Convergence as distribution dynamics
(with or without growth)

by

Danny T. Quah*
LSE Economics Department and CEP

CENTRE FOR ECONOMIC PERFORMANCE
DISCUSSION PAPER NO. 317
November 1996

This paper is produced as part of the Centre's Programme on National Economic Performance.

* I thank Fabio Canova, Daniel Cohen, and Lucrezia Reichlin for helpful discussions. All data analysis was performed using the author's econometrics shell tSrF.

Nontechnical Summary

Convergence is a catchy idea, but one that organizes serious thinking in areas as diverse as economic growth, theoretical econometrics, finance, European politics and monetary union, regional planning and geography, up through but not ending at entertainment and multi-media technology, and the software industry.

Some growth economists define convergence as a single economy approaching its theoretically-derived steady state growth path. Others translate this to whether poor economies are catching up with rich ones. Yet others think of these two statements as being identical, and thus of either indicating convergence.

At one extreme, econometricians and probabilists have found it useful to work with different notions of convergence of sequences of random variables. At a different extreme, economists and policy-makers in Europe have obsessed on the Maastricht convergence criteria. Finally, when high-tech, fast-growth market participants—people who actually create value in modern economies—get together, they too excitedly discuss convergence, but now between biological and machined products, or between communications, computers, and content. 1996). In every instance the term convergence is used with a different meaning—and rightly so.

That convergence suffers from this meaning-overload should not disguise its importance. This paper concerns convergence in the sense of poor economies catching up with the rich. If by economies, one means countries, then magnitudes of the numbers alone should already show why convergence is important. Some countries have been doubling per capita incomes every decade; yet others have been stagnant, with levels of per capita income a hundred times lower than those of the leading economies.

Such back-of-the-envelope facts are obvious and easy to obtain. But are there empirical regularities beyond them that could be useful for advancing economists' understanding of convergence? If the current situation continues, how will cross-country income distributions look in the

future? Will rich economies always remain in a "club" of rich countries, and similarly the poor? What possibilities are there for switches across relatively poor and relatively rich?

Consider the same question, but now with the economies being regions within countries: are poorer regions languishing, currently and forever behind richer ones, or do they face any possibility of catching up? As before, one seeks an empirical characterization on such dynamic possibilities sufficiently precise and apposite so that one can understand the implications of alternative scenarios.

Such questions concern the behavior over time of cross-section distributions of income (or output or welfare): the issues are, writ large to the scale of macroeconomies or regions, the same as those that have traditionally been the concerns of research on the dynamics of inequality and personal income distributions. Are the distributions collapsing, so that everyone shows a tendency to become equally well off? Instead, are the distributions increasing in dispersion, so that those relatively better off are getting more so? Or, are the distributions tending towards shapes that show clusters and subgroups, so that the population is polarizing and then stratifying into distinct classes? Questions like these are useful for appreciating patterns of cross-country growth, just as they are for understanding patterns of dynamics and mobility in individuals' incomes within societies.

Such questions also show that economic growth—while unquestionably important for welfare—need not be the only mechanism permitting understanding of convergence. In the framework adopted below, convergence might occur because of growth. Or, convergence might occur without growth. Pinning down a theory of growth is not essential to understanding whether poor economies will become rich, or whether they will remain poor. I should clarify that I am not referring here to the classical Kuznets question: What is the relation between income distributions *within* an economy and the aggregate growth path of that economy? Rather, I am concerned with the growth paths of many different aggregate economies and the implications those have for the dynamics of the income distribution across that same collection of economies. The more central question is, What determines the dynamics of cross-economy income distributions? Mechanisms of growth might be an important part of the answer; then again, they might not.

This paper provides an overview of recent research that takes this distribution-dynamics approach to analyzing convergence. It clarifies how this work differs from more traditional analysis. And, it points to where subsequent work is needed, and suggests alternative theoretical ideas to explore.

Convergence as distribution dynamics
(with or without growth)
by
Danny T. Quah
LSE Economics Department and CEP
November 1996

ABSTRACT

Convergence concerns the poor catching up with the rich—if not instantaneously, then at least having a tendency to do so. When poor and rich here refer to entire economies, then whether convergence occurs is traditionally viewed as just a side consequence of a more central question, namely that concerning the nature of economic growth. This paper argues instead that convergence itself is of direct interest. When convergence is made central and thus investigated, new theoretical issues and empirical insights emerge: this paper provides a brief overview of what those lessons are, and conjectures what next might be learnt.

Keywords: cross-country convergence, distribution dynamics, divergence, polarization, regional convergence, stratification

JEL Classification: C23, D30, E17, O57

Communications to: D. T. Quah, LSE, Houghton Street, London WC2A 2AE.

[Tel: +44-171-955-7535, Email: `dquah@exz.lse.ac.uk`]

[(URL) `http://cep.lse.ac.uk/homepage/dquah/`]

1. Introduction

Convergence is a catchy idea, but one that organizes serious thinking in areas as diverse as economic growth, theoretical econometrics, finance, European politics and monetary union, regional planning and geography, up through but not ending at entertainment and multi-media technology, and the software industry.

Some growth economists define convergence as a single economy approaching its theoretically-derived steady state growth path. Others translate this to whether poor economies are catching up with rich ones. Yet others think of these two statements as being identical, and thus of either indicating convergence.

At one extreme, econometricians and probabilists have found it useful to work with different notions of convergence of sequences of random variables. At a different extreme, economists and policy-makers in Europe have obsessed on the Maastricht convergence criteria. Finally, when high-tech, fast-growth market participants—people who actually create value in modern economies—get together, they too excitedly discuss convergence, but now between biological and machined products, or between communications, computers, and content (e.g., Kelly, 1994 and Tapscott, 1996). In every instance the term convergence is used with a different meaning—and rightly so.

That convergence suffers from this meaning-overload should not disguise its importance. This paper concerns convergence in the sense of poor economies catching up with the rich. If by economies, one means countries, then magnitudes of the numbers alone should already show why convergence is important. Some countries have been doubling per capita incomes every decade; yet others have been stagnant, with levels of per capita income a hundred times lower than those of the leading economies.

Such back-of-the-envelope facts are obvious and easy to obtain. But are there empirical regularities beyond them that could be useful for advancing economists' understanding of convergence? If the current situation continues, how will cross-country income distributions look in the future? Will rich economies always remain in a "club" of rich countries, and similarly the poor? What possibilities are there for switches across relatively poor and relatively rich?

Consider the same question, but now with the economies being regions within countries: are poorer regions languishing, currently and forever behind richer ones, or do they face any possibility of catching up? As before, one seeks an empirical characterization on such dynamic possibilities sufficiently precise and apposite so that one can understand the implications of alternative scenarios.

Such questions concern the behavior over time of cross-section distributions of income (or output or welfare): the issues are, writ large to the scale of macroeconomies or regions, the same as those that have traditionally been the concerns of research on the dynamics of inequality and personal income distributions. Are the distributions collapsing, so that everyone shows a tendency to become equally well off? Instead, are the distributions increasing in dispersion, so that those relatively better off are getting more so? Or, are the distributions tending towards shapes that show clusters and subgroups, so that the population is polarizing and then stratifying into distinct classes? Questions like these are useful for appreciating patterns of cross-country growth, just as they are for understanding patterns of dynamics and mobility in individuals' incomes within societies.

Such questions also show that economic growth—while unquestionably important for welfare—need not be the only mechanism permitting understanding of convergence. In the framework adopted below, convergence might occur because of growth. Or, convergence might occur without growth. Pinning down a theory of growth is not essential to understanding whether poor economies will become rich, or whether they will remain poor.[1] The more central question is, What determines the dynamics of cross-economy income distributions? Mechanisms of growth might be an important part of the answer; then again, they might not.

---

[1] I should clarify that I am not referring here to the classical Kuznets question: What is the relation between income distributions *within* an economy and the aggregate growth path of that economy? Rather, I am interested in growth paths of many different aggregate economies and the implications those have for the dynamics of the income distribution across that same collection of economies.

This paper provides an overview of recent research that takes this distribution-dynamics approach to analyzing convergence. It clarifies how this work differs from more traditional analysis. And, it points to where subsequent work is needed, and suggests alternative theoretical ideas to explore.

To clarify how such an approach to studying convergence differs from more traditional ones, Section 2 develops a series of theoretical models. It begins with models that underpin conventional understanding on growth and convergence, and progresses to models that focus on convergence as distribution dynamics, and where growth occurs almost mechanically. In this last set of models, despite the near-trivial specification of economic growth, intricate patterns of convergence and divergence nevertheless emerge.

Section 3 provides the empirical counterpart to the analyses in Section 2: once again, we study how configurations of convergence and divergence—or, alternatively, of stratification and polarization—appear in the cross section of economies. Finally, Section 4 briefly concludes.

2. Theoretical models

This section develops a series of models to make precise the notion of convergence across economies. I begin with a deterministic, classical growth model, and consider its convergence implications. These predictions are the ones traditionally taken to be useful for understanding the dynamics of rich and poor economies. In one interpretation, they are also the implications that have importantly influenced thinking on exogenous and endogenous growth. It is this framework that has traditionally tied together so closely discussions of growth on the one hand and convergence on the other.

Next, I provide a stochastic version of essentially the same growth model: this permits clarifying where the original interpretation of convergence is useful, and where it is not. Finally, I develop a model where convergence—catch up between rich and poor across a rich cross section of economies—becomes the central concern. In this last model, growth in each of the economies is fairly mechanical:

nevertheless, what emerges are intricate patterns of convergence and divergence, and of stratification and polarization.

2.1 Neoclassical deterministic growth and convergence

The development given here of deterministic, neoclassical growth theory's convergence predictions is well-known (e.g., Romer, 1994). Nonetheless, it will be useful as a starting point for the analyses that follow.

Assume that, in the representative economy, aggregate output is produced by a multiplicative-technology, constant returns to scale production function in capital and labor. Normalizing by labor, per worker output can be written as the product of technology A and a function f of per worker capital k (lower-case symbols denoting per worker quantities):

$$y = Af(k), \quad f \text{ differentiable and invertible.} \tag{1}$$

It is traditional to refer to y as productivity (in the sense of average per-worker output), and to A as the Solow or productivity residual. Taking growth rates on both sides of equation (1) gives

$$\dot{y}/y = \dot{A}/A + \left[\frac{f'(k)k}{f(k)}\right] \dot{k}/k. \tag{2}$$

Two points are immediate, both relating to the factor $f'(k)k/f(k)$ that multiplies $\dot{k}/k$ on the right-hand side. First, if the economy is one that compensates factor inputs according to marginal product, then this multiplicative term has an economic interpretation: it is physical capital's factor share in national income. Second, it follows that—regardless of how compensation actually occurs—that multiplicative term is bounded between 0 and 1. This therefore restricts how much capital deepening $\dot{k}/k$ can hope to contribute to productivity growth $\dot{y}/y$.

Such a bound applies regardless of how $\dot{k}/k$ is itself determined. That determination, however, is useful to develop explicitly. Call N the total quantity of labor. In our convention of lower- and upper-case symbols, the aggregate quantity of capital is $K = kN$ and aggregate output is $Y = yN$. Then,

$$\dot{k}/k = \dot{K}/K - \dot{N}/N.$$

Suppose that the (technology-adjusted) rate of aggregate capital accumulation is constant:

$$\sigma = \frac{\dot{K}}{Y/A} \in (0, 1).$$

Substituting this into the previous equation,

$$\begin{aligned} \dot{k}/k &= ((Y/A)/K)\sigma - \dot{N}/N \\ &= ((y/A)/k)\sigma - \dot{N}/N \\ &= \frac{(y/A)}{f^{-1}(y/A)}\sigma - \dot{N}/N, \end{aligned}$$

where equation (1) has been inverted to solve for k. Using this expression for $\dot{k}/k$ in equation (2) gives y's proportional growth rate as a function of, among other things, its level:

$$\dot{y}/y = \left(\dot{A}/A - \left[\frac{f'(k)k}{f(k)}\right]\dot{N}/N\right) + \left[\frac{f'(k)k}{f(k)}\right]\sigma \times \frac{(y/A)}{f^{-1}(y/A)}. \tag{3}$$

For analyzing convergence, what matters is the last levels term on the right-hand side. If the production technology in (1) has the power form

$$f(k) = k^{\gamma}, \quad 0 < \gamma \leq 1, \tag{4}$$

then that last term in levels becomes

$$\frac{(y/A)}{f^{-1}(y/A)} = (y/A)^{1-\gamma^{-1}},$$

where, by construction, $1 - \gamma^{-1} \leq 0$. This technology specification (4) also gives that the restricting factor $f'(k)k/f(k)$ is not just bounded (possibly varying) between 0 and 1 but actually constant at γ. Using these in equation (3), we have

$$\dot{y}/y = \left(\dot{A}/A - \gamma\dot{N}/N\right) + \gamma\sigma A^{-(1-\gamma^{-1})} \times y^{1-\gamma^{-1}} \tag{5}$$

$$= \xi_0 + \xi_1 \times y^{1-\gamma^{-1}} \tag{6}$$

(with ξ_0 and ξ_1 defined in the obvious way).

From a convergence perspective, equation (6) gives the theory's predicted relation between income growth $\dot{y}/y$, income levels y, and physical capital's productivity coefficient γ. Since $1 - \gamma^{-1} \leq 0$ and increases in γ, tending to 0 as γ goes to 1, one might read equation (6) to say that rich economies grow slower, while poor ones grow faster. Convergence is faster, the smaller is γ. Abstracting away from differences and scale effects in ξ_0 and ξ_1, the convergence situation is as in Fig. 1, where the different time paths illustrate different economies.

To see the algebra through completely, one might proceed as follows. Rearrange (5) to get

$$\dot{y}/y - \dot{A}/A = \gamma\left[\sigma\,(y/A)^{1-\gamma^{-1}} - \dot{N}/N\right]. \tag{7}$$

A steady state in y/A thus exists when $\dot{N}/N$ is positive and (eventually) constant. That steady-state level for y/A is given by the unique zero of (7) at

$$(y/A)_* = \left(\sigma^{-1}\dot{N}/N\right)^{(1-\gamma^{-1})^{-1}}.$$

Moreover, since $1 - \gamma^{-1}$ is negative, the ratio y/A is globally stable around its steady state state. (At that steady state, y of course grows at the same rate as A.)

How rapidly does convergence occur? Define $z \stackrel{\text{def}}{=} y/A$, and log-linearize (7) around steady state $z_* = (y/A)_*$ to get

$$\dot{z}/z = -(1-\gamma)\left(\dot{N}/N\right) \times (\log z - \log z_*). \tag{8}$$

This is simply a first-order differential equation in $\log z - \log z_*$. The larger is γ, the slower is the rate of convergence of $\log z$ to $\log z_*$.

Such a depiction of convergence gives rise to a research program with a number of features (see, e.g., Barro and Sala-i-Martin, 1992). First, the researcher might consider ξ_0 and ξ_1 in (6) or $\dot{N}/N$ and z_* in (8) to vary, plausibly, with differing structural characteristics of different economies. Examples of such characteristics might include democracy, political stability, tax regime, religious ethos, colonial heritage, and so on. The researcher then seeks "conditional convergence", where Fig. 1 obtains only after conditioning on those other characteristics. Such conditioning, however, leaves unchanged the basic message: those auxiliary variables, after all, entered the discussion only *after* the derivation of convergence in equations (1)–(6) and Fig. 1. Conditioning can affect neither the intuition nor the interpretation surrounding the convergence of Fig. 1.

Second, to investigate empirically equations (6) or (8) or Fig. 1, the researcher seeks a convergence regression: there, one is interested if growth (on the left-hand side) depends negatively on levels (on the right). That regression would formalize the intuition where as one looks across the different economies in Fig. 1, one expects to see richer ones growing relatively slower, and poorer economies, relatively faster. On the left-hand side of such a regression, one might proxy growth rates by an average of log first-differences. On the right-hand side, one might have a range of auxiliary conditioning variables, and then income levels. When negative, the coefficient on income levels denotes convergence (in the sense of Fig. 1); its magnitude measures the speed of convergence and varies negatively with physical capital's productivity coefficient. In this model, a zero coefficient on income levels would imply physical capital's productivity coefficient γ equal to 1.

If one were to interpret equations (6) and (8) as applying only over the long-run, then the averaged log first-differences appearing on the left-hand side of the regression should be averages taken over an appropriately long time horizon. Consequently, the regression would be estimated effectively only over the cross-section (and not time-series) dimension. As usual in such work, the researcher investigates the convergence regression equations (6) and (8) tacking on a stochastic residual at the end.

When one carries out this research program—either using explicit regression analysis (e.g., Barro and Sala-i-Martin, 1992) or informal comparison (e.g., Romer, 1994), one concludes that in the data the growth/levels relation implies γ should be close to 1. More significantly, however, one concludes that γ is much larger than would be implied by capital's factor share in national income accounts. This, by itself, has been an important spur to the development of growth models that feature externalities, increasing returns, endogenous technical progress, or other possibilities breaking the link between γ and capital's measured factor share. Researchers like Barro and Sala-i-Martin (1992), moreover, claim to find conditional convergence, and thus conclude that Fig. 1 is operative, and that the poor do catch up with the rich, if slowly.[2]

Throughout this discussion, we have only considered a single, representative economy. Thus, in this research program, one implicitly identifies convergence of that representative economy's z to its z_* also as cross-sectional convergence of different economies towards each other (as in Fig. 1).

2.2 Stochastic growth

Above, I described the empirical implementation of equations (6) and (8) in terms of tacking on a stochastic residual at the end. Since the equations cannot be expected to fit perfectly, something like this is obviously necessary. Two issues, however, then arise. First, there may be interesting economics omitted when

[2] An earlier literature, e.g., Grier and Tullock (1989), studied a similar regression equation with growth on the left-hand side and explanatory variables on the right. I distinguish this from the later work only because that research did not show the same preoccupation with convergence. It instead only investigated, using exploratory empirical techniques, the determinants of growth—an important question, certainly, but distinct from convergence.

one simply adds on a stochastic residual as an afterthought. And, second, the interpretation of Fig. 1 in terms of the poor catching up with the rich might no longer be appropriate. We consider these in turn.

Stochastic disturbances induce uncertainty: a rich cross section of economies might then not behave just as a collection of individualistic, autarkic elements, as implicitly assumed in the previous development. In this case, one fruitful approach might be to explore insurance arrangements across the distribution of aggregate economies (using ideas from, e.g., Bertola, 1995; Lucas, 1992; and Thomas and Worrall, 1990). Groupings of selected economies might endogenously emerge—depending on patterns of insurable and uninsurable disturbances across economies—with growth patterns then varying across different groupings. Because identification becomes important—which economies go in which groupings—such an outcome would depart from the "representative economy" analysis above. To keep within space restrictions, however, I will not analyze this further here: the model developed in the next subsection will carry similar predictions (although it will not be stochastic, and thus groupings will occur for other than insurance reasons).

A different, more direct possibility is that stochastic disturbances might, properly considered, break the link between convergence, capital's productivity coefficient γ, and factor income shares. This idea has been explored in den Haan (1995), Kelly (1992), and Leung and Quah (1996). I follow the last of these in developing this possibility, but keep as close as I can to the deterministic model previously studied. Some properties of the stochastic model will translate directly from the deterministic case; however, to be clear where the new subtleties are, it helps to be more explicit about economic behavior than I have been thus far.

In the cross section distribution, index economies by j; it will be convenient now to suppose that time is discrete. Variables specific to economy j at time t will have subscript j and be parenthesized t.

Assume that each economy behaves autarkically and that, in equilibrium, can be described by the following social planning problem: at time t_0, solve

$$\sup_{\{(c_j(t),k_j(t+1)):t\geq t_0\}} E_{t_0} \sum_{t=t_0}^{\infty} (1+\rho)^{-t} \log c_j(t), \quad \rho > 0 \tag{9}$$

$$\text{s.t. } y_j(t) = A(t) k_j(t)^{\gamma(t)} \tag{10}$$

$$c_j(t) \leq y_j(t) - k_j(t+1), \; t \geq t_0,$$

$$k_j(t_0) > 0 \text{ given.}$$

In (9), ρ describes the discount factor. The last two relations state, respectively, the capital accumulation constraint and that the time t_0-extant capital stock is fixed.

The stochastic elements introduced are encapsulated in the technology and productivity terms A and γ: these can now vary randomly through time. Assume that $\{\,(A(t), \gamma(t)) : \text{ integer } t\,\}$ is a jointly stationary vector process with all entries (almost surely) positive.[3] When this process is degenerate, equation (10) collapses to the production technology of the previous section. By contrast, with A and γ nondegenerate stochastic processes, we will see that they give rise to the stochastic disturbances that, in the previous section, had been simply tacked on at the end. Keeping to our earlier interpretation, A and γ are taken to be common across economies: thus, while they might vary over time, they are constant over the cross section.

[3] To simplify the technical exposition, I also assume that $\log A$ has bounded second moment. In neoclassical exogenous growth models, it is important that A be permitted to grow without bound, whereupon these stationary and bounded-moment assumptions would be violated. However, as should have already become clear from our earlier discussion, growth in A is not central to the current convergence discussion. At the cost of more extended exposition, everything here can be done taking deviations relative to an unboundedly growing A process. However, no additional insight obtains.

Assume that at time t every economy observes the same history,

$$\mathcal{F}(t) = \{\, A(s), \gamma(s), y_j(s), k_j(s) : \; s < t, \; \text{all } j \,\} .$$

Expectations conditioned on this history will be denoted $E_t = E(\,\cdot\,|\mathcal{F}(t))$, a notation already used in (9). By this timing assumption, in general, $E_t A(t) \neq A(t)$ and $E_t\gamma(t) \neq \gamma(t)$. Not allowing economies, at time t, to know time-t productivity disturbances is not crucial for the discussion but simplifies notation.

In equilibrium, in each time period t, economies behave as if a social planner chooses consumption and investment functions based on what is observable at time t. Under regularity conditions, those optimal decision rules lead to

$$k_j(t+1) = \left((E_t\gamma(t))(1+\rho)^{-1}\right) y_j(t), \tag{11}$$

so that from (10) output behaves as

$$y_j(t+1) = A(t+1)\left[\left((E_t\gamma(t))(1+\rho)^{-1}\right) y_j(t)\right]^{\gamma(t+1)} .$$

Taking logs and defining $\widetilde{y} \stackrel{\text{def}}{=} \log y$ give the first-order stochastic difference equation:

$$\begin{aligned} \forall t \geq t_0 : \quad \widetilde{y}_j(t+1) &= \log A(t+1) + \gamma(t+1)\log(E_t\gamma(t)) \\ &\qquad - \gamma(t+1)\log(1+\rho) + \gamma(t+1)\widetilde{y}_j(t) \\ &= \eta(t+1) + \gamma(t+1)\widetilde{y}_j(t) \end{aligned} \tag{12}$$

with initial condition

$$\widetilde{y}_j(t_0) = \log A(t_0) + \gamma(t_0)\widetilde{k}_j(t_0), \tag{13}$$

and where η in (12) is defined as:

$$\eta(t) = \log A(t) + \gamma(t)\log(E_{t-1}\gamma(t-1)) - \gamma(t)\log(1+\rho). \tag{14}$$

Immediate from the assumptions on $\{\, (A(t), \gamma(t)) :\ \text{integer } t\,\}$ we have that the stochastic process $\{\, \eta(t) :\ \text{integer } t\,\}$ is stationary and common across economies.[4]

All the model's dynamics—and hence all its convergence implications—are embedded in equations (12)–(14). To make those implications explicit, first take the case where $\gamma(t) = \gamma \in (0, 1]$: this has capital's productivity coefficient constant as in the previous section, but the Solow productivity residual A is stochastic. Then, (14) becomes

$$\eta(t) = \log A(t) + \gamma \log((1+\rho)^{-1}\gamma);$$

up to a shift in mean, η simply inherits all the stochastic properties of $\log A$. From (12) and (13), the distribution of output across economies evolves as

$$\widetilde{y}_j(t+1) = \gamma\widetilde{y}_j(t) + \eta(t+1) \quad \text{for } t \geq t_0, \tag{15}$$

$$\widetilde{y}_j(t_0) = \log A(t_0) + \gamma\widetilde{k}_j(t_0). \tag{16}$$

Iterating (15) forwards from the initial condition (16), we get

$$\widetilde{y}_j(t) = \gamma^{t-t_0}\widetilde{y}_j(t_0) + \sum_{s=0}^{t-1-t_0} \gamma^s \eta(t-s). \tag{17}$$

Since η is stationary, if γ is 1, then $\widetilde{y}_j$ for each j is an integrated (order 1) process.[5] If, further, A is iid through time, then equation (15) says that $\widetilde{y}_j$ (for each j) is a random walk with drift

$$E\eta = E \log A + \gamma \log((1+\rho)^{-1}\gamma)$$

[4] Leung and Quah (1996) observe that since η, the "residual", is common across economies, a cross-section convergence regression cannot actually be estimated consistently in this model. Relaxing the commonality of η, on the other hand, means that convergence as in Fig. 1 necessarily cannot occur.

[5] When $\gamma = 1$, then from (15), $\widetilde{y}_j$ is always one order of integration higher than η. Thus, should the Solow residual A itself tend to an integrated order 1 sequence, $\widetilde{y}_j$ will then be integrated order 2.

which could be positive, negative, or zero. Should the discount rate ρ be large (relative to the technology terms γ and $E \log A$), then $\widetilde{y}_j = \log y_j$ diverges to $-\infty$ (when $\gamma = 1$), implying then that y_j converges to 0, independent of initial conditions.

This case is, however, relatively uninteresting: it is when economic agents discount the future so heavily that accumulation doesn't occur and the economy thus collapses on itself. More relevant is when the drift $E\eta$ is non-negative. Then there are two possibilities: when γ is strictly less than 1, and when γ is exactly 1.

Take first the case with $\gamma < 1$. From (17), as t grows large, $\widetilde{y}_j$ (for each j) converges to a random variable having a unique stationary or invariant distribution given by the distribution of

$$\zeta(t) \stackrel{\text{def}}{=} \sum_{s=0}^{\infty} \gamma^s \eta(t-s)$$

for any integer t. (If, for example, η is distributed iid normal with mean $E\eta$ and variance $\text{Var}(\eta)$, then that invariant distribution is normal with mean $E\eta/(1-\gamma)$ and variance $\text{Var}(\eta)/(1-\gamma^2)$ identically across $\zeta(t)$'s).

That unique invariant distribution, by definition, must be attained independent of initial conditions. It is, moreover, typically nondegenerate. What happens, however, to the cross-section distribution of $\widetilde{y}_j$'s across j? Since the limiting $\zeta(t)$'s are the *same* for all j, the cross section distribution turns out to collapse to a single point: it behaves not at all like the time-series distribution of $\widetilde{y}_j$ for a fixed (but arbitrary) j.

This difference between cross section and time series dimensions arises for a quite trivial and easily identified reason: the model has taken A and γ, and thus η, to be common across economies.

The same insight applies to when γ is 1: once again, simply study equation (17). Although for each j, the sequence $\widetilde{y}_j(t)$ does not converge in t to any invariant distribution, the cross-section distribution across j's is completely stable, and does not diverge. That cross-section distribution just moves up and down, perturbed by a single common factor: no intra-distribution relative movements occur.

Again, this happens because A and γ have been assumed to be common across economies. It thus seems fairly trivial. The conceptual difficulty arises, however, when one asks, What assumption should replace this commonality property? *Any* assumption—especially an iid one—is going to be equally arbitrary, and more to the point, will give predictions for the cross section distribution as mechanically and simply as those just obtained. In this kind of "representative economy" analysis, economic theory plays almost no role in helping us understand the behavior of the cross-section distribution. Depending only on arbitrary assumptions on unobservable disturbances, the cross section distribution across economies on the one hand and the time-series distribution of any single economy in the cross section on the other can behave completely differently.

The point is emphasized if we return to the general case in equations (12)–(14). Then, iterating (12), we obtain the counterpart of equation (15):

$$\widetilde{y}_j(t) = \left[\prod_{s=1}^{t-t_0} \gamma(t_0+s)\right] \widetilde{y}_j(t_0) + \sum_{s=0}^{t-1-t_0} \left[\prod_{r=0}^{s-1} \gamma(t-r)\right] \eta(t-s). \tag{18}$$

Kelly (1992) and Leung and Quah (1996) prove that equation (18) converges to a degenerate cross-section distribution across economies, even when $\gamma(t)$ has expectation 1 and exceeds 1 with positive probability. Put another way, equation (12) could show a coefficient on lagged $\widetilde{y}_j$ that is large—the technology (10) could show, on average, physical capital having productivity coefficient 1—yet, convergence of the cross-section distribution to a degenerate point could still occur.

The stochastic model thus far has been used to argue that a convergence regression coefficient indicating no convergence—in the sense of Section 2.1—can be consistent with the cross-section distribution collapsing to a point. The opposite can also occur: a convergence regression coefficient indicating convergence is also consistent with the cross-section distribution expanding. An easy way to understand this is to use equation (17) with $\gamma < 1$, and allow η to be iid across economies. Then, by Glivenko-Cantelli Theorem, the cross-section and time-series distributions coincide. Since the (time-series) invariant distribution is unique, the cross-section distribution tends to it from wherever that cross-section distribution might happen to be. In particular, this must happen even when the cross-section distribution is already more tightly concentrated than the invariant distribution (see Fig. 2). Under such circumstances, the cross-section distribution will be seen to diverge, even when $\gamma < 1$, and convergence regressions indicate convergence.

What I have just described is an instance of the general message from Galton's fallacy reasoning (see, e.g., Friedman, 1992 and Quah, 1993). Here, one can usefully regard its lesson as the following: the dynamics of a representative or average economy in the cross section say little about the behavior of the entire cross section distribution.

To conclude, the message from this subsection can be stated in two different ways. First, in the kind of "representative economy" model studied thus far, it is the assumptions on exogenous stochastic disturbances that are dominant for the behavior of the cross section distribution: economic theory gives no useful guide to those dynamics. Second, even a complete characterization of the dynamic behavior of a representative economy tells little about the dynamics of the entire cross-section distribution.

These statements suggest how misleading is the apparently obvious message of Fig. 1—of poor catching up with rich provided only that technology coefficients take particular values or that certain regression coefficients take particular signs. To make progress on analyzing convergence, one needs an economic model that theorizes explicitly in terms of the cross section distribution.

2.3 Cross-section dynamics and convergence

Quah (1996e) develops a model of convergence for a rich cross section distribution of economies.[6] In the model, a balance between a force for consolidation and a force for fragmentation results in coalitions forming across different parts of the distribution of economies. Those coalitions then turn out to behave like convergence clubs (e.g., Baumol, 1986). The model explains the dynamics of the entire cross section distribution, and directly gives predictions on convergence.

The model is most naturally viewed as one where growth and convergence arise from human capital or the generation of ideas.[7] We therefore switch from the earlier models where accumulating physical capital k is important to one where it is accumulating human capital h that matters. Let $\mathcal{J}$ be the index set of economies, taken as fixed throughout the discussion. A *coalition* of economies is a subset $\mathcal{C}$ of $\mathcal{J}$. Each economy l in $\mathcal{J}$ is characterized by an economy-specific stock of human capital h_l. That stock is used in two nonrival ways: first, it represents the potential for generating ideas, and second, it produces non-storable output for consumption. Ideas that are further developed then increment the stock of human capital, thereby driving economic growth.

Production occurs from coalitions of economies forming to jointly produce a single nondurable consumption good. Denote the total output of coalition $\mathcal{C}$ by $Y_{\mathcal{C}}$. Assume that $Y_{\mathcal{C}}$ depends on the distribution of h_l across l in $\mathcal{C}$, and is increasing in each h_l. Assume also that out of the total coalition output, economy l in $\mathcal{C}$ gets $\psi(Y_{\mathcal{C}}, h_l)$, with ψ increasing in both arguments, and satisfying exact product exhaustion:

$$\sum_{l \text{ in } \mathcal{C}} \psi(Y_{\mathcal{C}}, h_l) = Y_{\mathcal{C}}.$$

[6] More general models with cross-sectional interaction (e.g., Benabou, 1996 and Durlauf, 1993) could also provide other useful insights here.

[7] This is a little bit of a misstatement as we will see that in equilibrium growth turns out to be fairly mechanical, while it is patterns of convergence and divergence that are interesting and intricate.

(Primitive assumptions sufficient for these properties would be first, compensation according to marginal product and second, the CES technology

$$Y_{\mathcal{C}} = \left[\sum_{l \text{ in } \mathcal{C}} h_l^{\theta} \right]^{1/\theta}, \quad 0 < \theta < 1$$

[with θ, describing the elasticity of substitution in the CES production function, giving isoquants between linear and Cobb-Douglas technologies]. Quah (1996e) gives the natural interpretation of these properties as economies of scale deriving from specialization.)

By the assumptions above, enlarging the coalition always increases total output Y. Then the compensation scheme ψ ensures that all economies unanimously agree to be in the single grand coalition comprising the entire cross section. This, therefore, is a force for consolidation.

Consider next how the distribution of h's evolves. In every instant of time, economy l generates ideas of average quality h_l. Assume that it first gets to use those ideas and then shares them with others in its coalition. Ideas propagate freely within coalitions, but do not transmit across them.[8] Call $H_{\mathcal{C}}$ the *average* quality of ideas generated in coalition $\mathcal{C}$, and suppose that human capital in economy l evolves as:

$$\dot{h}_l = \widetilde{\phi}(h_l, H_{\mathcal{C}}) \quad \text{for } l \text{ in } \mathcal{C},$$

with $\widetilde{\phi}$ increasing in both arguments and homogeneous degree 1. Dividing throughout by h_l we get:

$$\dot{h}_l / h_l = \widetilde{\phi}(1, H_{\mathcal{C}}/h_l) \stackrel{\text{def}}{=} \phi(H_{\mathcal{C}}/h_l).$$

[8] This might be because ideas or memes are like viruses and thus could be dangerous—members of different coalitions are not trusted. Or, members of a coalition are able to enforce intellectual property rights perfectly across coalitions. Quah (1996e) gives a more extended discussion of this. The general idea of memes, or ideas as genes, is discussed in Dawkins (1976) and Kelly (1994).

By construction ϕ is increasing in the ratio $H_{\mathcal{C}}/h_l$.

It is now easy to understand the force for fragmentation. Economies in higher average h coalitions have faster proportional growth rates. The problem with allowing a coalition to get too large—expanding below—is that the coalition thereby lowers its average $H_{\mathcal{C}}$: this would slow growth for all economies already in the coalition. Economies already in good coalitions would, ceteris paribus, refuse to admit economies that lower the coalition average $H_{\mathcal{C}}$.

The force for consolidation (the compensation $\psi(Y_{\mathcal{C}}, h_l)$) is a *level* effect—it affects current consumption. The force for fragmentation (the growth $\phi(H_{\mathcal{C}}/h_l)$) is a *slope* effect—it affects future consumption. Parameterizing economies' discount rates for intertemporal consumption allows calibrating the tradeoff across level and slope effects, and thus provides a theory of coalition formation. Quah (1996e) describes an equilibrium where nontrivial consecutive subsets of the cross section distribution of economies form coalitions. The distribution of income across economies within the same coalition converges towards equality; those across different coalitions separate and then diverge (Fig. 3).

In equilibrium rich economies converge towards each other, but remain rich; similarly the poor remain poor. The middle class eventually vanishes, and the income distribution stratifies. Those in the middle part of the income distribution might begin close to each other, but over time diverge apart: small differences here become magnified eventually into large disparities. By contrast, within extreme parts of the distributions, economies, over time, have their differences diminish. Because only two convergence clubs form in Fig. 3, it is natural to consider the dynamics here as showing an *emerging twin peaks* character. In general, multiple clubs may form, the number of which will then the number of emerging modes in the long-run cross-section distribution of incomes.

In this model, conditional convergence occurs. This is a conditional convergence, however, that depends critically on the coalition structure in equilibrium. Different rules for how coalitions form would lead, in general, to different distribution dynamics. Appreciating those rules provides understanding on the stratification and polarization that emerges across the cross section.

More generally, the model in this subsection draws attention to two features in Fig. 3 relevant for convergence. First is the abeyant emergence of peaks in the cross-economy income distribution. Second is the intra-distribution dynamics, where different economies in the cross section, over time, transit to different parts of the distribution.

3. Empirics

Conventional wisdom on cross-section regression analyses of convergence is that cross-economy convergence occurs, once appropriate conditioning is applied (Barro and Sala-i-Martin, 1992; Sala-i-Martin, 1996). In this conventional wisdom, such convergence is stable and uniform at 2% per year.

Quah (1996c) has argued that such empirical stability might be simply a statistical artifact. He provides Monte Carlo evidence that heterogeneous unit root data—that, by construction, contradict stable uniform convergence in the sense described above—could nevertheless generate a stable 2% convergence rate, given the sample sizes of observations typically used. In those Monte Carlo experiments, the cross-section income distributions diverge over time; yet, the 2% convergence regression finding is reproduced, on average.

But the theoretical analysis in Section 2 provides yet other reasons for doubting the usefulness of such results for analyzing convergence. Whether or not the 2% estimate is statistically reliable, it can provide *no* guide for convergence behavior across the cross section. The interesting features in Fig. 3 can certainly never be captured by cross-section (conditional or unconditional) convergence regression analysis.[9] However, it is exactly such dynamics that will shed light on convergence patterns across economies.

The most apposite empirical approach parallels the theoretical analysis of subsection 2.3 in modelling directly the dynamics of the entire cross section distri-

[9] Regression techniques that could potentially do so are the adaptive procedures in Ben-David (1994) and Durlauf and Johnson (1995).

bution. Note that such a motivation differs from the technical ones traditionally given of either exploiting simultaneously cross section and time-series dimensions of the data for great precision in estimation and inference or permitting "flexibility" in estimating time-varying and nonlinear regression or distribution functions.[10] Instead, the goal is to provide a picture of how the entire cross-section distribution evolves over time, and to understand the long-run or limiting behavior of that cross-economy income distribution.

Fig. 4 gives a first preliminary look at cross-economy distribution dynamics. The different panels in Fig. 4 show point-in-time snapshots of the density of normalized productivity across countries.[11] Thus, Fig. 4 cannot show intra-distribution or churning dynamics in the evolving distributions. It can, however, and evidently does show an emerging twin-peakedness.

The econometric task in studying convergence, therefore, is to formulate a model that (i) is capable of capturing the different possibilities in Fig. 3 (including, significantly, the intra-distribution dynamics as well as the shapes of the point-in-time distributions); (ii) accepts data in the form of distributions as in Fig. 4; and (iii) allows analysis of long-run or out-of-sample behavior in the distributions. Call a structure allowing (i)–(iii) a *model of explicit distribution dynamics* (or *medd*). Desdoigts (1994), Lamo (1996), Paap and van Dijk (1994), and Quah (1993a, b,

---

[10] Thus, the aim here differs from those emphasized in Canova and Marcet (1995), Forni and Reichlin (1995), Lee, Pesaran, and Smith (1995), Quah (1994), and Quah and Sargent (1993).

[11] The underlying data here are from Summers and Heston (1991), version 5.6. Productivity is per worker output; the normalization is with respect to average world productivity. For simplicity, below, I refer to *per capita income* or just *income* interchangeably with average worker productivity. All densities were obtained using a gaussian kernel with bandwidth selected automatically, as suggested in Silverman (1986) 3.4.2. I used a fast fourier transform to calculate the resulting kernel estimator; a reflection method (Silverman 1986, 2.10) took into account nonnegativity in the productivity data.

1996b) have studied convergence in terms of *medd* structures.[12] The underlying framework common to all this research is the following.

Let F_t denote the time t cross-economy income distribution. Associated with each F_t is a probability measure λ_t, where

$$\forall\ y \in \mathbb{R}: \quad \lambda_t((-\infty, y]) = F_t(y).$$

A stochastic difference equation describing distribution dynamics is then

$$\lambda_t = T^*(\lambda_{t-1}, u_t), \quad \text{integer } t, \tag{19}$$

where $\{\, u_t : \text{ integer } t\,\}$ is a sequence of disturbances, and T^* is an operator mapping the Cartesian product of probability measures with disturbances to probability measures. (Needless to say, the first-order specification in (19) is just a convenience for the discussion. Nothing substantive hinges on it, and the model easily generalizes to higher-order dynamics.)

For *medd* analysis, one is also interested in intra-distribution dynamics. Equation (19) therefore has to record more than just means and standard deviations or, more generally, a finite set of moments of the distribution sequence $\{\, F_0, F_1, \ldots\}$. Equation (19) takes values that are measures, rather than just scalars or finite-dimensioned vectors, and thus differs from the typical time-series model.

The structure of T^* reveals if dynamics like those in Fig. 3 occur. Estimated from observed data, T^* allows empirical quantification of those dynamics. Economic hypotheses restrict T^* in particular ways: they therefore provide predictions on how λ_t, and thus the distributions F_t, evolve over time.

---

[12] When one is interested only in a small subset of the entire cross-section distribution, then vector time-series methods (e.g., Bernard and Durlauf, 1995) would be informative for convergence properties. Univariate time-series models, however, never are: I discuss this more below. Also, *medd* structures need not be restricted only to convergence issues, but might also be used for modelling business cycle fluctuations across many different sectors or regions (e.g., Quah, 1994, 1996a).

Just as in time-series analysis, the researcher might seek to understand T^* by its "impulse response function": set the disturbances u to $\mathbf{0}$ (whatever $\mathbf{0}$ means here), and run out the difference equation:

$$\begin{aligned} T^*(\lambda_{t+s-1}, \mathbf{0}) &= T^*(T^*(\lambda_{t+s-2}, \mathbf{0}), \mathbf{0}) \\ &\vdots \\ &= T^*(T^*(T^* \ldots (T^*(\lambda_t, \mathbf{0}), \mathbf{0},) \ldots \mathbf{0}), \mathbf{0}), \end{aligned} \tag{20}$$

with the result being a proxy for λ_{t+s}. Then, convergence in country incomes to equality might be represented by (20) tending, as $s \to \infty$, towards a degenerate point mass. Alternatively, the world polarizing into rich and poor might be represented by (20) tending towards a two-point measure: the implied limit distribution $F_{t+s},\ s \to \infty$, would then be bimodal or twin-peaked. More generally, stratification into different convergence clubs might manifest in (20) tending towards a multi-point, discrete measure, or equivalently, a multi-modal distribution. How quickly a given initial distribution, F_0, evolves into the limiting distribution, $F_{t+s},\ s \to \infty$, can be read off T^*'s (spectral) structure.

Finally, T^* also contains information on intra-distribution dynamics. Exploiting that structure, one can quantify the likelihood of the poor catching up with the rich, and characterize the (random) occurrence times for such events.

While this framework borrows ideas from standard time-series analysis, certain conceptual differences are critical. To appreciate those differences, first consider when the researcher discretizes the underlying income state space so that distribution λ is given by just a probability vector. The researcher might be tempted to write (19) as

$$\lambda_t = M\lambda_{t-1} + u_t \tag{21}$$

(with M a square matrix representing the operator T^*), and then call (21) simply a vector autoregression (VAR) taking values on the unit simplex. This, however, would be incorrect. If (19) could be written as the VAR in (21), matrix M would be identified from (i.e., uniquely determined by) the sequence $\{\, \lambda_t : \text{ integer } t \,\}$.

However, for describing distribution dynamics, multiple M's in (21) can be associated with any given distribution sequence: as just one example, the distribution sequence $\lambda_t = (1/2, 1/2)$ for all t is fit perfectly in (21) by any of

$$M_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}, \qquad \text{or} \qquad M_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus, the difference equation (19) must contain strictly more information than the VAR formulation (21).

Next, a time-series researcher might ask, Since (19) describes the joint dynamics of a collection of individual time-series processes, why not just characterize the behavior of the underlying variables directly? One might do this by estimating univariate or small multivariate time-series representations for subsets of those underlying income series. Suppose the researcher estimated univariate representations for each of the underlying variables. What can the researcher learn about "emerging twin peaks" tendencies from those? The answer is nothing. Knowing univariate representations for each of the underlying income series can give no information on *comovements* across the cross section: establishing, say, that the first income series is a first-order autoregression with coefficient 0.8, while the second is a mixed moving average autoregression, and so on, can provide no understanding of how the different incomes covary. Thus, there might be a tendency towards twin peaks in the cross-section distribution; or there might not: one cannot tell from the collection of all univariate representations. Moreover, this negative conclusion extends from when the researcher is studying only univariate representations to when studying multivariate time-series models for subsets of the cross section: there too one learns nothing about "emerging twin peaks" possibilities for the entire cross-section distribution.[13]

[13] Note that *after* one has learnt about the dynamics of the entire distribution, it might be of interest to go back and study the underlying individual time-series representations. But, to be clear, such analysis only complements the study of distribution dynamics; it cannot be a substitute.

The time-series researcher might then consider building a model for the joint dynamics of a rich cross-section of individual time series (e.g., Canova and Marcet, 1995; Forni and Reichlin, 1995; or Quah and Sargent, 1993). It is tautological that a rich enough model must be able to capture all the features of interest in the original data set. The question then becomes what is the most direct and transparent way to capture characteristics of interest. In my view, when one is concerned with an "emerging twin peaks" property, then a model of explicit distribution dynamics of the form (19) is best: the distribution λ_t becomes the fundamental object of study.

Following the notation above, the estimated T^*'s in Desdoigts (1994), Lamo (1996), Paap and van Dijk (1994), and Quah (1993a, b, 1996b) all show an emerging twin-peaks character in the cross-country income distributions.[14] Because estimated T^*'s qualify as *medd*'s, they can shed light on that seductive intuition—the poor growing faster and thereby catching up with the rich—that growth-on-levels regressions wish to exploit. Quah (1996b) calculates, from an estimated T^*, the probability density of passage times from poor parts of the income distribution to rich parts. He finds that although growth miracles—the Hong Kongs, the South Koreas, and the Singapores—can happen with reasonable positive probability, the passage time from the bottom 5% percentile to the top, given the magnitude of the gap extant, averages in the hundreds of years. Thus, persistence and immobility characterize the world cross section of country incomes.[15]

[14] Bianchi (1995) sidesteps the dynamics in equation (19), and considers λ_t's in isolation. He too, however, finds the twin-peaks property of Fig. 3, as one might expect from the evidence in Fig. 4. Bianchi's work differs from those mentioned in the text as he seeks modal properties only in-sample; the others allow those properties to manifest out of sample by the extrapolation in equation (20). Bianchi's analysis is a statistical formalization of exactly Fig. 4.

[15] It is worth noting that similar twin-peaks features do not describe every such macro income distribution sequence. Compare what I have just said of the world cross section of countries with, for example, US and European regional behavior,

The evidence therefore suggests, in the large, keeping Fig. 3, but modifying it so that the arrows showing intra-distribution dynamics are allowed to cross. Why this crossing comes about is not yet well-captured in explicit distribution-dynamics models like those in Quah (1996b, e). However, these calculations can be viewed as the econometric formalization of the "miracles" (e.g., Lucas, 1993) that theory should continue to seek to explain.

The natural next step then is to ask, What explains the observed emerging twin-peaks patterns in cross-country income distributions? Of course, in a sense, models like those in Quah (1996b, e) already explain those patterns; what such a question must intend instead is, What *measured* variables empirically account for patterns like emerging twin peaks? When the fundamental object of study is an entire distribution, empirically accounting for the patterns of interest in it has a subtle interpretation. For one, it cannot mean just getting a high R^2 or significant t-statistics in a regression: a regression, at best, helps us understand conditional means; it does not account for the entire distribution.

The approach I take here—described in detail in Quah (1996d)—is analogous to constructing a conditional distribution from the unconditional distribution, in classical probability theory. In this framework, explaining features like emerging twin peaks means obtaining conditional distributions so that such features *no longer appear*.

Figures 5.d and 6.d show cross-country income distributions conditioning *on trade* and *on space*; Figures 5.s and 6.s show stochastic kernels transforming the original income distributions in Fig. 4 into the corresponding distributions in Figures 5.d and 6.d.

What I mean by conditioning is the following: instead of taking incomes relative to the world average—and then finding their cross-section distribution—take instead incomes relative to those of the economy's principal trading partners (conditioning on trade), or relative to those of the economy's physical neighbors

---

e.g., Quah (1996c, f). Quah (1996f) has also analyzed the behavior of cross-economy income distributions, conditioning on spatial effects.

(conditioning on space). By trade, I mean the sum of exports and imports; by principal trading partners, I mean those economies such that their total trade with a given economy exceeds 50% of that economy's total trade. Obvious variations on this are easy: one might look only at exports or at imports rather than their sum; one might look at the 3 (or n) largest trading partners, rather than a variable number of them; one might use 80% as the threshold rather than 50% for defining the group of trading partners; and so on.[16] Within a reasonable range of variation, however, perturbations along these lines do not dramatically alter the conclusions below.

To appreciate better what is involved here, take, for instance, Algeria: its trade-conditioned income is defined to be its income relative to the average of incomes in France, USA, Italy, and Morocco, with that average weighted by these economies' trade shares in Algeria. By contrast, Algeria's spatial-conditioned income is its income relative to the average of incomes in Mali, Mauritania, Morocco, Niger, and Tunisia, weighted by the number of workers in those economies. For Singapore, its trade-conditioned income is relative to Guinnea Bissau, Oman, the US, Japan, Malaysia, Maritius, and Qatar, while its spatial-conditioned income is relative to Malaysia and Indonesia.

As for spatial conditioning, if economies across the world formed a seamless web in physical geography, then one might expect the resulting (conditioned) distribution to be tightly concentrated about 1. After taking into account spatial factors, all economies would be just about average: all the poor economies in sub-Saharan Africa might be poor relative to the rest of the world, but not relative to each other. Similarly, the relatively rich economies—rich relative to the world—again would be just about average, relative to their also-rich neighbors.

The stochastic kernel shows how the unconditional distribution is transformed into a conditional one. If the mass of the kernel piled up on the 45-degree diagonal, then the transformation is one that leaves unchanged the original distribution's fea-

[16] These variations all constitute examples of what Quah (1996d) calls *conditioning schemes*.

tures. If, on the other hand, the kernel swings counter-clockwise—with most of its mass then lying in swathes parallel to the *Conditioned* axis—then that conditioning operation has successfully explained the original distribution's characteristic.

Figures 5 and 6 display exactly that counter-clockwise pivot. Thus, they show the importance of space and trade in explaining the "emerging twin peaks" of Fig. 4. Quah (1996d) explores these issues further.

4. Conclusion

This paper has argued that convergence—because it concerns poor economies catching up with rich ones—forces the researcher to study what happens to the entire cross sectional distribution of economies. What matters for convergence is *not* whether a single economy is tending towards its own, individual steady state. Instead, what matters is the behavior of the entire distribution.

Moreover, while understanding economic growth is undoubtedly important, convergence can be insightfully studied by itself. This paper shows that taking such an approach leads to theoretical and empirical analyses that, in turn, raise further interesting questions.

In section 2, a series of growth models progressively emphasized the importance of understanding the dynamics of the cross section distribution. Section 2 argued that "representative-economy" reasoning would be unlikely to shed much light on whether the poor can catch up with the rich. A model that departs usefully from such reasoning might then involve ideas about coalition and group formation, as analyzed above, but would help explain the dynamics of cross-economy income distributions.

Section 3 described empirical analyses that adopt this distribution-dynamics perspective. Cross-country empirical findings thus far have confirmed the importance of studying convergence in this way: they have revealed a range of behavior in the cross section that are hidden from "representative-economy", convergence regression analysis.

Statistical analysis of *medd*'s is, in economics, in its early stages. (Extensions to what I have described above are in, among others, Desdoigts, 1996; Magrini, 1995; Quah, 1995a, b; and Trede, 1995.)

Viewing convergence as distribution dynamics diverts focus away from economic growth as just boosting the arguments of a production function and towards understanding interaction across economies. Such interaction could take the form of ordinary merchandise trade. Or, as in Section 2.3 above, it could take the form of coalition formation. Theoretical models of individual cross-section interaction and group formation have also begun to be explored elsewhere (already mentioned above, for instance, are Benabou, 1996 and Durlauf, 1993). Much, however, remains to be done.

Finally, in this paper I have emphasized convergence only in the context of cross-economy behavior. Many of the same ideas and techniques, of course, apply directly to other types of distribution dynamics: a leading example would be the behavior of personal and family income distributions (e.g., Johnson and Reed, 1996). Convergence there—in the sense of dynamic inequality—is as important as convergence across economies. While, through Kuznets-based reasoning, dynamic inequalities in personal incomes could be related to economic growth, it does so in a way different than does cross-country convergence. Therefore, that growth matters for both might, in fact, be usefully ignored in their study.

References

Barro, Robert J., and Xavier Sala-i-Martin (1992) "Convergence," *Journal of Political Economy* 100(2), 223–251, April

Baumol, William J. (1986) "Productivity growth, convergence, and welfare," *American Economic Review* 76(5), 1072–85, December

Ben-David, Dan (1994) "Convergence clubs and diverging economies," Working Paper 922, CEPR, February

Bénabou, Roland (1996) "Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance," *American Economic Review* 86(3), 584–609, June

Bernard, Andrew B., and Steven N. Durlauf (1995) "Convergence in international output," *Journal of Applied Econometrics* 10(2), 97–108, April

Bertola, Giuseppe (1995) "Uninsurable shocks and international income convergence," *American Economic Review* 85(2), 301–306, May

Bianchi, Marco (1995) "Testing for convergence: A bootstrap test for multimodality," Working Paper, Bank of England, May

Canova, Fabio, and Albert Marcet (1995) "The poor stay poor: Non-convergence across countries and regions," Discussion Paper 1265, CEPR, November

Dawkins, Richard (1976) *The Selfish Gene* (Oxford: Oxford University Press)

den Haan, Wouter J. (1995) "Convergence in stochastic growth models: The importance of understanding why income levels differ," *Journal of Monetary Economics* 35(1), 65–82, February

Desdoigts, Alain (1994) "Changes in the world income distribution: A nonparametric approach to challenge the neoclassical convergence argument," PhD dissertation, European University Institute, Florence

———— (1996) "Determining development patterns and "club" formation using projection pursuit techniques," Working Paper, ECARE, ULB, August

Durlauf, Steven N. (1993) "Nonergodic economic growth," *Review of Economic Studies* 60(2), 349–366, April

Durlauf, Steven N., and Paul Johnson (1995) "Multiple regimes and cross-country growth behavior," *Journal of Applied Econometrics* 10(4), 365–384, October

Forni, Mario, and Lucrezia Reichlin (1995) "Dynamic common factors in large cross sections," Discussion Paper 1285, CEPR, London W1X 1LB, December

Friedman, Milton (1992) "Do old fallacies ever die?," *Journal of Economic Literature* 30(4), 2129–2132, December

Grier, Kevin B., and Gordon Tullock (1989) "An empirical analysis of cross-national economic growth, 1951–80," *Journal of Monetary Economics* 24(2), 259–276, November

Johnson, Paul, and Howard Reed (1996) "Two nations? The inheritance of poverty and affluence," Commentary 53, Institute for Fiscal Studies, London, January

Kelly, Kevin (1994) *Out of Control* (Addison Wesley)

Kelly, Morgan (1992) "On endogenous growth with productivity shocks," *Journal of Monetary Economics* 30(1), 47–56, October

Lamo, Ana R. (1996) "Cross-section distribution dynamics," PhD dissertation, LSE

Lee, Kevin, M. Hashem Pesaran, and Ron Smith (1996) "Growth and convergence in a multi-country empirical stochastic Solow model," Working Paper 9531, DAE, University of Cambridge, Cambridge, August

Leung, Charles, and Danny T. Quah (1996) "Convergence, endogenous growth, and productivity disturbances," *Journal of Monetary Economics,* December

Lucas Jr., Robert E. (1992) "On efficiency and distribution," *Economic Journal* 102(2), 233–247, March

———— (1993) "Making a miracle," *Econometrica* 61(2), 251–271, March

Magrini, Stefano (1995) "Income disparities among European regions," Working Paper, Geography Department, LSE, London, November

Paap, Richard, and Herman K. van Dijk (1994) "Distribution and mobility of wealth of nations," Working Paper, Tinbergen Institute, Erasmus University, October

Quah, Danny T. (1993a) "Empirical cross-section dynamics in economic growth," *European Economic Review* 37(2/3), 426–434, April

———— (1993b) "Galton's fallacy and tests of the convergence hypothesis," *The Scandinavian Journal of Economics* 95(4), 427–443, December

———— (1994a) "Exploiting cross section variation for unit root inference in dynamic data," *Economics Letters* 44(1), 9–19, January

———— (1994b) "One business cycle and one trend from (many,) many disaggregates," *European Economic Review* 38(3/4), 605–613, April

———— (1995a) "Coarse distribution dynamics for convergence, divergence, and polarization," Working Paper, Economics Department, LSE, July

———— (1995b) "International patterns of growth: II. Persistence, path dependence, and sustained take-off in growth transition," Working Paper, Economics Department, LSE, December

———— (1996a) "Aggregate and regional disaggregate fluctuations," *Empirical Economics* 21(1), 137–159

———— (1996b) "Convergence empirics across economies with (some) capital mobility," *Journal of Economic Growth* 1(1), 95–124, March

———— (1996c) "Empirics for economic growth and convergence," *European Economic Review* 40(6), 1353–1375, June

———— (1996d) "Empirics for growth and distribution: Polarization, stratification, and convergence clubs," Working Paper, Economics Department, LSE, September

———— (1996e) "Ideas determining convergence clubs," Working Paper, Economics Department, LSE, April

———— (1996f) "Regional convergence clusters across Europe," *European Economic Review* 40(3–5), 951–958, April

Quah, Danny T., and Thomas J. Sargent (1993) "A dynamic index model for large cross sections," In *Business Cycles, Indicators, and Forecasting,* ed. James Stock and Mark Watson, vol. 28 (Chicago IL: University of Chicago Press and NBER) chapter 7, pp. 285–306

Romer, Paul M. (1994) "The origins of endogenous growth," *Journal of Economic Perspectives* 8(1), 3–22, Winter

Sala-i-Martin, Xavier (1996) "Regional cohesion: Evidence and theories of regional growth and convergence," *European Economic Review* 40(6), 1325–1352, June

Silverman, Bernard W. (1986) *Density Estimation for Statistics and Data Analysis* (New York NY 10001: Chapman and Hall)

Summers, Robert, and Alan Heston (1991) "The Penn World Table (Mark 5): An expanded set of international comparisons, 1950–1988," *Quarterly Journal of Economics* 106(2), 327–368, May

Tapscott, Don (1996) *The Digital Economy* (New York: McGraw-Hill)

Thomas, Jonathan P., and Tim Worrall (1990) "Income fluctuation and asymmetric information: An example of a repeated principal-agent problem," *Journal of Economic Theory* 51(2), 367–90, August

Trede, Mark M. (1995) "The age profile of earnings mobility: Statistical inference for conditional kernel density estimates," Discussion Paper, Universitat zu Koln, January

Fig. 1: Deterministic neoclassical convergence.

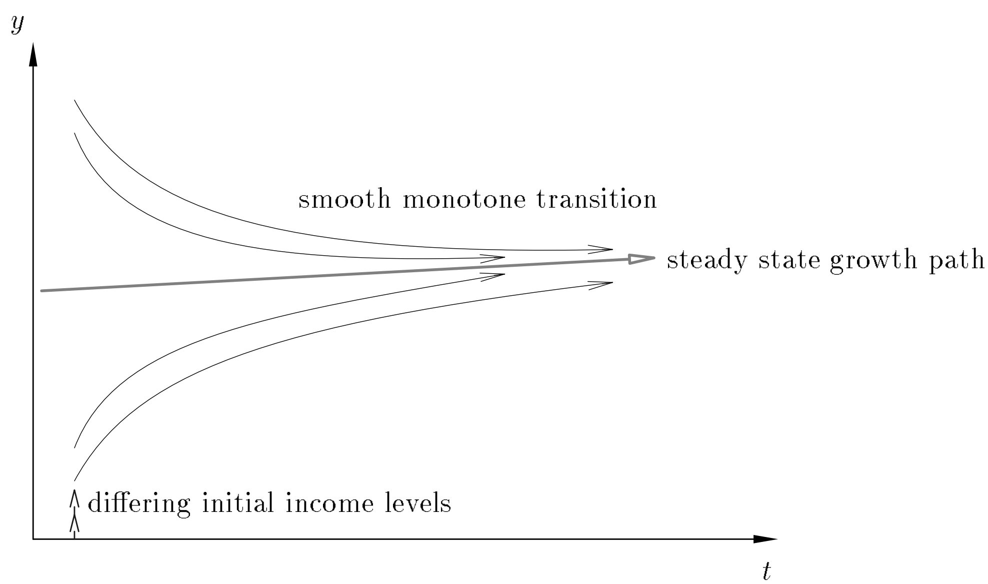


Fig. 2: Divergence towards nondegenerate steady-state invariant distribution.

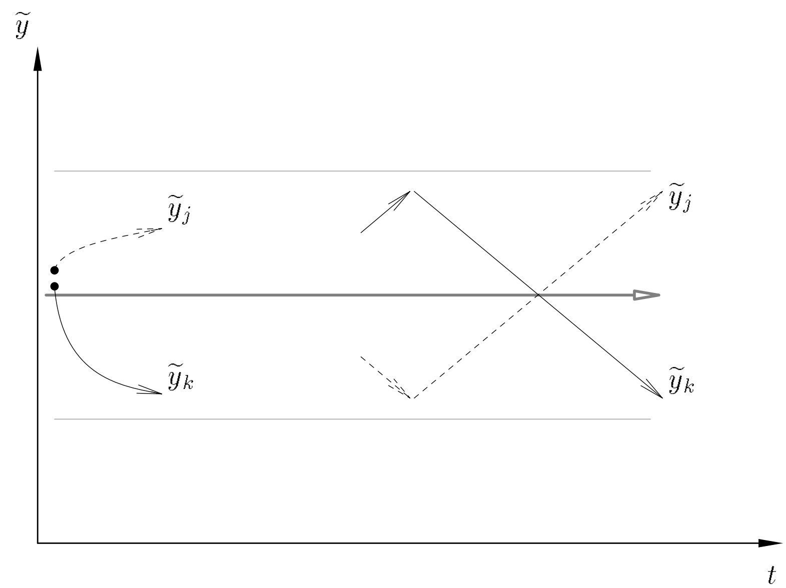


Fig. 3: Stratification, polarization, and convergence clubs.

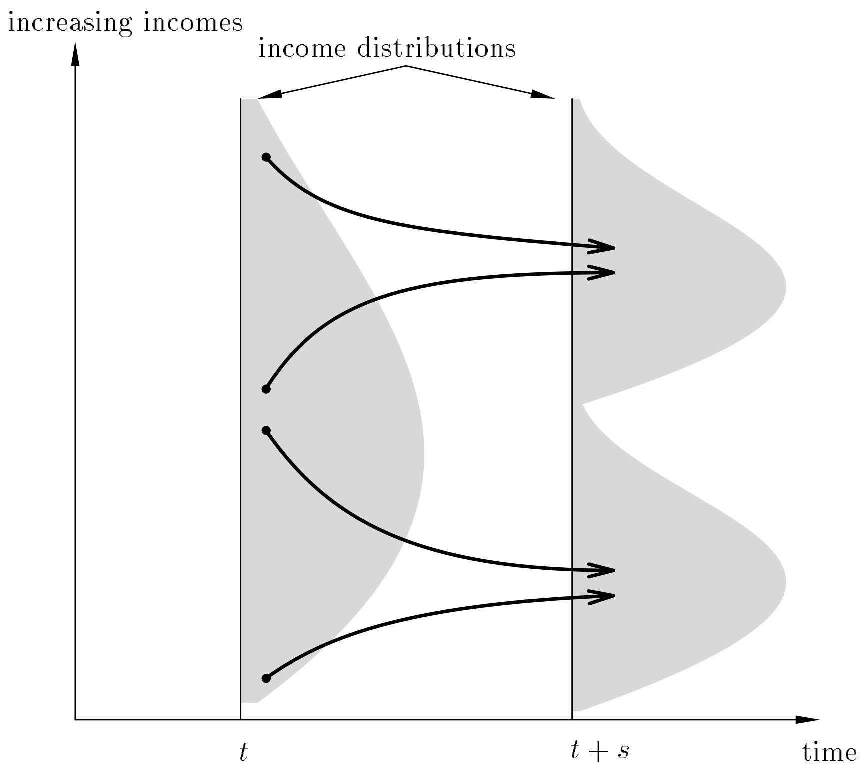
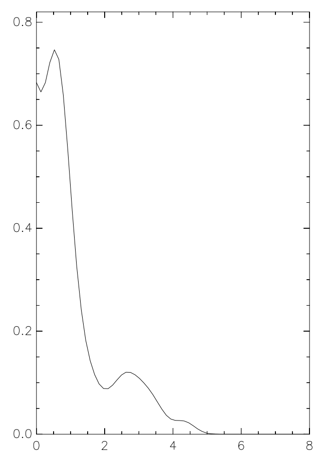
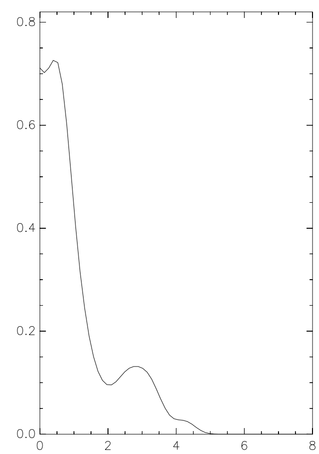


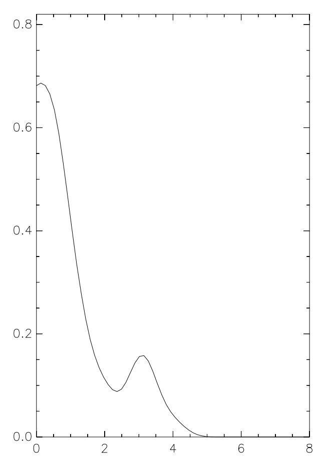
Fig. 4: Densities of normalized cross-country productivities.
Emerging twin peaks.



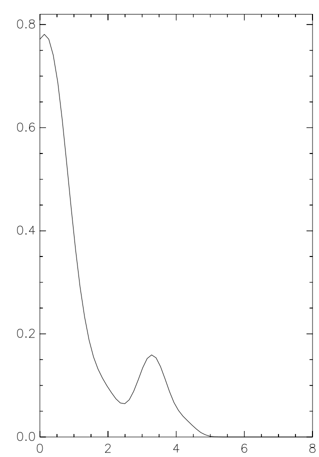
(a): 1961



(b): 1970

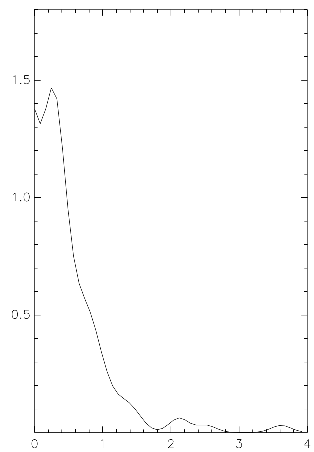


(c): 1980

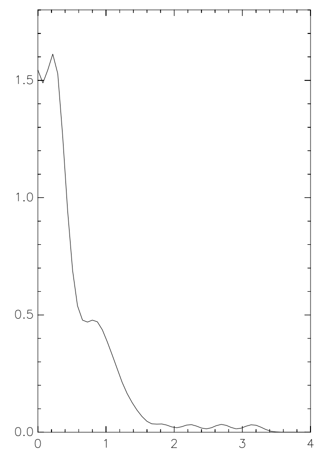


(d): 1988

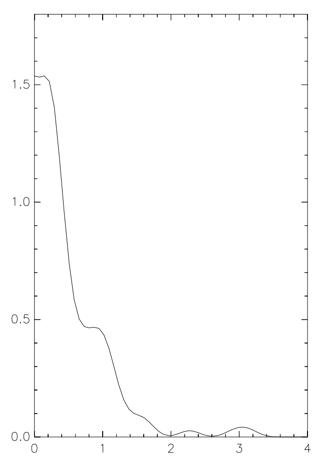
Fig. 5.d: Densities of normalized cross-country productivities.
Trade conditioning.



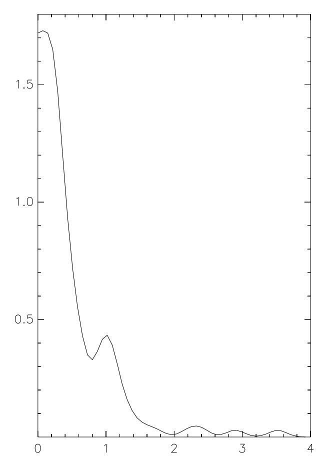
(a): 1961



(b): 1970

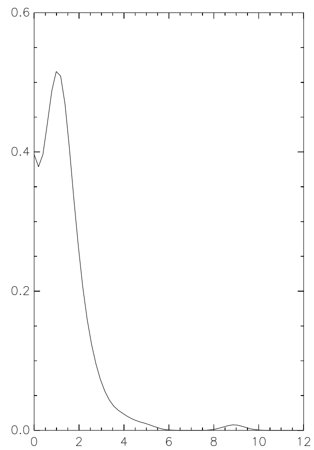


(c): 1980

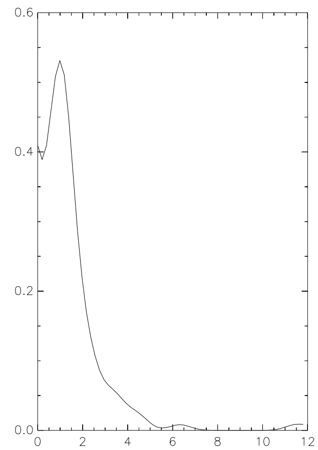


(d): 1988

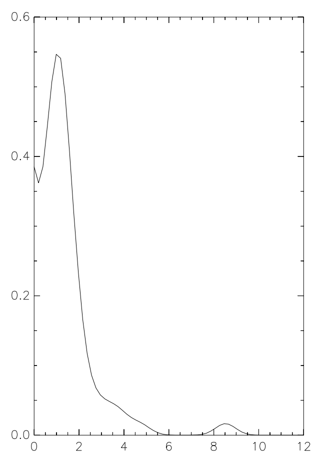
Fig. 6.d: Densities of normalized cross-country productivities.
Spatial conditioning.



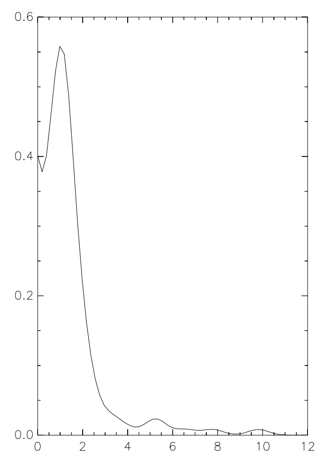
(a): 1961



(b): 1970



(c): 1980



(d): 1988

Fig. 5.s: Stochastic kernel, normalized cross-country productivities.
Trade conditioning.

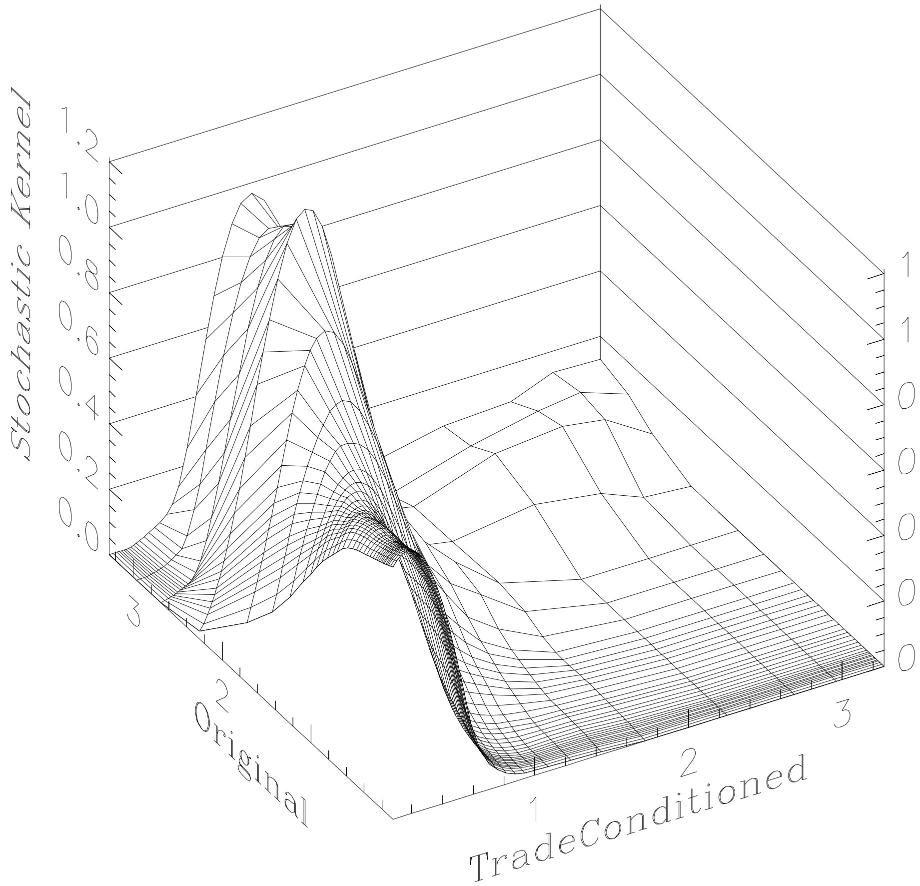


Fig. 6.s: Stochastic kernel, normalized cross-country productivities.
Spatial conditioning.