On the Power of Absolute Convergence Tests*

Rómulo A. Chumacero

Abstract

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

*I would like to thank William Easterly, Rodrigo Fuentes, Klaus Schmidt-Hebbel, Raimundo Soto, two anonymous referees, and participants on seminars and meetings at the Central Bank of Chile, LACEA, the Society for Computational Economics, the Society for Nonlinear Dynamics and Econometrics, and the European Economic Association for helpful comments and suggestions. Francisco Gallego provided able research assistance. Financial support from Fondecyt (N° 1030681) is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

With the possible exception of Mincerian regressions (Mincer, 1974), few subjects in applied economic research have been studied as extensively as the convergence hypothesis advanced by Solow (1956) and documented by Baumol (1986).[1] In simple terms, the hypothesis 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 reach the same income per capita.[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. However, 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 this hypothesis a "fair" chance. The paper 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. Section 4 discusses how likely would it be for time series generated from those models to reject absolute convergence. Finally, section 5 draws some conclusions.

2 Results from the Empirical Literature

This section presents a brief review of the main results of empirical growth analyses that test the convergence hypothesis.

2.1 Absolute Convergence is Strongly Rejected

The first stylized fact that appears uncontroversial is that whatever the type of data set used (a cross section of countries or panel data), the data strongly reject absolute convergence (Barro and Sala-i-Martin, 1995). The simplest test that can be devised to verify this claim using cross-sectional observations

[1] An 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), and Kocherlakota and Yi (1996,1997).

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

takes the form

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

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

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

When pooled data are used, tests for absolute convergence usually take the form

$$\Delta \ln y_{i,t} = \zeta + \vartheta \ln y_{i,t-1} + \varepsilon_{i,t}. \tag{2}$$

In both cases absolute convergence is said to be favored by the data if the estimate of ϑ is negative and statistically different from zero. 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 GDP per capita.

As Table 1 and Figure 1 show, the convergence hypothesis is strongly rejected by the data.[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 GDP per capita.

	Cross-section	Pooled Data
$\widehat{\vartheta}$	0.0047	0.0048
	(0.0014)	(0.0010)
Adjusted R^2	0.051	0.007
No. of countries	116	85
No. of observations	116	3,219

Table 1: Tests for absolute convergence. Standard errors consistent with heteroskedasticity are in parentheses.

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, equation (2) makes explicit that a test for (no) absolute convergence is fundamentally related to a test for a unit root on y. That

[3] All tests using panel data were conducted using the latest version of the Penn World Tables data set described in Summers and Heston (1991), with data for most variables ranging from 1960 to 1998. Cross-section regressions were conducted using the data set described in Sala-i-Martin et al (2004).

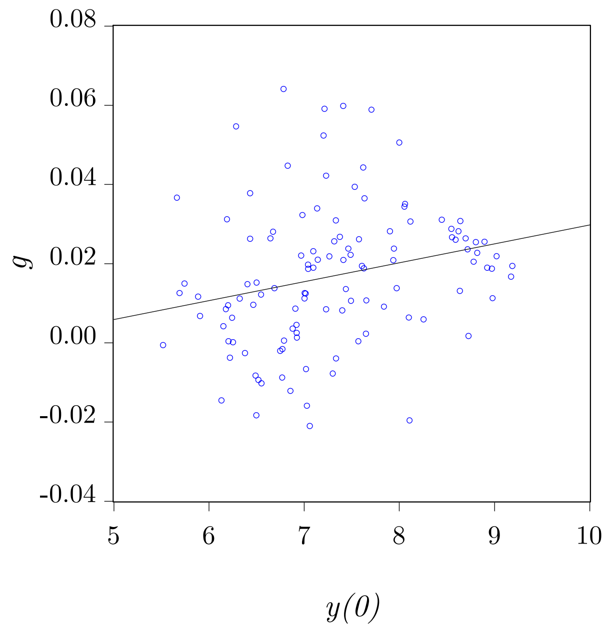

Figure 1: Growth rate from 1960 to 1998 versus 1960 GDP per capita

is, under the null hypothesis of a unit root in y, convergence is rejected. As abundantly documented elsewhere, these tests not only have nonstandard asymptotic properties, but also lack 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 coefficient 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 equation (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 levels of GDP per capita. Tests for conditional convergence using cross-sectional observations usually take the form

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

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

	Cross-section	Panel Data
$\widehat{\vartheta}$	-0.0154	-0.0456
	(0.0028)	(0.0062)
Adjusted R^2	0.811	0.181
No. of countries	79	85
No. of observations	79	2,552

Table 2: Tests for conditional convergence. Standard errors consistent with heteroskedasticity are in parentheses.

Table 2 presents the results of cross-sectional and panel regressions that include some of the usual candidates for specifications such as equation (3).[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, 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

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

and Renelt, 1992; Sala-i-Martin, 1997; and Sala-i-Martin et al, 2004). Sixth, several of the variables included in the x vector are fixed effects that cannot be modified; if these variables were actually long-run determinants of growth, convergence would never be achieved (even with $\vartheta < 0$).[5] Finally, the null hypothesis ($\vartheta = 0$) on equation (3) can be viewed as a unit root test with co-variates. Although this test has better power than univariate unit root tests, it still is a unit roots test with added non conventional asymptotic distribution (Hansen, 1995; Elliot and Jansson, 2003).[6]

2.3 Clubs

Durlauf and Johnson (1995) suggest that cross-sectional 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 GDP per capita on which each country converges depends on some initial condition (such as initial GDP per capita) and that, depending on this characteristic, some countries converge on one level and others on another. A common specification used to test this hypothesis considers a modification of equation (1) that takes the form

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

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

If equation (4) were the actual data-generating-process (DGP), results such as those in Table 1 could be easily motivated, given that if two regimes are present, with each regime converging to a different state and at a different rate, estimations based on a single regime might produce a nonsignificant estimate for the convergence parameter. On the other hand, equation (4) states that if the threshold variable (in this case, initial GDP per capita) is correlated with some of the x variables included in equation (3), results such as those reported in Table 2 are likely to be encountered, even if the x variables are

[5] A curious example of such a variable is "absolute latitude", which measures how far a country is from the Equator. When statistically significant, its coefficient is usually positive, implying that one way to enhance growth would be for a country to move its population toward the North or the South Pole.

[6] I would like to thank one of the referees for pointing out the link between conditional convergence and these tests.

not (necessarily) determinants of long-run growth.[7] However, equation (4) has an unequivocal implication in terms of the distribution of GDP per capita 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 (1993,1997) noticed that relative GDP per capita (defined as the ratio of the GDP per capita of country i with respect to average world GDP per capita, represented here by $\widetilde{Y}_{i,t}$) displays such bimodality. He conjectured that if clubs of convergence were present, even if the unconditional distribution of initial GDP per capita 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 one level of GDP per capita and another group to another, in which case twin peaks would arise.

Figure 2 presents kernel estimators of the unconditional density of relative GDP per capita 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 a permanently lower GDP per capita. On the other hand, Figure 3 presents surface and contour plots of the (log of) relative GDP per capita, 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 equation (4), no formal test of this theory can be provided with this visual evidence. Quah (1993) tried to formalize the twin peaks hypothesis by deriving the ergodic distribution of the transition matrix of relative incomes among countries. Table 3 presents estimates of the one-year transition matrix of $\widetilde{Y}$ 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 important, with the sample analyzed, the ergodic distribution does appear to be bimodal in the sense that (unconditionally) higher probabilities are associated with countries that have less than one-quarter of average world GDP per capita or more than twice this average.

However, this distribution is highly nonlinear and extremely noisy (Kremer et al, 2001). The resulting ergodic distribution is sensitive to the choice of thresholds for each category, the number of years used to compute the transition matrix, and the variable used to perform the comparisons.[8] More

[7] This would happen if, for example, y is persistent, x is correlated with initial income, and x is itself persistent (or a fixed effect).

[8] Kremer et al. (2001) consider that a better choice of variable for constructing the

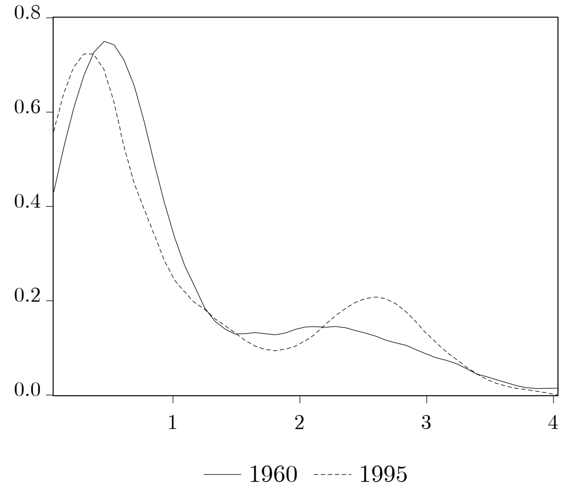

Figure 2: Densities of relative GDP per capita

	$\widetilde{Y}_{t+1} \leq \frac{1}{4}$	$\frac{1}{4} < \widetilde{Y}_{t+1} \leq \frac{1}{2}$	$\frac{1}{2} < \widetilde{Y}_{t+1} \leq 1$	$1 < \widetilde{Y}_{t+1} \leq 2$	$\widetilde{Y}_{t+1} > 2$
$\widetilde{Y}_t \leq \frac{1}{4}$	0.973	0.027	0	0	0
$\frac{1}{4} < \widetilde{Y}_t \leq \frac{1}{2}$	0.047	0.927	0.026	0	0
$\frac{1}{2} < \widetilde{Y}_t \leq 1$	0	0.035	0.948	0.017	0
$1 < \widetilde{Y}_t \leq 2$	0	0	0.018	0.949	0.033
$\widetilde{Y}_t > 2$	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

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 to the persistence of the GDP per capita level.

transition matrix is the ratio of each country's GDP per capita to the average GDP per capita of the five leading countries or the leading country.

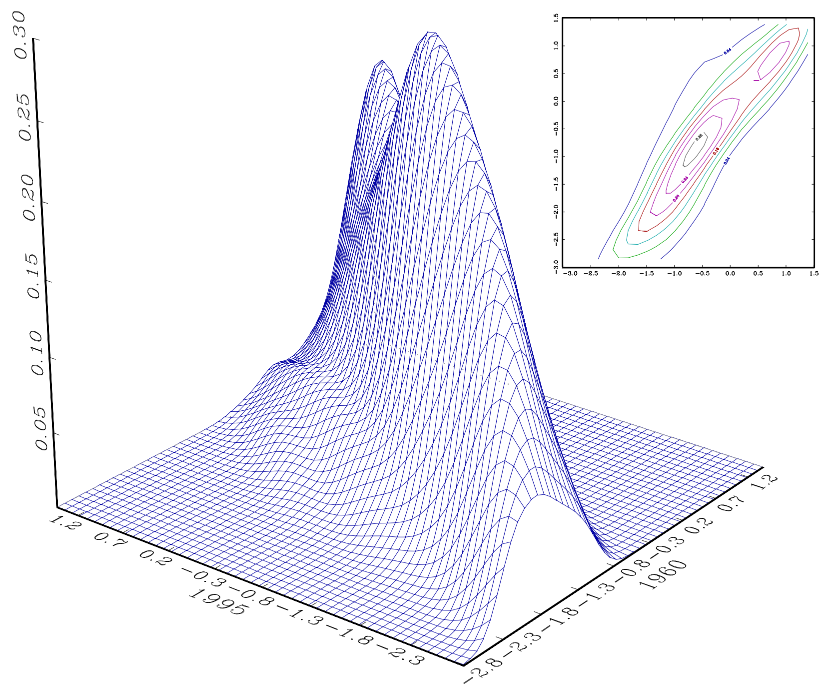

Figure 3: Surface and contour plots of (log of) relative GDP per capita

3 A Simple 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 consumption per capita,[9] $\gamma > 0$ is the Arrow-Pratt relative risk aversion coefficient, and $\mathcal{E}_t$ is the expectations operator conditional on information available for period t. There is no utility from leisure, and the labor force is equal to L_t.[10] Utility is maximized with respect to consumption per capita and the capital stock per

[9] Lower case letters denote per capita values, upper case totals, and a hat above a variable denotes that the value is per unit of effective labor.

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

capita, k_{t+1}, subject to the budget constraint:

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

where $0 < \alpha < 1$ is the compensation of 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.[11] In fact, one can transform the above economy to a stationary economy and obtain exactly the same solutions for $\widehat{k}_t$ and $\widehat{c}_t$. Such an economy can be characterized by the following maximization problem:

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

subject to

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

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), I include a simple channel to induce correlation between each economy's income. Specifically, I obtain correlated incomes by assuming that the law of motion of technology shocks in country i can be written as

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

where v_t and $w_{i,t}$ are independent $\mathcal{N}\left(0, \sigma_j^2\right)$ random variables (for $j = v, w$). If ϕ is equal to zero, all countries face the same aggregate shock; 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 I consider the case in which fertility is exogenous and has the following law of motion:

$$\ln\left(1+\eta_{i,t}\right) = \overline{\eta}(1-\tau) + \tau \ln\left(1+\eta_{i,t-1}\right) + n_{i,t}, \tag{8}$$

where $n_{i,t}$ is an independent $\mathcal{N}\left(0, \sigma_n^2\right)$ random variable, $\overline{\eta}$ is the long run rate of population growth and $0 < \tau < 1$.[12]

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.

---

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

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

4 Convergence Tests and the Model

The goal is to evaluate whether or not the tests for convergence presented in section 2 would be robust. That is, if time-series realizations were generated using a model in which convergence holds, would tests for convergence find convergence? 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
- twin peaks should not be present in the long run.

To clarify concepts, I next specialize the model of section 3, describe its properties, derive the DGP that $\ln y$ would obey, and ask whether the tests discussed in section 2 are really tests for convergence. To understand whether the tests discussed in Section 2 are useful in testing for convergence, I tailor the model to instances in which a closed-form expression for the DGP of the log of GDP per capita is available. This simplification imposes a 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 hypothesis.[13]

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

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

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

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

$$\ln \widehat{y}_t = z_t + \alpha \ln \widehat{k}_t, \tag{10}$$

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

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

[13]In particular, the parameterization chosen forces the time series representaion of log GDP per capita to be linear. If this were not the case, linear specifications that test for convergence would be misspecified and would have less power than the results presented below.

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

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

where B and D are constants.[14]

Four features of equation (12) are worth mentioning: First, as is typical of exogenous growth models, GDP per capita is trend stationary.[15] Second, given that the technology shock follows an AR(1) process, $\ln y$ follows an AR(2) process.[16] Third, even without exogenous growth ($\lambda = 0$), an AR(1) process for $\ln y$ such as equation (2) is consistent with equation (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 equation (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 GDP per capita, independent of their initial conditions and independent 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 ($\phi > 0$), GDP per capita would converge to a nondegenerate 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, I focus on the implications of two parameterizations of equation (12) for the convergence tests discussed in section 2.[17] For each parameterization I draw 2,000 artificial samples of time series of GDP per capita for 100 countries. Each sample begins with a bootstrapped sample (with replacement) of

---

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

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

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

[17] One of the referees rightly considers that a power study should include a richer parametrization than the cases analyzed. However, what is important here is to show that even the simplest parameterizations of the model already provide strong evidence of the lack of power of the convergence tests of the literature.

initial GDP per capita as the one observed in 1960.[18] Based on these initial conditions, values of $\ln y_{i,t}$ are simulated from equation (12) for a 36-year period.

4.1 Independently and Identically Distributed Shocks

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

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

Thus, independent of the initial distribution of GDP per capita and population growth rates, $\widehat{\vartheta}$ in equation (2) will consistently estimate the coefficient $\alpha - 1$, and convergence should occur.[19]

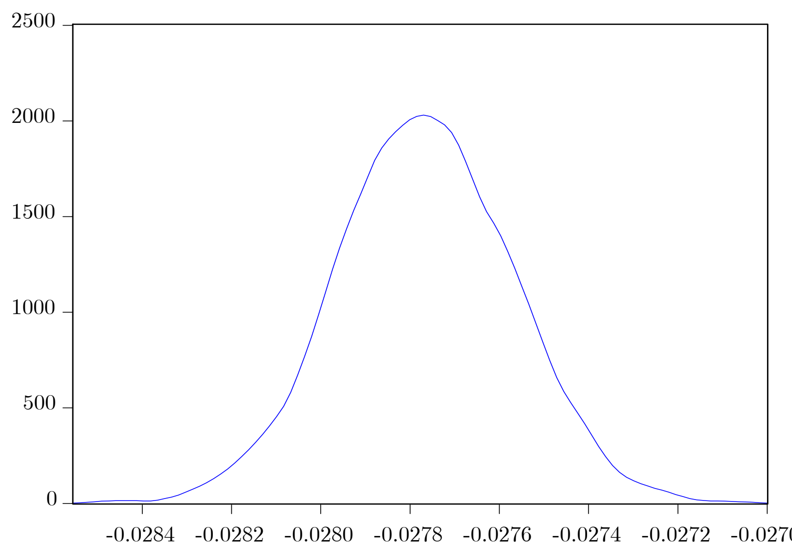

Figure 4: Distribution of $\widehat{\vartheta}$ from absolute convergence tests with i.i.d. shocks. Estimates obtained from 2,000 artificial samples for 100 countries.

[18] This allows us to generate artificial realizations of GDP per capita that are consistent with the initial bimodality observed in 1960 (Figure 2).

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

Figure 4 presents the empirical distribution of $\widehat{\vartheta}$, computed from artificial samples of countries. For each sample an estimate for ϑ was obtained by running a regression like equation (1).[20] Obviously, the probability of obtaining estimates of $\widehat{\vartheta}$ consistent with the results from section 2 is zero because even if the distribution of GDP per capita in 1960 is considered as the initial condition, i.i.d. shocks with realistic figures for α are unable to produce enough persistence in $\ln y$.

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 GDP per capita, whereas if some shocks are idiosyncratic, in the long run, GDP per capita converges to a nondegenerate distribution.

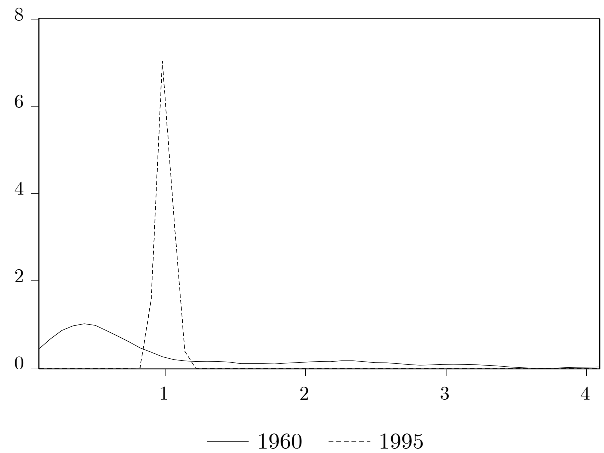

Figure 5: Densities of relative GDP per capita with i.i.d. shocks. Empirical densities for an artificial realization of 100 countries.

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

A main feature of this model is that once initial conditions have dissipated (which 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, if μ_t and b represent the limits of the mean and the variance of

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

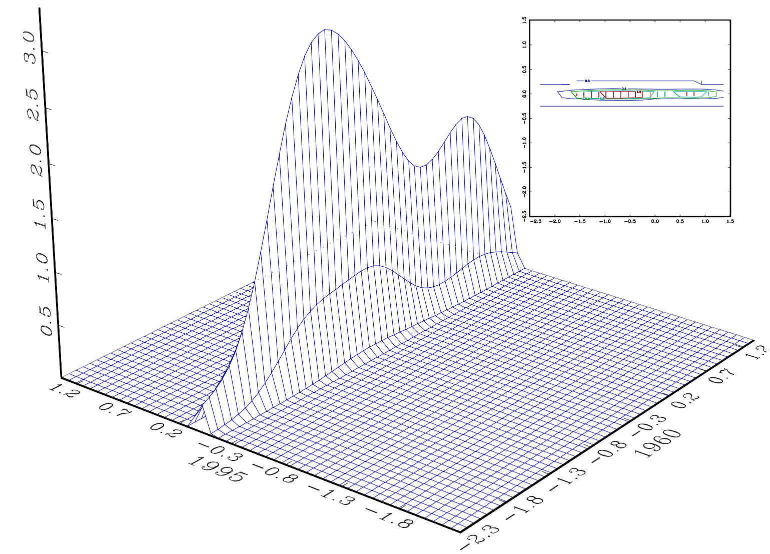

Figure 6: Surface and contour plots of (log of) relative GDP per capita for i.i.d. shocks. Results for an artificial realization of 100 countries.

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

$$\mathcal{E}\left(\widetilde{Y}_i\right) = 1, \;\; \mathcal{V}\left(\widetilde{Y}_i\right) = e^b - 1. \tag{14}$$

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

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

but

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

Thus, the probability that $\widetilde{Y}_i$ 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 $\widetilde{Y}_i$ and its ergodic probabilities depends solely on b, which in turn is a function of the volatility of technology shocks 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 $\widetilde{Y}_i$ appears to be both bimodal and strongly asymmetric, in the sense that (unconditionally) the median of $\widetilde{Y}_i$ is close to 0.5 and not to the mean (which is, by construction, one). Of course, the log-normal distribution is asymmetric; thus a simple way to verify whether 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

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

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

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

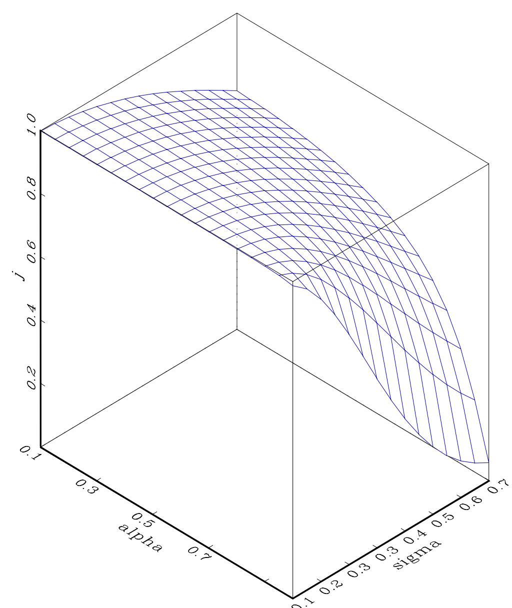

Figure 7: Median of $\widetilde{Y}$ for different values of α and σ_ε with i.i.d. shocks

Figure 7 shows that a median close to $\widetilde{Y} = 0.5$ can only be obtained with extremely volatile technology shocks ($\sigma_\varepsilon > 0.3$) or an unrealistic capital share in 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.

4.2 Persistent Shocks

Once we abandon the unrealistic setup of i.i.d. technology shocks, we can obtain significant persistence for $\ln y$ by choosing a value of ρ in (12) close to one. Persistence of technology shocks is routinely invoked in the Real Business Cycles 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.

One immediately notices that convergence tests such as equation (2) are misspecified. If pooled observations were used in equation (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 zero.

Figure 8 presents an exercise similar to that 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 nonnegligible probability (11 percent) that the estimated coefficient would indeed be positive (implying divergence).

Furthermore, as Figure 9 reveals, persistent technology shocks can replicate a bimodal joint distribution of the initial (log of) GDP per capita (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. 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:

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

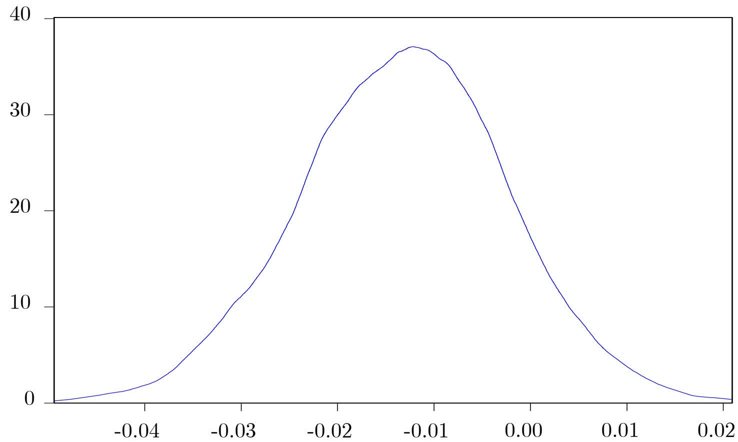

Figure 8: Absolute convergence tests with AR(1) shocks: empirical distribution of the $\widehat{\vartheta}$ coefficients obtained with 2,000 artificial samples for 100 countries.

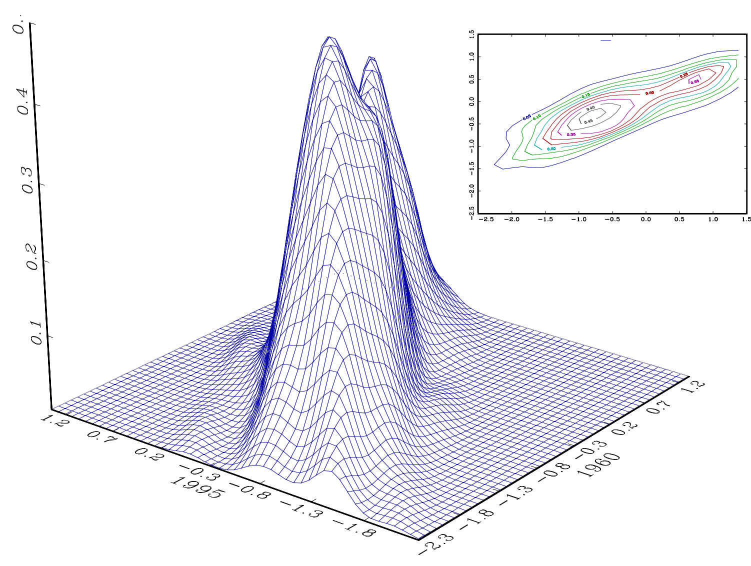

Figure 9: Surface and contour plots of (log of) relative GDP per capita for AR(1) shocks. Results for an artificial realization of 100 countries.

Thus the unconditional distribution of $\widetilde{Y}$ will still be log-normal with mean and variance given by equation (14), but b in this case is given by equation (16). We can conduct an experiment identical 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 $\widetilde{Y}$ can be set close to 0.5 with extremely persistent and moderately volatile technology shocks.

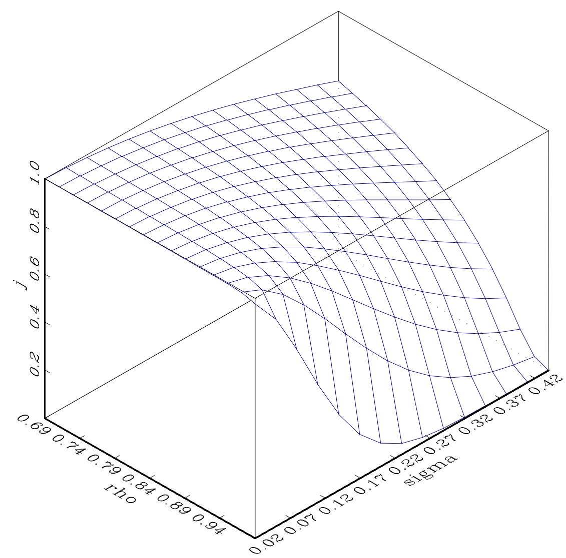

Figure 10: Median of $\widetilde{Y}$ for different values of ρ and σ_ε with AR(1) shocks

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 GDP per capita, 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 GDP per capita, even over relatively 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.

4.3 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) GDP per capita, persistence by itself could generate an illusion of bimodality for prolonged periods.

Furthermore, the models just discussed are among the simplest that can be generated from our theoretical model. In particular, if θ is different from one, the population growth rate becomes 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 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.

A case for conditional convergence could be made if, for example, distortionary taxes were included. If distortions were persistent (or permanent), countries with lower distortions would converge to higher income levels. However, according to this model and contrary to the endogenous growth literature, if the distortion were lifted, convergence would be achieved.

5 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. It shows 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. This is so because ultimately, tests against convergence are simply unit root tests, and have the power problems abundantly documented in the econometrics literature.

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 GDP per capita. Furthermore, there is a nonnegligible probability that misspecified econometric models will reject absolute convergence even when it is present.

Nevertheless, 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.

Furthermore, simple (and realistic) variations of the models presented, which 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 (with an incorrect law of motion of GDP per capita 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 GDP per capita should be achieved if these policies also converge.

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.

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.

Elliot, G. and M. Jansson (2003). "Testing for Unit Roots with Stationary Covariates," *Journal of Econometrics* 115, 75-89.

Hansen, B. (1995). "Rethinking the Univariate Approach to Unit Root Testing," *Econometric Theory* 11, 1148-71.

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 (2001). "Searching for Prosperity," *Carnegie-Rochester Conference Series on Public Policy* 55, 275-303.

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.

Sala-i-Martin, X., G. Doppelhofer, and R. Miller (2004). "Determinants of Long-term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review* 94, 813-35.

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.