THE ELUSIVE EMPIRICAL SHADOW OF GROWTH CONVERGENCE

By

Peter C.B. Phillips and Donggyu Sul

February 2003

COWLES FOUNDATION DISCUSSION PAPER NO. 1398

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

The Elusive Empirical Shadow of Growth Convergence*

Peter C.B. Phillips
Cowles Foundation, Yale University
University of Auckland & University of York

Donggyu Sul
Department of Economics
University of Auckland

January, 2003

Abstract

Two groups of applied econometricians have figured prominently in empirical studies of growth convergence. In terms of a popular caricature, one group believes it has found a black hat of convergence (evidence for growth convergence) in the dark room of economic growth, even though the hat may not exist (the task may be futile). A second group believes it has found a black coat of divergence (evidence against growth convergence) even though this object also may not exist (empirical reality, including the nature of growth divergence, is ever more complex than the models used to characterize it). The present paper seeks to light a candle to see whether there is a hat, a coat or another object of identifiable clothing in the room of regional and multi-country economic growth. After our examination, we find that the candle power of applied econometrics is too low to clearly distinguish a black hat in the huge dark room of economic growth. However, in our theory model, we find an important new role for heterogeneity over time and across economies in the transitional dynamics of economic growth; and, in our empirical work, these transitional dynamics reveal an elusive shadow of the conditional convergence hat in both US regional and inter-country OECD growth patterns.

Keywords: Convergence Parameter, Conditional Convergence, Economic Growth, Growth Convergence, Heterogeneity, Neoclassical Economics, Transition measures.

JEL Classification Numbers: 040; C32; C33.

First Draft: August, 2000

*The authors thank Steven Durlauf and Feng Zhu for comments on the original version of the paper and Phillips thanks the NSF for support under Grant No. SES 00-92509.

1 Introduction

For the last two decades, numerous studies have searched for empirical evidence in favour or against growth convergence. The task is important in helping to shape our understanding of the economic growth trajectories of nations and in judging which, if any, theory of economic growth fits the evidence. Cross-sectional, time series, and panel studies have all been used to examine growth convergence using several concepts of convergence. However, when these conceptualizations are incorporated in econometric specifications, the null hypotheses that are tested do not necessarily imply growth convergence or divergence. This is because heterogeneity in the speed of convergence can discombobulate the evidence. For instance, the concept of 'beta-convergence' that is commonly implemented in cross-sectional tests can be justified only when there is homogeneity in the speed of convergence over time across individual units. Moreover, conventional cointegration tests applied to cross country income are also valid only under homogeneity of the speed of convergence.

Growth convergence has been used to evaluate the fundamental question of whether poor countries can eventually catch up to rich countries in terms of their overall standard of living and whether that is actually happening. Empirical studies to address this question commonly start with the notion of transition to the steady state. A simple formulation of the transition dynamics of log per capita real income in neoclassical growth theory with labor augmented technological progress consists of the following three terms

$$\log y_i(t) = g_i(0) + b_i e^{-\beta_i t} + x_i t, \tag{1}$$

where i and t stand for the cross-sectional unit and time, β_i represents the speed of convergence and is a function of the growth rate of technological progress x_i. The first term, $g_i(0)$, in (1) represents initial conditions and steady state levels, while the second term b_i stands for the distance between the steady state and initial condition (see equation (2) below). Growth convergence between economies i and j does not necessarily require homogeneity of β_i and x_i across units, but asymptotic homogeneity of x_i over time is required. That is, if x_i changes over time, then its limit needs to be identical across units as t becomes large. Existing empirical studies of growth convergence typically assume homogeneity of x_i over time and across countries, so that technology is effectively treated as a public international good bestowed on all economies. This assumption of homogeneity in technology growth is often accompanied by the assumption of homogeneity in the speed of convergence also, so that the only source of heterogeneity across economies lies in the initial condition $g_i(0)$. Empirical work on growth convergence within the framework of these homogeneity restrictions include both cross sectional study and time series/panel studies (for example, Barro and Sala-i-Martin, 1992, Mankiw, Romer and Weil, 1992, Bernard and Durlauf, 1995, 1996, Evans and Karras, 1996, and Quah, 1996).

Cross-sectional homogeneity is often assumed for empirical convenience, because if x_i differs across economies during the transitional period, then the growth paths of the transitional economies can diverge and testing convergence becomes a very difficult task. The assumption

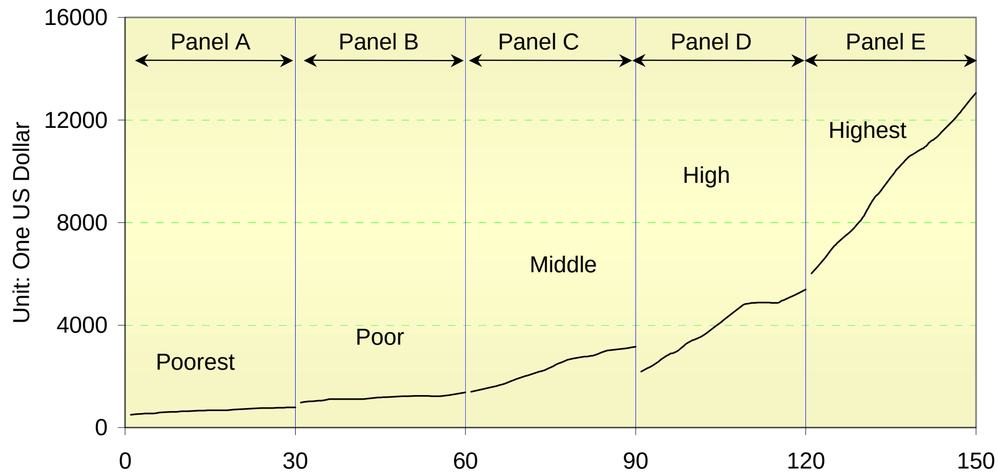

Figure 1: Average Real per Capita Income over the 30 year period 1960-1989 with Country Groupings based on Initial Income Orderings.[2]

of homogeneity of x_i across economies seems much more plausible near the steady state.

One effect of the assumption of homogeneity in x_i on empirical growth is the appearance of conditional convergence. Fig. 1 shows five groupings of cross sectional averages of per capita real income for 120 countries (from the Penn World Table) over 1960 to 1989. The groups (which consist of 25 countries) are selected according to initial income level and the averages are shown over the five successive panels in the figure covering the same 30 year period. For instance, the initial period average income for the poorest country group in 1960 was around \$503 and the last period average income in 1989 for the same group was \$788. So, for this group there was around 1.9% growth on average over the 30 year period. For the richest group, the corresponding initial period and the final period figures for average income were \$6,012 and \$13,065, giving an annual growth of 4% for this group.

Fig. 2 shows quantiles of the bootstrap distribution (based on 5,000 replications) of real per capita income for the same country groupings (the groups are formed using initial income ordering). Fig. 3 shows the cross section average trajectories against the maximum and the minimum, using actual (rather than bootstrapped) data. These curves give some idea of the variability in the actual growth trajectories over time within these groupings. The outcomes in Fig. 3 indicate that some members of each group have substantial prospects of transitioning into higher growth groups over the 30 year period. Table A in the Appendix gives numerical

[2] Each panel in Fig. 1 (and subsequently Figs. 2-3) refers to the same 30 year period. For example, Panel C tracks average real per capita income in the middle income category over the same 30 year period as the other panels. Dating on the horizontal axis is reset at the beginning of each panel. Thus, 60 in Panel B refers to the end of the 1960-1989 period for the poor income category while 61 refers to the beginning of the period for the middle income category in Panel C.

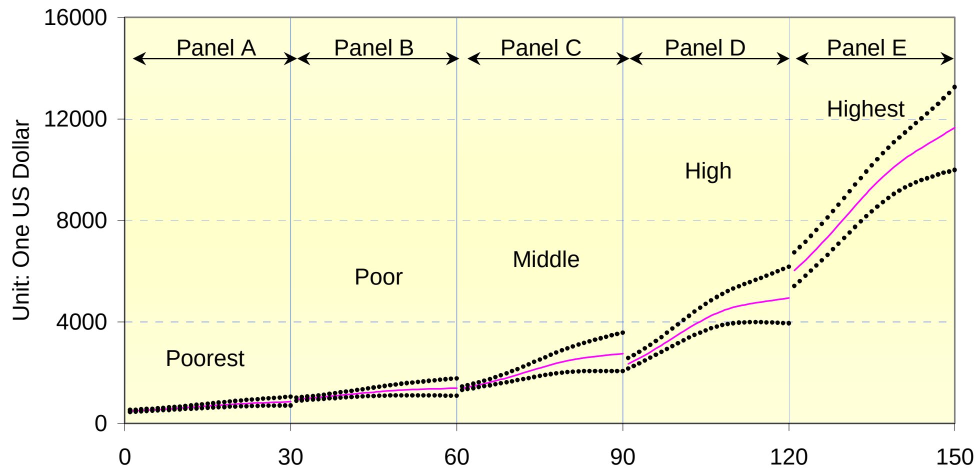

Figure 2: 2.5%, 50% and 97.5% Quantiles of Bootstrap Distribution of Real per Capita Income over 1960-1989.

details of the distributions displayed in these three figures.

Assuming homogeneity of x_i and β_i rules out the possibility of transitional divergence and focuses attention on convergence to a common steady state from different initial conditions. In consequence, empirical studies commonly treat poor (rich) countries as poor (rich) because of a state of low (high) initial income, leading to the view that there is evidence of growth convergence when poor countries grow faster. This view generates two conclusions. First, those poor countries that have grown slower than rich countries over the last 30 years will never catch up to the rich countries because homogeneous growth rates over time are assumed. Second, countries in similar income groupings often show evidence of convergence, suggesting that convergence may be conditional.

An alternate perspective is that poor countries are poor partly because of their low initial levels of technology, whereas rich countries are rich in part because of higher initial levels of technology. In that event and using simple neoclassical growth theory, what circumstances would give rise to the high degree of cross-sectional divergence that we observe during transitional periods while ultimately permitting growth convergence in the steady state? First, we consider the case where there is homogeneity in x_i over time across units but heterogeneity in the speed of convergence. In this case we find that neoclassical economies can generate a high degree of divergence during transitional periods but only under rather unrealistic values for the speed of convergence. A poor country (e.g. one with $800 of initial per capita income) may have a very low value of β, but without allowing for heterogeneity in x_i over time, such a country will need more than two thousand years to attain 90% of the income of a rich country in which per capita income is $4000.

Only when we allow heterogeneity in x_i over time and across units (while at the same time

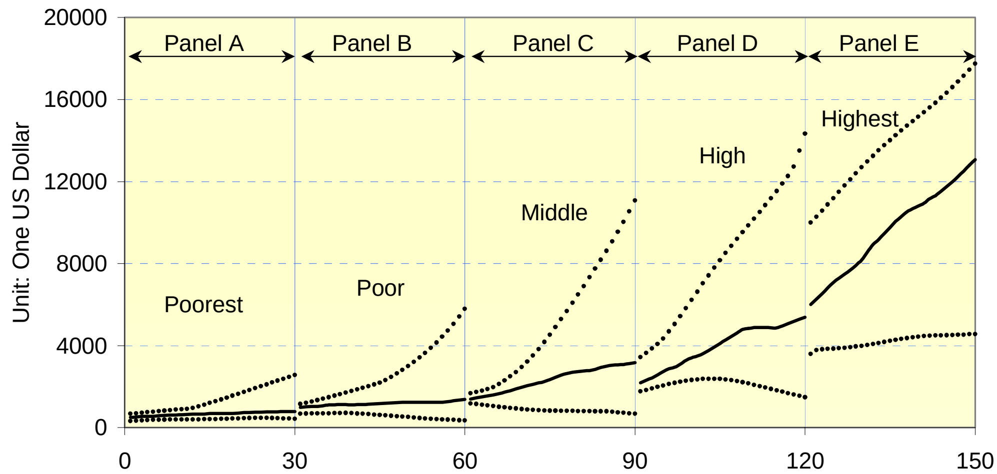

Figure 3: Average, Minimum and Maximum trajectories of Actual Distribution of Real per Capita Income over 1960-1989

requiring that the x_i converge to a common constant over time to ensure convergence) do we obtain more realistic results covering transitional periods and the long term. In this case a poor country may grow faster because its speed of technological learning (or technological transfer) is faster than the speed of technological creation in a rich country. When the speed of learning is lower than the speed of creation, the growth rate may be lower than that of a rich country and it may be much lower when the poor country's learning speed is very slow. In such cases, divergence in growth paths may occur in the early transitional period.

We ask whether there is evidence in the data for such behavior. Unfortunately, growth convergence tests based on time series or panel cointegration methods are not really adequate empirical tools for eliciting this information from the data because these econometric methods search for long run patterns in the data and therefore presume that the data is already around the steady state. In cases where there are some cross sectional units with long transitional regimes, such tests can be expected to bias results in favor of divergence. Panel tests that take account of cross sectional dependence, such as those developed in Phillips and Sul (2002) and Moon and Perron (2002), are known to improve the size and power properties of panel unit root tests in general and may mitigate the effects of bias from transitional regimes because of the averaging effects over correlated units in different stages of transition and convergence. These tests are applied to determine whether there is growth convergence in the 48 contiguous U.S. States from 1929 to 1998. They provide evidence only in support of certain regional (or conditional) growth convergence among the U.S. States.

Next, we take a different approach to testing growth convergence. Since we allow the speed of convergence as well as the growth of technological progress to be heterogenous over time and across units, we use filtering techniques to extract estimates of a 'transition parameter' over time and examine the evidence for growth convergence by testing whether this transition parameter diverges or converges. In the case where the transition parameter converges monotonically, economic growth follows the now famous rule that 'initially poorer countries grow faster' ($\beta-$convergence) and 'cross-sectional income distribution should decrease over time' ($\sigma-$convergence). With data for the 48 contiguous U.S. States from 1929 to 1998 as well as data for 21 OECD countries from 1950 to 1992, we find evidence for growth convergence on this criterion.

The transition parameter approach is also applied to the Penn World Table (PWT) panel from 1960 to 1989 across 120 countries. When economies have different speeds of convergence, different growth rates for technological progress, and heterogeneous initial income, neither $\beta-$convergence nor $\sigma-$convergence is possible. Instead, we rely on the following descriptive characteristics. First, among similar initial period income countries, the transition parameter should diverge when there is heterogeneity in β_i and x_i. Second, among similar final period income countries, the transition parameter should converge. The evidence for convergence from this approach cannot be clearly distinguished from what is sometimes called conditional β-convergence. This is because conditional $\beta-$convergence can also generate divergence in the early stages among countries in different convergence groups. Moreover, conditional $\beta-$convergence generates convergence within countries in a convergence group. Thus, if conditional β-convergence holds for each group in Fig. 1, one implication is that a poor group will become relatively poorer in comparison to rich groups over time. Some of these comparative issues are confronted at the end of the paper, although we do not claim to have fully resolved them.

The plan of the paper is as follows. Section 2 studies some of the conflicts between theory and evidence on growth convergence, giving an important new role to heterogeneity in the growth convergence process. Section 3 considers the bridge between theory and empirical models setting up some econometric specifications that are useful for considering convergence, divergence and the measurement of convergence. Section 4 discusses panel unit root and cointegration analyses of convergence and provides some new empirical evidence that takes account of cross section dependence. Section 5 examines techniques for estimating the transition parameter and gives some associated tests. Section 6 concludes and a short appendix follows. To make the paper more accessible to a general audience, we make extensive use of graphs in the text and discussion.

2 Theory, Evidence and the Role of Heterogeneity in Convergence

We start from the neoclassical theory of growth convergence and attempt to build some connections between the theoretical formulations and observed empirical regularities. In this respect, our discussion draws on the wide ranging overview of the subject by Durlauf and Quah (1999), who document some of the empirical characteristics of cross country growth For instance, rich countries grow slower than some of the newly developed countries, such as the Asian Dragons and some other rapidly growing developing countries like China. The remaining countries appear to grow at similar rates or slower rates than the rich countries.

Write the production function in the neoclassical theory of growth with labor augmented technological progress as $Y = F(K, LHA)$ and define

$$\tilde{y} = f(\tilde{k}),\ \tilde{y} = Y/LHA,\ \tilde{k} = K/LHA,\ y = \tilde{y}HA = \tilde{y}A$$

where Y is total output, L is the quantity of labor input, H is the stock of human capital (here normalized to unity), A is the state of technology, K is physical capital, and $\tilde{y}$ is output per effective labor unit. Using a Cobb-Douglas technology, the conventional transitional growth path for country i is given by

$$\log y_i(t) = \log \tilde{y}_i^* + [\log \tilde{y}_i(0) - \log \tilde{y}_i^*]\, e^{-\beta_i t} + \log A_i(0) + x_i t \tag{2}$$

where y_i is per capital real income, $\tilde{y}_i^*$ is the corresponding steady state level,

$$\beta_i = f(\underset{-}{\alpha_i}, \underset{+}{\delta_i}, \underset{+}{v_i}, \underset{+}{x_i}, \underset{+}{\gamma}), \tag{3}$$

α_i is the technology parameter in the Cobb-Douglas function, δ_i is the rate of depreciation, v_i is the population growth rate, and x_i is the growth rate of technical progress for country i. Appropriate sign effects are indicated beneath these parameters in (3). It is assumed that technological progress for country i follows

$$A_i(t) = A_i(0)e^{x_i t}. \tag{4}$$

We follow the definition of growth convergence proposed by Bernard and Durlauf (1995, 96) given by[3]

$$\lim_{k\to\infty} (\log y_i(t+k) - \log y_j(t+k)) = 0 \tag{5}$$

That is, growth convergence requires that log per capital real income in country i should be the same as that in county j in the long run. We consider the following three cases that arise within our framework (2).

Case 1: $x_i \neq x_j$ if $i \neq j$. When this condition is sustained over time, growth convergence simply cannot hold. For, even when the other parameters in (2) are homogenous across i, as long as $x_i \neq x_j$ is sustained, the two economies will diverge. Hence, the only possibility for growth convergence in this case is convergence of $x_i = x_i(t)$ over time, viz.

$$\lim_{t\to\infty} x_i(t) \to x. \tag{6}$$

In other words, if x_i is time dependent[4], then it needs to converge to a common value as $t \to \infty$ for convergence to apply. There are, in fact, several ways to specify stochastic or

[3] We ignore the conditional expectation operator in equation (5) since at least for the time being all variables are assumed to be non-stochastic.

[4] Of course, when x_i is dependent on time t, the transitional technology (4) is more complex than simple exponential growth.

deterministic growth in technological progress in place of (4). However, as Durlauf and Quah (1999) point out, such adjustments are ad hoc and do not reconcile directly with neoclassical theory. For the time being, we rule out this case and assume that $x_i = x$ over time, focusing on the initializations of $\log y_i(0)$ and $\log A_i(0)$, and the term $e^{-\beta_i t}$. The case of time varying movement of x_i will be considered in a later section.

Case 2: $x_i = x$, $A_i(0) \neq A_j(0)$. This case provides a key element in the discussion of Durlauf and Quah (1999). Since the initial conditions of technology are not necessarily the same across i, we have

$$\lim_{t\to\infty} [\log y_i(t) - \log y_j(t)] = \log \frac{\tilde{y}_i^*}{\tilde{y}_j^*} + \log \frac{A_i(0)}{A_j(0)}. \tag{7}$$

Then, if $A_i(0)$ is randomly distributed, pairs of economies do not converge according to (5). In this event, neoclassical growth theory does not provide any meaningful information about convergence. Hence, for the time being at least, we also rule out this case. We note, however, that when β_i and x_i are both time dependent, this problem can be resolved and we address this issue below. For now, we simply assume $A_i(0) = A(0)$ for all i.

Case 3: $x_i = x$, $A_i(0) = A(0)$, $\beta_i \neq \beta_j$. Note that the steady state level $\tilde{y}_i^*$ can be written as a function of other parameters, viz., $\tilde{y}_i^* = g((\delta_i + v_i + x)^{-1} s_i)$ where s_i is the saving rate, and $g(\cdot)$ is a continuous function. If δ_i, v_i and s_i are identical across i or if $(\delta_i + v_i + x)^{-1} s_i$ is the same across i, then it implies that the steady state value of $\tilde{y}_i(t)$ is identical over i, and the heterogeneity of β_i comes from different parameters α_i. The analysis in the following subsection does not depend on the heterogeneity of $\tilde{y}_i^*$, so we simply assume $\tilde{y}_i^* = \tilde{y}^*$. Hence, there are only two heterogenous parameters in this case – initial income and the speed of convergence.

We examine this case in detail to see how cross-sectional heterogeneity may be linked to certain empirical regularities.

2.1 Cross-Sectional Heterogeneity of Initial Income and Speed of Convergence

After imposing homogeneity restrictions on $x_i, \tilde{y}_i^*$, and $A_i(0)$, the transition dynamics of real per capita income have the form

$$\log y_i(t) = \log \tilde{y}^* + [\log \tilde{y}_i(0) - \log \tilde{y}^*] e^{-\beta_i t} + \log A(0) + xt \tag{8}$$

Accordingly, the per capita real income differential between two economies i and j is given by

$$\log y_i(t) - \log y_j(t) = [\log \tilde{y}_i(0) - \log \tilde{y}^*] e^{-\beta_i t} - [\log \tilde{y}_j(0) - \log \tilde{y}^*] e^{-\beta_j t}$$

If $\beta_i > 0$ for all i, $\lim_{t\to\infty} e^{-\beta_i t} = 0$. Hence, as long as $\beta_i > 0$, growth convergence holds regardless of the value of $\log y_i(0)$.

Barro and Sala-i-Martin (1992) impose a homogeneity restriction on β. If $\beta_i = \beta$, then

$$\log \frac{y_i(t)}{y_i(t-1)} - \log \frac{y_j(t)}{y_j(t-1)} = (1 - e^{\beta}) e^{-\beta t} [\log \tilde{y}_i(0) - \log \tilde{y}_j(0)],$$

so that

$$\log \frac{y_i(t)}{y_i(t-1)} < \log \frac{y_j(t)}{y_j(t-1)} \iff \log \tilde{y}_i(0) > \log \tilde{y}_j(0), \tag{9}$$

giving the so called $\beta-$convergence condition.

However, even when $\beta_i \neq \beta_j$, the $\beta-$convergence condition can be obtained as long as[5]

$$\beta_i \leq \beta_j \iff \log \tilde{y}_i(0) > \log \tilde{y}_j(0) \tag{10}$$

Equation (10) is, in fact, a more general condition than (9). Since the growth rate of technology is assumed to be identical across economies, heterogeneity in the convergence speed is attributed to the parameter α_i. One economic interpretation of α comes from international trade theory. If $\alpha_i > \alpha_j$, country i has a relatively capital intensive production technology. From equation (3), the speed of convergence is faster the lower is α. In other words, a country characterized by labor intensive production grows faster than other countries with more capital intensive production when the technological growth rate is the same across economies. So, condition (10) implies that initially poor countries can grow faster if they produce a more labor intensive product than richer countries. However, in spite of the additional flexibility in (10) and its underlying economic interpretation, neither (10) nor (9) explains the observed cross-country income differences that are presented in the PWT.

Next, consider the case when $\beta_i \geq \beta_j$ but $\log \tilde{y}_i(0) > \log \tilde{y}_j(0)$. The growth rate differential between the two countries, g is given by

$$\begin{aligned} g &= \log \frac{y_i(t)}{y_i(t-1)} - \log \frac{y_j(t)}{y_j(t-1)} \\ &= (1-e^{\beta_i})e^{-\beta_i t}\left[\log \tilde{y}_i(0) - \log \tilde{y}^*\right] - (1-e^{\beta_j})e^{-\beta_j t}\left[\log \tilde{y}_j(0) - \log \tilde{y}^*\right]. \end{aligned}$$

The sign of the growth rate difference depends on time in the following manner

$$\text{sign}(g) \gtreqless 0 \Leftrightarrow sign\left[\frac{\log \tilde{y}_i(0) - \log \tilde{y}^*}{\log \tilde{y}_j(0) - \log \tilde{y}^*}\frac{e^{\beta_i}-1}{e^{\beta_j}-1}e^{-(\beta_i-\beta_j)t}\right] \gtreqless 1. \tag{11}$$

The first factor in the square bracket of (11) is less than one if $\log \tilde{y}^* > \log \tilde{y}_i(0) > \log \tilde{y}_j(0)$. If $\beta_i < \beta_j$, then the second factor is also less than one. Finally, if $\beta_i > \beta_j$ then the third factor in (11) is less than unity. Combining these factors, we have the following potential growth scenario. At the beginning of the period (or when t is small), the growth rate of an initially poor country may be lower than the growth rate of an initially rich country and divergence can occur during the early period of transition, while ultimate convergence can hold in the long run. Figure 4 shows a stylized illustration of how such an 'Asian Miracle' can happen in the neoclassical world with heterogeneous $\beta's$. Economy 3 is a rich economy with higher initial income while economies 1 and 2 are initially poor countries. Economy 2 has a higher β than economy 3 while economy 1 has the lowest β. Then, between economies 2 and 3, β-convergence

[5] Suppose that at time $T < \infty$ we have $\log y_i(T) = \log y_j(T)$. Then $\left[\log \tilde{y}_i(0) - \log \tilde{y}^*\right] e^{-\beta_i T} = \left[\log \tilde{y}_j(0) - \log \tilde{y}^*\right] e^{-\beta_j T}$, and hence, $\beta_i \leq \beta_j$.

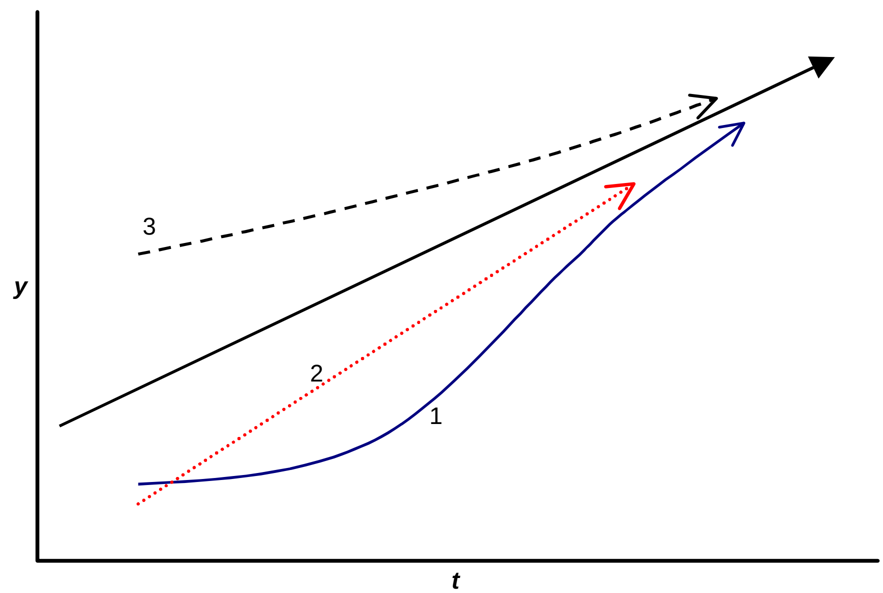

Figure 4: Transitional Divergence and Ultimate Convergence

holds since $\beta_2 > \beta_3$ but $\log \tilde{y}_2(0) < \log \tilde{y}_3(0)$. Economy 1 has a similar initial endowment as economy 2, but this country's growth diverges initially from both economy 2 and 3. This divergence arises because $\beta_1 < \beta_3 < \beta_2$.

The next question is how many years does it take for convergence to occur between economies 1 and 3. For example, let $\beta_3 = 0.03$, while $\beta_1 = 0.005$.[6] Initial income for economy 3 is assumed to be $\log y_3(0) = 8.3$ ($\$e^{8.3} \approx \$4,000$), while that for economy 1 is $\log y_1(0) = 6.9$ ($\$e^{6.9} \approx \$800$). Recall that $\log y_i(0) = \log \tilde{y}_i(0) + \log A(0)$. Since $\log \tilde{y}_i(0) > 0$, we take $\log A(0) \in (0.1, 6.8)$. Let $c = \log \tilde{y}^* - \log \tilde{y}_3(0)$. We allow c values in the range $c \in (1, 10)$, which implies a difference between $\tilde{y}_3(0)$ and $\tilde{y}^*$ of \$3 and \$22,000 for $c = 1$ and 10, respectively. Finally we compute the time value t^* for which per capita real income of the poor country reaches $(k \times 100)\%$ of the per capita real income of the rich country. Then,

$$\log \frac{y_3(t^*)}{y_1(t^*)} = \log \frac{1}{k},$$

and

$$\log \tilde{y}^* \left[e^{-\beta_1 t^*} - e^{-\beta_3 t^*} \right] + \log \tilde{y}_3(0) e^{-\beta_3 t^*} - \log \tilde{y}_1(0) e^{-\beta_1 t^*} = -\log k.$$

[6] It is worth noting that a number of empirical studies on growth convergence have centered on the estimation of β and the range of resulting estimates varies from positive to negative depending on the choice of the cross section units in the panel (particularly, whether developing countries are included or not; see Mankiw, Romer and Weil, 1992). Empirical studies for industrialized country panels suggest a range of β values between 0.01 and 0.04. Rappaport (2000) reports that the values of β in poor country panels are between -0.004 and 0.009 while that in rich country panel lies between 0.04 and 0.045.

We found $t^* \in (450, 910)$ for $k = 0.9$, and $t^* \in (107, 534)$ for $k = 0.5$. In the worst case scenario, it takes more than 900 years for a poor country to catch up with a rich country.

We consider another experiment with different values of β_3. When $\beta_3 = 0.01$, $t^* \in (296, 527)$ for $k = 0.9$, and $t^* \in (55, 266)$ for $k = 0.5$. When $\beta_3 = 0.001$, $t^* \in (2970, 5049)$ for $k = 0.9$, $t^* \in (366, 2669)$ for $k = 0.5$. When the initial income differential is increased, the time period t^* gets larger still.

Such low values of β for developing countries and such long convergence times are both questionable. The next section provides a possible solution to these empirical issues.

2.2 Time Series Heterogeneity of Speed of Convergence

Barro and Sala-i-Martin (1992) assume that $x_i = x_i(t)$ converges to a common steady state value x, adding a condition to the formulation (4). Accordingly, we rewrite the transitional dynamics of $\log y_i(t)$ as

$$\log y_i(t) = \log \tilde{y}_i^* + [\log \tilde{y}_i(0) - \log \tilde{y}_i^*] \, e^{-\beta_i(t)t} + \log A_i(t), \tag{12}$$

where

$$\log A_i(t) = \log A(0) + x_i(t)t,$$

and where $\beta_i = \beta_i(t)$ as a function of time because it depends on $x_i = x_i(t)$. The assumption of a common steady state value implies that

$$\lim_{t\to\infty} \beta_i(t) = \beta_i,$$

and

$$\lim_{t\to\infty} A_i(t) = \begin{cases} A < \infty & if \ \lim_{t\to\infty} x_i(t)t = \text{const} \\ \infty & if \ \lim_{t\to\infty} x_i(t) = x \end{cases}.$$

We assume the other parameters are homogenous across i except for α_i and x_i, so that

$$\lim_{t\to\infty} \tilde{y}_i^* = \tilde{y}^*.$$

However

$$\lim_{t\to\infty} \log y_i(t) = \infty, \text{ if } \lim_{t\to\infty} x_i(t) = x > 0.$$

Hence, during transition, relative growth rates depend on $x_i(t)$ both directly and indirectly, viz.,

$$\log \frac{y_i(t)}{y_i(t-1)} = [\log \tilde{y}_i(0) - \log \tilde{y}^*] \left(e^{-\beta_i(t)t} - e^{-\beta_i(t-1)t} e^{\beta_i(t-1)} \right) + [x_i(t) - x_i(t-1)] \, .$$

Suppose we have a panel data set covering time q to time T, and let the first observation be the initial condition. If there is no time series heterogeneity in β and x, this initial condition does not matter in the analysis. In contrast, under time series heterogeneity of β and x, it does matter. To see this, write

$$\log y_i(q) = \log \tilde{y}_i(q) + \log A_i(q).$$

Under time series homogeneity of x_i, β_i is time invariant and

$$\begin{aligned}\log \tilde{y}_i(q+t) &= \log \tilde{y}^* + [\log \tilde{y}_i(0) - \log \tilde{y}^*]\, e^{-\beta_i(q+t)} \\ &= \log \tilde{y}^* + [\log \tilde{y}_i(q) - \log \tilde{y}^*]\, e^{-\beta_i t}\end{aligned} \tag{13}$$

However, when β_i changes over time, this relation no longer holds. Let β_i^1 be the value of β_i at time $q+t$, β_i^0 be the value at time q, and define the differential as

$$\beta_i^1 = \beta_i^0 + \Delta_i. \tag{14}$$

Under time invariance of the other parameters in (3), the change in β_i over time is due to the change in technological progress. We can then write (3) as $\beta_i(t) = g\,(x_i\,(t))$ and, if g is analytic, $\beta_i(t+q)$ has a power series representation about $\beta_i(t)$, and we can write the differential (14) in the form

$$\Delta_i = \varphi\,(x_i(t+q) - x_i(q)) = \varphi\,(\Delta_{x_i})\,,\ \text{say,}$$

where $\Delta_{x_i} = x_i(q+t) - x_i(q)$.

In place of (13), the transition now has the form

$$\begin{aligned}\log \tilde{y}_i(q) &= \log \tilde{y}^* + [\log \tilde{y}_i(0) - \log \tilde{y}^*]\, e^{-\beta_i^0 q}, \\ \log \tilde{y}_i(q+t) &= \log \tilde{y}^* + [\log \tilde{y}_i(0) - \log \tilde{y}^*]\, e^{-\beta_i^1 q} e^{-\beta_i^1 t} \\ &= \log \tilde{y}^* + [\log \tilde{y}_i(q) - \log \tilde{y}^*]\, e^{\beta_i^0 q} e^{-\beta_i^1 q} e^{-\beta_i^1 t} \\ &= \log \tilde{y}^* + [\log \tilde{y}_i(q) - \log \tilde{y}^*]\, e^{-\beta_i^+ t},\end{aligned}$$

where

$$\beta_i^+ = \beta_i^0 + \Delta_i \frac{t+q}{t}.$$

Take the case where $i = \{poor, rich\}$ and suppose that $\beta_{poor}^0 = \beta_{rich}^0$ at time q. If the speed of technological creation (which we take as the relevant factor in the rich country) is faster than the speed of learning or technological transfer (the relevant factor for the poor country) over the period $(q, t+q)$, then the inequality $\Delta_{rich} > \Delta_{poor}$ holds and we have $\beta_{poor}^+ < \beta_{rich}^+$. When the speed of creation is slower than the speed of learning, then $\beta_{poor}^+ > \beta_{rich}^+$.

The differential in the rate of technological progress is $\Delta_{x_i} = x_i(q+t) - x_i(q)$ and we have

$$\begin{aligned}\log A_i(q+t) &= \log A_i(0) + x_i(q)q + \Delta_{x_i} q + x_i(q+t)t \\ &= \log A_i(q) + \Delta_{x_i} q + x_i(q+t)t.\end{aligned}$$

The i'th country's transitional growth path is then given by

$$\log y_i(q+t) = \log \tilde{y}^* + \log A_i(q) + \Delta_{x_i} q + [\log \tilde{y}_i(q) - \log \tilde{y}^*]\, e^{-\beta_i^+ t} + x_i(q+t)t. \tag{15}$$

The log per capital income differential between two countries i and j now has three components:

$$\log y_i(q+t) - \log y_j(q+t) = B(t) + C(t) + C(t), \tag{16}$$

where

$$
\begin{aligned}
B(t) &= \log A_i(q) + \Delta_{x_i} q - \log A_j(q) - \Delta_{x_j} q \\
C(t) &= \left[\log \tilde{y}_i(q) - \log \tilde{y}_i^*\right] e^{-\beta_i^+ t} - \left[\log \tilde{y}_j(q) - \log \tilde{y}_j^*\right] e^{-\beta_j^+ t} \\
D(t) &= x_i(q+t)t - x_j(q+t)t
\end{aligned}
\tag{17}
$$

Growth convergence therefore requires the following conditions to hold:

C-1 $\lim_{t\to\infty} B(t) = 0 \iff \lim_{t\to\infty} \left(\Delta_{x_j} - \Delta_{x_i}\right) q = \log A_i(q) - \log A_j(q)$

C-2 $\lim_{t\to\infty} C(t) = 0 \iff \beta_i^+ > 0$

C-3 $\lim_{t\to\infty} D(t) = 0 \iff x_i(q+t) - x = o\left(\frac{1}{t}\right)$.

Conditions (C-2) is obvious and has previously been considered in the growth literature. A necessary condition for (C-3) is that there is some ultimate rate x for which $\lim_{t\to\infty} x_i(q+t) = x$, although this alone is not sufficient because the convergence rate must be at least $1/t$. Condition (C-1) says that, whatever the initial technological conditions, ultimate growth convergence can take place provided there is appropriate adjustment in the rate of technological transfer or technological creation. Thus, if $\log A_i(q) - \log A_j(q) > 0$, and $\lim_{t\to\infty} \left(\Delta_{x_j} - \Delta_{x_i}\right) = d > 0$ with $d = (\log A_i(q) - \log A_j(q))/q$, then the technological deficit is made up and growth convergence will be ultimately achieved. On the contrary, if x_i is not time varying, growth convergence does not occur whenever initial technological conditions differ across units, c.f., equation (7). In this sense, time series heterogeneity can compensate for heterogeneity in initial conditions.

Now consider the time to convergence when allowance is made for time series heterogeneity. Suppose that economy 3 is in steady state at time q in the sense that $x_3(q) = x$ and $\beta_3(q) = \beta$, while for economy 1, $x_1(q) < x_3(q)$. This inequality results in $\beta_1(q) < \beta_3(q)$ and $y_1(q) < y_3(q)$. The poor country is poor here because of its low technology. Set $q = 0$ without loss of generality and write (16) as

$$
\log y_3(t) - \log y_1(t) = B_{31}(t) + C_{31}(t) + D_{31}(t),
$$

using the explicit indices 3 and 1 in place of i and j. If $x_1(t) \to x_3 = x$, then $\beta_1(t) \to \beta_3 = \beta$. The speed of growth convergence is heavily dependent on the speed of adjustment in the rate of technological progress. Convergence time can be substantially reduced if the convergence speed of $x_1(t)$ is fast enough to make up for initial differentials in technology.[7]

3 Econometric Specifications

The previous section showed how the passages to a steady state can be complex and varied across individual units, so that simple cross-sectional tests for growth convergence can easily

[7] Howitt and Mayer-Foulkes (2002) illustrate such a situation. They consider three groups of countries. For one group (the rich countries), $x_i = x$. For another group (those catching up with the rich countries), $x_i(t) > x$. For the last group, $x_i(t) < x$.

discombobulate the evidence and may therefore have low discriminatory capability in applications. This section builds econometric specifications suitable for testing growth convergence using panel data and also examines time series convergence tests that are based on cointegration analysis. We start by setting up an empirical specification under the null hypothesis of growth convergence and then consider the alternative of divergence. We denote $y_i(t)$ as y_{it}, in keeping with discrete time modeling convention.

3.1 Growth Convergence

Under homogeneity, the transition path of log per capita real income can be written as

$$\begin{aligned}\log y_{it} &= \log \tilde{y}^* + \log A_0 + xt + [\log \tilde{y}_{i0} - \log \tilde{y}^*]\, e^{-\beta t} \\ &= a_t + k_i c_t = k_i(a_t/k_i + c_t) = k_i g_{it},\end{aligned}$$

and, allowing for heterogeneity, we have the same general representation

$$\begin{aligned}\log y_{it} &= \log \tilde{y}^* + \log A_{i0} + x_{it}t + [\log \tilde{y}_{i0} - \log \tilde{y}^*]\, e^{-\beta_{it} t} \\ &= a_{it} + k_i c_{it} = k_i(a_{it}/k_i + c_{it}) = k_i g_{it}.\end{aligned}$$

This path can be written in the equivalent form

$$\log y_{it} = b_{it}\mu_t, \tag{18}$$

where μ_t has the interpretation of a common growth component across economies ($g_{it} = \gamma_{it}\mu_t$, say), and b_{it} $(= k_i\gamma_{it})$ is the proportional deviation of individual i's growth from the common component. During the transitional period, b_{it} is a function of the speed of convergence β_i, the growth rate of technological progress x_i and the initial growth component level y_{i0}.

We start by considering the following two empirical models:

$$\begin{aligned}\text{Model A} &: \quad \log y_{it} = b_i\mu_t + \varepsilon_{it}, \quad \varepsilon_{it} = a_i + \rho_i\varepsilon_{it-1} + u_{it}, &(19)\\ \text{Model B} &: \quad \log y_{it} = b_{it}\mu_t + \varepsilon_{it}, \quad \varepsilon_{it} = a_i + \rho_i\varepsilon_{it-1} + u_{it}, &(20)\end{aligned}$$

where $u_{it} \equiv iid\ (0, \sigma_u^2)$, and b_i and b_{it} can be either stochastic or deterministic. For the moment, we will take b_{it} to be deterministic (and use ordinary limiting operations), but there is no serious loss of generality in doing so and stochastic cases will be considered later. We deliberately keep the time series structure of ε_{it} simple in recognition of the fact that the number of time series observations is often quite limited.

The conditions for growth convergence are given by

Model A

$$\begin{aligned}\text{C1} &: \quad b_i = b \ \text{ for all } i \text{ and } t. \\ \text{C2} &: \quad |\rho_i| < 1 \ \text{ for all } i.\end{aligned}$$

$$\begin{array}{ccl} & & \text{Model B} \\ \text{C1} & : & \lim_{t\to\infty} b_{it} = b \;\; \text{for } all \; i \\ \text{C2} & : & |\rho_i| < 1 \;\; \text{for } all \; i. \end{array}$$

It is straightforward to show that Model B satisfies the definition of growth convergence and Model B clearly nests Model A. We have the following proposition.

Proposition 1: *Under conditions C1 and C2 for each model, both A and B satisfy the definition of growth convergence*

Model A assumes that all economies are in steady state or near steady state. In other words, the growth rate of technological progress is constant during the time period over which the convergence hypothesis is being tested. Model B does not assume either that the economies are near steady state, or that they are in transition. Model B states only that if we have enough time series data and all economies converge, then b_{it} will converge to a single value, b, which is the growth rate in the steady state. Section 3.3 discusses this transition in more detail. The speed of convergence of b_{it} in this case should be faster than the speed of divergence of μ_t, i.e., $b_{it} - b = o\left(\mu_t^{-1}\right)$ if $\mu_t \to \infty$ as $t \to \infty$. Then,

$$b_{it}\mu_t = b\mu_t + (b_{it} - b)\mu_t = b\mu_t + o\,(1)\,, \quad \text{as } t \to \infty. \tag{21}$$

If the growth component μ_t is stochastic and $\mu_t = O_p(t^a)$ as $t \to \infty$, then (21) holds with $o\,(1)$ replaced by $o_p\,(1)$ provided $b_{it} - b = o(t^{-a})$.

Model A requires no such side condition because $b_{it} = b$ for all i and t and the difference between per capita income for economies i and j is $O_p(\mu_t^{-1})$.

Within this framework for growth convergence, most time series and panel empirical studies have relied on model A to test convergence. The approach is particularly inviting with panel data sets because, under model A, panel unit root tests can be used in a straightforward way to assess the evidence for growth convergence (GC). When A is the true data generating process (dgp), it is easy to see that the following equivalences hold:

$$\log y_{it} - \log y_{jt} = \begin{cases} I(0) \text{ for } all \; i \text{ and } j & \Longleftrightarrow \quad GC \text{ holds} \\ I(1) \text{ for } any \; i \text{ and } j & \Longleftrightarrow \quad GC \text{ not hold} \end{cases},$$

and

$$\log y_{it} - \frac{1}{N}\sum_{i=1}^{N} \log y_{it} = \begin{cases} I(0) \text{ for } all \; i & \Longleftrightarrow \quad GC \text{ holds} \\ I(1) \text{ for } any \; i & \Longleftrightarrow \quad GC \text{ not hold} \end{cases}.$$

On the other hand, if Model B is the true dgp, then for sample data with $t = 1, ..., T$ and fixed T,

$$\log y_{it} - \log y_{jt} = \begin{cases} I(0) \text{ for } all \; i \text{ and } j & \Longrightarrow \quad GC \text{ holds} \\ I(1) \text{ for } some \; i \text{ and } j & \not\Rightarrow \quad GC \text{ not hold} \end{cases},$$

and

$$\log y_{it} - \frac{1}{N}\sum_{i=1}^{N}\log y_{it} = \begin{cases} I(0) \text{ for } \textit{all } i. & \Longrightarrow \quad GC \text{ holds} \\ I(1) \text{ for } \textit{some } i & \not\Rightarrow \quad GC \text{ not hold} \end{cases}.$$

Obviously, the difference between models A and B arises because we may well have different trend growth behavior for finite t (i.e., $b_{it} \neq b_{jt}$ for finite t). In that case, $\log y_{it} - \log y_{jt}$ may test as $I(1)$ for some i and j, even though as t becomes large the $I(1)$ effect diminishes because of condition C1. On the other hand, if $\log y_{it} - \log y_{jt}$ tests as $I(0)$ for all i and j over the sample data, then C1 ensures that $I(0)$ behavior obtains over the long run and so GC holds.

Model B appears more realistic than model A in that B allows for heterogeneity in β and some degree of divergence for finite t, but requires convergence in the limit. Model A is much simpler and permits easy testing of growth convergence, but may be too restrictive in practical applications.

3.2 Growth Divergence

There are many ways of characterizing divergence[8]. We consider two specific alternatives here. In the first, log per capita real income follows an individual random walk with drift. That is,

$$\log y_{it} = \xi_{it},\ \xi_{it} = c_i + \xi_{it-1} + \varepsilon_{it}. \tag{22}$$

where $\varepsilon_{it} \equiv iid\ (0, \sigma_\varepsilon^2)$ across t. We may allow ε_{it} to be correlated across i provided the random walks have full rank N, so there is no cointegration between ξ_{it} and ξ_{jt} for $i \neq j$. Since $\log y_{it}$ diverges and there is growth divergence between $\log y_{it}$ and $\log y_{jt}$ for all $i \neq j$. Suppose we now write (22) in the form of (18), viz.

$$\log y_{it} = b_{it}\mu_t.$$

More specifically, we can write (22) as

$$\log y_{it} = c_i t + \sum_{s=1}^{t}\varepsilon_{is} = \left[c_i + \frac{1}{t}\sum_{s=1}^{t}\varepsilon_{is}\right]t, \tag{23}$$

so that $\mu_t = t$ and $b_{it} = c_i + \frac{1}{t}\sum_{s=1}^{t}\varepsilon_{is} \to_{a.s.} c_i$ as $t \to \infty$. Clearly, the same result will hold if ε_{it} is stationary and ergodic.

A second possibility is that log per capita real income follows a trend stationary process whose deterministic part is given by

$$\log y_{it} = a_i + c_i t = \left[\frac{a_i}{t} + c_i\right]t \tag{24}$$

[8] The ultimate requirement for convergence is $x_i(t) \to x$. Under this condition, log per capita real income will be trend stationary in the steady state. However, during transition, it could well be stochastically nonstationary depending on the manner in which x_{it} and β_{it} evolve over time. For instance, x_{it} could follow a martingale that converged almost surely to a constant value x.

where the stationary error component is ignored. As long as $c_i \neq c_j$ for $i \neq j$, $\log y_{it} - \log y_{jt}$ is divergent as $t \to \infty$. There is an observational equivalence problem in transition if c_i is the inverse of a_i. To see this, suppose $c_i = 1/a_i$, and observe that

$$\log y_{it} - \log y_{jt} = a_i - a_j + \frac{a_j - a_i}{a_i a_j} t,$$

giving

$$\log y_{it} = \log y_{jt}, \text{ at } t^* = a_i a_j.$$

Accordingly, $\log y_{it}$ and $\log y_{jt}$ may seem to converge as the economies evolve toward time t^* (which will be large if a_i and a_j are large) and diverge subsequently. In fact, the growth curves cross and have different slopes because $c_i = 1/a_i \neq 1/a_j = c_j$. So this case can be excluded.

3.3 A Transition Parameter to Assess Convergence

Irrespective of whether economies converge, the growth component of log per capita real income for country i can be written as

$$\log y_{it} = b_{it} \mu_t. \tag{25}$$

We can add in a stochastic component to this equation to capture cyclical and random components in the actual data and define $\log w_{it}$ to be actual observed log per capita real income

$$\log w_{it} = \log y_{it} + v_{it}, \tag{26}$$

where v_{it} is an additive shock.

Suppose $\log y_{it}$ is known or observable to the econometrician and (25) holds. Taking ratios to cross-sectional averages gives the standardized quantity

$$h_{itN} = \frac{\log y_{it}}{\frac{1}{N}\sum_{i=1}^{N} \log y_{it}} = \frac{b_{it}}{\frac{1}{N}\sum_{i=1}^{N} b_{it}}. \tag{27}$$

Under the null hypothesis of growth convergence and fixed N we have

$$\lim_{t\to\infty} b_{it} = b, \quad \lim_{t\to\infty} \frac{1}{N} \sum_{i=1}^{N} b_{it} = b, \tag{28}$$

so that

$$\lim_{t\to\infty} h_{itN} = 1. \tag{29}$$

In the steady state, log per capita real income follows the trend $b\mu_t$. This implies the limit of b_{it} determines the growth rate in the steady state (it is the growth rate when $\mu_t = t$). If the cross-sectional average of b_{it} is b as $N \to \infty$, then h_{itN} can be regarded as the degree of departure from the steady state. Hence we call h_{itN} a 'transition parameter' because its behavior as $t \to \infty$ reflects the convergence of b_{it} in (28).

Asymptotic Behavior of the Transition Parameter

We first consider the asymptotic behavior of h_{itN} as $t \to \infty$ with N fixed. Let $\sigma_{tN}^2 = \frac{1}{N}\sum_{i=1}^{N}(h_{itN}-1)^2$ be the sample cross section variation of h_{itN}. Under convergence, the following properties hold, the first being (29) above.

(i) $\lim_{t\to\infty} h_{itN} = 1$;

(ii) $\lim_{t\to\infty} \sigma_{tN}^2 = 0$.

If the alternative hypothesis of divergence holds, we have $\lim_{t\to\infty} b_{it} = c_i < \infty$, with $c_i \neq c_j$ for $i \neq j$, in place of (28). Define

$$\bar{c}_N = \frac{1}{N}\sum_{i=1}^{N} c_i, \quad \overline{c^2}_N = \frac{1}{N}\sum_{i=1}^{N} c_i^2.$$

With N fixed, we now have:

(iii) $\lim_{t\to\infty} h_{itN} = \phi_{iN} := c_i/\bar{c}_N$; and

(iv) $\lim_{t\to\infty} \sigma_{tN}^2 = \frac{1}{N}\sum_{i=1}^{N}(\phi_{iN}-1)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{c_i-\bar{c}_N}{\bar{c}_N}\right)^2 = \frac{\overline{c^2}_N-\bar{c}_N^2}{\bar{c}_N^2} = d_N > 0.$

Result (ii) is immediate. For (iii) and (iv), it is convenient to show the result using the previous two examples – a stochastic trend with drift, and a linear trend. In the case of a unit root process with drift, we have (c.f. (23))

$$b_{it} = c_i + \frac{1}{t}\sum_{s=1}^{t}\varepsilon_{is} \tag{30}$$

and by the strong law of large numbers

$$b_{it} \to_{a.s.} c_i, \quad \text{as } t \to \infty, \tag{31}$$

which is the slope coefficient of the linear drift. In the case of a linear trend, we have (c.f. (24))

$$\lim_{t\to\infty} b_{it} = \lim_{t\to\infty}\left[\frac{a_i}{t} + c_i\right] = c_i, \tag{32}$$

and again the limit is the slope coefficient of the linear trend. In both cases, we get

$$h_{itN} = b_{it}/\frac{1}{N}\sum_{i=1}^{N} b_{it} \to_{a.s.} c_i/\bar{c}_N = \phi_{iN}, \tag{33}$$

giving (iii) and (iv). The limit d_N in (iv) is non zero under divergence, representing the heterogeneity in the coefficients c_i which, in turn, reflects the heterogeneity in the fundamental parameters β_i and x_i.

Next consider the properties of h_{itN} for fixed (large) t and asymptotically large N. We confine our attention to the case (30), allowing ε_{is} to be cross sectionally correlated but requiring that laws of large numbers apply as $N \to \infty$. In particular, we require that the sample moments $(\bar{c}_N, \overline{c^2}_N)$ have finite limits, viz.,

$$\bar{c}_N \to \bar{c}, \quad \overline{c^2}_N \to \overline{c^2} \quad \text{as } N \to \infty, \tag{34}$$

and assume that

$$\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}, \frac{1}{N}\sum_{i=1}^{N} c_i\varepsilon_{is} \to_{a.s.} 0, \quad \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{is}^2 \to_{a.s.} \sigma^2, \quad \text{as } N \to \infty. \tag{35}$$

We further require that

$$E\left(\sum_{s=1}^{T} \varepsilon_{is}\right)^2 = T\omega_i^2 + o\left(T\right), \quad \text{uniformly in } i \tag{36}$$

where $\omega_i^2 = 2\pi f_{\varepsilon_i}(0)$ is the long run variance of ε_{is} and f_{ε_i} is its spectral density, and that

$$\frac{1}{N}\sum_{i=1}^{N} \omega_i^2 \to \omega^2. \tag{37}$$

Primitive conditions under which limits such as these hold are provided in Phillips and Moon (1999). Under the additional requirements (34) - (37), it is easily seen that

(v) $\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} b_{it}^2 = \overline{c^2} + \frac{\omega^2}{t} + o\left(\frac{1}{t}\right)$;

(vi) $\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} h_{itN}^2 = \frac{\overline{c^2} + \frac{\omega^2}{t} + o\left(\frac{1}{t}\right)}{\bar{c}^2}$; and

(vii) $\text{plim}_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} (h_{itN} - 1)^2 = \frac{\overline{c^2} - \bar{c}^2 + \frac{\omega^2}{t} + o\left(\frac{1}{t}\right)}{\bar{c}^2}$.

We may use property (vii) to compare the sample variation, σ_{tN}^2, of h_{itN} for large N at two points in time $t = T_1 = [Tr_1]$ and $t = T_2 = [Tr_2]$ with $r_2 > r_1$. In this case, we have the approximate relation

$$\frac{\sigma_{T_2N}^2}{\sigma_{T_1N}^2} \simeq \frac{\overline{c^2} - \bar{c}^2 + \frac{\omega^2}{T_2}}{\overline{c^2} - \bar{c}^2 + \frac{\omega^2}{T_1}} < 1, \quad \text{for } T_2 > T_1 \tag{38}$$

suggesting that the sampling variability σ_{tN}^2 may decrease as t increases, but not to zero when there is growth divergence.

The finite sample properties of the transition parameter h_{itN} depend on the distribution of initial income, technological progress, and the speed of convergence. Testing convergence using the transition parameter involves estimating h_{itN} and assessing evidence in support of the above properties.

4 Empirical Panel Unit Root Analysis with Cross Section Dependence

The following are two popular panel regression models used to test economic convergence based on Model A.

$$A1 : \log w_{it} - \log w_{\cdot t} = \alpha_i + \rho_i(\log w_{it-1} - \log w_{\cdot t-1}) + \sum_{s=1}^{p_i} \omega_{is}\Delta \left\{\log w_{it-s} - \log w_{\cdot t-s}\right\} + u_{it}$$

$$A2 : \log w_{it} - \log w_{jt} = \alpha_i + \rho_i(\log w_{it-1} - \log w_{jt-1}) + \sum_{s=1}^{p_i} \lambda_{ijs}\Delta \left\{\log w_{it-s} - \log w_{jt-s}\right\} + u_{it}$$

Here w_{it} is as defined in (26) and panel unit root tests are frequently used to study convergence. Regression A1 is studied by Evans and Karras (1996) and Evans (1998) while Regression A2 is examined by Bernard and Durlauf (1995). None of these studies account for cross-sectional dependence in the residuals, which seems to be an important feature of the data in empirical work. We therefore seek to take this dependence into account. The empirical analysis that follows uses this panel unit root approach to examine the evidence for regional convergence among the 48 contiguous US States over 1929 to 1998, a context where cross section dependence is to be expected. We report empirical results only for Regression A1 since the results are similar for Regression A2.

As the degree of cross-sectional dependence becomes greater, it is known (Phillips and Sul, 2002) that panel unit root tests that fail to take account of cross-sectional dependence quickly become very unsatisfactory. A conventional way of dealing with cross-sectional dependence is to include a common time effect, but this approach is also unsatisfactory when the common time effect has a different impact on different series[9]. To allow for some extra generality, therefore, we suppose the regression residuals have the following parametric cross-sectional structure (used in Barro and Sala-i-Martin, 1992, and Phillips and Sul, 2002)

$$u_{it} = \delta_i\theta_t + e_{it}, \tag{39}$$

where the e_{it} are iid and independent of θ_t, and δ_i is nonstochastic. Under this form of cross sectional dependence, conventional panel unit root tests fail because of nuisance parameter dependencies. Phillips and Sul (2002) proposed an orthogonalization procedure (OP) to eliminate the idiosyncratic loading parameters δ_i in the model (39), which makes it possible to conduct asymptotically similar tests for a panel unit root. A related approach was suggested independently by Moon and Perron (2002).

The null hypothesis for both regressions A1 and A2 is a common panel unit root, i.e., $H_0^A : \rho_i = 1$ for ALL i. With a common panel unit root, all economies diverge so that rejection

[9] Regression A includes a common time effect. However, the common time effects work only when the cross-sectional dependence is identical. For example, if $u_{it} = \delta_i\theta_t + e_{it}$ as in (39), then $u_{it} - u_{\cdot t} = (\delta_i - \bar{\delta})\theta_t + e_{it} - e_{\cdot t}$. So, cross-sectional dependence is not eliminated by using a common time effect when the idiosyncratic loadings δ_i differ across i. The consequences of using a common time effect when the errors u_{it} have different factor loadings are investigated in Phillips and Sul (2002).

of H_0^A does not itself imply overall convergence. To allow for the fact that some units may converge and others may not, subgroup analyses can be used to confirm overall convergence and check for conditional convergence. This approach raises the question of how to construct subgroups. Three somewhat crude classifications are used here to determine subgroupings. The classifications involve cross sectional correlation[10], geographical distance, and the level of income[11]. Alternative subgrouping methods are possible, some of which are data based. For example, Sun (2001) considers a clustering algorithm to find a small group of so-called convergence clubs; and Hobijin and Franses(2000) rely on individual p-value of the KPSS test statistic to construct regional convergence groups.

Table 1: Panel Unit Root Evidence on Regional Convergence among US States over 1929-1998

	G^+	Z	% of $\hat{\rho}_{emu} = 1$
P-values			
All (48)	0.032	0.003	40
Subgroupings According to Income Level			
High (10)	0.282	0.259	33
Mid (17)	0.003	0.003	20
Low (21)	0.090	0.055	34
Subgroupings According to Cross-Sectional Error Correlation			
High (25)	0.361	0.071	100*
Mid (11)	0.005	0.019	27
Low (12)	0.262	0.136	43
Subgroupings According to Broad Regional Specification			
Northeast (16)	0.024	0.019	18
West (18)	0.000	0.004	17
South (14)	0.000	0.001	13

(*) signifies use of a seemingly unrelated median unbiased estimator (see Phillips and Sul, 2002).

[10] We consider the following iterative strategy for the selection of sub-panels based on cross section correlation. This is a likelihood-based algorithm for producing state clusters. First, we construct k-state subpanels where k is a small number like 3. Second, we estimate ρ by using seemingly unrelated median unbiased estimator proposed by Phillips and Sul (2002) and calculate the corresponding imputed likelihood. Drop those cross sectional units that fall in the lowest 1 % of likelihood values. Finally, we increase k and construct another set of sub-panels from the smaller overall group. Then repeat first 2 steps as long as the lowest 1% and highest 1% are distinct, dropping the lowest 1% until the lowest and highest 1% are no longer distinct. At this point, collect all the States that have been dropped - these constituted the lowest 1%. The highest 1% are those that stay in the highest 1% classification until the final iteration. The remaining States constitute a middle group. The lowest, average, and highest cross sectional correlations are (0.85, 0.92, 0.99), (0.74, 0.84, 0.95) and (0.56, 0.79, 0.95) for the high, middle, and low cross-sectional panel, respectively.

[11] Those States for which the average income over the 70 years is more than \$14,000 are classified as the high income group. States with average income between \$14,000 and \$12,000 are classified as mid-income, and those below \$12,000 as low income.

Table 1 summarizes the empirical results. The first two columns show the p-values of Phillips and Sul's (2002) two test statistics (G^+ and Z) for panel unit roots allowing for cross-sectional error dependence of the form (39) in regression A1.

For all 48 States, the p-values are lower than 5%, so the null hypothesis of panel unit roots (i.e., all States diverge) is rejected strongly. As discussed earlier, rejection does not imply that all States converge, only that some States converge. To confirm whether rejection implies overall convergence here, we estimate ρ_i individually using an exact median unbiased (EMU) estimator to avoid small sample bias (see Andrews, 1991, and Phillips and Sul, 2002) and calculate how many EMU estimates, $\hat{\rho}_{emu}$, are equal to unity (Note that EMU estimates can equal to unity with positive probability in finite samples). We use these numbers as weak supplementary evidence about the convergence properties in a mixed panel, i.e, a panel whose elements contain both non-stationary and stationary time series. This value for regression A1 is displayed in the third column. About 20 estimates of ρ_i out of 48 (43%) in these regressions equal unity. However, the interpretation of this result is not that 43% of the States diverge or that 57% converge. For, even if the true $\rho_i = 1$, then only 50% of the sampled EMU estimates would be equal to unity because the estimator is median unbiased. The correct interpretation of the number 43% is that it is highly likely that some of the ρ's are unity or very close to unity.

Interestingly, the empirical results shown in Table 1 seem to tell us that high income States diverge, and the same is so (at least marginally) for low-income States, but mid-income States may converge, although the fractional percentage of $\hat{\rho}_{emu}$'s equal to unity is still 20% or greater in this latter case. A subpanel analysis based on cross-sectional error correlation confirms these findings. For the highly dependent States, there is little evidence of economic convergence. Nor is there for States with low dependence. With subgroupings based on broad regional groups, however, there appears to be evidence for economic convergence, in contrast to the other classifications.

5 Fitting the Transition Parameter

Fitting h_{itN} from data using a regression approach inevitably encounters difficulties because the number of unknowns is the same as the number of observations. Smoothing methods offer an obvious alternative approach to extracting h_{itN} from $\log w_{it}$. Accordingly, we employ the Whittaker-Hodrick-Prescott (WHP) filter[12] to estimate the trend component $b_{it}\mu_t$. The procedure is flexible, requires only the input of a smoothing parameter, and does not require prior specification of the number of common trends μ_t in $\log w_{it}$. Having computed the WHP estimate

$$\hat{f}_{it} = \widehat{b_{it}\mu_t} \tag{40}$$

[12] Whittaker (1923) first suggested this penalized method of smoothing or 'graduating' data. The approach has been used regularly in empirical work in time series macroeconomics since Hodrick and Prescott (1982/1997), although the link to the earlier work by Whittaker and the large literature on smoothing (e.g. see Kitagawa and Gersch, 1996) seems not to have been generally acknowledged in the economics literature.

of $f_{it} = b_{it}\mu_t$, we take cross-sectional averages as in (27), leading directly to estimates,

$$\hat{h}_{it} = \frac{\hat{f}_{it}}{\frac{1}{N}\sum_{i=1}^{N}\hat{f}_{it}}$$

of h_{itN}. For notational simplicity and because N is given, we use $\hat{h}_{it}$ rather than $\hat{h}_{itN}$ here and in what follows. Writing

$$\hat{f}_{it} = f_{it} + e_{it} = \left[b_{it} + \frac{e_{it}}{\mu_t}\right]\mu_t,$$

it is clear that under growth convergence (where $b_{it} \to b$ as $t \to \infty$) we have

$$\hat{h}_{it} = \frac{\left[b_{it} + \frac{e_{it}}{\mu_t}\right]}{\frac{1}{N}\sum_{i=1}^{N}\left[b_{it} + \frac{e_{it}}{\mu_t}\right]} \to_p 1, \quad \text{uniformly in } i \text{ as } t \to \infty, \tag{41}$$

if $\frac{e_{it}}{\mu_t} = o_p(1)$ [13] uniformly in i as $t \to \infty$. Under the same conditions

$$\hat{\sigma}_t^2 = \frac{1}{N}\sum_{i=1}^{N}(\hat{h}_{it} - 1)^2 \to_p 0, \quad \text{as } t \to \infty, \tag{42}$$

giving empirical versions of the earlier properties (i) - (ii) of h_{itN}. In effect, under conditions of convergence, when the parameters β_i and x_i take similar cross section values and b_{it} tends to a common value over time, the cross sectional variance of $\hat{h}_{it}$ will decline to zero over time.

Under the alternative of no convergence we have

$$\hat{h}_{it} \to_p \frac{c_i}{\bar{c}}, \quad \lim_{t\to\infty}\hat{\sigma}_t^2 > 0,$$

corresponding to the earlier properties (iii) - (iv) of h_{itN}, which provides evidence on divergent behavior. In effect, when the β_i and x_i are heterogeneous and there is heterogeneity over b_{it} as $t \to \infty$, the cross sectional variance of $\hat{h}_{it}$ will not decline to zero. Similarly, empirical estimates $\hat{\sigma}_t^2/\hat{\sigma}_q^2$ of the variance ratio for $t > q$ and its behavior as t increases can be compared to that of the asymptotic approximation of $\sigma_{tN}^2/\sigma_{qN}^2$ given by (38).

6 Some Empirical Findings on Convergence

We provide two different empirical illustrations. The first involves two separate panels in each of which there appear to be similar speeds of convergence: the 48 contiguous U.S. States panel; and a panel of national output of 21 OECD countries. The second involves the full PWT (version 5.6) panel where there is clearly heterogeneity over time in convergence rates.

[13] Since e_{it} is the error in the WHP filter estimate of f_{it} and μ_t is the common trend across units $i = 1, ..., N$, the condition $\frac{e_{it}}{\mu_t} \to_p 0$ uniformly in i seems reasonable. Primitive conditions under this holds will depend, inter alia, on the choice of the smoothing parameter in the WHP filter and its asymptotic behavior as the sample size increases. Phillips and Jin (2002) provide some asymptotic theory for the WHP filter under various assumptions about the smoothing parameter in the filter.

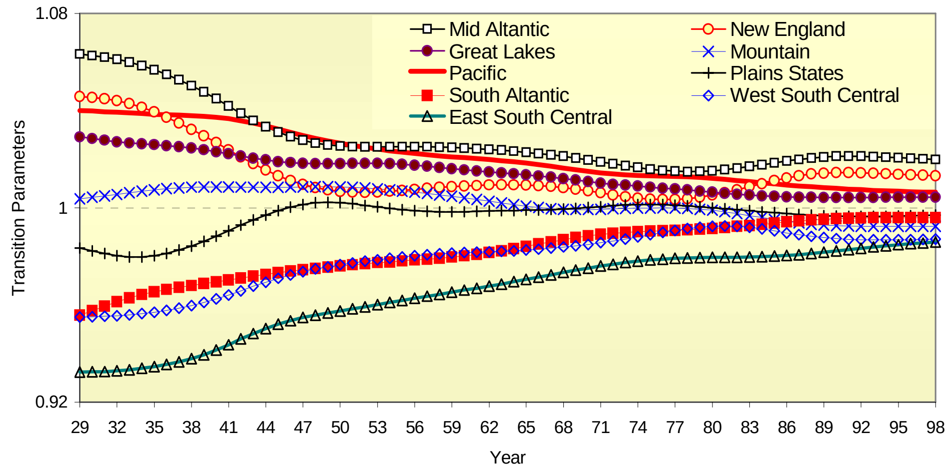

Figure 5: Time Profile of Regional Averages of Transition Parameter Estimates: 48 Contiguous U.S. States.

6.1 Panels with Similar Speed of Convergence

Fig. 5 shows the cross-sectional average of $\hat{h}_{it}$ over 9 regions for the contiguous U.S. States data, the importance of regional groupings of the US States data having shown up earlier in Table 1. The time profile of these cross sectional averages shows clear evidence that the h_{itN} have been converging over the last 70 years. Figure 6 shows the corresponding time profiles of $\hat{h}_{it}$ for the national output of 21 OECD countries taken from the Summers and Heston data set (PWT 5.6) from 1950 to 1992. Evidently, these two data sets tell much the same story – that these regional groupings and national economies are apparently converging over time.

The cross-sectional variance of $\hat{h}_{it}$ about the sample mean for given t is

$$\hat{\sigma}^2_{tN} = \frac{1}{N}\sum_{i=1}^{N}(\hat{h}_{it} - \frac{1}{N}\sum_{i=1}^{N}\hat{h}_{it})^2.$$

Under the null hypothesis of growth convergence, $\hat{\sigma}^2_{tN} \to 0$ as $t \to \infty$, while under the alternative, $\hat{\sigma}^2_{tN} \to d > 0$. Graphs of the empirical estimates of $\hat{\sigma}^2_{tN}$ over time are shown in Figs. 7 and 8 for the US States and OECD data. An approximate 95% confidence band for the path is shown in each case and is computed by bootstrapping[14] the estimated f_{it}. In both cases there

[14] The WHP filtered data is $\hat{f}_{it}$ from (40). For each t there are N cross-sectional units. The nonparametric bootstrap procedure is as follows:

1. Draw N random samples with replacement from $\hat{f}_{it}$, giving bootstrap data $\hat{f}^*_{it}$
2. Construct the bootstrap sample $\hat{h}^*_{it}$ and calculate the bootstrap sample variance $\hat{\sigma}^{*2}_t$.
3. Repeat the above steps 5,000 times and construct upper and lower 2.5% confidence intervals from the resulting bootstrap distribution of $\hat{\sigma}^{*2}_t$.

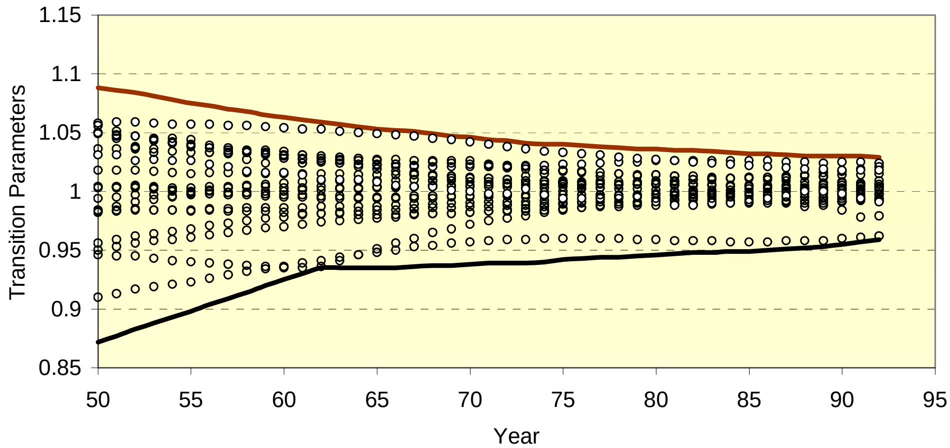

Figure 6: Time Profile of Transition Parameter Estimates: 21 OECD Countries.

is evidence of steady decline in $\hat{\sigma}^2_{tN}$ over time, as would be expected under convergence. The confidence bands do not include the origin and this is the result of the bootstrap construction, which uses an empirical distribution of sample variances, all of which have positive values.

6.2 Dealing with Time Heterogeneity

Fig. 9 shows the maximum, median and minimum estimates $\hat{h}_{it}$ computed as above using the WHP filter for 120 countries in the PWT Panel[15]. Country choice is based on data availability. Evidently the pattern in Fig. 9 is indicative of lack of convergence over the time frame of the data (1960 to 1989). But this profile does not necessarily imply that the 120 countries are diverging. As showed earlier, when there is cross-sectional and time series heterogeneity of x_i (and β_i), during transition we can expect some degree of divergent behavior. Hence, Fig. 9 does not of itself provide unequivocal evidence of divergence.

Our approach to assessing the empirical evidence for divergence is to compare the observed behavior of the transition parameter estimates $\hat{h}_{it}$ to the asymptotic properties known to hold for h_{it} from Section 3.3. Under divergence, from property (iii), $\hat{h}_{it}$ should not converge to the same value, nor should its sampling variability decline to zero. It is of particular interest to look at the behavior of these transition estimates in certain subpanels where we may expect to find more evidence of convergence. The following characterization of the subgroupings is useful.

Take a subgroup for which initial income levels are similar. If initial income is independent of β_i and x_i, then the speed of growth convergence within the subgroup will depend totally on

[15] We also used panel data from 1960 to 2000 for 97 countries from the World Bank Development Indicator (WBDI). The empirical results discussed in the remainder of this Section are generally stronger with the WBDI data. The results are available on request from the authors.

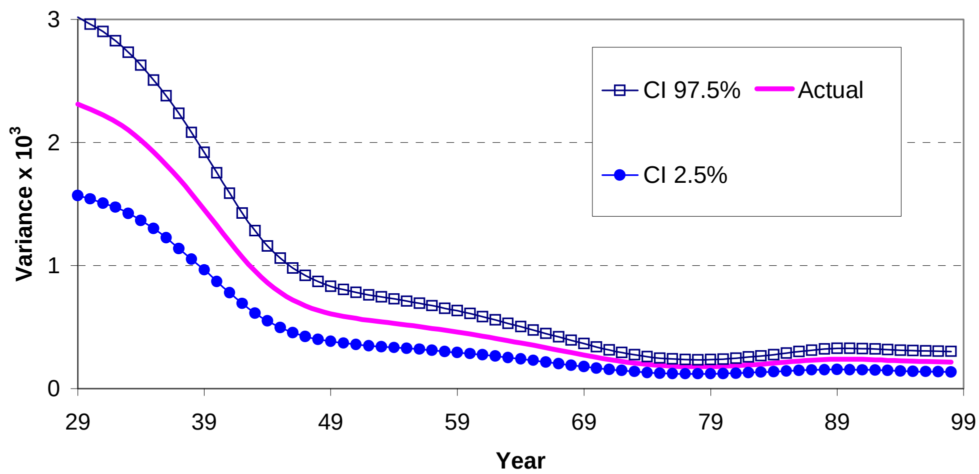

Figure 7: Time Path of $\hat{\sigma}_t^2$ with Bootstrap Confidence Interval for the 48 US States Panel.

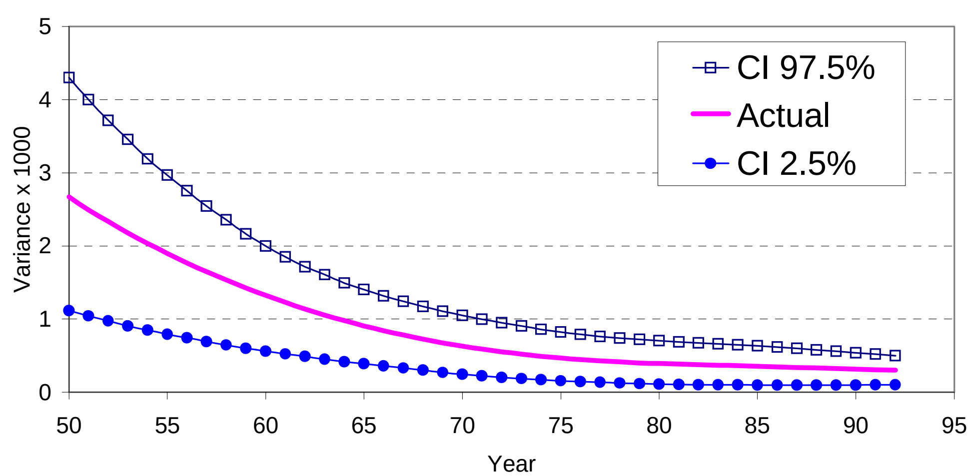

Figure 8: Time Path of $\hat{\sigma}_t^2$ with Bootstrap Confidence Interval for the 21 country OECD Panel

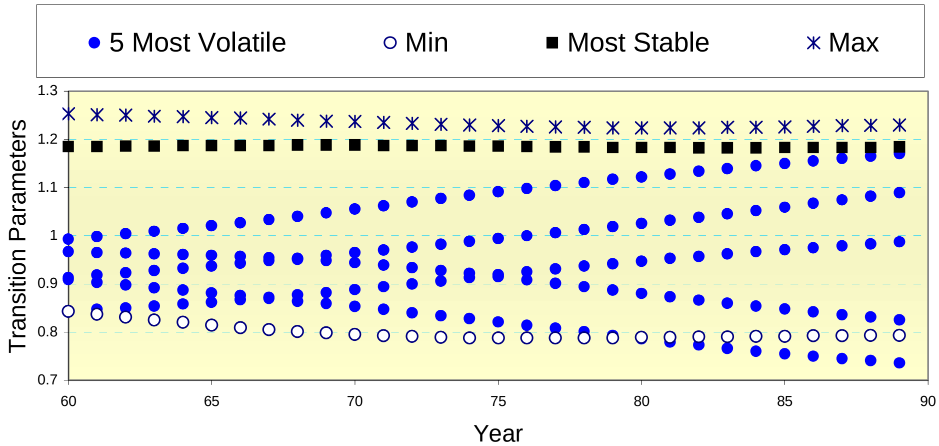

Figure 9: Transition Parameters in PWT. (From 1960 to 1989 over 120 Countries)

β_i and x_i. Provided $\beta_i \neq \beta_j$, log per capita real incomes in this subgroup should diverge over time. Hence, when panel clustering is based on initial income, the following behavior might be expected. In view of (33), even though $\log y_{it}$ diverges over time, the transition parameter h_{it} is ultimately constant. Accordingly, for a subpanel in which initial incomes are similar, the transition parameters should stabilize over time regardless the level of initial income. But when $\beta_i \neq \beta_j$ and there is growth divergence, there will be also be heterogeneity in the constant levels to which the transition parameters stabilize and this will be reflected in the cross section variance estimates.

Subgroup analysis based on an ordering of final period income is also helpful to detect heterogeneity in β_i and x_i. The followings outcome might be expected when the panel clustering is based on final period income. Among countries for which final period incomes are similarly high, the transition parameter should converge over time relatively faster according as initial period income is more heterogenous.

We examine the evidence for these theory-based claims using data for 120 countries from the PWT (version 5.6) which covers the period from 1960 to 1989. We split the 120 countries into five groups. Each group contains 24 countries based on a sequential income ordering. That is, the first 24 poorest countries (classified here according to both initial period and final period income orderings) become the members of the first group, the next 24 poorest countries form the next group, and so on. We also considered the ad hoc income ordering followed by Durlauf and Quah (1999) and found that the particular selection rule for the subgrouping did not alter the results in any major way. We report here only the results based on sequential income ordering[16].

Figs. 10 and 11 display some supportive evidence for growth convergence among the 120

[16] The ad hoc income ordering results can be obtained from the authors on request.

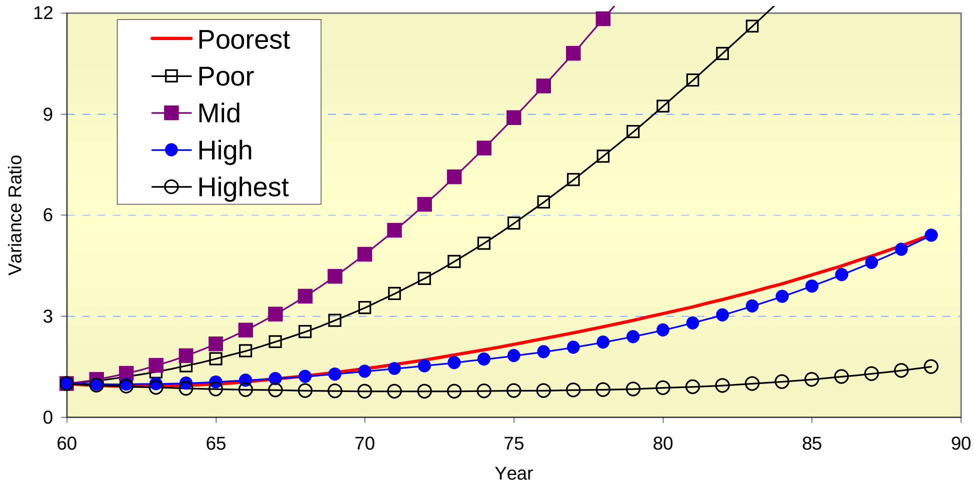

Figure 10: Variance Ratio based on Initial Period Income Ordering: (Variance Ratio = $\hat{\sigma}^2_{h,t}/\hat{\sigma}^2_{h,1}$)

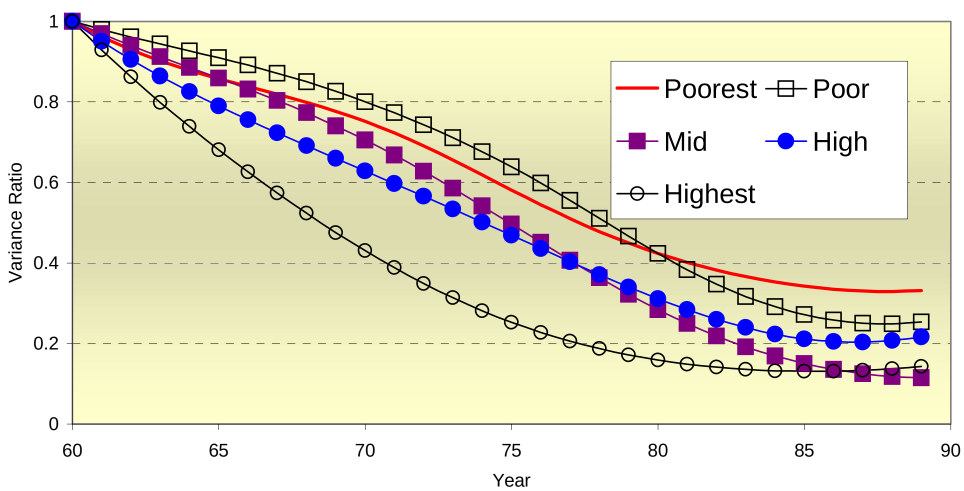

Figure 11: Variance Ratio based Final Period Income Ordering (Variance Ratio = $\hat{\sigma}^2_{h,t}/\hat{\sigma}^2_{h,1}$).

countries based on the above criteria. Fig. 10 shows the ratio of the variance of the latest period transition parameter $\hat{h}_{it}$ to that of the first value $\hat{h}_{i1}$ when the data are clustered according to initial income. If the transition parameters converge (diverge) over time, then the variance ratio should decrease (increase) over time. As neoclassical theory suggests under cross-sectional and time series heterogeneity, when initial incomes are similarly high, the transition parameters do not converge or diverge. But, on the contrary, when initial incomes are similar and moderately low, the $\hat{h}_{it}$ seem to diverge rapidly - as seen in Fig. 10. The economic implication of Fig. 10 is simply that early divergence of the transition parameters can occur in situations where there may also be ultimate growth convergence.

Similarly, Fig. 11 shows the ratio of the variance of the latest period transition parameter $\hat{h}_{it}$ to that of the first value $\hat{h}_{i1}$ when the data are clustered according to final period income. In all cases, it appears that the variance ratio declines, and the rate of decay seems to be greatest for clusters with the highest income.

Do these results provide confirmatory evidence of growth convergence in the PWT across so many nations? Unfortunately, the evidence is not persuasive enough to support such as strong conclusion. It turns out that similar behavior can obtain under conditional convergence, as we proceed to show in the next section.

6.3 Conditional $\beta-$Convergence

Suppose that there is no cross-sectional heterogeneity in β_i or time series heterogeneity of x_i. Then the transition dynamics of growth convergence should rely only on initial incomes. Suppose, in addition, that we successfully picked five conditional convergence groups. The question then is whether figures 10 and 11 could be generated under conditional $\beta-$ convergence.

Use the notation in section 2 and assume there are two different steady states involving two groups of countries, G_1 and G_2. The number of countries in G_1 and G_2 are N_1 and N_2, respectively, with $N = N_1 + N_2$. Observe that

$$\log y_i(t) = \log \tilde{y}_s^* + [\log \tilde{y}_i(0) - \log \tilde{y}_s^*]\, e^{-\beta t} + \log A_s(0) + x_s t, \text{ for } i \in G_s \text{ and } s = 1, 2. \quad (43)$$

$$\lim_{t\to\infty} b_{it} = b_s \text{ for } i \in G_s \text{ and } s = 1, 2, \quad (44)$$

and

$$\lim_{t\to\infty} h_{itN} = \begin{cases} c_1 = b_1/\left(kb_1 + (1-k)b_2\right) & if\ i \in G_1 \\ c_2 = b_2/\left(kb_1 + (1-k)b_2\right) & if\ i \in G_2 \end{cases} \quad \text{for } k = N_1/N$$

Suppose that we take the two groups together and estimate the transition parameters. During transition, the transition parameters in G_1 approach c_1 while those in G_2 approach c_2. Figure 12 shows 6 hypothetical economies with two conditional convergence groups. The per capita real income in economies 1, 2 and 3 converge to each other but diverge from those in economies 4, 5 and 6. If $\beta-$convergence holds in the two groups, final period income ordering results in convergence among economies $\{1, 2, 3\}$, or $\{4, 5, 6\}$. Conditional $\beta-$ convergence

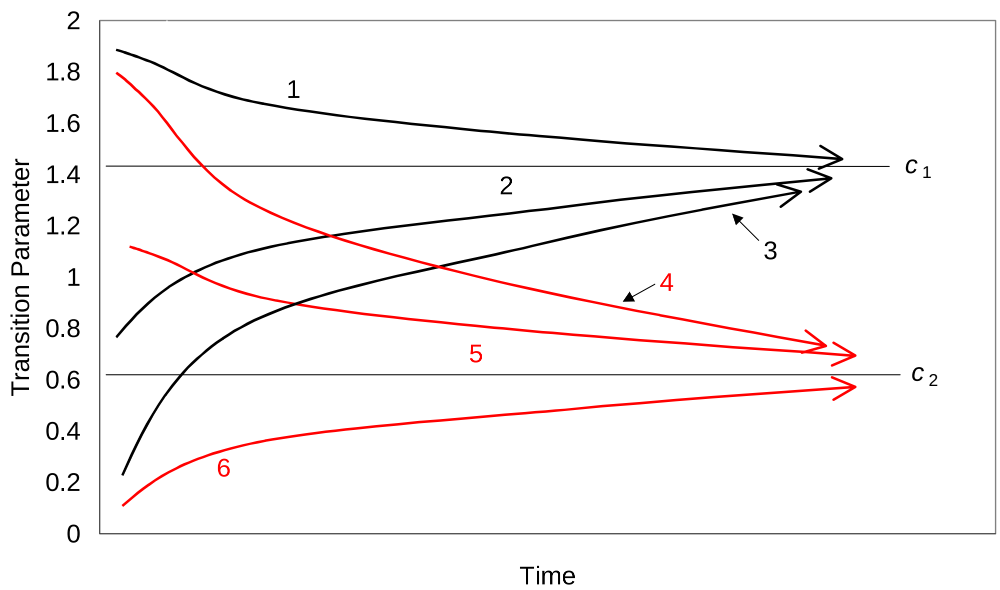

Figure 12: Conditional β-Convergence

then explains how the transition parameters can converge over time faster when initial period income is more heterogenous.

By contrast, initial income ordering yields divergence between economies 1 and 4, 2 and 5, or 3 and 6. Also, some countries (e.g. 3 and 6) with similarly low initial incomes diverge, as do some countries (e.g., 1 and 4) with similarly high initial incomes. The paths of the transition parameters for these countries will not converge, reflecting that divergence.

Therefore, absolute convergence with cross-sectional and time series heterogeneity or conditional $\beta-$ convergence (possibly with initial income variance differentials across groups) can both generate the observed divergence among these 6 economies in transition.

The example illustrates how elusive the shadow of growth convergence can be. In effect, the transitional time paths that arise under absolute convergence where there is cross-sectional and time series heterogeneity can have observational similarities to those that hold under conditional $\beta-$ convergence. Yet there is a huge difference between the long term economic implications of absolute convergence and conditional convergence. For, if conditional $\beta-$ convergence holds, some poor countries will remain poor and may never catch up with the rich countries or the Asian Dragons. Whereas under absolute convergence, all poor countries will eventually catch up.

The final question bears directly on this issue. Can we distinguish empirically between these two hypotheses about convergence. In particular, is it possible to distinguish whether the 120 countries in the PWT are on growth trajectories consistent with absolute convergence or conditional $\beta-$ convergence? With only 30 time series observations on these economies, the task of making this empirical distinction seems almost impossibly elusive. Nonetheless, it is possible to characterize some relevant changes that have been taking place over this 30 year period that do bear on this issue.

Suppose we sort the countries into groups according to per capita real income at an intermediate (middle) time observation, say t_M. Transition parameters in each group should converge if countries are en route to convergence. If conditional $\beta-$ convergence holds in the group, then the transition parameters should converge further after t_M. If absolute convergence holds because there is cross-sectional and time series heterogeneity that accomplishes this after t_M, then the transition parameters should correspondingly reveal this heterogeneous behavior after t_M.

We select the middle year, $t_M = 1974$, in our panel data set and sort the countries accordingly. Fig. 13 shows the results of this exercise giving the ratio of the variance of the latest period transition parameter $\hat{h}_{it}$ to that of the first value $\hat{h}_{i1}$ when the data are clustered according to middle year income. For all groups, conditional $\beta-$ convergence does not seem to hold because the variance ratios of the transition parameters evidently starts to diverge following 1975 for these particular groups.

There is, of course, the possibility that conditional $\beta-$ convergence holds for only $k\%$ of the countries in a group. In that case, the divergence observed in Fig. 12 may arise from the behavior of the remaining $(1-k)\%$ of the countries in the group. We examined this possibility by selecting subgroups in a mechanical way by adding and subtracting countries from each group according to their initial income levels[17]. This process showed little evidence for conditional $\beta-$convergence within the given country groupings.

7 Conclusion

The paper seeks to adopt an agnostic position on the issue of growth convergence and to let the data speak on the issue in a manner that is not overly restrictive. In particular, we allow for the presence of heterogeneity across countries and regions as well as over time when we study patterns of growth to investigate convergence or divergence. By doing so, we catch an

[17] Details of the method used are as follows. We chose the first K countries in a group based on their income level at t_M. If there were a group of countries for which conditional $\beta-$convergence held, then the variance ratio of $\hat{\sigma}^2_{1989}/\hat{\sigma}^2_{1974}$ (or $\hat{\sigma}^2_T/\hat{\sigma}^2_{T_M}$) should be less than unity, or at most near unity. Let the critical value for the variance ratio be 'V_c'. If the variance ratio is less than V_c, then we drop the first low income country and add the next higher income country in 1975 to maintain K. We repeat this recursion until there is no remaining country unconsidered. For given K, the total number of cases is $T-K+1$. We set $V_c = 1$, the minimum K to be 10, and the maximum K to be 40. With this selection process, we were not able to find any case where the variance ratio was less than V_c out of 2945 cases. Increasing V_c up to 1.2 still did not yield any case.

We set $t_M = 1979$ and repeated this whole exercise. If the distribution of initial income in a particular conditional $\beta-$convergence group is wide, then selecting a larger t_M should help to identify the convergence group. In this case, we found one subgroup with $V_c = 1.1$ and $K = 13$ while another subgroup with $V_c = 1.2$ and $K = 33$.

Overall these exercises indicate that there is little evidence for conditional $\beta-$convergence within the given country groupings.

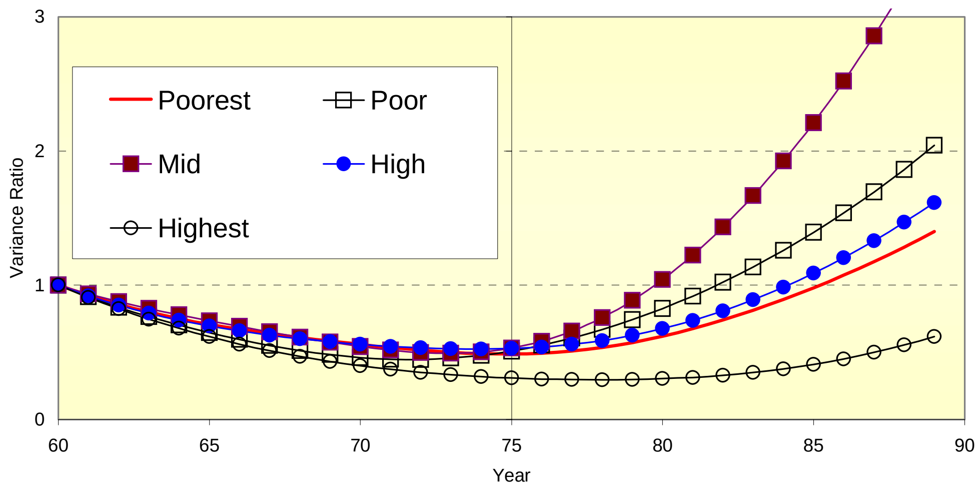

Figure 13: Sorted by the middle observation. $t_M = 1975$

empirical glimmer of absolute growth convergence among the 120 nations of the Penn Tables. Assuming that growth rates of technological progress converge to a constant in the long run, we can generate the cross-sectional divergence that is present in the data. To do so is important because when the growth rate of technological progress in a poor country is lower than it is in a rich country, temporary divergence in the transitional growth paths can occur. Reconciling this divergence in transitional paths within an overall pattern of growth convergence is a challenge that stretches existing panel data to its limits and explains our characterization of convergence as an elusive shadow.

These findings corroborate some of the arguments made in Lucas (2002), where it is suggested that the world economy diverges during transition and then starts to converge. Lucas's conjecture is based on heterogeneity in the speed of human capital accumulation. While it is not a sequential growth path but instead represents the average growth of clusters of countries with widely different incomes, our Fig. 1 is similar in shape to the 1,500 year growth path plotted by Lucas (2002). In the early stages of an economy's growth, the cumulative human capital stock is low and growth is slow. As an economy learns, imports or creates technology, the economy grows faster. These stages of growth are reflected in Lucas's growth path and in Fig. 1.

One implication of the analysis in this paper is that technological progress and its dependence on time can play an important role in growth convergence, affecting the speed of convergence as well as the final steady state. However, the growth rate of technological progress and its time path are assumed to be exogenous and unknown. In consequence, like many other studies of economic growth and convergence, this paper fails to cast light on the growth path of technological progress, which continues to be in large part mysterious. In the absence of light on this path, our understanding of economic growth and convergence is inevitably very limited. As such, the black hat of convergence and the coat of divergence remain somewhat elusive empirical shadows.

References

[1] Andrews, D. W. K. (1993), "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models." *Econometrica*, 61, 139-165.

[2] Barro, R.J., and Sala-i-Martin (1992), "Convergence", *Journal of Political Economy* 100 (2): 223-251.

[3] Barro, R.J., and Sala-i-Martin (1995), *Economic Growth.* New York: McGraw-Hill.

[4] Bernard, A.B, and S.N. Durlauf (1995), "Convergence in International Output", *Journal of Applied Econometrics* 10(2): 97-108.

[5] Bernard, A.B, and S.N. Durlauf (1996), "Interpreting Tests of the Convergence Hypothesis", *Journal of Econometrics*, 71(1-2): 161-174.

[6] DeLong, J.B (1988), "Productivity Growth, Convergence, and Welfare: A Comment", *American Economic Review* 78(5):1138-55.

[7] Den Haan, W.J. (1995), "Convergence in Stochastic Growth Models: the Importance of Understanding Why Income Levels Differ", *Journal of Monetary Economics* 35(1): 65-82.

[8] Durlauf, S.N. and D.T. Quah (1999), "The New Empirics of Economic Growth", Ch. 4 in Handbook of Macroeconomics 1A, (Elsevier)

[9] Evans, P (1998), "Using Panel Data to Evaluate Growth Theories", *International Economic Review* 39(2): 295-306.

[10] Evans, P. and G. Karras (1996), "Convergence Revisited," *Journal of Monetary Economics*, 37, 249-265.

[11] Hodrick, R. and E. Prescott (1997). "Post war business cycles: An empirical investigation". *Journal of Money Credit and Banking*, 29, 1-16.

[12] Howitt, P. and D. Mayer-Foulkes (2002), "R&D, Implementation and Stagnation: A Schumpeterian Theory of Convergence Clubs," mimeo, Brown University.

[13] Kitagawa, G. and W. Gersch (1996). *Smoothness Priors Analysis of Time Series*. New York: Springer Verlag

[14] Lee, K., M.H. Pesaran and R.P. Smith (1997), "Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model", Journal of Applied Econometrics 12(4): 141-163.

[15] Lucas, R.E. Jr. (2002) "The Industrial Revolution: Past and Future," in Lucas, R. E. Jr *Lectures on Economic Growth*. Cambridge, Massachusetts: Harvard University Press.

[16] Mankiw, N.G., D. Romer, and D.N. Weil (1992), "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics,* 107, 407-438.

[17] Moon, H. R and B. Perron (2002). "Testing for a Unit Root in Panels with Dynamic Factors," USC, *mimeographed*.

[18] Nerlove, M. "Growth rate convergence, fact or artifact? An essay on panel data econometrics", in J. Krishnakumar and E. Ronchetti (eds.) *Panel Data Econometrics. Papers in Honour of Professor Pietro Balestra*. Amsterdam: North Holland.

[19] Phillips, P.C.B. and H.R. Moon (1999) : Linear Regression Limit Theory for Nonstationary Panel Data, *Econometrica*, 67, 1057–1111.

[20] Phillips, P.C.B. and D. Sul, (2002) "Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence", Cowles Foundation Discussion Paper No. 1362, Yale University.

[21] Phillips, P. C. B. and S. Jin, (2002). "Limit Behavior of the Whittaker (HP) Filter". Yale University, mimeographed.

[22] Quah, D. (1993), "Galton's Fallacy and Tests of the Convergence Hypothesis," *Scandinavian Journal of Economics* 95, 427-443.

[23] Quah, D. (1996), "Empirics for Economic Growth and Convergence," *European Economic Review* 40, 1353-75

[24] Rappoport, J. (2000), "Is the Speed of Convergence Constant?" FRB of Kansas, Working Paper RWP 00-10.

[25] Sun, Y. (2001), "Catching up, Forging ahead, and Falling Behind: A Panel Structure Analysis of Convergence Clubs". Yale mimeographed.

[26] Whittaker, E. T. (1923). On a new method of graduation. Proceedings of the Edinburgh Mathematical Association, 78, 81-89.

Table A: World Income Distribution from Poor to Rich[18] (unit: US$)

Ordering	T=1960			T=1989		
Initial Income Ordering with Actual Data						
	Min	Mean	Max	Min	Mean	Max
Poorest	315	503	673	424	788	2,567
Poor	673	977	1,151	350	1,377	5,793
Middle	1,174	1,393	1,673	676	3,164	11,077
Rich	1,763	2,193	3,434	1,479	5,389	14,333
Richest	3,594	6,012	9,997	4,563	13,065	17,739
Last Income Ordering with Actual Data						
	Min	Mean	Max	Min	Mean	Max
Poorest	344	586	1,223	350	637	879
Poor	315	1,016	1,788	879	1,379	2,168
Middle	420	1,427	2,772	2,169	3,002	3,672
Rich	824	2,154	6,657	3,683	5,729	8,229
Richest	1,483	5,830	9,997	8,597	13,569	17,739
Initial Income Ordering: Confidence Intervals from Boostrap						
	2.5%	50%	97.5%	2.5%	50%	97.5%
Poorest	452	490	530	706	855	1,051
Poor	889	947	1,007	1,090	1,394	1,775
Middle	1,316	1,383	1,455	2,063	2,753	3,575
Rich	2,158	2,350	2,573	3,949	4,947	6,177
Richest	5,414	6,031	6,734	9,984	11,660	13,245
Last Income Ordering: Confidence Intervals from Boostrap						
	2.5%	50%	97.5%	2.5%	50%	97.5%
Poorest	550	639	749	576	630	687
Poor	738	886	1,047	1,252	1,381	1,530
Middle	1,133	1,361	1,605	2,685	2,869	3,068
Rich	1,824	2,201	2,691	5,127	5,677	6,288
Richest	4,513	5,476	6,426	12,096	13,060	14,027

[18] Actual data for Fig. 1 and Fig. 3 and bootstrap data for Fig. 2 based on 5,000 replications of the observed trajectories.