Convergence Empirics Across Economies with (Some) Capital Mobility

by

Danny T. Quah*
LSE Economics Department and CEP

CENTRE FOR ECONOMIC PERFORMANCE
DISCUSSION PAPER NO. 257
August 1995

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

---

* This paper is substantially revised from "Dependence in growth and fluctuations across economies with mobile capital." The change in title reflects the revisions, and emphasizes the paper's central message. I thank M. Arellano, J. Dolado, Z. Eckstein, O. Galor, G. Koop, R. E. Lucas Jr., X. Sala-i-Martin, A. Wood, referees, members of the LSE Macro Breakfast Group, and seminar participants at CEP, CORE, Essex, ESSIM, Sussex, and Tel Aviv for useful comments and discussions.

Convergence Empirics Across Economies
with (Some) Capital Mobility

by

Danny T. Quah
LSE Economics Department and CEP

November 1995

ABSTRACT

This paper uses a model of growth and imperfect capital mobility across multiple economies to characterize the dynamics of (cross-country) income distributions. This allows convenient study of the convergence hypothesis, and reveals, where appropriate, polarization and clumping within subgroups. The data show little cross-country convergence; instead, the important features are persistence, immobility, and polarization, exemplified by "convergence club" or "twin peaks" dynamics.

Keywords: convergence club, distribution dynamics, polarization, stochastic kernel, twin peaks

JEL Classification: C23, F43, O47

Communications to: D. T. Quah, LSE, Houghton Street, London WC2A 2AE.
[Tel: +44-171-955-7535, Fax: +44-171-831-1840, Email: `dquah@lse.ac.uk`]

1. Introduction

Compare labor productivities and incomes (per capita) across countries and ask, Are poorer countries catching up with richer ones? Are they likely to in the future? Or, are countries converging only within "clubs"? If so, are these clubs of the very rich and the very poor, or is most of the world becoming only middle class? Answers to these questions—on catch-up and convergence—are basic for thinking about economic growth: they can be viewed either as checks on different growth models or as empirical regularities to be explained by theory. More fundamentally, they provide direct measurements on the dynamics of relative well-being and income mobility across economies—interest in this is the same as interest in income distributions and mobility across people within an economy.

This paper provides a new empirical method for addressing such questions. Of course, a vast literature (e.g., Barro and Sala-i-Martin (1992), Baumol (1986)) already tackles similar issues. This paper's approach differs from those earlier ones in a number of important ways; all the differences, however, stem from one simple insight. The analysis here recognizes that to address questions of catch-up and convergence, one needs to model explicitly the dynamics of *the entire cross-country distribution* of incomes. By contrast, the traditional approach of modelling only the behavior of an average or representative economy (e.g., Barro and Sala-i-Martin (1992)) sheds little light on catch-up and convergence.

Thus, the traditional approach is silent on some interesting questions. It is silent on how an economy, initially among the poorest 10% of world economies, will catch up with the richest 5%, or will converge to within the median 20%. It can say nothing on whether the poorest economies will stagnate, permanently distant from the richest ones: it is silent on patterns of stratification and polarization. The traditional approach does clarify if a particular economy will converge to its own steady state: this, however, is not catch-up, and is arguably a less interesting notion of convergence.

To summarize, while growth economics presupposes an interest in growth behavior across the entire range of economies—are poorer ones catching up with those richer—standard empirical analyses have only looked at the growth behavior of a single (representative) economy. Traditional analyses are thus unable to properly address the original questions of interest.

Going from studying a representative economy to studying a small, select number of economies—as done in vector time-series studies (e.g., Bernard and Durlauf (1995))—moves in the right direction. Proceeding from there to analyzing the dynamics of the entire cross-country distribution is the logical next step. This paper takes that step. It refines earlier such analyses (e.g., Friedman (1992), Kirman and Tomasini (1969), Laursen and Paldam (1982), Parente and Prescott (1993)), and studies growth and convergence directly in terms of the dynamics of the cross-country income distribution.

In this, the paper connects to a number of rich literatures. First, the approach draws inspiration from studies of individual income distribution and social mobility (e.g., Atkinson (1970), Shorrocks (1978)). However, instead of taking individuals as the primitive unit of observation, this paper takes whole macro-economies. Second, the focus here is on the global dynamics of the entire distribution, not on regression coefficients describing average behavior. Thus, this paper is in the spirit of early sociological studies of longitudinal data, not the later, albeit related econometric work.[1] A similar concern drives macroeconomic VAR analysis, where interest lies not in regression coefficients per se, but in dynamic, global properties of the data (e.g., Sims (1980)). Third, recent theoretical work in growth has considered phenomena like convergence clubs, polarization, and poverty traps (among many others, Azariadis and Drazen (1990), Baumol (1986), Ben-David (1994), Esteban and Ray (1994), Galor and Zeira (1993), Quah (1995b)); this paper can be viewed

[1] Contrast the focus of, e.g., Singer and Spilerman (1976) and Brandon Tuma, Hannan, and Groeneveld (1979) with that in, e.g., Heckman and Singer (1985) and Lancaster (1990). Integrating both sets of concerns is, of course, possible (e.g., Lillard and Willis (1978)); however, this paper adopts the more direct modelling strategy.

as seeking empirical verification for such effects.[2]

Similar reasoning had earlier motivated Quah (1993a, 1993b); see also Desdoigts (1994) and Lamo (1995). Section 3 details how the work here improves on those earlier studies. Three key refinements, though, can be mentioned now. First, the current work allows for explanatory variables (importantly, physical capital investment and schooling) before examining convergence properties. Second, earlier analyses constructed arbitrary, discrete partitions of incomes to analyze their evolving distributions. The current work removes this arbitrariness by analyzing income distribution dynamics directly on a continuous state space. We will see, however, that these refinements preserve the principal conclusions in Quah (1993a, 1993b): the world cross section of countries appears to be polarizing into convergence clubs of rich and poor. Moreover, the conditioning analysis below, different from that traditionally performed, indicates endogeneity of savings relative to growth: the two are jointly determined. Third, the current work calculates passage-time distributions to analyze "growth miracles," such as observed in Singapore or South Korea.

The remainder of this paper is organized as follows. Section 2 provides an explicit model of economic growth and dynamically evolving distributions. In the model, growth is due to capital accumulation; under certain assumptions on capital mobility, the cross-section distribution of countries polarizes into rich and poor. This section has two goals: first, it clarifies how traditional analyses can incorrectly interpret patterns of growth. Second, it motivates the empirical analysis in Section 3. Some of the results in Section 3 are only suggestive—as expected with new methodology—and others more technically precise. The empirical analysis reveals polarization in the world cross section of countries: over time, convergence clubs at

---

[2] The concern in Galor and Zeira (1993) is explicitly about *personal* income distribution; however, this paper will, in section 2, reinterpret that analysis to obtain predictions on cross-country income distributions. Esteban and Ray (1994), on the other hand, clearly intend their analysis to apply both to people and to entire economies.

high and low ends of the income distribution appear, and the middle income class vanishes. Passage-time estimates, to calibrate the possibilities for growth miracles such as Singapore or South Korea, show that the observed speed and magnitude of these occur with reasonable (5%) likelihood. Finally, Section 4 concludes and describes extensions.

2. Capital mobility, polarization, and convergence

I present here a simple model of growth in a cross-section of economies. The theoretical model has structure inspired by Galor and Zeira (1993), although emphases, details, and interpretations differ.[3]

In the model, growth can potentially occur through accumulating physical and human capital. (In the sequel, "skill" and "human capital" are used interchangeably.) Skills are specific to economies; physical capital, however, is mobile, although only imperfectly, across economies. Economies are distinguished by the amount of human and physical capital their populations own, and by their distance from (or inversely, their degree of capital markets integration with) other economies. Economies are otherwise identical. There is one commodity; thus, exchange across economies occurs exclusively through interest-bearing capital loans.

[3] A referee has emphasized that the model here is no more than illustrative for the empirics that follow: *any* model with cross-sectional heterogeneity and multiple steady states would make the same general points. Those with strong intuition on these, and interested only in this paper's empirical contribution, can proceed directly to Section 3, after looking over Subsection 2.5 below—the empirics can be useful across different motivations. However, specific points discussed in 2.5, e.g., the statements on conditional convergence regressions, *are* model-specific—thus, the model is not without content.

2.1. Technology

Production takes time: inputs expended this period yield output only next period. Output produced can be either consumed or costlessly transformed into physical capital and transferred across time.

The production technology accepts two inputs: physical capital (K) and skills embodied in labor. Physical capital is used up in production (this is inessential, and can be relaxed at the cost of more notation). Skills take only two values, 0 for unskilled and 1 for skilled; increasing skill values is costly. Because by assumption skills are bounded, ongoing growth occurs only from accumulating K.

Assume the following formalization for the description just given. With skill level 0, input K this period produces in the next period

$$Y^{(n)}(K) = KF^{(n)} + \delta^{(n)}, \qquad \text{for constants } F^{(n)}, \delta^{(n)} > 0;$$

the $^{(n)}$ superscript denotes "no skill". With skill level 1, input K produces similarly in the next period

$$KF^{(s)} + \delta^{(s)}, \qquad \text{for constants } F^{(s)}, \delta^{(s)} > 0;$$

the $^{(s)}$ superscript denotes "skill". But because skill acquisition costs resources s, net output here is only

$$Y^{(s)}(K) = \begin{cases} (K - s)F^{(s)} + \delta^{(s)}, & \text{if } K \geq s \\ 0 & \text{otherwise} \end{cases}$$

(remember s has to be paid upfront, before production can proceed). The 0 in the second branch of $Y^{(s)}$ indicates that skilled labor is unavailable without expending s. It is intuitively plausible to take $F^{(s)} > F^{(n)}$ and $\delta^{(s)} > \delta^{(n)}$: below, assumptions ($T_1$) and ($T_2$) state precisely how large $F^{(s)}$ and $\delta^{(s)}$ are relative to their unskilled versions. One can thus think of the technology as comprising two separate processes: the first (unskilled-labor) is simply linear in inputs; the second (skilled-labor) can be run only after costly skill acquisition, following which it too is linear in inputs.

Because of the assumed time pattern of production, it is natural to interpret $F^{(n)}$ as a gross interest rate. Thus, I take $F^{(n)}$ to be at least 1—this is convenient but not necessary for the discussion below. Coefficients F can, however, also be viewed as marginal products of physical capital. Coefficients δ are net outputs achievable with zero physical capital. That F and δ are positive constants independent of K is inessential: curvature can be allowed without affecting the principal results, but at the cost of complicating the calculations; with curvature, δ's could also be taken to be zero, provided assumptions (T_1) and (T_2) below are appropriately replaced.

Assume that the skilled process dominates the unskilled process in that $F^{(s)}$ and $\delta^{(s)}$ are jointly sufficiently higher than $F^{(n)}$ and $\delta^{(n)}$:

$$(T_1) \qquad F^{(s)} > F^{(n)} + \delta^{(n)} s^{-1}$$

$$(T_2) \qquad \delta^{(s)} > sF^{(s)} + \delta^{(n)}.$$

Assumption (T_1) says that, abstracting from schooling costs, even were $\delta^{(s)}$ zero, the skilled process would still dominate the unskilled process for $K \geq s$. Assumption (T_2) says that, even were the unskilled process to improve its marginal product to match the skilled process, the latter would still dominate for $K \geq s$ by its intercept $\delta^{(s)}$ being sufficiently large.

Assumption (T_1) gives $F^{(s)} > F^{(n)}$; together, (T_1) and (T_2) imply

$$\delta^{(s)} > sF^{(n)} + \delta^{(n)} > \delta^{(n)}.$$

The assumed disparity between $^{(n)}$ and $^{(s)}$ values involves schooling costs s: the larger is s, the larger must be $\delta^{(s)} - \delta^{(n)}$, although the smaller need be $F^{(s)} - F^{(n)}$. It follows also that at $K = s$, the skills-using technology $Y^{(s)}$ dominates $Y^{(n)}$, and does so increasingly as K increases. Figure 1 summarizes the discussion thus far.

2.2. People

Two-period-lived overlapping generations populate every economy. Each person has exactly one offspring so that population size is constant through time. I assume each generation comprises a single agent—this only saves notation and having to say "per capita" repeatedly below. (Inconsistently, however, I will refer to "the young" or "the old" in the plural.)

People are born unskilled but receive bequest K of physical capital from their parents. In the first period of their lives, they do not consume, but only decide whether to remain unskilled or to acquire skills. Production occurs between the first and second periods. At the start of the second period of their lives, these now-old receive income W from that production; they consume C and provide bequests K' to their offspring. Those young who, once again, are born unskilled receive that K', and then decide on skill acquisition. The process repeats.

As in Galor and Zeira (1993), I assume the representative person has preferences $U(C, K')$; that person solves, in the second period of life, the program:

$$\max_{C,K'} \quad U(C, K')$$
$$\text{subject to} \begin{cases} C + K' \leq W, \\ C \geq 0, \ \ K' \geq 0. \end{cases}$$

In the first period of life, that person only needs to decide on skill acquisition which subsequently determines W. Thus decisions occur in two stages: first, conditional on K, maximize W; then choose C and K' optimally. Assumptions on U will be discussed below.

Thus far the model reinterprets and simplifies Galor and Zeira (1993); it is the subsequent analysis on capital mobility where differences manifest.

2.3. Cross-country interaction

Countries interact through borrowing and lending in physical capital. Let subscripts j and k denote two different countries; define R_{jk} to be the gross one-period interest rate paid by economy j towards economy k on a loan extended from k to j.[4] I say "towards" rather than "to" because—and this is where imperfect capital mobility enters the discussion—I assume that economy k receives not R_{jk} but only a smaller $R_{\cdot k}$ on such loans.

The difference between $R_{\cdot k}$ and R_{jk} has a number of interpretations. Likely the easiest of these is the "melting-iceberg" transportation costs used in regional economics and trade: moving capital involves wastage. To see how this works, suppose that transporting from country j to k loses fraction ϵ_{jk} in transit; similarly, transporting from k to j, ϵ_{kj}. These ϵ's might be increasing in geographical distance between j and k; they need not equalize in opposite directions if sending and receiving technologies differ. If economy j wants to use 1 unit of physical capital, economy k needs to ship it $(1-\epsilon_{kj})^{-1}$. Economy k does so on condition that it receives $(1-\epsilon_{kj})^{-1}R_{\cdot k}$ in the next period. To ensure that, economy j needs to ship back $(1-\epsilon_{jk})^{-1}(1-\epsilon_{kj})^{-1}R_{\cdot k}$ after its use of the 1 unit it received. Thus, economy j pays a gross interest rate of $R_{jk} = (1-\epsilon_{jk})^{-1}(1-\epsilon_{kj})^{-1}R_{\cdot k} > R_{\cdot k}$. In this interpretation, all that is needed is that some ϵ be different from zero. Moreover, ϵ's can differ across country pairs so that, fixing any capital-rich center, some countries can be viewed as geographically closer to it, others further away.

A less literal interpretation is that the difference between R_{jk} and $R_{\cdot k}$ reflects lack of integration in capital markets: complete integration gives $R_{\cdot k} = R_{jk}$. The less integrated are capital markets, the greater the disparity between R_{jk} and $R_{\cdot k}$. In this interpretation, the previous paragraph's ϵ's still measure "distance" between countries, but now that distance is no longer a physical concept. (As an example, consider that capital markets are likely better integrated between the US and Japan, say, than between China and Japan.) Also, distance, as used here, clearly need not be related to income levels in the countries concerned.

If there are many capital-rich lending countries acting competitively, then for

---

[4] The subscripts order in R follows that used in Markov chain analysis, and so should be easily remembered. Note that lower case k indexes economies while, from earlier, upper case K denotes physical capital.

them $R_{\cdot k}$—the interest rate received—will equal $F^{(s)}$, as $F^{(s)}$ is the return on capital's alternative use in production. Therefore, in general, interest rates paid by capital-poor borrowing countries will strictly exceed $F^{(s)}$.

2.4. Optimal decisions and dynamic equilibrium

Economy j can borrow from abroad at gross interest rate

$$R_{j\cdot} \stackrel{\text{def}}{=} \min_k\{R_{jk}\} > \min_k\{R_{\cdot k}\} \geq F^{(s)} > F^{(n)}.$$

Thus, the young in economy j will borrow from abroad, if at all, only to top up their inheritance K to the amount s. Borrowing less doesn't get one the benefits from schooling and is unprofitable since $R_{j\cdot} > F^{(n)}$; borrowing more is unprofitable since $R_{j\cdot} > F^{(s)}$. If K already exceeds s, the young never borrow, again, since $R_{j\cdot} > F^{(s)}$. Thus, borrowing, if it occurs at all, is always in the amount $(s - K)$.

Figure 2 shows the opportunity set available to the young in economy j. The segments marked $Y^{(s)}$ and $Y^{(n)}$ are directly from figure 1. There is, however, now an additional line emanating downwards with slope $R_{j\cdot}$ from point $(s, \delta^{(s)})$; this *loans line*,

$$W = \delta^{(s)} - (s - K)R_{j\cdot},$$

gives net second-period income as a function of initial K, conditional on borrowing $(s - K)$. Call $K^{(p)}$ the abscissa of the intersection of this line with $Y^{(n)}$; this exists and is positive whenever $R_{j\cdot}$ exceeds $[\delta^{(s)} - \delta^{(n)}]s^{-1}$. The $^{(p)}$ superscript is to suggest *participation* in international capital markets. If K is less than $K^{(p)}$, then that economy does better by shutting itself off from capital markets and using only the unskilled process $Y^{(n)}$. Gross interest payments—if that economy borrowed—would exceed the gain in output from going to the skilled process $Y^{(s)}$.

Thus, provided $R_{j\cdot}$ is sufficiently large, values for K partition into three regions $[0, K^{(p)})$, $[K^{(p)}, s)$, and $[s, \infty)$. [This partitioning is similar to that in Galor and Zeira (1993).] The first two regions are determined by the participation threshold $K^{(p)}$ which, in turn, depends on $R_{j\cdot}$. Fix the loans line at point $(s, \delta^{(s)})$ and vary its slope $R_{j\cdot}$: it becomes obvious that $K^{(p)}$ is increasing in $R_{j\cdot}$. Threshold $K^{(p)}$ reaches its minimum value of 0 when $R_{j\cdot} = [\delta^{(s)} - \delta^{(n)}]s^{-1}$, and it tends to a maximum equal to s when $R_{j\cdot}$ increases without bound.

Therefore, the overall optimization program that each generation in economy j solves is:

$$\max_{C,K'} \quad U(C, K')$$
$$\text{subject to} \begin{cases} C + K' \leq W(K, R_{j\cdot}) \\ C \geq 0, \;\; K' \geq 0 \end{cases}$$

where

$$W(K, R_{j\cdot}) \stackrel{\text{def}}{=} \begin{cases} F^{(n)}K + \delta^{(n)} & \text{for } K \text{ in } [0, K^{(p)}) \\ R_{j\cdot}K + \delta^{(s)} - sR_{j\cdot} & \text{for } K \text{ in } [K^{(p)}, s) \\ F^{(s)}K + \delta^{(s)} - sF^{(s)} & \text{for } K \text{ in } [s, \infty). \end{cases}$$

Net income $W(K, R_{j\cdot})$—optimally determined by the first-period decision on skill-acquisition—explicitly records its dependence on $R_{j\cdot}$. The effect of $R_{j\cdot}$ on W appears not just as the slope in the middle segment loans line, but also in determining the threshold $K^{(p)}$.

Under standard assumptions on preferences U, optimal decisions are interior, and can be written as $C(W)$ and $K'(W)$; moreover, maximized welfare is monotone increasing in W. If, further, U is Cobb-Douglas,

$$U(C, K') = (1 - \beta) \log C + \beta \log K', \qquad 0 < \beta < 1,$$

then the optimal decision rules are

$$C = (1 - \beta) \times W(K, R_{j\cdot}) \qquad \text{and} \qquad K' = \beta \times W(K, R_{j\cdot}).$$

The second of these is a difference equation in K; its graph is simply a β-scaled version of figure 2.

If U were not Cobb-Douglas, then transitions in K can still be determined from something analogous to an appropriate (non-uniform) scaling of figure 2—the results, however, will be more difficult to characterize. Thus, I follow Galor and Zeira (1993) once again, and assume U Cobb-Douglas. In addition, I strengthen the assumption on β to

$$s/\delta^{(s)} < \beta < 1/F^{(s)};$$

assumption (T_2) guarantees that this restriction is meaningful. The coefficient β measures, roughly, concern for future generations; my assumption therefore asserts that the old care neither too little nor too much for their immediate offspring.[5] The β-scaled version of figure 2 can then be graphed as figure 3.

In addition to altering the vertical scale, figure 3 adds a number of other useful features to figure 2. To understand them, first note that $\beta > s/\delta^{(s)}$ implies that the point $(s, \beta\delta^{(s)})$ lies above the 45-degree line through the origin. (Hereafter, "45-degree line" always refers to that through the origin.) Next, notice that $K^{(\infty)}$, defined by the intersection of the 45-degree line with $\beta \times W(\cdot, \cdot)$ on K in $[s, \infty)$, exists, exceeds s, and is finite, by $\beta F^{(s)} < 1$.

Then, define $K_{(\infty)}$ by the intersection of the 45-degree line with $\beta \times Y^{(n)}$. As the notation suggests, $K^{(\infty)}$ and $K_{(\infty)}$ will be upper and lower limit points for the growth process. Since $\delta^{(n)}$ is positive, and $\beta < [F^{(s)}]^{-1}$ implies $\beta F^{(n)} < 1$, a positive, finite $K_{(\infty)}$ necessarily exists. Moreover, because

$$K_{(\infty)} = \left[1 - F^{(n)}\beta\right]^{-1} \delta^{(n)}\beta,$$

[5] As emphasized in Galor and Zeira (1993), this interpretation is strictly correct only when it is next generation's utility, not K', that is the second argument in U. In that case, however, the analysis becomes harder, but does not substantively change the conclusions below, provided there is sufficiently heavy discounting of the future.

and

$$\beta < \left[F^{(s)}\right]^{-1} < \left[F^{(n)} + \delta^{(n)}/s\right]^{-1} = \frac{s}{F^{(n)}s + \delta^{(n)}}$$

$$\Longrightarrow \delta^{(n)}\beta < \left[1 - F^{(n)}\beta\right] s \Longrightarrow \left[1 - F^{(n)}\beta\right]^{-1} \delta^{(n)}\beta < s,$$

we have that $K_{(\infty)} < s$.

Thus far, we have established that $K^{(\infty)}$ and $K_{(\infty)}$ exist, and satisfy $K_{(\infty)} < s < K^{(\infty)}$. Turn to $K^{(p)}$ and $K^{(c)}$; the first of these we introduced above as the participation threshold; the second will turn out to be a polarization *cutoff* level, hence the $^{(c)}$ superscript. Whenever $K^{(p)} \geq K_{(\infty)}$, i.e., whenever $R_{j\cdot}$ is sufficiently high—I will make this precise by calculating the lower limit value R^* below—define $K^{(c)}$ by the intersection of the 45-degree line with the middle segment of $\beta W(K, R_{j\cdot})$; otherwise, for $K^{(p)} < K_{(\infty)}$, set $K^{(c)} = K^{(p)}$. What determines $K^{(p)}$ and $K^{(c)}$ is, of course, the interest rate $R_{j\cdot}$. Call R^* the critical value for the gross interest rate such that $K^{(p)} = K_{(\infty)}$: when $R_{j\cdot} < R^*$, then $K^{(c)} = K^{(p)} < K_{(\infty)}$; when $R_{j\cdot} > R^*$, then $K^{(c)} > K^{(p)} > K_{(\infty)}$.

From figure 3, it is apparent that $R^* > [\delta^{(s)} - \delta^{(n)}]s^{-1} > 0$. To provide sharper intuition, first obtain from their definitions that:

$$K_{(\infty)} = \left[1 - F^{(n)}\beta\right]^{-1} \delta^{(n)}\beta$$

and

$$K^{(p)} = s\left[\frac{R_{j\cdot} - \left[\delta^{(s)} - \delta^{(n)}\right] s^{-1}}{R_{j\cdot} - F^{(n)}}\right].$$

Equating these at $R_{j\cdot} = R^*$ gives

$$R^* = \frac{\left[\delta^{(s)} - \delta^{(n)}\right] - F^{(n)}\beta\delta^{(s)}}{s - \left[F^{(n)}s + \delta^{(n)}\right]\beta},$$

which can be shown to exceed $\left[\delta^{(s)} - \delta^{(n)}\right] s^{-1}$, as expected. The explicit expression for R^* involves parameters of both technology and preferences; it is, moreover, independent of the particular economy j. Apart from these observations, the expression appears to have little economic interpretation. Dynamic implications from it, however, are interesting.

For the sequel, denote the capital stock in economy j at time t by $K_j(t)$. Figure 3 shows transitions in K_j over a single period; by iteration, it gives the entire subsequent time path for K_j, as indicated by the arrows at the bottom.

Proposition 2.1: *If capital mobility is sufficiently low, i.e., $R_{j\cdot}$ exceeds R^*, then there exists a cutoff level $K^{(c)}$ between $K^{(p)}$ and s such that*

$$\text{at any } t \quad \begin{cases} K_j(t) > K^{(c)} & \Longrightarrow \lim_{t' \to \infty} K_j(t+t') = K^{(\infty)}, \\ K_j(t) < K^{(c)} & \Longrightarrow \lim_{t' \to \infty} K_j(t+t') = K_{(\infty)}, \\ K_j(t) = K^{(c)} & \Longrightarrow \forall t' > 0 \quad K_j(t+t') = K^{(c)}. \end{cases}$$

The proposition establishes the limiting behavior of K from different starting points; its proof is immediate from figure 3.[6]

Since the cutoff level $K^{(c)}$, like $K^{(p)}$, varies with $R_{j\cdot}$, where an economy tends over time depends on both its current capital stock and the interest rate it faces. Economies having identical physical capital but having different "distances" from capital-rich centers—and hence different $R_{j\cdot}$'s—will, in general, have different

[6] Earlier readers have remarked, though, on what appears to be a slight puzzle. Why do forward-looking, utility-maximizing agents allow their physical capital stock to deteriorate, as happens when $K(t) \downarrow K_{(\infty)}$ for $K^{(p)} < K(t) < K^{(c)}$? For initial K in this range, utility improves over remaining unskilled by borrowing and thereby acquiring skills. However, after payment of the (high interest) loan, the utility-maximizing bequest K' turns out to be lower than the original inheritance K—that bequest is, nevertheless, higher than if skill-acquisition had *not* occurred. The correct comparison, therefore, is between two possible paths: (i) acquiring skills and allowing K to decline towards $K_{(\infty)}$ and (ii) not acquiring skills and, again, allowing K to decline towards $K_{(\infty)}$. Path (i) clearly gives higher welfare than (ii).

dynamic behavior: one economy could converge to $K_{(\infty)}$, and the other to $K^{(\infty)}$, thus resulting in the two diverging from each other. Overtaking is also possible: economy j could start out with a higher capital stock than economy k, and thus appear richer, but if economy k is sufficiently close to a capital-rich center—or, put differently, is in an appropriate club or cluster of economies—then over time economy k will overtake j.

Under Proposition 2.1, interest rates in rich and poor economies, converging towards $K^{(\infty)}$ and $K_{(\infty)}$ respectively, need not be very different. Thus, rates of convergence—determined by $F^{(s)}$ and $F^{(n)}$ through $K' = \beta W(K, \cdot)$—could be similar, although the convergence is to different limit points.

Consider now equilibrium as $R_{j\cdot}$ falls, perhaps because of better integration of capital markets and thus higher mobility in K. With perfect capital mobility the lower bound for $R_{j\cdot}$ is $F^{(s)}$. Important critical values as $R_{j\cdot}$ falls, however, are first R^* (when $K^{(p)} = K_{(\infty)}$) and then $\left[\delta^{(s)} - \delta^{(n)}\right] s^{-1}$ (when $K^{(p)} = 0$). Under assumption (T_2), the smaller of these two critical values, $\left[\delta^{(s)} - \delta^{(n)}\right] s^{-1}$, always exceeds $F^{(s)}$; thus, so does R^*.

Proposition 2.2: *If $R_{j\cdot} = R^*$, then $K^{(p)} = K^{(c)} = K_{(\infty)}$ but the statements on the dynamic behavior of K_j in Proposition 2.1 remain true.*

Again, the proof is immediate from figure 3 by driving $K^{(p)}$ and $K^{(c)}$ towards $K_{(\infty)}$, or equivalently, driving the gross interest rate down to R^*. Extending to when $R_{j\cdot}$ falls *below* R^*, figure 3 gives the following.

Proposition 2.3: *If $R_{j\cdot} < R^*$, then K_j converges to $K^{(\infty)}$, independently of the initial value.*

Under Proposition 2.3, even when the initial capital stock is less than $K^{(p)}$, the economy still converges to the upper limit point $K^{(\infty)}$. Its time path, moreover, is interesting: the economy experiences first low interest rates ($F^{(n)}$), then high ($R_{j\cdot}$), and then low again ($F^{(s)}$). Thus, there is at first slow growth, then a speeding up when K first exceeds $K^{(p)}$ and the economy starts to participate in international loan markets, and finally a slowing down again when K finally exceeds s. (These rates of growth and convergence can be read directly off the different branches of $K' = \beta W(K, R)$.)

The dynamics of Proposition 2.3—convergence to a unique point—can apply even for $R_{j\cdot} > F^{(s)}$. Thus, perfect capital mobility is not necessary for convergence, although it is clearly sufficient. With it, $R_{j\cdot} = F^{(s)}$ is less than $\left[\delta^{(s)} - \delta^{(n)}\right] s^{-1}$; then, convergence from everywhere to a unique point occurs at identical uniform rates.

2.5. Concluding comments and observable implications

This section has presented a simple model of accumulation and imperfect capital mobility. Under reasonable assumptions, the model has two distinct stable limit points—$K^{(\infty)}$ and $K_{(\infty)}$ in figure 3—independent of the interest rate.

If interest rates $R_{j\cdot}$ are sufficiently low, i.e., if capital is sufficiently mobile, then $K_{(\infty)}$ is not observable: all economies converge to $K^{(\infty)}$. Perfect capital mobility is not necessary for this, only that there be *enough* capital mobility, relative to the distribution of K extant.

With imperfect capital mobility, poor economies sufficiently distant from capital-rich centers remain poor. Capital poverty, however, is relative: two apparently identical economies could diverge away from each other, and end up converging to different points, depending on their respective distances from capital-rich centers.

Some of the model's assumptions and predictions are interesting to examine further; others, less so. Figure 4 shows one set of observable (and, in my view, the interesting) implications. At time t_0 there is some initial distribution of K's across the entire cross section of economies.[7] Geography (or history) determines the

[7] Contrary to some readers' intuition, figure 4 and the word "distribution" do not say that random disturbances have, somehow, been added to the model. Given a set of economies, one can always *define* the distribution of K's across that cross section of economies—regardless of whether K is stochastic or deterministic. "Distribution" is used here in the same sense as "income distribution" across

pattern of capital markets integration across the distribution. Working through figure 3 and Propositions 2.1–2.3, over time some economies become better off, others worse off. Convergence clubs form; the distribution tends towards a bimodal distribution at time t_1, with the two modes centering on $K^{(\infty)}$ and $K_{(\infty)}$. As the figure indicates, overtaking could occur (conditions for this have been described above).

For brevity, I will refer to the range of behavior depicted in figure 4 as *twin peaks dynamics*. Such dynamics warn on potential misinterpretations of conditional convergence regressions.[8] Recall that in that work, a researcher attempts to understand the behavior of incomes across economies by estimating a cross-section regression, "controlling" for human capital and other observable variables. What will that researcher find if the model here is at work?

Initially, without accounting for different skill levels across countries, the researcher concludes divergence in the distribution of incomes: that researcher notes that middle-income countries, initially close to each other at t_0 in figure 4, grow apart from one another. But at t_1 in figure 4, those economies clustering around the higher mode will have higher skills in the labor force; those around the lower mode, lower. [To understand this just look back at figures 2 or 3.] Thus, taking into account different skill levels, the researcher sees two facts: (i) convergence for countries in subgroups having similar skill levels, and (ii) richer countries also having a higher-skilled labor force. The researcher thus concludes: first, there is no unconditional convergence; second, that conditional convergence occurs once one conditions on human capital; and third, that human capital explains cross-country patterns of growth.

Such conclusions mislead. Instead, it is patterns of cross-country capital market integration—how high R_j.'s are—that explain everything: human and physical capital stocks are only responding endogenously to extant integration patterns in

---

individuals, e.g., Atkinson (1970).

[8] In the model, measured income is just an increasing function of K, and so all statements about physical capital extend immediately to income.

R. The polarization into rich and poor, which would vanish if capital markets were integrated, is inappropriately interpreted as conditional convergence, explained by human capital. More important, however, is that such conditional-convergence analysis fails altogether to detect the rich-poor polarization.

A similar, but more subtle, fallacy arises in correlating economic growth with interest rates. In the model, the fastest growing economies could be those experiencing the highest interest rates—see, e.g., the discussion after Proposition 2.3. Does this mean that a high interest rate—low capital mobility—is good for growth? Clearly not: in the model, economies escape poverty (i.e., prevent convergence downwards to $K_{(\infty)}$) through greater access to capital loans, not less. It is simply that the fastest-growing economies are those who find it welfare-maximizing to take high-interest loans, *despite* those interest rates being high.

These potential pitfalls in interpreting cross-section (conditional) convergence regressions add to those previously given in Bernard and Durlauf (1996) and Quah (1993b, 1996b). Because the current pitfalls are specific to a theoretical model, they are more directed, and thus less general. Taken together, however, these criticisms form an important argument for circumspection in interpreting cross-section growth and convergence regressions.

Turning back to the current model, what useful empirics does it suggest? More generally, what are useful empirics for studying growth and convergence? I take the model's key predictions to be on the dynamics of the cross-country income distribution—exemplified in the twin peaks dynamics of figure 4. It is this behavior that should be empirically studied. Such analysis allows directly examining convergence—in the sense of poor economies catching up with rich ones. It allows directly examining the formation of convergence clubs or clusters. Is the cross-section distribution, over time, collapsing to a single point? Or is there clumping around multiple modes?[9] Do parts of the distribution remain where they

[9] I say "clumping" rather than clustering, to minimize confusion with cluster analysis in statistics. Some readers have suggested that the model here is not about clumping within distributions but instead about correlation within sub-

began? Do parts transit from low to high, and vice versa?

These questions motivate looking at convergence empirics in a way that departs from standard cross-section regression analysis. They also suggest that the interesting empirical results are not tests of particular, tightly-specified restrictions. Instead, what is interesting is to document the dynamics of the evolving cross-country income distributions.[10]

3. Convergence empirics

To study these distribution dynamics, take each period's observation to be the cross-country distribution of (per capita or per worker) incomes. The empirical model will analyze how this distribution evolves, tracking intra-distribution dynamics, clumping, and long-run tendencies.

Such focus differs from that in standard cross-section or panel data econometric models that study the behavior of a representative or average unit. There,

---

groups of countries. In a deterministic setup, the appropriate interpretation is unclear. However, suppose one were to fix R and to introduce additive iid disturbances on the left hand side of the equation defining W. Then, in equilibrium, capital stocks are discrete-time Markov processes within each country, but are *independent* across countries, regardless of whether those countries clump about $K_{(\infty)}$ or $K^{(\infty)}$. Clumping can thus occur without correlation. I conclude that correlation is an incorrect interpretation; clumping or convergence-club formation is the robust prediction.

[10] Merely documenting, rather than testing more rigorously, is of course, only a stopgap analysis. Taking that next step, however, is hard. This will become clearer in the next section, but technically, what will be of interest are global, "shape" properties in an infinite-dimensional operator describing transitions of measures. Performing appropriate inference thus means first getting a stochastic process characterization for the distribution of that operator, and then integrating over appropriate subsets of the (infinite-dimensional) operator space. While the way to proceed is conceptually clear, the details are hard and remain to be investigated.

the cross sectional averaging yields a (conditional) representative, whose behavior need not be revealing for the entire distribution's. The current focus also differs from that in standard time series models, as here each period's observation is not just a scalar or a finite-dimensional vector but a distribution.

Standard econometric analysis, thus, does not readily provide a convenient tractable model of distribution dynamics. For that, I exploit a duality property from Markov process theory.

In the part of this theory most used by economists, the researcher observes a scalar (or finite-dimensioned vector) stochastic process. The researcher then infers the implied unobservable sequence of probability distributions associated with that process. This hypothesized distribution sequence is *dual* to the original observed process.

In the current work, it is instead a sequence of distributions that is observed, while its dual—the scalar process—is implied but never observed.[11] The same mathematics works, of course, independent of whether it is the scalar process or the distribution sequence that is primal.

Transition probability functions describe the dynamics of the scalar process. Dual to this, *stochastic kernels* describe the law of motion of the sequence of distributions (see, e.g., Chung (1960), Futia (1982), Stokey and Lucas (Ch. 8, 1989)).

Denote by λ_t the measure corresponding to the cross-country income distribution at time t. The stochastic kernel describing the evolution of λ_t to λ_{t+1} is a mapping M_t to $[0,1]$ from the Cartesian product of income values and Borel-measurable sets such that:

$$\text{(1)} \qquad \forall \text{ Borel-measurable } A: \qquad \lambda_{t+1}(A) = \int M_t(y, A)\, d\lambda_t(y).$$

[11] Although artificial, with properties potentially different from those of any single element in the distribution, the associated scalar process could still be useful in interpretation. Quah (1996a) has exploited this in studying co-movements between aggregate and disaggregate fluctuations.

Such an M_t encodes how the distribution at time t evolves into one at time $t+1$; it contains information on both intra-distribution dynamics and the external shapes of the distributions. Knowing about intra-distribution dynamics sheds light on a range of interesting events: two countries, initially comparable, over time diverging so that one transits to the rich part of the income distribution, the other to the poor part; initially poor countries catching up with the rich; initially rich countries falling behind others originally poor—in brief, the events described in figure 4.

If M_t were time-invariant and equation (1) were augmented with a "disturbance" term, then (1) becomes analogous to a standard time-series first-order vector autoregression. Only, equation (1) takes values that are measures (or distributions), rather than just scalars or finite-dimensioned vectors. Maintaining time-invariance in M but suppressing the disturbance—as done in VAR impulse response analysis—equation (1) can be written as the convolution

$$\lambda_{t+1} = M * \lambda_t. \tag{2}$$

Iterating (2) gives (a predictor for) future cross-section income distributions

$$\lambda_{t+s} = (M * M * \cdots * M) * \lambda_t = M^s * \lambda_t.$$

Taking this to the limit as s gets arbitrarily large then characterizes the long-run distribution of incomes. Is λ_{t+s} eventually invariant to λ_t, so that the long-run distribution is also ergodic? Does λ_{t+s} tend towards a degenerate point measure, so that there is convergence towards equality? Does λ_{t+s} tend towards a bimodal measure—as in figure 4—so that the world is polarizing into rich and poor? Cross-sectional mobility and the speed of convergence of the evolving distributions can be studied from the spectral characteristics of (the infinite-dimensional operator implied by) M. Variants of equation (1) therefore allow answering a wealth of interesting questions about cross-sectional income dynamics.

Previous applications of these ideas to convergence empirics (Quah (1993a, 1993b)) have noted that the stochastic kernel conveniently and simultaneously informs on four characteristics of dynamically evolving distributions: (I) their changing external shapes; (II) intra-distribution dynamics; (III) long-run behavior; and (IV) the speed of convergence to that long run. However, that empirical work discretized the space of income values into a finite grid: the measures λ_t are then probability vectors; the stochastic kernel is a transition probability matrix; and the integral in (1) becomes a matrix product. Even with these simplifications, that previous work concluded twin-peaks behavior for the cross-country distribution of incomes.

While continuing to maintain M time-invariant, this section improves that earlier work in three ways. First, it allows the space of income values to be continuous, eschewing (necessarily arbitrary) discretization; thus it estimates the infinite-dimensional stochastic kernel nonparametrically.[12] Second, this section analyzes a fifth characteristic obtainable from the stochastic kernel, namely (V) the distribution of first-passage times for movements between different parts of the income distribution. This then allows calibrating the likelihood of spectacular growth successes, such as post-War Singapore or South Korea. Finally, whereas earlier work gave only unconditional income distribution dynamics, this section provides distribution dynamics, both unconditional and conditioned on auxiliary variables, including physical and human capital investment.

In the following I study a 117-economy subset of the Summers-Heston (1991) national incomes data. As in Parente and Prescott (1993), I take the basic data to be the log of per worker productivity relative to the US, thus giving 116 cross-sectional units. (I will at times refer to this also as per capita income; although different in details, the overall movements of normalized productivity and income are generally similar.) Normalizing by US leaves unaltered how countries differ

[12] Such refinement is useful beyond just spurious generality. It is well-known (e.g., Chung (1960)) that discretization can remove the Markov property from an otherwise well-behaved Markov process: important features of the evolving distributions—intra-cell movements within a particular grid, for instance—could be inappropriately hidden in a discretized stochastic kernel.

from each other, but is a convenient way to remove some of the overall trend in the cross section.[13]

3.1. Conditioning

It is often unclear which auxiliary conditioning variables are appropriate in studying convergence. In some well-known studies (e.g., Barro and Sala-i-Martin (1992), Sala-i-Martin (1994)) additional right-hand-side variables in a convergence regression are justified, not by a precise theoretical model, but instead only by a rough intuition that those variables should affect long-run growth possibilities. Examples of this include measures of democracy, industrial/agricultural mix, religion, or continent dummy variables.

Similarly, the model in Section 2 does not yield one unambiguous conditioning regression to estimate; this is common to all models that focus on distributional implications (e.g., Durlauf (1992), Galor and Zeira (1993), and Quah (1995b)). In one interpretation, the model's predictive content lies entirely in its sequence of dynamically evolving distributions. To examine those predictions coherently, the analysis should not bring into the analysis additional variables because such variables would fall in one of two categories. First, they can be like physical and human capital that develop endogenously in the model, and so are not usefully viewed as explanatory variables. Second, they come from outside the model; but if deemed important they should already have earlier been in the theoretical investigation. The alternative view, implicit in Barro and Sala-i-Martin (1992) and Sala-i-Martin (1994), is that convergence should only be sought after the obvious reasons potentially preventing it are removed. In this reasoning, one asks if convergence obtains not in the original income distribution, but in the conditional one, conditioning on those "obvious reasons."

[13] The appropriate way to do this in large cross sections that have large, unknown trend dynamics remains an important, unresolved question for future study (see, e.g., Quah and Sargent (1993)). Treating this more rigorously here, however, would take us too far afield. The data appendix explains the choice of data sample.

While the first approach seems truer to the theoretical analysis, the empirical method here certainly allows conditioning. Thus, some of the results below will condition on schooling enrollment, physical capital investment, and a dummy for the African continent.[14] The first two variables appear in the model of section 2; the third brings in a commonly-used dummy variable.

Unlike in standard convergence regressions, however, here the time-varying conditioning variables are not assumed to be exogenous. Instead, conditioning proceeds by first regressing growth rates on a two-sided distributed lag of the time-varying conditioning variables and then extracting the fitted residuals for subsequent analysis. This procedure yields, in large samples, an appropriate conditional distribution regardless of the exogeneity of the right-hand-side variables.[15]

To fix ideas, begin with just investment in physical capital, and ask, Can this be taken exogenous in a regression explaining income or productivity growth rates? Table 1 reports Granger causality tests for bivariate VARs in productivity (per worker output) growth rates and investment shares, all relative to the US. The table indicates significant dynamic inter-dependence between growth and investment. While investment does help to predict future growth, it is also itself incrementally predicted by lagged growth. Thus, investment cannot be taken exogenous in a productivity growth equation.

The results in Table 1 are obtained by OLS, pooling cross-section and time-series observations. Unlike in standard panel data application (e.g., Holtz-Eakin, Newey, and Rosen (1988)) individual effects are not allowed in these regressions. Permitting those would be equivalent to leaving permanent differences in growth rates unexplained—but it is exactly those differences that we are ultimately trying to understand here. The projection of growth on investment, not allowing for

[14] I thank David Weil for kindly providing the enrollment data, earlier used in Mankiw, Romer, and Weil (1992). Those data do not exactly match the sample here, and so some adjustment was needed: see the data appendix for details.

[15] Such two-sided distributed lag regressions are common in Granger causality analysis, e.g., Sims (1972).

individual effects, is precisely the best linear predictor (Chamberlain (1984)), and thus correctly gives residuals that are the components unexplained by (or, more correctly, orthogonal to) investment.

Table 1 suggests that regressions of growth rates on investment—current and lagged, or even on just current investment—have no interesting structural interpretation. Such regressions, therefore, do not characterize the distribution of output growth conditional on investment the way economists usually imagine. Instead, a more appropriate conditional distribution obtains by conditioning on current, lagged, and future investment. This is what Table 2 presents.

Looking across the columns of Table 2, we notice a marked stability in the coefficients of the two-sided projections. Fit, as measured by R^2, does not increase dramatically with increasing lag lengths. The heteroskedasticity-robust t-statistic on investment led 3 years exceeds 2 in the fourth-order two-sided projection, but not in the third-order. In all projections, investment at lead 1 through lag 2 appear significant for predicting growth, but other leads and lags not consistently. In the analysis that follows, I present results using the residuals from the second-order two-sided projection (a compromise)—the conclusions remain unchanged were one to take instead residuals from the other projections. By the same reasoning, I condition also on two leads and lags of schooling, and a dummy variable for the African continent.[16]

[16] Details on this are, again, in the data appendix. The analysis to follow could be presented for differing subsets of regressors—doing so, however, quickly taxes the reader's patience. I have chosen here to use just investment, schooling, and the African continent dummy as the conditioning variables. The results are almost always the same when one uses only investment in physical capital. Thus, wherever "conditioning information" appears, little changes if this is taken to mean either all the conditioning variables or just physical capital.

3.2. (Nonparametric) Stochastic kernels

In the discrete stochastic kernel analyses in Quah (1993a, 1993b), characteristics (I)–(IV) derive from manipulating a transition probability matrix. That matrix, in turn, is estimated from probabilities of transiting through an appropriate grid in incomes space; each row of the transition probability matrix is a (conditional) probability vector.

As already described above, in the continuous case, the transition probability matrix becomes an infinite-dimensional operator on an appropriate space of measures. Alternatively, it can be viewed as a non-negative function defined on current and future incomes, where, holding current income constant, the function is a probability density over future incomes.

Figures 5–6 ((a) and (b)) show stochastic kernels describing fifteen-year-horizon evolutions of the distribution of relative productivity.[17] Each figure (a) shows a three-dimensional plot of the stochastic kernel; each figure (b) shows a contour plot of the function in (a). Axes marked *Period* t and *Period* $t+k$ measure the log of relative productivity at different time periods; the vertical axis in (a) graphs the kernel. (The Technical Appendix contains details on this estimation.)

To interpret these graphs, think of them as continuous versions of a Markov transition probability matrix. From any point on the axis marked *Period* t, looking in the direction parallel to the other axis traces out a probability density describing transitions to different parts of the income distribution. Thus, a ridge in the kernel piled up on the (positive sloped) diagonal shows high persistence and immobility: different parts of the income distribution remain roughly where they begin. On the other hand, the kernel being equal-valued as one moves parallel to the *Period* $t+k$ axis indicates low persistence: location in the future income distribution is independent of current status. A different extreme where piling up

[17] An early working paper version (Quah (1994)) of this paper also gives transitions over one-year horizons. Following a referee's suggestion, only the longer-horizon results are presented here, to conserve space and because they are more informative. The one-year results do not alter any substantive statements.

occurs along the negatively-sloped diagonal shows economies dynamically overtaking one another. Distinct peaks along the diagonal indicate "convergence club" behavior—economies within a particular income class tend to remain in that class; but, more importantly, over time every economy becomes attached to precisely one such income class.[18] Finally, convergence—the poor growing faster and the rich slowing down so that all eventually collect together—would manifest in the kernel accumulating on a single ridge parallel to the *Period* t axis. In this last case, income levels eventually equalize, regardless of whether an economy began rich or poor. Following the earlier discussion on conditioning, when stochastic kernels are given for fitted residuals from a first-stage regression, the convergence properties should be read as conditional on those auxiliary variables.

Do these figures suggest convergence? Evidently not. The dominant characteristic in these kernels is a ridge along the main diagonal, indicating persistence and immobility.[19] Moreover, figure 5 shows evidence for polarizing convergence clubs: twin peaks are directly evident in the stochastic kernel at high and low portions of the main diagonal; these are particularly clear in the contour plot figure 5 (b).

---

[18] This seems to me more flexible and revealing than Durlauf and Johnson's (1994) regression tree method for studying such dynamics, although both techniques have their relative advantages and disadvantages. Ben-David (1994) gives empirical analyses with motivations similar to mine, although models and methods differ substantially.

[19] The subsequent discussion focuses only on the point estimates displayed in figures 5–6. It is possible to develop point-wise standard errors around the estimates (see, e.g., Silverman (1986)); however, such standard errors aren't informative for the current discussion, and could instead mislead. As pointed out earlier in footnote 10, current interest lies in global, not local, features of the estimated stochastic kernels. Knowing only the standard errors at particular points does not allow us to assess the uncertainty associated with, for instance, the statement that "the dominant characteristic is a ridge along the main diagonal." Appropriate calibration of such uncertainty remains an important item for future research.

Stronger evidence on this dynamic polarization obtains from the shape of the implied ergodic distribution, i.e., the stochastic kernel's limit point.[20] Explicitly characterizing that limit is difficult in this general, continuous model, although the sequence of densities given in figure 6 of Quah (p. 436, 1993b)—tending towards bimodality—is suggestive. If, instead one were to discretize the stochastic kernel, then the ergodic distribution can be found as the eigenvector corresponding to the leading eigenvalue of the stochastic kernel. Quah (1993a, 1993b) has calculated exactly that, and shown that the ergodic distribution displays peaks at rich and poor extremes with the middle portion vanishing.[21]

Comparing unconditional and conditional kernels (figures 5 and 6) one sees that fine details differ, but the global dynamics of the distribution remain roughly unchanged. There are the same polarization, persistence, and immobility features in both. While the conditioning variables do affect the behavior of productivities in each country, they do not affect the dynamics of the entire distribution. (This message will re-emerge when passage times are studied below.)

To conclude, the empirics here suggest polarization and divergence across the cross section, the opposite of the poor catching up with the rich. The evidence is graphical and high-dimensional, however, and it might be useful to provide some

---

[20] Earlier referees objected to my use of the terms "ergodic" here and "distribution" earlier, because—according to those referees—the model implicitly being discussed [the evolving cross sections of figure 4] bears no uncertainty: how is it possible to discuss ergodicity in a deterministic model and, worse, without checking if the time series are integrated? Above, I have already discussed why "distribution" is appropriate. Use of the term "ergodic", once one has a model for how distributions evolve, and one can characterize the limit point of that dynamic scheme (from, say, Chapman-Kolmogorov equations), is standard in classical probability theory; see, e.g., Chung (1960), Doob (1953), or especially Feller (Ch. X, 1971). This has nothing to do with whether the model is deterministic or random.

[21] Bianchi (1995), applying point-in-time tests to these data, finds bimodality of the cross section in later years, but not in earlier ones—thus supporting the dynamic claims made in the text.

simple summary statistics. On the other hand, doing so necessarily hides details: e.g., the twin peaks and clustering dynamics need no longer be as evident. Passage times, studied next, are a convenient compromise.

3.3. Passage times

From the stochastic kernels, it is possible to infer the speed of first passage, for an economy to move from one part of the income distribution to another. Such a transition speed is comparable to the rate of convergence studied in e.g., Barro and Sala-i-Martin (1992) and Sala-i-Martin (1994): only here the concept recognizes our interest in how an economy moves against a background of the dynamic cross-sectional distribution of all other economies.

To understand the statistics to follow on first passage times, it is convenient to proceed in three stages. First, recognize that the continuous case involves no relevant ideas beyond those present in the discrete case: the additional subtleties are mostly measure-theoretic, and will not be discussed further in this paper.

Second, since all the interesting issues already manifest in the discrete case, a detailed but brief description for that case will be worthwhile: I will provide this below. Finally, in footnotes, I explain how the calculations are performed for the continuous case—these details contain no substantive interest, but might be useful for other researchers wishing to perform similar analyses.

Consider a Markov chain with a discrete state space and stationary transition probabilities. Denote by $\phi_{j,k}^{(t)}$ the probability that the chain first enters state k in t steps conditional on the current state being j. For concreteness, one might think of state j as corresponding to an economy being in the j-th decile of the income distribution; thus, one might be interested in the likelihood that that economy progresses (or declines) to the k-th decile in however many years. For fixed but arbitrary j and k, the infinite sequence $\{\phi_{j,k}^{(t)} : t = 1, 2, \ldots\}$ is the probability density of *first-passage times* from j to k. When $j = k$, the sequence in t is more accurately called the probability density of *recurrence times* for state j; then when $\sum_{t\geq 1} \phi_{j,j}^{(t)} < 1$, the state j is said to be *transient*. In words, such a state is eventually not observed. (In a world of convergence clubs, the region outside those clubs would be transient.) Again, for fixed but arbitrary j and k, the mean first-passage time between j and k can be found as $\sum_{t\geq 1} t\phi_{j,k}^{(t)}$: this might, of course, be infinite. The more probability the density $\phi_{j,k}$ places on high values of t, the fewer transitions will be seen from state j to state k, and thus the lower the intra-distribution mobility.

Earlier parts of this section have discussed distribution dynamics entirely in terms of the stochastic kernel. How does ϕ—the first-passage time density—relate to the stochastic kernel? In the discrete case, as already pointed out, the stochastic kernel becomes equivalent to a matrix of transition probabilities. Let Q denote that matrix so that $[Q^t]_{jk}$ denotes the (j,k) entry of its t-th power. Then

$$\forall t \geq 1, \text{ and } j,k: \qquad \left[Q^t\right]_{jk} = \sum_{s=1}^{t} \phi_{j,k}^{(s)} \times \left[Q^{t-s}\right]_{kk}. \tag{3}$$

This states the following. Suppose that the chain starting from j has its first passage through k occurring at the s-th step, and, following that, the chain lands again in k after another $t-s$ steps: that journey is exactly one from j to k in t steps. This path occurs with probability $\phi_{j,k}^{(s)} \times [Q^{t-s}]_{kk}$. Taking all possible first passage times s between 1 and t gives a union of disjoint events that is precisely the event where the chain moves from state j to state k in t periods; but this last event has probability $[Q^t]_{jk}$.

Equation (3) gives a simple recursion by which one can calculate the entire distribution of first passage times:

$$\phi_{j,k}^{(1)} = [Q]_{jk}$$

$$\phi_{j,k}^{(t)} = \left[Q^t\right]_{jk} - \sum_{s=1}^{t-1} \phi_{j,k}^{(s)} \times \left[Q^{t-s}\right]_{kk}, \qquad t = 2,3,4,\ldots.$$

This calculation works for transition dynamics between discrete states in a Markov chain. In our study, however, interest lies in transition dynamics where first, the analogue of matrix Q is an (infinite-dimensional) stochastic kernel, and second, the probability of transiting to any single value of the state space is, by continuity, zero. To get around the second difficulty, I consider only transitions into a (positive-measure) subset of income values. For the first difficulty, no purely analytic solution is available; I use stochastic simulation from the estimated stochastic kernel representation.[22]

The choice of transition events to consider is necessarily arbitrary; here, I take the event of interest to be "growth miracles", defined as passage from the 10th percentile of the cross-country distribution to anywhere above the 90th percentile. To be explicit about what this means, without conditioning this event is equivalent in 1965 to Myanmar (3.5% of US per capita income) becoming at least as rich as West Germany (56.4%). With conditioning, this event is equivalent in 1965 to Honduras (22.2% of US per capita income) becoming at least as rich as Luxembourg (210%). (After conditioning, many countries turned out to have higher per capita income than the US.)

The income distance to traverse in the unconditional case exceeds that in the conditional. Despite this, the unconditional kernel in figure 5 implies a mean first-passage time of 201 years (5th percentile: 75; 95th percentile: 435), *smaller* than the mean first-passage time of 760 years (5th percentile: 150, 95th percentile: 1980) implied by the conditional kernel in figure 6.[23] Thus, although with conditioning,

---

[22] To keep the simulations well-behaved, I combined those small subsets of income values that represented zero-probability, pathological events.

[23] The percentile figures do not measure the uncertainty associated with these estimates. Remember that an entire distribution of passage times is studied here; the percentile figures are simply points along that single, fixed distribution. To check robustness, I have verified that passage times implied by kernels estimated for different transition horizons—not presented here—retain the same rankings and orders of magnitude as those described in the text. Finally, it is interesting to observe that these mean first-passage times are comparable to those from studies of personal income distributions (Durlauf (1992)).

income distances between countries become smaller, intra-distribution mobility also falls.

One shouldn't unnecessarily emphasize the exact numbers obtained here. Instead, what is important is that the first-passage time estimates are, simply put, large. Nevertheless, "growth miracles" are predicted to occur with positive probability: in the unconditional case, with 5% probability, those spectacular events can occur within the space of (roughly) three generations.[24]

Three conclusions emerge from this passage time analysis. The first confirms an earlier message from Figures 5 and 6: while individual countries' productivities are importantly affected by conditioning variables, the global dynamics of the entire distribution are not. If anything, they only amplify conclusions available from the unconditional distributions: persistence, immobility, and polarization remain the important characterizations. Thus the conditioning variables leave unexplained why rich countries remain rich and poor ones, poor. Second, because the mean first-passage times are large, on average, growth miracles are unusual. Third, despite this, growth miracles over relatively short time spans, do occur with reasonable (5%) probability.

4. Conclusions

This paper has examined the convergence hypothesis using an empirical model of dynamically evolving distributions. A theoretical model of growth through accumulation, but with imperfect capital mobility, motivated the empirical analysis.

---

[24] I have also calculated first passage times for other events; I do not present those estimates, however, as they are not particularly informative for the questions of interest here. For instance, one might consider "growth disasters," or transitions from top to bottom of the income distributions. Disasters turn out to have distributions different from miracles, but the exact numbers don't add further insight. Also, one might consider less spectacular events such as transitions from either top or bottom to the middle. Again, however, the exact numbers, don't seem to add to the discussion in the text.

That theoretical model shows how "conditional convergence" regressions can mislead on patterns of growth; more interesting, however, it produces, in equilibrium, twin-peaks dynamics—a form of polarization across countries. The paper's empirics finds, consistent with these theoretical predictions, that the dominant features of cross-country income dynamics are *persistence*, *immobility*, and *polarization*. According to these empirics, spectacular growth miracles are expected to occur with some regularity. This set of findings contrast starkly with the uniform 2% rate of convergence that has been emphasized in other work (Barro and Sala-i-Martin (1992), Sala-i-Martin (1994)).

The tools used to document these characteristics may be of independent interest. The empirical methods here differ from those common in cross section, panel data, and time-series econometrics. They are methods well-designed to uncover phenomena like clumping, stratification, and polarization. Examples where these are relevant include industry evolutions; economic geography, location dynamics, and regional business cycles; consumption risk sharing; asset market comovements; personal income distributions and intergenerational income mobility; and disaggregate price inflations.

The current analysis points to where further theoretical and empirical analyses are useful. For one, take more general theoretical models of stratification and polarization: this paper has suggested one way to study those effects, and has illustrated their importance in cross-country incomes. Related theoretical work includes Durlauf (1992), Esteban and Ray (1994), and Quah (1995b); the robustness and empirical relevance of their conclusions need to be studied. Empirically, the stochastic kernels are no more than point estimates and remain (nonparametric and) unstructured. Calibrating precision of the estimates; providing interpretable structure to the continuous stochastic kernels; and parameterizing transition intensities in a semi-Markov generalization—thus relaxing the Markov assumption—are worthwhile, and currently under investigation.[25]

[25] Quah (1995a, 1995c) studies these extensions.

Technical Appendix

The stochastic kernels in the paper are estimated as follows: First, use an Epanechnikov kernel to nonparametrically estimate the joint density of log relative incomes at dates t and $t + k$, choosing window width optimally, as suggested in Silverman (4.3.2, 1986). That estimated joint density implies a current-period marginal density; calculate this by integration, and then divide the joint density by the implied marginal. The result is the stochastic kernel graphed in the text. For presentation, the kernels have been given with greater detail wherever the corresponding marginals are higher; see, e.g., figure 6 (a) where the grid lines become more finely spaced in different parts of the income space. Contour plots in the (b) figures are obtained by projecting vertically onto the floor of the (a) figures—the contour levels were chosen after experimentation to be informative of some of the fine structure in the (a) figures.

The literature on large-sample properties for density estimation is enormous; the reader is referred to Silverman (3.7, 1986), and references given there, for more discussion. Under assumptions giving consistency for the joint density estimator, the implied marginal is also consistently estimated. Provided then that the true marginal is bounded away from zero, the stochastic kernel is consistently estimated.

All the graphs and calculations here were performed using the econometrics shell tsrf.

Data Appendix

The data derive from that given in Summers and Heston (1991). Real per worker output is taken to from `RGDPW` (Real GDP per worker, 1985 international prices). Countries in the sample were selected by first disallowing those not having continuously available data on these two variables for the period 1960–1985. I then also excluded Kuwait—a 3-dimensional graph of the variables easily shows the Kuwait observation to dominate every other feature of the data. The remaining 117 countries are listed below (integers immediately before the country names are the indexes in the Summers-Heston database). Since I always normalized relative to the US, the constant ratio of 1 for the US observation is excluded in the calculations.

Investment in Section 3 above refers to investment share of GDP, or series `I` in Summers and Heston (1991). Again, this is normalized relative to the US.

For schooling, I used the series secondary school enrollment rate in Mankiw, Romer, and Weil (1992). That is, however, only available at five-year intervals, over 1960–1985. To compute the two-sided projections in the paper, one can, of course, simply calculate the ratio of cross-spectra and spectra, averaging over the cross-section. But given the smooth time-paths one would expect for schooling, that will likely give much the same projections as just smoothly interpolating the available data. Thus I used the latter, restricting the between-observed data to be on a straight line. Schooling is again taken to be relative to the US, and so there is no logical necessity for it to be bounded between zero and one. Given the actual realizations, however, there were few observations where these exceeded one by much. When schooling is used as an additional conditioning variable in the paper, five economies from the list below were excluded for lack of data. These were: (8) CapeVerdeIs, (53) DominicanRep, (83) China, (103) Taiwan, and (130) Yugoslavia.

Finally, the African continent dummy variable is directly from the Summers-Heston database.

The 117 economies included in the analysis are:

1	(1)	Algeria	2	(2)	Angola
3	(3)	Benin	4	(4)	Botswana
5	(6)	Burundi	6	(7)	Cameroon
7	(8)	CapeVerdeIs	8	(9)	CentralAfrR
9	(10)	Chad	10	(12)	Congo
11	(13)	Egypt	12	(14)	Ethiopia
13	(15)	Gabon	14	(16)	Gambia
15	(17)	Ghana	16	(18)	Guinea
17	(20)	IvoryCoast	18	(21)	Kenya
19	(22)	Lesotho	20	(23)	Liberia
21	(24)	Madagascar	22	(25)	Malawi
22	(26)	Mali	24	(27)	Mauritania
25	(28)	Mauritius	26	(29)	Morocco
27	(30)	Mozambique	28	(31)	Niger
29	(32)	Nigeria	30	(33)	Rwanda
31	(34)	Senegal	32	(36)	SierraLeone
33	(37)	Somalia	34	(38)	SouthAfrica
35	(39)	Sudan	36	(40)	Swaziland
37	(41)	Tanzania	38	(42)	Togo
39	(43)	Tunisia	40	(44)	Uganda
41	(45)	Zaire	42	(46)	Zambia
43	(47)	Zimbabwe	44	(49)	Barbados
45	(50)	Canada	46	(51)	CostaRica
47	(53)	DominicanRep	48	(54)	ElSalvador
49	(56)	Guatemala	50	(57)	Haiti
51	(58)	Honduras	52	(59)	Jamaica
53	(60)	Mexico	54	(61)	Nicaragua
55	(62)	Panama	56	(65)	TrinidadTobag
57	(66)	USA	58	(67)	Argentina
59	(68)	Bolivia	60	(69)	Brazil
61	(70)	Chile	62	(71)	Colombia
63	(72)	Ecuador	64	(73)	Guyana

65	(74)	Paraguay	66	(75)	Peru
67	(76)	Suriname	68	(77)	Uruguay
69	(78)	Venezuela	70	(79)	Afghanistan
71	(81)	Bangladesh	72	(82)	BurmaMyanmar
73	(83)	China	74	(84)	HongKong
75	(85)	India	76	(87)	Iran
77	(88)	Iraq	78	(89)	Israel
79	(90)	Japan	80	(91)	Jordan
81	(92)	KoreaSouthR	82	(94)	Malaysia
83	(95)	Nepal	84	(97)	Pakistan
85	(98)	Philippines	86	(99)	SaudiArabia
87	(100)	Singapore	88	(101)	SriLanka
89	(102)	Syria	90	(103)	Taiwan
91	(104)	Thailand	92	(107)	Austria
93	(108)	Belgium	94	(109)	Cyprus
95	(110)	Denmark	96	(111)	Finland
97	(112)	France	98	(113)	GermanyWest
99	(114)	Greece	100	(116)	Iceland
101	(117)	Ireland	102	(118)	Italy
103	(119)	Luxembourg	104	(120)	Malta
105	(121)	Netherlands	106	(122)	Norway
107	(124)	Portugal	108	(125)	Spain
109	(126)	Sweden	110	(127)	Switzerland
111	(128)	Turkey	112	(129)	UK
113	(130)	Yugoslavia	114	(131)	Australia
115	(132)	Fiji	116	(133)	NewZealand
117	(134)	PapuaNGuinea			

Section 3 uses two-sided conditioning regressions for growth, and then analyzes their unexplained residual components. Call $X_j(t)$ the j-th economy's period t log relative income, i.e., $\log(Y_j(t)/Y_0(t))$. The unexplained, residual components in X are calculated as follows: take fitted values from the two-sided projections, and accumulate them, country by country, to get the time-varying trend paths, $g_j(t)$, explained by the accumulation of the conditioning variables, physical capital, schooling, the African continent dummy. This determines up to an additive constant level the component for each economy unexplained by physical capital. To get that level, recognizing that the resulting location must be related to the conditioning variables for each country, we solve the minimization program:

$$\min_{a,b,c} \sum_j \sum_t \left[X_j(t) - (a \cdot I_j + b \cdot S_j + c \cdot \mathit{dummyAfrica} + g_j(t))\right]^2$$

where I_j and S_j are the (time-)average investment and secondary schooling for economy j. (Coefficients b and c can be set to zero if schooling and the African continent dummy are omitted.) Define our basic data to be the difference between actual and fitted time paths, i.e., $X_j(t) - a \cdot I_j - b \cdot S_j - c \cdot \mathit{dummyAfrica} - g_j(t)$.

The procedure I have just described seems to me one natural, convenient way to calculate that component of a country's per capita income log level explained by its (accumulation of) conditioning variables. As one might expect from studies of time-detrending (e.g., Nelson and Kang (1981)) the absolute location of the explained component is crucial in the whole exercise. Here, I have used the cross-section variation in I_j, S_j, and *dummyAfrica* to tie down that absolute level.

The general problem here is one of decomposing a time series (per worker output) into a growth component that can be explained by a set of conditioning variables and one that cannot. That question might well be subject to the identification difficulties as discussed in Lippi and Reichlin (1994) and Quah (1992)—here, however, I have chosen one particular identification scheme. Other identification schemes might give different answers, but for reasons of space, I will have to leave that for future research.

References

Atkinson, Anthony B. (1970) "On the measurement of inequality," *Journal of Economic Theory* 2(3), 244–263, September

Azariadis, Costas, and Allan Drazen (1990) "Threshold externalities in economic development," *Quarterly Journal of Economics* 105(2), 501–526, May

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, London W1X 1LB, February

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

——— (1996) "Interpreting tests of the convergence hypothesis," *Journal of Econometrics*. Forthcoming

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

Chamberlain, Gary (1984) "Panel data," In *Handbook of Econometrics vol. II,* ed. Zvi Griliches and Michael D. Intriligator (Amsterdam: Elsevier North-Holland) chapter 22, pp. 1247–1318

Chung, Kai Lai (1960) *Markov Chains with Stationary Transition Probabilities* (Berlin: Springer-Verlag)

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

Doob, Joseph L. (1953) *Stochastic Processes* (New York NY: John Wiley)

Durlauf, Steven N. (1992) "A theory of persistent income inequality," Working Paper, University of Wisconsin-Madison, October

Durlauf, Steven N., and Paul Johnson (1994) "Multiple regimes and cross-country growth behavior," Working Paper, University of Wisconsin, May

Esteban, Joan-María, and Debraj Ray (1994) "On the measurement of polarization," *Econometrica* 62(4), 819–851, July

Feller, William (1971) *An Introduction to Probability Theory and Its Applications*, 2nd ed., vol. 2 (New York: John Wiley)

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

Futia, Carl (1982) "Invariant distributions and the limiting behavior of Markovian economic models," *Econometrica* 50(1), 377–408, January

Galor, Oded, and Joseph Zeira (1993) "Income distribution and macroeconomics," *Review of Economic Studies* 60(1), 35–52, January

Heckman, James J., and Burton Singer, eds (1985) *Longitudinal Analysis of Labor Market Data* Cambridge University Press

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen (1988) "Estimating vector autoregressions with panel data," *Econometrica* 56(6), 1371–1395, November

Kirman, Alan P., and Luigi M. Tomasini (1969) "A new look at international income inequalities," *Economia Internazionale* 22(3), 3–27, August

Lamo, Ana R. (1995) "Cross-section distribution dynamics," PhD dissertation, LSE, London. In progress

Lancaster, Tony (1990) *The Econometric Analysis of Transition Data* (Cambridge University Press)

Laursen, Karsten, and Martin Paldam (1982) "The dynamics of the world's income distribution 1955–2000," In *Economic Essays in Honor of Jørgen H. Gelting,* ed. Svend Andersen, Karsten Laursen, P. Nørregaard Rasmussen, and J. Vibe-Pedersen (Copenhagen: Danish Economic Association) pp. 135–146

Lillard, Lee A., and Robert J. Willis (1978) "Dynamic aspects of earning mobility," *Econometrica* 46(5), 985–1012, September

Lippi, Marco, and Lucrezia Reichlin (1994) "Diffusion of technical change and the decomposition of output into trend and cycle," *Review of Economic Studies* 61(1), 19–30, January

Mankiw, N. Gregory, David Romer, and David N. Weil (1992) "A contribution to the empirics of economic growth," *Quarterly Journal of Economics* (429), 407–437, May

Nelson, Charles R., and Heejoon Kang (1981) "Spurious periodicity in inappropriately detrended time series," *Econometrica* 49(3), 741–751, May

Parente, Stephen L., and Edward C. Prescott (1993) "Changes in the wealth of nations," *Federal Reserve Bank of Minneapolis Quarterly Review* 17(2), 3–16, Spring

Quah, Danny (1992) "The relative importance of permanent and transitory components: Identification and some theoretical bounds," *Econometrica* 60(1), 107–118, January

——— (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

——— (1994) "Convergence empirics across economies with (some) capital mobility," STICERD Working Paper EM/94/275, Economics Department, LSE, London, April

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

——— (1995b) "Ideas determining convergence clubs," Working Paper, Economics Department, LSE, London, August

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

——— (1996a) "Aggregate and regional disaggregate fluctuations," *Empirical Economics.* Forthcoming

——— (1996b) "Empirics for economic growth and convergence," *European Economic Review.* Forthcoming

Quah, Danny, 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

Sala-i-Martin, Xavier (1994) "Cross-sectional regressions and the empirics of economic growth," *European Economic Review* 38(3/4), 739–747, April

Shorrocks, Anthony F. (1978) "The measurement of mobility," *Econometrica* 46(5), 1013–1024, September

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

Sims, Christopher A. (1972) "Money, income, and causality," *American Economic Review* 62(4), 540–552, September

——— (1980) "Macroeconomics and reality," *Econometrica* 48(1), 1–48, January

Singer, Burton, and Seymour Spilerman (1976) "The representation of social processes by Markov models," *American Journal of Sociology* 82(1), 1–54, January

Stokey, Nancy L., and Robert E. Lucas Jr. (with Edward C. Prescott) (1989) *Recursive Methods in Economic Dynamics* (Cambridge MA: Harvard University Press)

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

Tuma, Nancy Brandon, Michael T. Hannan, and Lyle P. Groeneveld (1979) "Dynamic analysis of event histories," *American Journal of Sociology* 84(4), 820–854, January

Table 1: Bivariate VAR
Exclusion Tests on Other Variable in Bivariate VAR
χ^2 statistic (marginal significance levels in parentheses)[a]

VAR: Per worker output growth rates (relative to US); investment share of GDP (relative to US); includes a constant in each equation.

Sample: 116 economies, $1963 + m$ through 1985, where m is the VAR lag length.

Left hand Variable	VAR lag length					
	2		3		4	
Growth	11.1	*	17.5	*	26.7	*
	4.5	(0.10)	9.3	(0.02)	18.0	*
Investment	60.2	*	60.3	*	75.0	*
	24.1	*	18.9	*	26.2	*

[a] The first row for each left-hand side variable gives the χ^2 statistic (and implied marginal significance level) for the exclusion test using the standard OLS estimated covariance matrix. The second row uses White's heteroskedasticity-consistent covariance matrix estimator. An entry $*$ (in place of the marginal significance level) indicates a value less than 0.005.

Table 2: Conditioning Regressions (Two-sided Projections)[a]

Dependent Variable: per worker output growth rates (percentage, relative to US).

Sample: 116 economies, 1963–1985 (truncating time dimension for leads and lags).

Conditioning: investment share of GDP (percentage, relative to US).

Investment		Coefficients in Two-sided Projections					
Lead	4					−0.39	(0.7/0.8)
	3			1.47	(0.6/0.8)	1.89	(0.8/0.9)
	2	−0.71	(0.6/0.9)	−1.61	(0.8/1.0)	−1.42	(0.8/1.1)
	1	4.93	(0.8/1.4)	4.62	(0.8/1.5)	4.55	(0.9/1.6)
	0	6.90	(0.8/1.4)	6.90	(0.8/1.4)	7.35	(0.9/1.5)
Lag	1	−6.13	(0.8/1.2)	−6.00	(0.9/1.2)	−6.09	(0.9/1.4)
	2	−3.36	(0.6/0.8)	−2.87	(0.9/1.1)	−2.88	(1.0/1.2)
	3			−1.16	(0.7/0.8)	−0.37	(1.0/1.3)
	4					−1.07	(0.7/1.1)
Constant		−0.61	(0.3/0.4)	0.00	(0.3/0.4)	−0.16	(0.3/0.4)
Sum of Coeffs.[b]		1.63		1.25		1.57	
R^2		0.15		0.15		0.17	

[a] Each column reports estimates of the coefficients in the projection. Numbers in parentheses are first the OLS and then White heteroskedasticity-consistent standard errors. It is possible to calculate serial correlation-robust standard errors as well; I have not done so here only because first, these projections aren't of interest in themselves, and second, per capita worker growth rates are already close to being serially uncorrelated. See, e.g., the working paper version of Quah (1993a) (to save space there, the published version excluded results on growth rates).

[b] Sum of coefficients on (leads and lags of) investment.

Figure 1: Technology

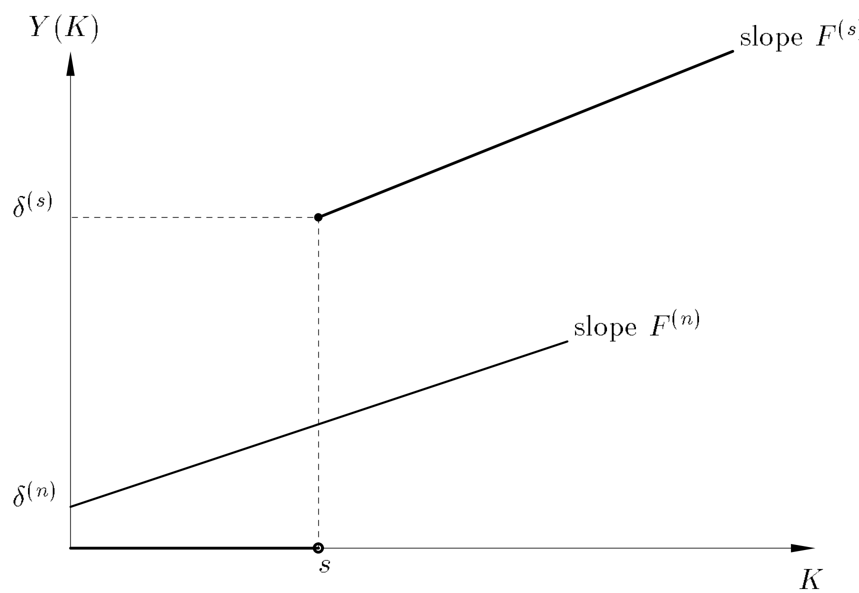


Figure 2: Opportunity set

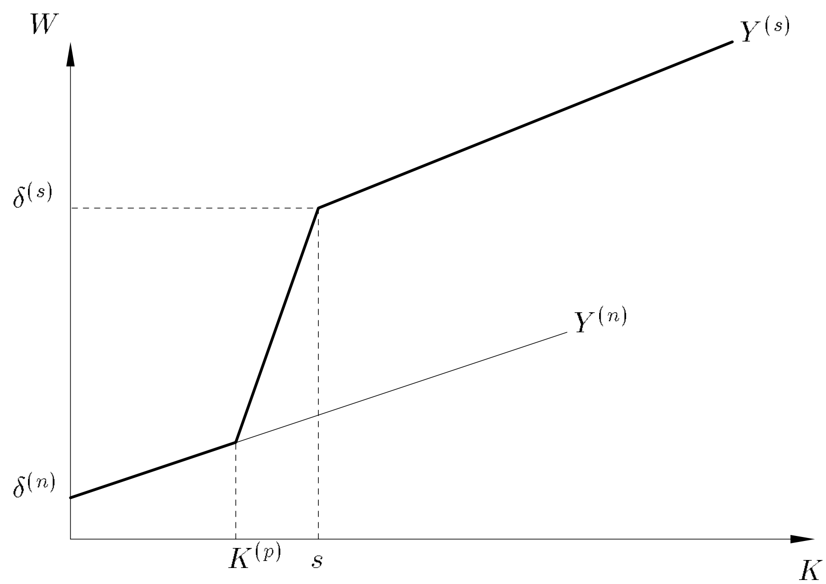


Figure 3: Dynamics

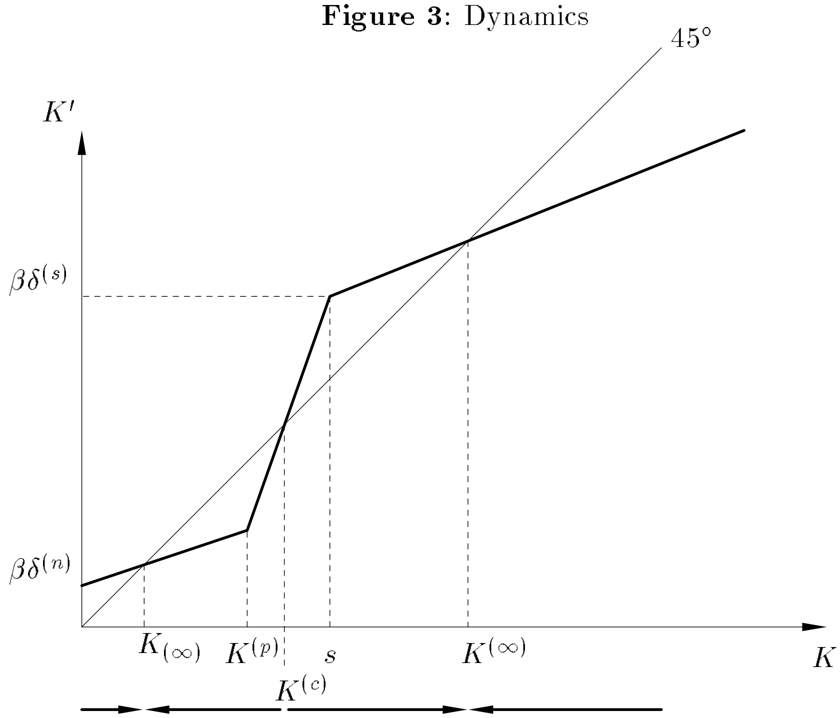


Figure 4: Evolving distributions, tending towards bimodal

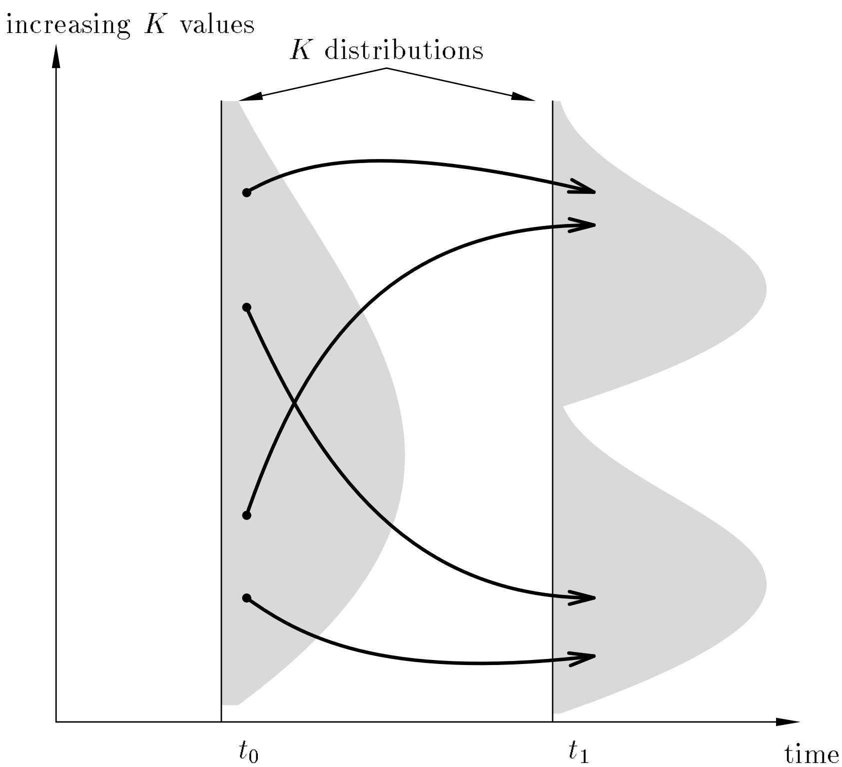


Figure 5(a): Stochastic Kernel, 3d plot
Dynamics of the Cross-country distribution of Output Per Worker
15-year transitions

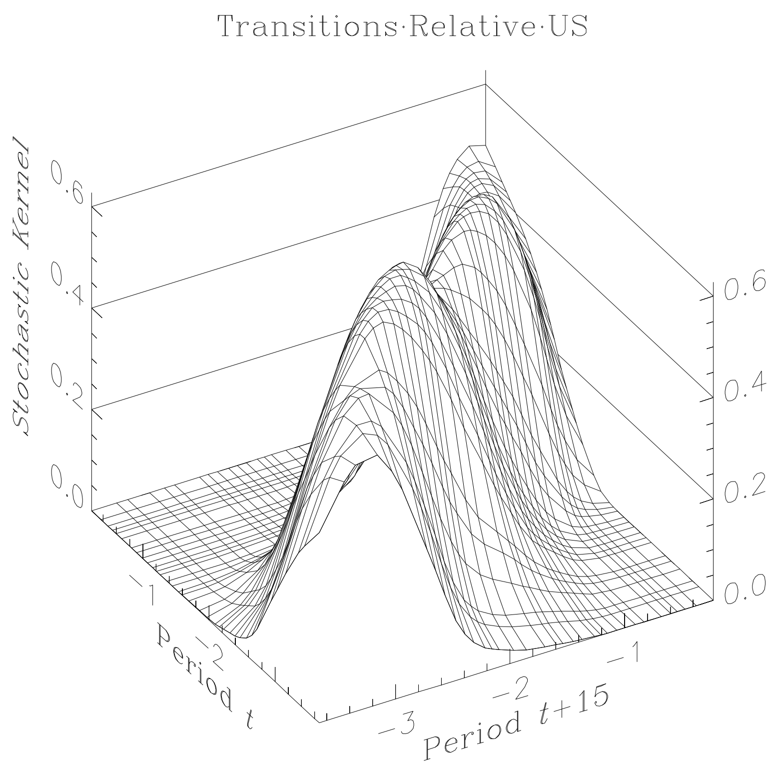
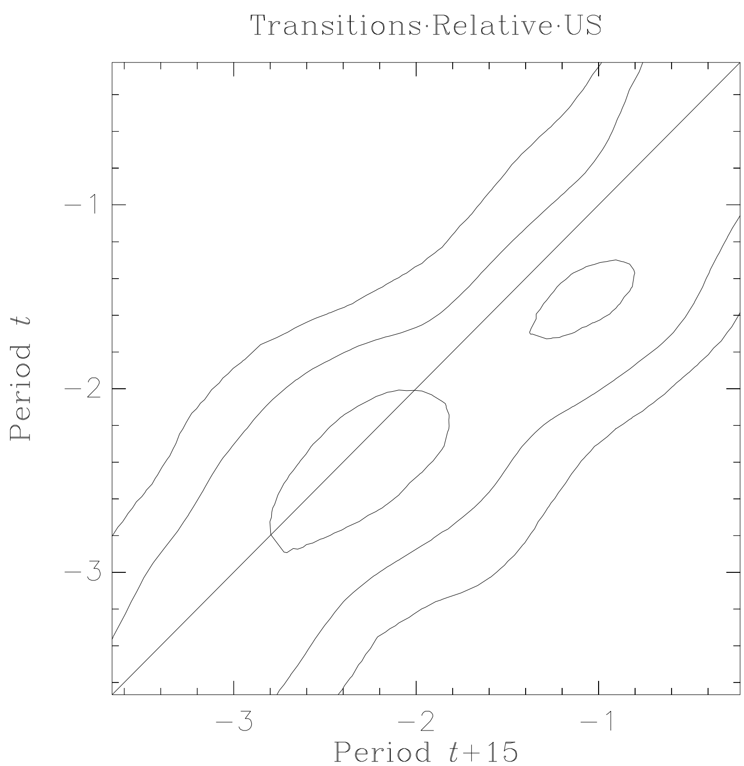


Figure 5(b): Stochastic Kernel, Contour plot
Dynamics of the Cross-country distribution of Output Per Worker
15-year transitions



Probability contours at 0.1, 0.3, 0.55

Figure 6(a): Stochastic Kernel, 3d plot
Dynamics of the Cross-country distribution of Output Per Worker
Unexplained by K, H, and D
15-year transitions

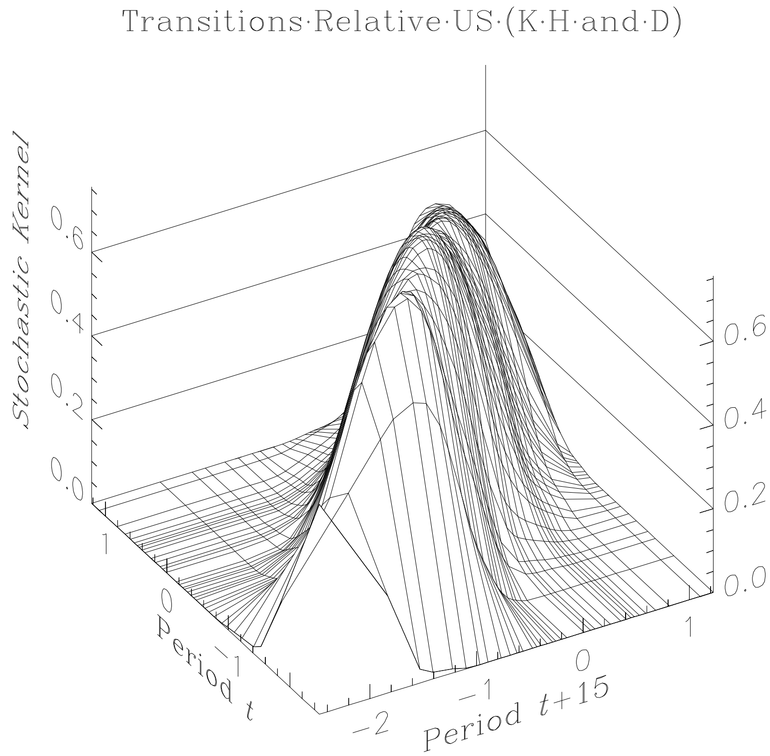
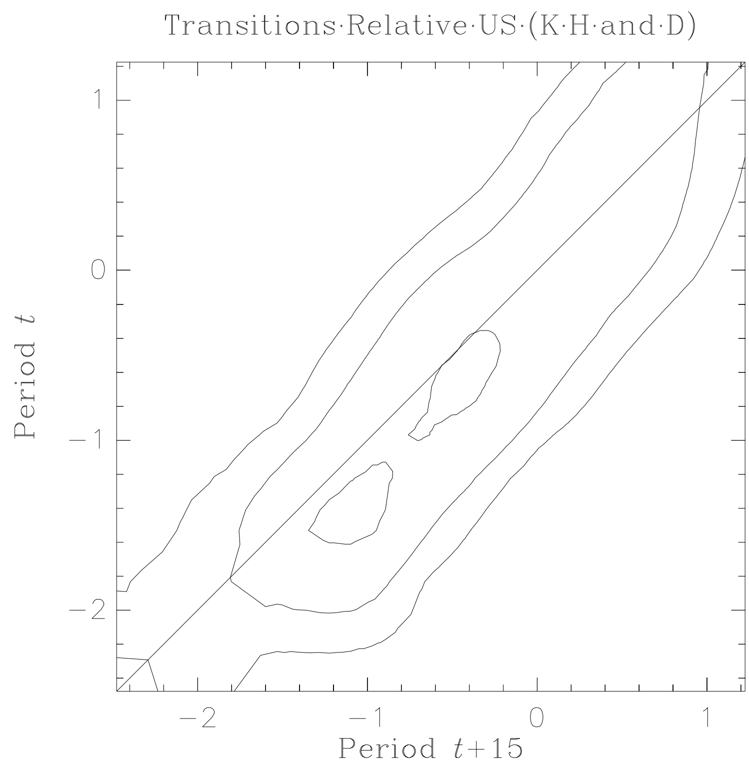


Figure 6(b): Stochastic Kernel, Contour plot
Dynamics of the Cross-country distribution of Output Per Worker
Unexplained by K, H, and D
15-year transitions



Probability contours at 0.2, 0.4, 0.7