form

$$\Delta \ln y_{i,t} = \zeta + \vartheta \ln y_{i,t-1} + \varepsilon_{i,t} \quad (2)$$

In both cases, absolute convergence is said to be favored by the data if the estimate of ϑ is negative and statistically different from 0. If the null hypothesis ($\vartheta = 0$) is rejected, we would conclude that not only do poor countries grow faster than rich countries, but also that they all converge to the same level of per capita GDP.

As Table 1 and Figure 1 show, the convergence hypothesis is strongly rejected.³ In fact, if these results are taken seriously, the evidence appears to favor divergence instead of convergence. That is, the countries that grew faster were those that had a higher initial per capita GDP.

	Cross-section	Pooled Data
$\hat{\vartheta}$	0.0047 (0.0014)	0.0048 (0.0010)
\overline{R}^2	0.051	0.007
Observations	$N=116$	3219 ($N=85$)

Table 1: Tests for Absolute Convergence. \overline{R}^2 =Adjusted R². N =Number of countries. Standard errors consistent with heteroskedasticity in parenthesis.

A major weakness of these tests is that given that the null hypothesis being tested in both cases is that ϑ is equal to zero versus the alternative that it is negative, (2) makes explicit that a test for absolute convergence is essentially a test for a unit

³In the case of panel data, all tests were conducted using the latest version of the Penn World Table data set described in Summer and Heston (1991), with most variables ranging from 1960 to 1998. In the case of cross-section regressions, the tests were conducted using the data set described in Doppelhofer et al (2000).

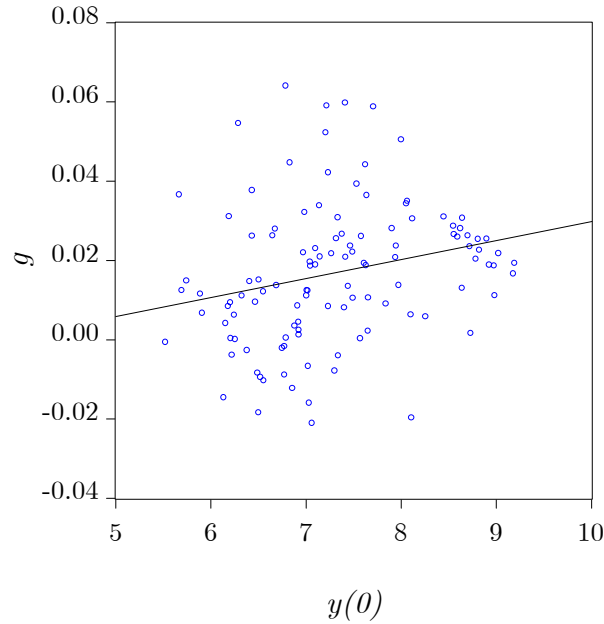


Figure 1: Growth rate from 1960 to 1998 versus 1960 per capita GDP

root on y . As is abundantly documented, these tests not only have non-standard asymptotic properties, but also lack of power.

In fact, if a traditional (Augmented Dickey-Fuller) unit root test on $\ln y$ were performed for each country, none would reject the null, at standard significance levels. Moreover, the first order autocorrelation of $\ln y$ for each country ranges from 0.610 to 0.999, with an average value of 0.947. These results suggest that even if a unit root were not present, $\ln y$ is extremely persistent, and initial conditions would take a long time to dissipate.

2.2 The Perils of Conditional Convergence

In light of the above results, Barro (1991) considered a modification of (1) in which even when convergence is still understood as the situation where poor countries grow faster than rich countries (unconditionally), their growth rate may be influenced by other factors that may prevent convergence in the levels of per capita GDP.

Tests for conditional convergence using cross-section observations usually take the form

$$g_i = \zeta + \vartheta \ln y_{i,0} + \varphi' x_i + \varepsilon_i \quad (3)$$

where x is a k -vector of variables that may influence growth. Given that the x variables are different for each country, even if ϑ were negative, levels might never converge.

Table 2 presents the results of running a regression that included some of the usual candidates for specifications such as (3) using both cross-section and panel data regressions.⁴

As noted by Durlauf (2001), serious problems plague this strategy. First, as economic theory is usually silent with respect to the set of x variables to be included,

⁴The model that uses cross-sectional observations included the following x variables: life expectancy in 1960 (+), equipment investment (+), years of open economy (+), a ‘rule of law’ index (+), a dummy variable for Sub-Sahara African countries (-), and fraction of people that profess the Muslim (+), Confucian (+), and Protestant (-) religions. The model that uses panel data was estimated using fixed effects and considered the following x variables: investment to GDP ratio (+), growth rate of the population (-), exports plus imports to GDP ratio (+), liquid liabilities to GDP ratio (-), inflation rate (-), and government consumption to GDP ratio (-).

	Cross-section	Panel Data
$\hat{\vartheta}$	-0.0154 (0.0028)	-0.0456 (0.0062)
\overline{R}^2	0.811	0.181
Observations	$N=79$	2552 ($N=85$)

Table 2: Tests for Conditional Convergence. \overline{R}^2 =Adjusted R². N =Number of countries. Standard errors consistent with heteroskedasticity in parenthesis.

empirical studies have often abused in terms of the potential candidates used; Durlauf and Quah (1999) report that as of 1998, over 90 different variables had appeared in the literature, despite the fact that no more than 120 countries are available for analysis in the standard data sets. Second, important biases in the results may be due to the endogeneity of most of the control variables used (Cho, 1996). Third, the estimated coefficients of the “convergence” parameter (ϑ) are rather small, suggesting that even after controlling for the x variables, $\ln y$ continues to be extremely persistent. Fourth, as a corollary of the previous observation, initial conditions may play a crucial role in the results. Fifth, the robustness of results in terms of the potential “determinants” of long-run growth is subject to debate (see, for example, Levine and Renelt, 1992; Sala-i-Martin, 1997; and Doppelhofer et al, 2000). Finally, several of the variables included in the x vector are fixed effects that can not be modified; if these variables were actually long-run determinants of growth, convergence would never be achieved (even with $\vartheta < 0$).⁵

⁵A curious example of a variable that satisfies this characteristic is “absolute latitude”, which measures how far a country is from the Equator. When statistically significant, its coefficient is usually positive, thus implying that a growth enhancer would be for a country to move its population to the North or South Pole.

2.3 Clubs

Durlauf and Johnson (1995) suggest that cross-section growth behavior may be determined by initial conditions. They explore this hypothesis using a regression tree methodology, which turns out to be a special case of a threshold regression (Hansen, 2000). The basic idea is that the level of per capita GDP on which each country converges depends on some initial condition (such as initial per capita GDP) and that, depending on this characteristic, some countries converge on one level and others converge on another. A common specification that is used to test this hypothesis considers a modification of (1) that takes the form:

$$g_i = \begin{cases} \zeta_1 + \vartheta_1 y_{i,0} + \varepsilon_i & \text{if } y_{i,0} < \varkappa \\ \zeta_2 + \vartheta_2 y_{i,0} + \varepsilon_i & \text{if } y_{i,0} \geq \varkappa \end{cases} \quad (4)$$

where \varkappa is a threshold that determines whether or not country i belongs to the first or second regime. In this case, convergence would not be achieved if the whole sample is taken into consideration, but it would be achieved between members of each group.

If (4) were the actual Data-Generating-Process (DGP), results such as the ones obtain in Table 1 could be easily motivated, given that if two regimes were present, with each regime converging to a different state and at a different rate, estimations based on a single regime may produce a non-significant estimate for the convergence parameter.

On the other hand, (4) states that if the threshold variable (in this case, the initial

per capita GDP) is correlated with some of the x variables included in (3) results such as those reported in Table 2 are likely to be encountered, even if the x variables are not (necessarily) determinants of long-run growth.

However, (4) has an unequivocal implication in terms of the distribution of per capita GDP across countries; if the parameters that characterize each regime are different, a threshold process should be consistent with a bimodal distribution for $\ln y$.

Quah (1993) and Quah (1997) noticed that the relative per capita GDP (defined as the ratio of the per capita GDP of country i with respect to average World per capita GDP; which we represent by $\tilde{Y}_{i,t}$) displayed such bimodality. He conjectured that if “clubs” of convergence were present, even if the unconditional distribution of the initial per capita GDP were unimodal, the existence of such clubs would imply that countries would not converge to a degenerate distribution in the long run (as absolute convergence would seem to imply) but that one group may converge to a level of per capita GDP and another group to other, in which case twin peaks would arise.

Figure 2 presents kernel estimators of the unconditional density of relative per capita GDP in 1960 and 1995. Consistent with Quah’s claim, twin peaks are present in 1995; however, a bimodal distribution also appears to be present in 1960. If Quah were right, rich countries would converge to one distribution while initially poor countries would never be able to catch-up and would converge to a distribution with

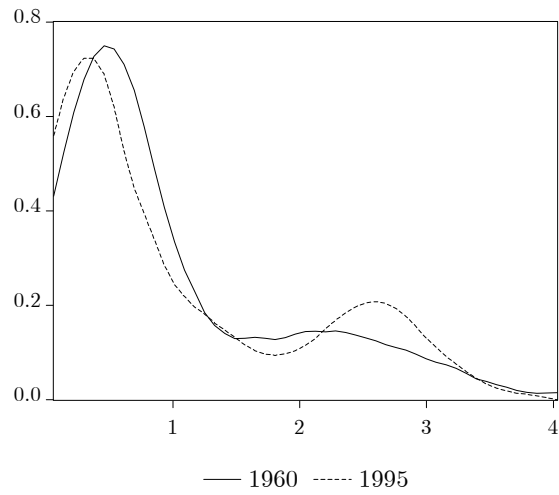


Figure 2: Densities of relative per capita GDP

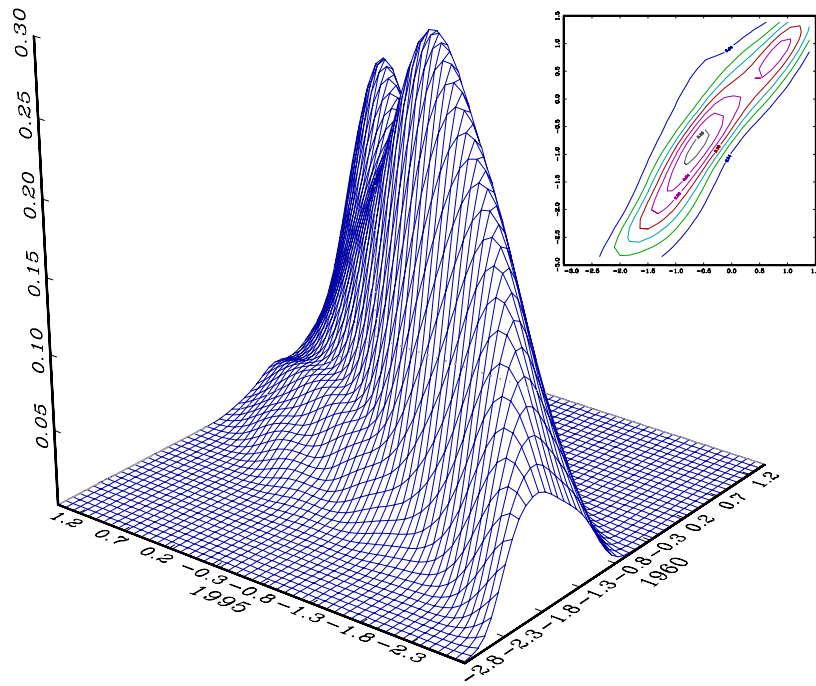


Figure 3: Surface and contour plots of (log of) relative per capita GDP

a permanently lower per capita GDP. On the other hand, Figure 3 presents surface and contour plots of the (log of) relative per capita GDP, which shows that a bimodal joint density does indeed appear to be consistent with the data.

A problem with this approach is that in contrast to (4), no formal test of this theory can be provided with this visual evidence. Quah (1993) tries to formalize the twin peak hypothesis by deriving the ergodic distribution of the transition matrix of relative incomes among countries.

	$\tilde{Y}_{t+1} \leq \frac{1}{4}$	$\frac{1}{4} < \tilde{Y}_{t+1} \leq \frac{1}{2}$	$\frac{1}{2} < \tilde{Y}_{t+1} \leq 1$	$1 < \tilde{Y}_{t+1} \leq 2$	$\tilde{Y}_{t+1} > 2$
$\tilde{Y}_t \leq \frac{1}{4}$	0.973	0.027	0	0	0
$\frac{1}{4} < \tilde{Y}_t \leq \frac{1}{2}$	0.047	0.927	0.026	0	0
$\frac{1}{2} < \tilde{Y}_t \leq 1$	0	0.035	0.948	0.017	0
$1 < \tilde{Y}_t \leq 2$	0	0	0.018	0.949	0.033
$\tilde{Y}_t > 2$	0	0	0	0.017	0.983
Ergodic	0.312	0.177	0.133	0.127	0.251

Table 3: One-year transition matrix and ergodic distribution: 1960-1995

Table 3 presents estimates of the one-year transition matrix of \tilde{Y} and its ergodic distribution. The results indicate the high persistence of the series, given that the main diagonal has transition probabilities that always exceed 0.9. More importantly, with the sample analyzed, the ergodic distribution does appear to be bimodal in the sense that (unconditionally) higher probabilities are attached for countries that have less than one-quarter of average world per capita GDP or more than twice this average.

However, this distribution is highly nonlinear and extremely noisy (Kremer et al, 2000). The resulting ergodic distribution is sensitive to the choice of thresholds for

each category, the number of years to compute the transition matrix, and the variable used to perform the comparisons.⁶ More fundamentally, given that the initial distribution is also bimodal, it is difficult to assess whether or not the bimodal distribution obtained is due to the presence of twin peaks or if it arises because of the persistence of the per capita GDP level.

3 A Simple Model with Absolute Convergence

This section presents a simple exogenous growth model in which absolute convergence holds, and asks whether or not the tests for convergence presented in the previous section would be robust. That is, if time series realizations were generated using a model in which convergence holds, would tests for convergence be consistent with it? Simply put, the models that we will discuss imply that:

- countries should converge to a stationary distribution,
- countries with initially lower GDP should grow faster,
- and no twin peaks should be present in the long run.

To clarify concepts, we next present the type of model that we will use, describe its properties, and the DGP that $\ln y$ would obey, and ask whether the tests discussed in the previous section are really tests for convergence.

⁶Kremer et al. (2000) consider that a better choice of variable for constructing the transition matrix is the ratio of each country's per capita GDP to the average per capita GDP of the five leading countries or the leading country.

3.1 The Model

The representative, infinitely-lived household maximizes

$$\mathcal{U}_0 = \mathcal{E}_0 \sum_{t=0}^{\infty} \beta^t L_t^\theta \frac{c_t^{1-\gamma} - 1}{1-\gamma}$$

where $0 < \beta < 1$ is the subjective discount factor, c_t ($=C_t/L_t$) is per capita consumption,⁷ $\gamma > 0$ is the Arrow-Pratt relative risk aversion coefficient, and \mathcal{E}_t is the expectation operator conditional on information available for period t . There is no utility for leisure and the labor force is equal to L_t .⁸ Utility is maximized with respect to per capita consumption, and per capita capital stock, k_{t+1} , subject to the budget constraint:

$$K_{t+1} + C_t = e^{z_t} K_t^\alpha [(1 + \lambda)^t L_t]^{1-\alpha} + (1 - \delta) K_t$$

where α is the compensation for capital as a share of GDP. In this economy, technological progress is labor-augmenting and occurs at the constant rate λ . Note that production is affected by a stationary productivity shock z_t . It is straightforward to show that capital and consumption per unit of effective labor, \widehat{k}_t and \widehat{c}_t are stationary.⁹ In fact, we can transform the economy above to a stationary economy and

⁷Lower-case letters denote per capita; upper-case total; and a hat above a variable denotes per unit of effective labor.

⁸The parameter $0 \leq \theta \leq 1$ is included, because this feature allows us to consider dynastic agents with endogenous fertility decisions (see Barro and Becker, 1989; Becker et al, 1990; or Razin and Sadka, 1995).

⁹ $\widehat{k}_t = k_t / (1 + \lambda)^t$ and $\widehat{c}_t = c_t / (1 + \lambda)^t$.

obtain exactly the same solutions for \widehat{k}_t and \widehat{c}_t . Such an economy can be characterized by the following maximization problem:

$$\max_{\{\widehat{k}_{t+1}, \widehat{c}_t\}} \mathcal{E}_0 \sum_{t=0}^{\infty} [\beta(1+\lambda)^{1-\gamma}]^t L_t^\theta \frac{\widehat{c}_t^{1-\gamma} - 1}{1-\gamma} \quad (5)$$

subject to

$$(1 + \eta_{t+1}) (1 + \lambda) \widehat{k}_{t+1} + \widehat{c}_t = e^{z_t} \widehat{k}_t^\alpha + (1 - \delta) \widehat{k}_t \quad (6)$$

where η_t is the rate of population growth for period t .

Given that this model will be used to compare the dynamics of different economies, following den Haan (1995), we include a simple channel to induce correlation between each economy's income. Specifically, we obtain correlated incomes by assuming that the law of motion of technology shock in country i can be written as

$$z_{i,t} = \rho z_{i,t-1} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} = (1 - \phi) v_t + \phi w_{i,t} \quad (7)$$

where the v_t and $w_{i,t}$ are independent $\mathcal{N}(0, \sigma_i^2)$ random variables (for $i = v, w$). If ϕ is equal to zero, then all countries face the same aggregate shock; and if ϕ is equal to one, each country faces only an idiosyncratic shock.

In order for the model to be fully characterized, a stance regarding the rate of population growth has to be taken. Here we will consider the case in which fertility is

exogenous and has the following law of motion:

$$\ln(1 + \eta_{i,t}) = \bar{\eta}(1 - \tau) + \tau \ln(1 + \eta_{i,t-1}) + n_{i,t} \quad (8)$$

where $n_{i,t}$ is an independent $\mathcal{N}(0, \sigma_n^2)$ random variable.¹⁰

Once values for the preference and technology parameters are chosen, this dynamic programming problem can be solved using numerical methods to generate artificial realizations of the variables of interest. In our case, we are interested in generating realizations of per capita GDP for several samples of “countries” and applying the convergence tests discussed in Section 2. As we will show below, this model implies convergence (in a sense to be defined below). Our goal is to evaluate how likely is it for the tests to conclude otherwise, even though the main feature of this model is that countries converge.

3.2 Convergence Tests and the Model

In order to understand if tests discussed in Section 2 are useful to test for convergence, we tailor our model to instances in which a closed form expression for the DGP of the log of per capita GDP is available. We argue that this simplification imposes a very rigid structure on the theoretical model, and makes it harder for its realizations to present the features considered signs of rejection of the absolute convergence

¹⁰If fertility is considered as endogenous, (8) can be ignored, and (5) may be used in order to consider dynastic models as in Razin and Sadka (1995).

hypothesis.

If $\gamma = 1$, $\theta = 1$, and $\delta = 1$, the dynamic programming problem maximizing the objective function (5) has logarithmic preferences subject to a Cobb-Douglas constraint (6), in which case an analytical expression for the capital stock policy function is available and is expressed as:

$$\ln \widehat{k}_{t+1} = \ln(\alpha\beta) - \ln(1 + \lambda) + \ln \widehat{y}_t \quad (9)$$

where $\widehat{y}_t = e^{z_t} \widehat{k}_t^\alpha$ is the per unit of effective labor GDP.

Because $\ln \widehat{y}_t$ can be expressed as:

$$\ln \widehat{y}_t = z_t + \alpha \ln \widehat{k}_t \quad (10)$$

we can replace (7) and (9) in (10) to obtain a simple expression for \widehat{y}_t :

$$\ln \widehat{y}_{i,t} = A + (\alpha + \rho) \ln \widehat{y}_{i,t-1} - \alpha \rho \ln \widehat{y}_{i,t-2} + \varepsilon_{i,t} \quad (11)$$

where $A = \alpha(1 - \rho)[\ln(\alpha\beta) - \ln(1 + \lambda)]$. Recalling that $\widehat{y}_{i,t}(1 + \lambda)^t = y_{i,t}$ we can use (11) to obtain a compact representation of the DGP of per capita GDP as follows:

$$\ln y_{i,t} = B + Dt + (\alpha + \rho) \ln y_{i,t-1} - \alpha \rho \ln y_{i,t-2} + \varepsilon_{i,t} \quad (12)$$

with B and D being constants.¹¹

Four features of (12) are worth mentioning: First, as is typical of exogenous growth models, per capita GDP is trend stationary.¹² Second, given that the technology shock follows an AR(1) process, $\ln y$ follows an AR(2) process.¹³ Third, even without exogenous growth ($\lambda = 0$), an AR(1) process for $\ln y$ such as (2) is consistent with (12) only if white-noise technology shocks ($\rho = 0$) are present. Finally, this model suggests that convergence on growth rates and GDP levels should eventually be achieved. The type of convergence on GDP levels would depend on the characteristics of the aggregate and idiosyncratic shocks that are present in (7). In particular, if the only source of variation in technology shocks is the aggregate shock ($\phi = 0$), all countries should eventually converge on the same per capita GDP level, independently of their initial conditions and independently of the persistence of z . On the other hand, if at least part of the variation in technology shocks is due to the idiosyncratic component ($\phi > 0$), per capita GDP levels would converge to a non-degenerate distribution that does not display a mass point. That is, $\ln y$ would converge to a normal distribution with positive variance; in which case, the probability of observing identical levels of y would be zero.

Next, we focus on the implications of different parameterizations of (12) for the convergence tests discussed in Section 2.

¹¹More precisely, $B = \alpha(1 - \rho) \ln(\alpha\beta) + \rho(1 - \alpha) \ln(1 + \lambda)$ and $D = (1 - \alpha)(1 - \rho) \ln(1 + \lambda)$.

¹²In fact, a case for divergence can only be made when $\ln y$ has a unit root. For that to be the case, we need either $\rho = 1$ (a unit root in the technology shock) or $\alpha = 1$ (a model of endogenous growth of the AK type).

¹³In general, if the productive shocks follow an AR(j) process, $\ln y$ follows an AR($j + 1$) process.

3.2.1 Independently and Identically Distributed Shocks

The only instance in which an absolute convergence test such as (2) is correctly specified is when the technology shocks are i.i.d., given that in that case (12) reduces to

$$\ln y_{i,t} = \alpha \ln(\alpha\beta) + (1 - \alpha) \ln(1 + \lambda)t + \alpha \ln y_{i,t-1} + \varepsilon_{i,t} \quad (13)$$

Thus, independently of the initial distribution of per capita GDP levels and population growth rates, $\hat{\vartheta}$ in (2) will consistently estimate the coefficient $\alpha - 1$ and convergence should occur.¹⁴

Figure 4 presents the empirical distribution of $\hat{\vartheta}$, computed from artificial samples of countries. Each sample consists of 100 countries and the initial per capita GDP is obtained from bootstrapping realizations of per capita GDP in 1960. Based on these initial conditions, values of $\ln y_{i,t}$ are simulated from (13) for a 36-year period. Finally, for each sample an estimate for ϑ was obtained by running a regression like (1).¹⁵ Obviously, the probability of obtaining estimates of $\hat{\vartheta}$ consistent with the results from Section 2 is 0. This is because even if we take per capita GDP distribution in 1960 as the initial condition, i.i.d. shocks with realistic figures for α are unable to produce enough persistence in $\ln y$.

¹⁴That is, $\hat{\vartheta}$ should be negative and statistically different from zero, provided that $0 < \alpha < 1$. Of course, (2) should also include a deterministic trend.

¹⁵The parameter values for this model were set as follows: $\alpha = 0.35$, $\beta = 0.96$, $\lambda = 0$, $\phi = 1$, and $\sigma_w^2 = 0.05^2$.

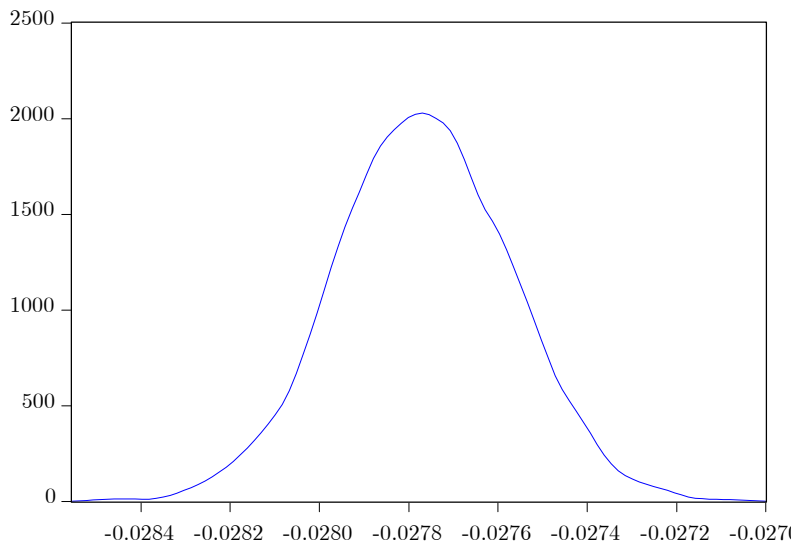


Figure 4: Absolute convergence tests with i.i.d. shocks: empirical distribution of the $\widehat{\vartheta}$ coefficients obtained with 2000 artificial samples for 100 countries.

Furthermore, the precise nature of absolute convergence will be dictated by ϕ . If $\phi = 0$, in the long-run, countries would converge (in probability) to the same per capita GDP; while if some shocks are idiosyncratic, in the long run, per capita GDP converges to a nondegenerate distribution.

Figures 5 and 6 reveal another characteristic of i.i.d. productivity shocks; even when they begin with a bimodal distribution for the initial per capita GDP, as y is not persistent enough, the bimodality quickly disappears. In fact, after 36 years, per capita GDP would not feature twin peaks.

A main feature of this model is that once initial conditions have dissipated (something that will occur rapidly in this case), $\ln y_{i,t}$ will be normally distributed. It turns out that in this case, distribution moments can be derived analytically. In particular,

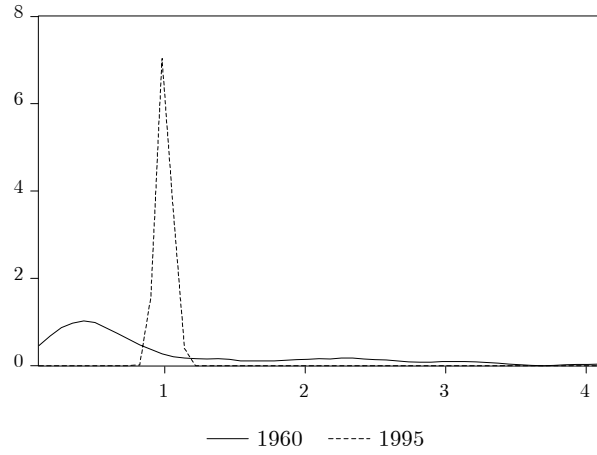


Figure 5: Densities of relative per capita GDP with i.i.d. shocks: empirical densities for an artificial realization of 100 countries.

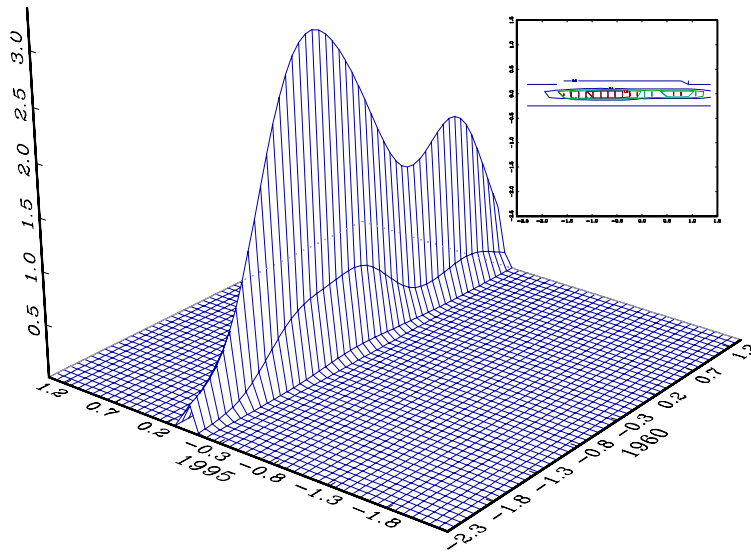


Figure 6: Surface and contour plots of (log of) relative per capita GDP for i.i.d. shocks: results for an artificial realization of 100 countries.

if μ_t and b represent the limits of the mean and variance of $\ln y_{i,t}$ we have

$$\mu_t = \frac{\alpha \ln(\alpha\beta) + (1 - \alpha) \ln(1 + \lambda) t}{1 - \alpha}, \quad b = \frac{\sigma_\varepsilon^2}{1 - \alpha^2}$$

Thus, given that $\ln y_{i,t}$ is normal, $y_{i,t}$ will be log-normal with $\mathcal{E}[y_{i,t}] = \exp(\mu_t + 0.5b)$.

Furthermore, \tilde{Y}_i (the ratio between y_i and $\mathcal{E}[y_{i,t}]$) will be unconditionally log-normal and its first two moments will be:

$$\mathcal{E}(\tilde{Y}_i) = 1, \quad \mathcal{V}(\tilde{Y}_i) = e^b - 1 \quad (14)$$

Obtaining the unconditional (ergodic) probabilities of \tilde{Y}_i for each of the categories described on Table 3 can be accomplished by noticing that

$$\Pr[\tilde{Y}_i \leq j] = \Pr[\ln \tilde{Y}_i \leq \ln j] = \Pr\left[\frac{\ln \tilde{Y}_i + 0.5b}{\sqrt{b}} \leq \frac{\ln j + 0.5b}{\sqrt{b}}\right]$$

but

$$\frac{\ln \tilde{Y}_i + 0.5b}{\sqrt{b}} \xrightarrow{D} \mathcal{N}(0, 1)$$

thus, the probability that \tilde{Y}_i does not exceed j can easily be computed by evaluating $\Phi\left(\frac{\ln j + 0.5b}{\sqrt{b}}\right)$; where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal variable. Thus, with i.i.d. shocks, the shape of the unconditional distribution of \tilde{Y}_i and its ergodic probabilities depend solely on b , which in turn is a function of

technology shock volatility and the persistence of $\ln y_i$ (which is α , capital's share of total output).

As Table 3 proves, given the one-year transition matrix estimated with the available data, the ergodic distribution of \tilde{Y}_i appears to be both bimodal and strongly asymmetric, in the sense that (unconditionally) the median of \tilde{Y}_i is close to 0.5 and not to the mean (which is, by construction, 1). Of course, the log-normal distribution is asymmetric, thus a simple way to verify if i.i.d. shocks are able to display such a degree of asymmetry is, given a value for b , to solve for the value of j that satisfies

$$\Phi\left(\frac{\ln j + 0.5b}{\sqrt{b}}\right) = \frac{1}{2} \quad (15)$$

But, as $\frac{\ln j + 0.5b}{\sqrt{b}}$ is asymptotically normal, and $\Phi(0) = \frac{1}{2}$, the value of j that solves (15) is

$$j = \exp\left(-\frac{b}{2}\right) = \exp\left(-\frac{\sigma_\varepsilon^2}{2(1-\alpha^2)}\right)$$

Figure 7 shows that a median close to $\tilde{Y} = 0.5$ can only be obtained with extremely volatile technology shocks ($\sigma_\varepsilon > 0.3$) or an unrealistic capital share over total GDP ($\alpha > 0.7$). In conclusion, i.i.d. shocks are inconsistent with the data, and if actual economies resembled this characterization, the probability of observing the evidence documented in Section 2 would be virtually nil.

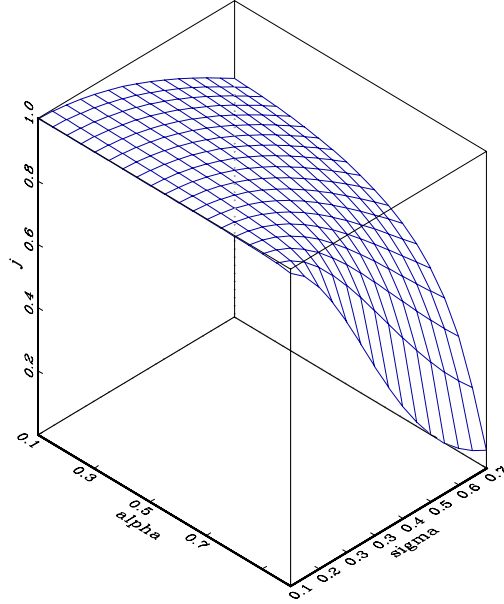


Figure 7: Median of \tilde{Y} for different values of α and σ_ε with i.i.d. shocks

3.2.2 Persistent Shocks

Once we abandon the unrealistic set-up of i.i.d. technology shocks, we can obtain significant persistence for $\ln y$ by choosing a value of ρ close to 1. Persistence of technology shocks is routinely invoked in the RBC literature and is broadly consistent with key stylized facts of modern economies. Once persistence in $\ln y$ is obtained, without having to resort to unrealistic values of α , the conclusions we reach regarding i.i.d. shocks change radically.

Remember that the law of motion of the univariate representation for $\ln y_{i,t}$ is

expressed by (12), that is,

$$\ln y_{i,t} = B + Dt + (\alpha + \rho) \ln y_{i,t-1} - \alpha\rho \ln y_{i,t-2} + \varepsilon_{i,t}$$

One immediately notices that convergence tests such as (2) are misspecified. Furthermore, as demonstrated by den Haan (1995), the estimated value of ϑ in (1) will be inconsistent and biased towards 0. That is, even if the model implied convergence, the estimated value of ϑ would be biased towards the rejection of this hypothesis. Furthermore, if pooled observations were used in (2), we would find that

$$\widehat{\vartheta} \xrightarrow{p} \psi - 1 = -\frac{(1 - \alpha)(1 - \rho)}{1 + \alpha\rho}$$

where $\psi = (\alpha + \rho) / (1 + \alpha\rho)$ is the first order autocorrelation of $\ln y$. This implies that the more persistent the technology shocks, the closer the probability limit of $\widehat{\vartheta}$ will be to 0.

Figure 8 presents a similar exercise to the one reported in Figure 4 for the i.i.d. case. Here, we consider exactly the same parameterization, but now we set $\rho = 0.97$. The difference is that even when the model implies convergence, the results of estimating equation (1) by bootstrapping the initial distribution of $\ln y$ that was observed in 1960 presents a non negligible probability (11%) that the estimated coefficient would indeed be positive (implying divergence).

Furthermore, as Figure 9 reveals, persistent technology shocks can replicate a

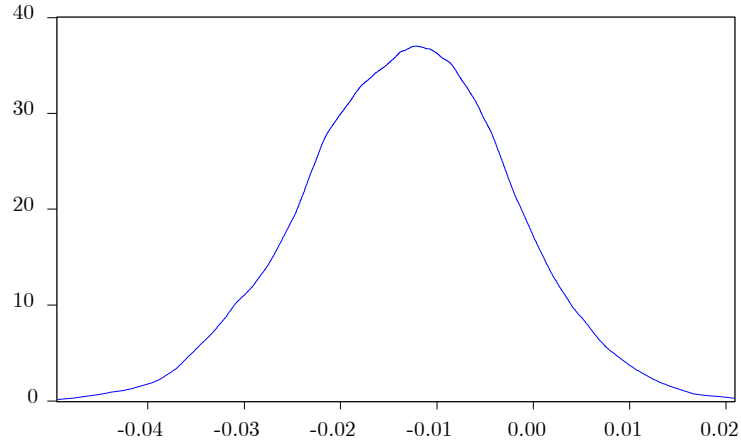


Figure 8: Absolute convergence tests with AR(1) shocks: empirical distribution of the $\hat{\vartheta}$ coefficients obtained with 2000 artificial samples for 100 countries.

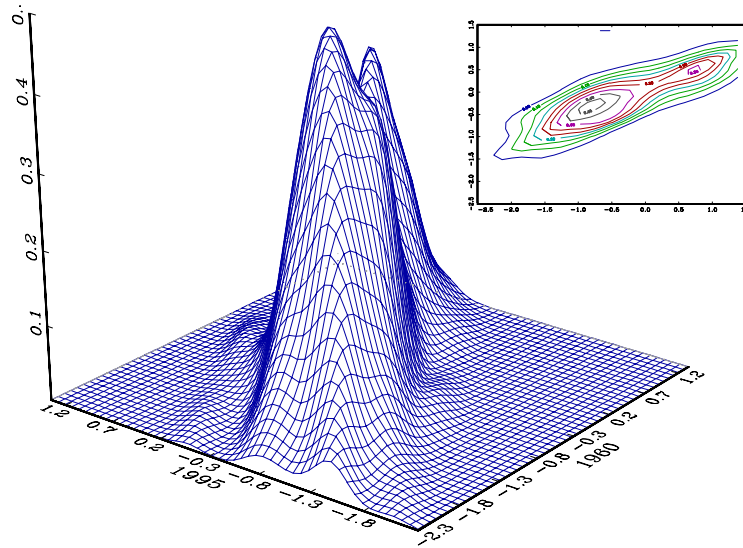


Figure 9: Surface and contour plots of (log of) relative per capita GDP for AR(1) shocks: results for an artificial realization of 100 countries.

bimodal joint distribution of the initial (log of) per capita GDP (consistent with the one observed in 1960) and the figures that would be obtained 35 years later. As initial conditions do not dissipate as fast as in the i.i.d. case, an initially bimodal distribution would persist even over long periods of time. Thus, bimodality in the “short run” is not inconsistent with a model that displays convergence in the long run.

As this model also displays convergence, $\ln y_{i,t}$ will be normal with the following mean and variance:

$$\mu_t = \frac{B + Dt}{(1 - \alpha)(1 - \rho)}, \quad b = \frac{\sigma_\varepsilon^2 (1 + \alpha\rho)}{(1 - \alpha\rho)(1 - \alpha - \rho + \alpha\rho)(1 + \alpha + \rho + \alpha\rho)} \quad (16)$$

Thus, the unconditional distribution of \tilde{Y} will still be log-normal with mean and variance given by (14), but b in this case is given by (16). We can conduct an identical experiment to the one reported in Figure 7, but now we set the value of α to 0.35 and let ρ and σ_ε vary. The results of this exercise are presented in Figure 10, which shows that the median of the unconditional distribution of \tilde{Y} can be set close to 0.5 with extremely persistent and moderately volatile technology shocks.

In summary, persistent technology shocks can be broadly consistent with the evidence reported in Section 2, in the sense that whatever the initial conditions of the distribution of per capita GDP are, they will fade slowly. In particular, this simple model, which displays convergence to a unimodal distribution in the long run, will be consistent with twin peaks in the distribution of per capita GDP, even over relatively

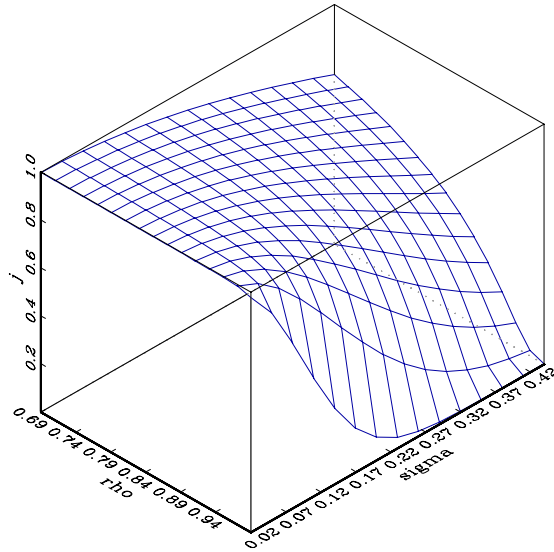


Figure 10: Median of \tilde{Y} for different values of ρ and σ_ε with AR(1) shocks

prolonged horizons. Furthermore, the asymmetry in the ergodic probabilities derived from the one-year transition matrix is characteristic of any log-normal distribution and is not (by itself) a proof of divergence.

3.2.3 The Model and Conditional Convergence

Once persistent shocks are allowed, even the simplest of the exogenous growth models can display several of the features that are considered evidence of divergence or club convergence. Thus, given an initially bimodal distribution of (the log of) per capita GDP, persistence by itself could generate an illusion of bimodality for prolonged periods.

However, the models just described are not consistent with evidence of conditional convergence. This is so because a few lags added to an equation like (2) would become

sufficient statistics for $\ln y$, and no other variable in the econometrician's information set should be informative. Nevertheless, the results of conditional convergence (statistically significant x variables) can be found when a misspecified law of motion for $\ln y$ is considered. In particular, if some x variables are correlated with the initial distribution of y , models that do not include as many lags of the variable as necessary can easily be found to be significant.

Furthermore, the models just discussed are among the simplest that can be generated from our theoretical model. In particular, if θ is different from 1, the population growth rate would become a determinant of $\ln y$; in such a case, even if $\ln \eta$ is stationary (a fact supported by the data), its exclusion from growth regressions could generate results consistent with conditional convergence, provided that technology shocks and population growth are persistent and that the x variables chosen correlate with initial conditions. In fact, as we stressed in Section 2, most of the "robust" x variables that are included in growth regressions are both persistent and strongly correlated with initial conditions.

Of course, if the economy is better characterized using parameters that do not allow for an analytical solution for the law of motion of $\ln y$, equations (1) and (2) can, at best, be viewed as linear approximations. The more nonlinear the model, the more inaccurate this approximation will be, and any nonlinear terms omitted may be approximated by any x variable that is correlated with the initial conditions.

4 Concluding Remarks

This paper takes issue with the interpretation of cross-country growth models that contend that the convergence hypothesis is strongly rejected by the data. We show that even the simplest exogenous growth model that displays absolute convergence in the long run can present several features that are argued to be evidence against convergence.

In particular, if persistent and moderately volatile productivity shocks are allowed, exogenous growth models can display features such as bimodality and asymmetries in the unconditional distribution of relative per capita GDP. Furthermore, there is a non-negligible probability that misspecified econometric models reject absolute convergence even when it is present.

Nevertheless, it is important to mention that persistence of technology shocks is not enough to generate these results. In this case, persistence implies that initial conditions will eventually dissipate, and if bimodality were present in a given period, it would not dissipate for long periods of time.

Furthermore, simple (and realistic) variations of the models presented, that ultimately imply convergence, can be made consistent with conditional convergence results, provided that the “determinants of growth” chosen are correlated with initial conditions and that the models being tested are misspecified (incorrect law of motion of per capita GDP or omission of nonlinearities).

It is only fair to mention that this paper does not explain the initial bimodality

that appears to be present in the data. It may well be the case that apparently relevant policy variables in conditional convergence regressions have something to do with this. In line with McGrattan and Schmitz (1999), distortionary policies may be behind this, but this model implies that if distortions are at fault, convergence to an ergodic distribution of per capita GDP should be achieved if these policies also do.

References

- Aghion, P. and P. Howitt (1997). *Endogenous Growth Theory*. The MIT Press.
- Barro, R. (1991). “Economic Growth in a Cross-Section of Countries,” *Quarterly Journal of Economics* 106, 407-43.
- Barro, R. and G. Becker (1989). “Fertility Choice in a Model of Economic Growth,” *Econometrica* 57, 481-501.
- Barro, R. and X. Sala-i-Martin (1992). “Convergence,” *Journal of Political Economy* 100, 223-51.
- Barro, R. and X. Sala-i-Martin (1995). *Economic Growth*. McGraw Hill.
- Baumol, W. (1986). “Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show,” *American Economic Review* 75, 1072-85.
- Becker, G., K. Murphy, and R. Tamura (1990). “Human Capital, Fertility, and Economic Growth,” *Journal of Political Economy* 98, S12-S37.
- Bernard, A. and S. Durlauf (1996). “Interpreting Tests of the Convergence Hypothesis,” *Journal of Econometrics* 71, 161-73.
- Cho, D. (1996). “An Alternative Interpretation of Conditional Convergence Results,” *Journal of Money, Credit, and Banking* 28, 669-81.

- den Haan, W. (1995). "Convergence in Stochastic Growth Models. The Importance of Understanding Why Income Levels Differ," *Journal of Monetary Economics* 35, 65-82.
- Doppelhofer, G., R. Miller, and X. Sala-i-Martin (2000). "Determinants of Long-term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *Working Paper 7750*, National Bureau of Economic Research.
- Durlauf, S. (2001). "Manifesto for a Growth Econometrics," *Journal of Econometrics* 100, 65-9.
- Durlauf, S. and P. Johnson (1995). "Multiple Regimes and Cross-Country Growth Behavior," *Journal of Applied Econometrics* 10, 365-84.
- Durlauf, S. and D. Quah (1999). "The New Empirics of Economic Growth," in J. Taylor and M. Woodford (eds.) *Handbook of Macroeconomics*. North Holland.
- Easterly, W., M. Kremer, L. Pritchett, and L. Summers (1993). "Good Policy or Good Luck? Country Growth Performance and Temporary Shocks," *Journal of Monetary Economics* 32, 459-83.
- Hansen, B. (2000). "Sample Splitting and Threshold Estimation," *Econometrica* 68, 575-603.
- Jones, C. (1995). "Time Series Tests of Endogenous Growth Models," *Quarterly Journal of Economics* 110, 495-525.

- Kocherlakota, N. and K. Yi (1996). "A Simple Time Series Test of Endogenous vs. Exogenous Growth Models," *Review of Economics and Statistics* 78, 126-34.
- Kocherlakota, N. and K. Yi (1997). "Is There Endogenous Long-Run Growth? Evidence from the United States and the United Kingdom," *Journal of Money, Credit, and Banking* 29, 235-62.
- Kremer, M., A. Onatski, and J. Stock (2000). "Searching for Prosperity," *Manuscript*. Harvard University.
- Levine, R. and D, Renelt (1992). "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review* 82, 942-63.
- Mankiw, G. D. Romer, and D. Weil (1992). "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107, 407-37.
- McGrattan, E. and J. Schmitz (1999). "Explaining Cross-Country Income Differences," in J. Taylor and M. Woodford (eds.) *Handbook of Macroeconomics*. North Holland.
- Mincer, J. (1974). *Schooling, Experience and Earnings*. National Bureau of Economic Research.
- Quah, D. (1993). "Empirical Cross-Section Dynamics in Economic Growth," *European Economic Review* 37, 426-34.

Quah, D. (1997). "Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs," *Journal of Economic Growth* 2, 27-59.

Razin, A. and E. Sadka (1995). *Population Economics*. The MIT Press.

Sala-i-Martin, X. (1997). "I Just Ran Four Million Regressions," *Working Paper* 6252, National Bureau of Economic Research.

Solow, R. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70, 65-94.

Summers, R. and A. Heston (1991). "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988," *Quarterly Journal of Economics* 106, 327-68.