International Patterns of Growth:
II. Persistence, Path Dependence, and Sustained Take-Off in Growth Transition

by

Danny Quah*
LSE Economics Department
October 1992

* I thank for helpful comments seminar participants at an LSE/FMG Conference, the NBER Summer Institute, and the Institute for Empirical Macroeconomics.

International Patterns of Growth:
II. Persistence, Path Dependence, and Sustained Take-Off
in Growth Transition

by

Danny Quah
LSE Economics Department

October 1992

ABSTRACT

This paper studies transition dynamics in the cross section of world economies. I formalize notions of persistence, path dependence, and sustained take-off precisely enough to allow their analysis in data that have significant cross-section and time-series variation.

Keywords: polarization, stochastic kernel, stratification

JEL Classification: C23, C41, D31, O47

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

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

[(URL) http://econ.lse.ac.uk/~dquah/]

1. Introduction

Do initially poor countries tend to become as wealthy as initially richer ones? Whether or not countries converge in incomes and productivity levels has recently been intensively studied, both theoretically and empirically: as Lucas [1988] and Romer [1989] have emphasized, this has important implications for economists' views on growth, development, and macroeconomic policy. Following recent terminology, we call a positive answer to this question the *convergence hypothesis*. A number of interesting theoretical ideas surrounding the convergence hypothesis have been formulated to explain its potential failure. This paper makes two main points: first, previous tests for the convergence hypothesis are unable to reveal the effects of interest; second, the data, properly considered, confirm the convergence hypothesis in growth rates.

Whether or not there is convergence would suggest where next to turn to understand economic growth. As an economic proposition, it suggests a characterization of the behavior of a broad cross-section of economies over an extended period of time. As a measurement (i.e., econometric) issue, it suggests statistical study of as wide and as long a data panel as possible. (Just as economists ask, Are stock markets efficient?, not Were stock markets efficient between July and August 1987?, the convergence hypothesis should be viewed not to ask whether a particular country, say Indonesia, is catching up with the richest countries, but rather to ask whether, in a broad cross-section of countries, there are discernible patterns of convergence and divergence over time.) This paper takes seriously these observations in its approach to analyzing the income dynamics of an extensive cross-section of countries. It seeks to establish new stylized facts on the convergence hypothesis in a form that addresses simultaneously the dynamic and cross-section predictions of different theoretical models. Put another way, the empirical work here is motivated by the fact that recent economic growth theories generate interesting testable predictions along both dynamic and cross-section dimensions in the data; thus, variation along both dimensions should be studied together. Previous empirical work has focused on analyzing only dynamic or cross-section variation in isolation. This paper provides econometric tools and empirical results that go one step further.

The remainder of this paper is organized as follows. Section 2 makes explicit some difficulties inherent in empirical analysis of the convergence hypothesis. Section 3 then establishes a framework to overcome those problems. Using the Summers-Heston data, Section 4 discusses some preliminary empirical results within this framework; this section also points out difficulties in interpreting these findings. Section 5 provides a way around those difficulties and gives formal structure to notions of persistence, path dependence, and sustained take-off in data where there is significant cross-section and time-series variation. The empirical findings from implementing those ideas are discussed in Section 6, and Section 7 concludes.

2. Empirical Analysis of the Convergence Hypothesis

The Wold decomposition is a convenient empirical framework for studying the dynamics of a scalar stochastic process. Thus, Beveridge and Nelson [1981], Campbell and Mankiw [1987], Cochrane [1988], Nelson and Plosser [1982], and Watson [1986] have usefully exploited this representation to address interesting issues about the time series properties of aggregate output. When a variable under study is related to other observable variables, a multivariate Wold decomposition (vector autoregression) can be used. These techniques are well-understood and their usefulness and failings well-known. The convergence hypothesis, however, poses new problems on how best to analyze the relevant data. We will see that standard methods are not particularly revealing for the questions of interest in this work.

The convergence hypothesis can be formalized in at least four different ways:

(a) Countries originally richer (faster growing) than average are more likely to turn below average eventually, and vice versa—the cycle repeats.

(b) Whether a country income is eventually above or below the cross-section average is independent of that economy's original position.

(c) Income disparities between countries are not persistent, i.e., disparities across countries have neither unit roots nor diverging deterministic time trends.
(d) Each country eventually becomes as rich as all the others—the cross-section dispersion diminishes over time.

Each of these descriptions involves—in a fundamental way—comparison of cross-section and time series behavior. Notice that statements like "rich economies tend to grow faster"—or their contrary—do not distinguish among (a)-(d); nevertheless, for empirical work we should bear in mind that these four characterizations are logically and empirically distinct. The main body of this paper below will address (a) and (b). Another paper (Quah [1992a]) considers (c) since that involves quite different econometric issues. I now briefly discuss (d); more precise and extensive statements of some of the ideas here can be found in Quah [1992b].

To assess the validity of (d), one might first examine the cross-section regression of average growth rates on initial income levels. It is natural to think that a negative regression coefficient here indicates convergence in the sense of (d), whereas a positive regression coefficient indicates the opposite. However, such a regression suffers from Galton's fallacy (Quah [1992b]) and, in fact, sheds no light on (d)-convergence: A non-degenerate time-invariant distribution of cross-country incomes is consistent with arbitrary signs on those regression coefficients, depending only on the univariate time-series properties of individual country incomes.

In response to this last observation, one might ask instead whether any measure of cross-section disperson is declining over time. It is easy, however, to construct examples where this is misleading. For instance, suppose the world is comprised of only two economies. Assume the first economy has income process Y_1 described by $Y_{1,t} = 1$ if t is odd whereas $Y_{1,t} = 2$ if t is even. Let the second economy have income process Y_2 given by $Y_{2,t} = Y_{1,t-1}$. Then no measure of cross-section dispersion shows decline over time whereas, from identifying the different economies in the cross-section, it is also clear that no one economy is permanently richer than another. In other words, studying just a time-series of cross-section dispersion doesn't reveal how much "mixing" or "turnover" occurs in the cross-section over time. But it is precisely this mixing property that ought to shed light on the convergence hypothesis.

As another alternative, one might simply assert that—regardless of cross-section dispersion behavior—the validity of the convergence hypothesis can be understood in terms of the univariate dynamics of individual country incomes. Thus, one might ask if, within each country, income is an integrated process. Riezman and Whiteman [1989] have taken this approach. In that work, cross-section information might be used to improve precision of the estimates, but the underlying probability model and its economic interpretation rely in no essential way on the cross-section income variation. A moment's reflection indicates then that whether or not countries are approaching each other (according to any criterion) cannot be ascertained by studying only their univariate dynamics. For instance, each country income could be a random walk: if, however, those incomes were equal across countries, then, in every sense, the convergence hypothesis holds despite income in each country following an integrated process. More generally, country incomes could be individually integrated but, at the same time, be jointly cointegrated across countries. Disparities between countries are then not persistent, and interpreting the unit root in each country income as evidence against the convergence hypothesis would be incorrect. This joint cointegration characterization—while a valid way to examine the convergence hypothesis—is not implementable in practice. In the typical data sets used in this work, there are more observation units in the cross-section dimension than there are time series observations. A vector time series model is not useful for studying dynamics in such structures. (Bernard and Durlauf [1990] have, however, used exactly this cointegration idea but by focusing only on small subsets of countries.)

We conclude from the discussion above that convergence cannot be usefully studied without simultaneously considering both cross-section and dynamic behavior. Consequently, this falls outside the standard framework of multivariate econometric analysis: Time series methods treat the case when the number of cross-section observation units is small relative to the time length (e.g., aggregate consumption and income observed over the post-War period); by contrast, panel data methods consider the case when the number of time points is small (taking the cross-section at each time point to follow a different equation). In the current work, dynamics is an important aspect of the analysis, but so is cross-section variation; the models of interest here imply restrictions on both dynamic as well as cross-country properties of the data.

3. Distribution Dynamics

This section establishes the mathematical structure for the empirical work to follow. Although the discussion initially borrows elements from stochastic process and Markov chain theory (as described in Futia [1982], Stokey and Lucas [1989], and elsewhere), the interpretations will differ significantly. At that level of generality, however, little can be done in empirical analysis. The subsequent econometric implementation therefore uses simplifications inspired by, among others, the work of Singer and Spilerman [1976].

We model the evolution over time of the empirical distribution of incomes (or income growth rates), observed as realizations of a random element in a space of distributions. (To omit needless words, we will simply say *incomes* to indicate both per-capita incomes and income growth rates when no ambiguity is possible. The empirical analysis below will consider these two cases separately as well as the transition behavior of country ranks (**not done yet**).) In doing this, recall we wish to record not just aggregative statistics—such as means and variances—of the distributions, but also information on the mobility of individual economies within the distribution.

Let (S, S) denote a measurable space, taken to be the underlying state space. In the general discussion below, this can be thought of as (R, R), i.e., the real line R together with the collection R of its Borel sets: S is the collection of points in which measured incomes fall. Let $\mathbf{B}(\mathsf{S}, \mathsf{S})$ denote the Banach space of bounded finitely additive set functions on (S, S), endowed with total variation norm:

$$\forall \mu \text{ in } \mathbf{B}(\mathsf{S}, \mathsf{S}): \qquad |\mu| = \sup \sum_j \mu(A_j),$$

with the supremum taken over all finite partitions of S. Cumulative distribution functions on S can be identified with probability measures on (S, S) which, in turn, are simply countably additive elements in $\mathbf{B}(\mathsf{S}, \mathsf{S})$ that assign value 1 to the entire space S. Let B denote the (Borel) σ-algebra generated by the open subsets (relative to norm topology) of $\mathbf{B}(\mathsf{S}, \mathsf{S})$, so that $(\mathbf{B}, \mathsf{B})$ is a measurable space.

In our application, with $\mathsf{S} = \mathsf{R}$, a data point at time t is the period t-observed cross-section empirical distribution F_t of per capita incomes Y_{jt}:

$$\forall y \text{ in } \mathsf{R}: \qquad F_t(y) = \text{ fraction of countries } j \text{ such that } Y_{jt} \leq y.$$

Let $(\Omega, \mathsf{F}, \Pr)$ denote the underlying probability space. We model that measure associated with the cdf F_t as a realization of some F/B-measurable map (random element) $\phi_t : (\Omega, \mathsf{F}) \to (\mathbf{B}, \mathsf{B})$. In the sequel, we study only ϕ_t's taking values that are probability measures in $\mathbf{B}$ (with Pr-probability 1).

The econometric problem now is to quantify the dynamic properties of the $\mathbf{B}$-valued stochastic process $\{\phi_t,\ t \geq 1\}$. Further, recall that we wish to do this in a way that records transitions across different portions of the underlying state space S. The construct of stochastic kernels serves precisely this purpose.

DEFINITION 3.1: *Let λ and μ be elements of $\mathbf{B}$ that are probability measures on (S, S). A* **stochastic kernel** *is a mapping $M_{(\lambda,\mu)} : \mathsf{S} \times \mathsf{S} \to [0, 1]$ satisfying:*

(i) $\forall y$ in S, $M_{(\lambda,\mu)}(y, \cdot)$ is a probability measure;
(ii) $\forall A$ in S, $M_{(\lambda,\mu)}(\cdot, A)$ is S-measurable;
(iii) $\forall A$ in S, $\lambda(A) = \int M_{(\lambda,\mu)}(y, A)\, d\mu(y)$.

To see why this is useful, first consider condition (iii). In an initial period, for given y, there is some fraction $\mu(dy)$ of countries with incomes close to y. Count up all countries in that group who turn out to have their incomes subsequently fall in a given S-measurable subset $A \subseteq \mathsf{S}$. When normalized to be a fraction of the total number of countries, this count is precisely $M(y, A)$ (where the (λ, μ) subscript can be omitted without loss of clarity). Fix A, weight the count $M(y, A)$ by $\mu(dy)$, and sum over all possible y's, i.e., evaluate the integral $\int M(y, A)\, d\mu(y)$. This gives the fraction of countries that end up in state A regardless of their initial income levels. If this equals $\lambda(A)$ for all (measurable) subsets A, then λ must be the measure associated with the subsequent income distribution. In other words, the stochastic kernel M is a complete description of transitions from state y to any other portion of the underlying state space S.

Conditions (i) and (ii) then simply guarantee that this interpretation is valid. By (ii), the right-hand-side of (iii) is well-defined as a Lebesgue integral. By (i), the right-hand-side of (iii) is a weighted average of probability measures $M_{(\lambda,\mu)}(y, \cdot)$, and therefore is itself a probability measure.

This stochastic kernel framework should recall analyses of the stochastic dynamics of a single time series. There, the measures λ and μ would be associated with the (unobserved) cdf of an observed variable in any given time period. In the current work, however, we are using this framework to model not just a single time-series but an entire collection of possibly heterogeneous and correlated variables. Thus, here, λ and μ refer to the actual observed distributions of the incomes of a cross-section of economies. Despite this substantive difference, however, the two views are formally identical, as we can always construct a hypothetical "representative economy" such that its income process has conditional distributions (over time) that match the actual cross-section distributions (across different economies, over time). We will use whichever language is more convenient in different parts of what follows.

To summarize, understanding the dynamics of the cross-section distribution of country incomes is equivalent to understanding the structure of the stochastic kernel M. The aim of the econometric exercise here is to uncover and interpret such structure.

There are three immediate difficulties in the analysis: first, M is infinite-dimensional and cannot be directly estimated with only finite data. Second, the formulation up until now describes a deterministic evolution of one cross-section distribution to another. There is thus no notion of randomness or error and so an econometric model is not yet explicit. Third, and closely related to the second, there is no guarantee that a single kernel M will apply to all time period transitions, $\phi_t \rightarrow \phi_{t+1}$; in general, there won't exist such a single time-invariant stochastic kernel. To repair these deficiencies, we develop some additional concepts and specialize the framework. (**Note:** This currently promises more than I've been able to deliver. More work needed.)

Let $\mathbf{b}(\mathsf{S}, \mathsf{S})$ be the Banach space under sup norm of bounded measurable functions on (S, S). Fix a stochastic kernel M and define the operator T mapping $\mathbf{b}(\mathsf{S}, \mathsf{S})$ to itself by

$$\forall f \text{ in } \mathbf{b}(\mathsf{S}, \mathsf{S}),\ \forall y \text{ in } \mathsf{S}: \qquad (Tf)(y) = \int f(x)\, M(y, dx).$$

Since $M(y, \cdot)$ is a probability measure, the image Tf can be interpreted as a forward conditional expectation. For example, if all countries in the cross-section begin with incomes y, and we take f to be the identity map (ignoring unboundedness), then $(Tf)(y) = \int x\, M(y, dx)$ is next period's average income in the cross-section, given that all countries had income y in the current period. Clearly, T is a bounded linear operator. Denote the adjoint of T by T^*. By Riesz representation, the dual space of $\mathbf{b}(\mathsf{S}, \mathsf{S})$ is just $\mathbf{B}(\mathsf{S}, \mathsf{S})$ (our original collection of bounded finitely additive set functions on S); thus, the adjoint T^* is a bounded linear operator mapping $\mathbf{B}(\mathsf{S}, \mathsf{S})$ to itself. It turns out that T^* is also exactly the mapping in (iii) of **3.1**, i.e.,

$$\forall \mu \text{ in } \mathbf{B}: \qquad (T^*\mu)(A) = \int M(y, A)\, d\mu(y).$$

(This follows from writing the left hand side as

$$(T^*\mu)(A) = \int \mathbf{1}_A(y)(T^*\mu)(dy) = \int (T\mathbf{1}_A)(y)\, d\mu(y)$$
$$= \int \left[\int \mathbf{1}_A(x) M(y,dx)\right] d\mu(y) = \int M(y,A)\, dy,$$

with $\mathbf{1}_A$ the indicator function for A.) Thus, instead of studying the stochastic kernel M directly, we can investigate the structure in T^*, the adjoint operator to the forward conditional expectation.

To simplify further, assume a countable state space $\mathsf{S} = \{S_1, S_2, S_3, \ldots\}$ for income levels, and a finite state space $\mathsf{S} = \{S_1, S_2, \ldots, S_r\}$ for income growth rates. The latter implies that $\mathbf{B}(\mathsf{S}, \mathsf{S}) = \mathbf{b}(\mathsf{S}, \mathsf{S}) = \mathsf{R}^r$ so that M is just an $r \times r$ stochastic matrix (i.e., a matrix with all elements non-negative and having row sums equal to 1). From the analogue of (iii) in **3.1**, one element of $\mathbf{B}$ is transformed into another by:

$$\lambda, \mu \in \mathbf{B}: \qquad \lambda = M'_{(\lambda,\mu)}\mu.$$

Further, for f in $\mathbf{b}$, we have $\mu f = f \cdot \mu = \sum_{k=1}^{r} f_k \mu_k = \mu \cdot f$. But then the operator T satisfies:

$$(Tf)(S_k) = [M]_{(k\text{–th row})} f \text{ for all } k = 1, 2, 3, \ldots \iff Tf = Mf \text{ or } T = M,$$

while the adjoint T^* obeys:

$$\lambda = T^*\mu = M'\mu \implies T^* = M'.$$

We recognize M to be simply a "transition probability matrix" in this case.[1] All relevant information on turnover in the distribution is therefore embedded in a finite matrix M.

---

[1] Recall again the caveat above on interpreting an evolving cross-section distribution in the language of scalar stochastic processes.

4. Preliminary Empirical Results

The discussion of Sections 2 and 3 suggests that transition probabilities are a useful way to understand the convergence hypothesis. Table 1 presents some estimated transition probability matrices for country growth rates. The underlying data have been drawn from Summers and Heston [1988]: First, I excluded the principal oil-exporting countries: Oman, the United Arab Emirates, Iran, Iraq, Saudi Arabia, Kuwait, Yemen, and Venezuela. Data for Afghanistan are unavailable from 1979, data for Indonesia are unavailable before 1962; thus, those economies were also excluded. I then constructed a series for world per capita income from data on population and real per capita GDP (measured in 1980 international prices) for the remaining 109 countries. Individual income disparities from this series were then taken to be the primitive data used to analyze the convergence hypothesis. Because data for many countries (particularly the lesser-developed economies) are not available prior to 1960, that year was taken to be the starting point for the subsequent analysis; the data extend through 1985.

I consider 3- and 4-state discretizations of the state space for income growth rates. The 3-state case uses cut-off points at -0.02 and 0.02 (corresponding to -2% and 2% annual growth): that is, growth rates are partitioned into three classes: less than -0.02 (S_1), between -0.02 and 0.02 (S_2), and greater than 0.02 (S_3). For the beginning of the sample, i.e., for growth from 1960 through 1961, this division gives 42 economies in S_1, 35 in S_2, and 32 in S_3, respectively; out of the total number of 109×25 growth observations over the entire sample, this division gives approximately one-third in each state. A similar partition is used with the 4-state model with the cut-off points taken to be -0.03, 0, and 0.03: for the beginning of the sample, this gives 34 economies in the lowest growth state S_1, 32 in S_2, 21 in S_3, and 22 in S_4. Of the total collection of growth observations over the entire sample, approximately one-fourth falls in each of the states.

Table 1 presents estimates of the unrestricted transition probabilities underlying T_t for t equal to 1, 13, and 24. In words, these describe the actual changes in observed growth rates for the entire cross-section of economies from 1960/61 up through 1961/62 (for T_1), from 1960/61 up through 1973/74 (T_{13}), and from 1960/61 up through 1984/85 (T_{24}). Considering the 3-state model, for instance, we see that over the one-year horizon, 33% of those economies that began with S_1 low growth transited to S_3 high growth in 1961/62. Over the 24-year horizon, only 17% of the same low-growth economies made that same transition. In both 3-state and 4-state models, the one-year horizon transition behavior shows a tendency for economies to remain in their original states—the diagonals are uniformly the highest entries in each row. Over the 24-year horizon, low- and medium-growth economies in each of the models show similar persistence. Notice, however, that over the one-year horizon there is nonetheless significant mobility across low- and high-growth states in both T_1's—the off-diagonal entries do not vanish.

Instead of calculating these point-in-time t-period transitions, a researcher might estimate transition probabilities by averaging over time. For example, the one-year transition probability might be estimated as an average over 24 one-year transition periods. This will give more precise estimates of transition probabilities—if one assumes the 24 different transitions were simply the same object observed with noise. In this work, however, it is precisely the possible non-stationarity in the cross-section distribution that is of interest. Therefore, we don't use any such averaging in the empirical analysis.

In contrast to T_1, for both models, the 13-year horizon transition behavior shows a marked tendency for economies to accumulate in the high-growth cell. Since this end point is 1973/74, this may at first seem a little surprising. A number of large economies did experience declines in per capita incomes then—notably India, the United Kingdom, and the United States—thus reducing world per capita income. However, many countries actually displayed positive growth over this year, among them Austria, Belgium, Brazil, Canada, Finland, France, West Germany, Iceland, Italy, Korea, Malaysia, Singapore, Sweden, and Switzerland. The outstanding feature of T_{13} then is not the dominant diagonal, but rather a tendency for high growth to be an attracting state.

Finally, for both models, T_{24} suggests opposite tendencies to those displayed in T_{13}: over this long horizon, relatively more clustering occurs in the low- and zero-growth states, regardless of the initial growth state. In this limited sense then, there is an apparent convergence property at work: regardless of initial conditions, economies are predominantly tending towards low growth over a long horizon.

It is clear, however, that the estimates in Table 1 should be interpreted with caution. Without imposing structure on the sequence $\{ T_t \mid t \geq 1 \}$—in particular, without somehow relating T_t's across different t horizons—we cannot infer any of the systematic properties of the data. The piling up observed in T_{13}'s high-growth columns could indicate a tendency for all economies eventually to grow rapidly, regardless of initial conditions, or it could simply indicate a transitory upwards movement in growth rates that happened to coincide across countries at that particular 13-year horizon. The next section therefore investigates a number of different parametrizations for T-sequences designed to get at the underlying structure in these transition probabilities.

5. Alternative T_t Structures

We turn to interpreting the transition kernels $\{M_t \stackrel{\text{def}}{=} M_{\phi_0,\phi_t},\ t \geq 1\}$ and their associated $\{T^*_t,\ t \geq 1\}$ and $\{T_t,\ t \geq 1\}$. From the discussion of Section 3, we are free to place structure on any of M, T, or T^*; restrictions on any one have implications on the other two that are easily worked out. It turns out to be convenient to work with T_t: recall that as described in the previous section, this has the interpretation of a t-period forward conditional expectation. While the discussion can again be formulated in terms of infinite-dimensional operators, nothing essential is lost if we now specialize to the finite case.

We now develop some empirically tractable stochastic kernels. The particular kernels we consider are directed at the following issues:

(i) What kind of heterogeneity of income dynamics is there in the cross-section?

(ii) Does low growth (income) imply a tendency for continued low growth (income) subsequently?

(iii) Do economies "age" in interesting ways? More concretely, relative to some starting point, does the cross-section of economies settle into an immobile growth pattern? Or, does the opposite occur, where an initially stagnant panel of economies increases in cross-section income mobility over time?

(iv) Does a steady state distribution of income levels or growth rates exist? And if so, what is the shape of that long run distribution? Does the steady-state distribution display interesting modal patterns, as predicted, for example, by the growth models of King and Robson [1989] and Tamura [1989]? Growth models typically imply restrictions on long-run dynamics, and so it is precisely these properties that are of interest.

Before beginning the discussion, it is important to note Heckman's criticisms (1981) of earlier work by Singer and Spilerman (1976), related to that reported here. Heckman was interested in transition across employment-unemployment states: in that situation, it is useful to model the evolution of each unit (a worker) in the distribution, possibly by a regression equation. Estimation then exploits the cross-section variation in the distribution, and that empirical investigation allows us to understand the factors underlying movement in and out of unemployment. The focus here, by contrast, is on the evolution of an entire distribution, and not on the behavior of any particular unit (a country) in the distribution. In addition, much of Heckman's criticisms regarding other methods of identifying heterogeneity and spurious state dependence stem from the view that, up to conditioning on observable exogenous variables, regression residuals should be iid across observations. In macroeconometric work, such a possibility is much less attractive—one, whether useful exogenous variables can be found is doubtful, and two, even if they existed, it is difficult to imagine using all the possible exogenous variables in a regression to induce uncorrelated conditional expectations errors across the individual country observations.[2] By modeling instead the entire empirical distribution

[2] One could, of course, construct such a conditional expectation but, in implementation, that construction would involve estimating as many parameters as

of observations, we don't impose or require strict independence across countries. The cross-section empirical distribution *becomes* the basic unit of observation.

In the following, let P and Q denote square stochastic matrices, not necessarily equal throughout. We will take all arrays denoted by Q to have only zeroes on the diagonal: thus Q describes transition behavior with zero conditional probability of remaining in a given state for longer than one period. To begin, consider the following Markov structure on T_t:

$$\text{For all } t \geq 1, \quad T_t = P^t. \tag{5.1}$$

This imposes stationarity on the transition dynamics so that the aggregate behavior (in conditional expectations) between periods t and $t+1$ is the same as that between $t-1$ and t. A single time-invariant stochastic kernel P regulates behavior over all single-period transitions.

We develop richer and more realistic models by progressively complicating this basic Markov structure. The first such model preserves the time-stationary properties of (5.1); it provides, however, more explicitly interpretable structure to the kernel P:

$$\text{For all } t \geq 1, \quad T_t = \exp(t\lambda[Q - I]), \qquad \lambda > 0. \tag{5.2}$$

For any positive number λ and stochastic matrix Q, there exists a model (5.1) representing that model (5.2): just take $P = \exp(\lambda[\,Q - I\,])$. It is not true, though, that every P in (5.1) can be written in the form (5.2). (This is known as the *embedding problem*; it is discussed in Geweke, Marshall, and Zarkin [1986], Singer and Spilerman [1976], and elsewhere.) Further, without the zero-diagonal restriction on Q, more than one (λ, Q) pair will generate the same T_t. To see this, notice that (λ_1, Q_1) and (λ_2, Q_2) are observationally equivalent exactly when

$$Q_2 = I + (\lambda_1/\lambda_2)[Q_1 - I]. \tag{5.3}$$

---

there are observations.

Suppose the Q arrays were not restricted to have zeroes on the diagonals but instead were only required to be stochastic matrices. Fix (λ_1, Q_1); for arbitrary $\lambda_2 \neq \lambda_1$, a suitable $Q_2 \neq Q_1$ can always be found, using (5.3), such that (λ_1, Q_1) and (λ_2, Q_2) are observationally equivalent. It is easy to verify that any Q_2 constructed from (5.3) is always a stochastic matrix. But when (5.3) holds and Q_1 is a stochastic matrix with zeroes on the diagonal, then Q_2 satisfies the same restrictions if and only if $\lambda_2 = \lambda_1$, the last implying Q_2 is the same as Q_1.

The argument above, however, does clarify that requiring Q to have zeroes on the diagonals is not necessary to identify (λ, Q) pairs: fixing the diagonal entries at *arbitrary* values in $[0, 1]$ would do just as well. The motivation for our identifying assumption derives instead from interpreting (5.2) in terms of *movers* and *stayers*.[3]

We now turn to such an interpretation. Recall the discussion of section 3: although we are interested in a cross-section of different economies, because we have the entire distribution of income characteristics, we can construct a fictitious *representative economy* whose random (unobserved) income has a distribution that matches exactly the observed cross-section distribution. Now suppose that that representative economy remains in a given state for a random waiting time $\tau > 0$. A *move exposure (event)* occurs after this time, whereupon the economy transits across states in a manner described by the (conditional expectations) operator Q. Since Q has zeroes on its diagonal, there is no ambiguity about the occurrence of a move exposure. Otherwise, a move exposure event could occur, but because there is always positive conditional probability of remaining in any given state, the occurrence of the event is indistinguishable from its non-occurrence. In this interpretation, therefore, it is natural to restrict Q to have all zero diagonal entries.

Next, assume move exposures arrive as a stationary Poisson process with parameter $\lambda > 0$ so that in any time interval of length t, the number of X move

---

[3] This terminology is from the labor economics and sociology literature on mobility (see e.g. Spilerman [1974]). The zero diagonal restriction either is not made explicit in that literature or instead obtains from formulating the model in continuous time. We do not use continuous time reasoning in the current work.

exposure events follows:

$$\Pr_t(X = k) = e^{-\lambda t}(\lambda t)^k/k!, \quad k = 0, 1, 2, \ldots .$$

From this, the waiting time τ is that random length of time until the first event and is distributed exponential (λ). This distribution has mean λ^{-1}, so that the larger is λ, the faster (on average) do move exposures arrive, and thus, other things equal, the greater is the cross-section mobility.[4] From these assumptions, average transition behavior over the time interval $[\,0, t\,]$ satisfies:

$$T_t = \sum_{k=0}^{\infty} \Pr_t(X = k)\, Q^k = e^{-\lambda t} \sum_{k=0}^{\infty} (\lambda t Q)^k/k! = \exp(t\lambda[\,Q - I\,]),$$

which is precisely (5.2).

The models below further generalize (5.2) by allowing different notions of heterogeneity, state and duration dependence, and aging; for tractability, we do this by focusing on the parameter λ, the arrival rate for move exposures.

Our first modification replaces the hypothesis of a constant λ by assuming instead some non-degenerate distribution G of λ's in the cross-section of economies. Then aggregate dynamic behavior satisfies:

$$\text{For all } t \geq 1, \quad T_t = \int \exp(t\lambda[\,Q - I\,])\, dG(\lambda). \tag{5.4}$$

In the empirical work, it is convenient to parametrize the heterogeneity by choosing G to be a gamma distribution with (unknown) parameters α and β. A straightforward calculation then shows (5.4) to simplify to:

$$\text{For all } t \geq 1, \quad T_t = \left(\frac{\beta}{\beta + t}\right)^{\alpha} \left(I - \frac{t}{\beta + t} Q\right)^{-\alpha}, \qquad \alpha, \beta > 0. \tag{5.5}$$

[4] Note we are using "mobility" here in a sense different from that in the income distribution literature, e.g. Geweke, Marshall, and Zarkin [1986]. In particular, we hold Q invariant in this informal description.

To see this, recall that if G is the gamma (α, β) distribution, it has density given by:

$$\text{pdf}\,(x) = \Gamma(\alpha)^{-1}\beta^{\alpha}x^{\alpha-1}e^{-\beta x} \quad \text{for } x > 0,$$

with Γ denoting Euler's gamma function: $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$. Thus,

$$T_t = \int_0^\infty \exp(t\lambda[\,Q - I\,]) \cdot \text{pdf}\,(\lambda)\,d\lambda = \Gamma(\alpha)^{-1}\beta^{\alpha}\int_0^\infty \lambda^{\alpha-1}\exp(\lambda[tQ - (\beta+t)I\,])\,d\lambda.$$

By analogy with the scalar case, take the change of variable $x = [(\beta + t)I - tQ\,]\lambda$. For (the eigenvalues of) $Q \leq (\beta + t)/t$, which is always satisfied, this becomes

$$T_t = \Gamma(\alpha)^{-1}\beta^{\alpha}[(\beta + t)I - tQ\,]^{-\alpha}\int_0^\infty x^{\alpha-1}e^{-x}\,dx = \left(\frac{\beta}{\beta + t}\right)^{\alpha}\left(I - \frac{t}{\beta + t}Q\right)^{-\alpha},$$

as in (5.5).

Model (5.5) can also be seen to relate to (5.2) in the following way. Since the gamma (α, β) distribution has mean α/β and variance α/β^2, (5.5) approaches (5.2) as α and β grow indefinitely large while $\alpha/\beta \to \lambda > 0$. For finite α and β, however, the cross-section shows positive heterogeneity in the form of differing λ's.

An alternative to assuming heterogeneity that is intrinsic to different individual economies is to assume instead heterogeneity intrinsic to different dynamic states, i.e., suppose the growth dynamics of otherwise identical economies differ if their past growth histories differ. And further, assume this difference in growth dynamics depends systematically on the particular growth path that a given economy had earlier taken. (This model is one formalization of ideas on path dependence that may be found in e.g. Krugman [1989].) This can be captured by parameterizing transitions as:

$$\text{For all } t \geq 1, \quad T_t = \exp\left(t\Lambda[\,Q - I\,]\right), \qquad \Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_r\}, \tag{5.6}$$

so that a different λ_j applies to each state. In words, the move exposure arrival rate varies with the state: smaller λ_j-states display greater persistence in that economies already in that state are slower to leave. Notice, again, that model (5.6) is identified due to the zero restriction on the diagonal entries of Q—for exactly the same reasons described in the discussion surrounding (5.2). By the same token, it follows that (5.6) forms a strict superset of (5.2): (5.6) collapses to (5.2) when $\lambda_j = \lambda_{j+1}$ for all j, but it is easy to make up models of the form (5.6) that cannot be represented as (5.2).

The constant λ assumption can alternatively be relaxed by allowing the move exposure arrival rate to vary over time. The next two models allow for this time-heterogeneity in different ways. First, suppose λ changes over time in a way that allows for "aging":

$$\begin{gathered}\text{For all } t \geq 1, \quad T_t = \exp(\,\rho(t)[\,Q - I\,]), \\ \text{with } \rho(t) = \int_0^t h(u)\,du, \text{ and } h(u) = \frac{g(u)}{1 - G(u)},\end{gathered} \tag{5.7}$$

where G is a cdf and g is its density. Then h in (5.7) can be interpreted as a hazard rate. It is immediate that (5.7) strictly contains the basic model (5.2) as that special case when $h(u) = \lambda$ for all u. Such a constant hazard would follow if G were assumed to be exponential with parameter λ. More general choices for G, however, will imply a non-constant hazard and consequently more interesting time-variation in $\rho(t)$.

To see the significance of this, it is useful again to use the representative economy interpretation: suppose that, for the representative economy, waiting time until a move exposure is distributed G. Given that no move exposure has yet arrived by t, the conditional probability that a move exposure will occur in the next dt units of time is:

$$h(t)\,dt = (1 - G(t))^{-1}\,g(t)\,dt.$$

Thus when h is monotone increasing (say), the conditional probability of a move exposure increases over time: holding Q fixed, this means increasingly greater potential mobility, over time, in the distribution of economies. Conversely, if h is monotone decreasing, potential cross-section mobility falls as time progresses. If h is constant at λ, then $\rho(t) = t\lambda$ and (5.7) is precisely (5.2).

For empirical implementation, it only remains to select waiting time distributions G that are tractable and that imply monotone hazard functions. It is convenient to note here that the integrated hazard function ρ equals the negative of the log of the survivor function $1 - G$. In the work below, we use G distributions that are distributed either gamma$\,(\alpha, \beta)$ or Weibull$\,(\lambda, \kappa)$. Note that while the gamma distribution earlier given under (5.5) described heterogeneity in the parameter λ in a cross-section, that given here describes a waiting time distribution for the representative economy. It is evident from (5.5) and (5.7) that neither contains the other as a special case, although their intersection includes (5.2).

When G is a gamma$\,(\alpha, \beta)$ distribution, $\alpha = 1$ implies that G is exponential with parameter β. It is easy to see that $\alpha > 1$ implies the hazard $h(u)$ increases monotonically from 0 to β as u increases from 0 to ∞; conversely, when $\alpha < 1$, the hazard $h(u)$ decreases monotonically from ∞ to β over the same range in u (see e.g. Cox and Lewis [1966]). Over sufficiently long horizons, therefore, regardless of parameter values, the gamma hazard becomes “close” to a nonzero constant: its behavior becomes marginally indistinguishable from that where waiting times are exponentially distributed.

By contrast, if G is Weibull$\,(\lambda, \kappa)$, $G(t) = 1 - \exp\left(-(t\lambda)^\kappa\right)$, then the implied hazard is $h(t) = \kappa\lambda^\kappa t^{\kappa-1}$. Thus, when $\kappa = 1$, the Weibull waiting time is exactly an exponential random variable, and the hazard is again constant. When $\kappa > 1$, however, the hazard increases monotonically from 0 to ∞; conversely, when $\kappa < 1$, the hazard decreases monotonically from ∞ to 0.

Either gamma or Weibull parametrizations in (5.7) allow considering whether, over time, the world economies are settling into patterns of relative mobility or immobility. No parametrization of G in (5.7), however, can capture a renewal phenomenon, where an economy alters its dynamic characteristics upon exiting a state. Such an effect would be one empirical formalization of “take-off into sustained growth": imagine an economy is trapped in a low growth state because it is currently experiencing low growth. But should that economy somehow exit that low growth state (from a "big push," say), it then acquires greater propensity to continue experiencing high growth. Such "renewal upon transition" properties are captured in the following parametrization:

$$\text{For all } t \geq 1, \quad T_t = \sum_{k=0}^{\infty} \left[G^{(k)}(t) - G^{(k+1)}(t) \right] Q^k, \tag{5.8}$$

where $G^{(k)}$ denotes the k-fold convolution of a cdf G (for $k \geq 1$), and $G^{(0)}(t)$ is defined to equal 1 for all $t > -\infty$, and 0 otherwise. To anticipate the interpretation of (5.8) to follow, notice that if $G^{(k)}(t) - G^{(k+1)}(t)$ is $e^{-\lambda t}(\lambda t)^k/k!$, then (5.8) is simply (5.2). More generally, though, (5.8) can be viewed as follows: Let τ_m denote the waiting time between move exposures $m-1$ and m. Assume $\{\, \tau_m \mid m = 1, 2, 3, \ldots \,\}$ are iid with cdf G; then $G^{(k)}$ is the cdf of the accumulation $\sum_{m=1}^{k} \tau_m$. From this, it follows that $G^{(k)}(t) - G^{(k+1)}(t)$ is the probability that $\sum_{m\geq 1} \tau_m$ first exceeds threshold t on the $(k+1)$-st step. (This can be seen explicitly by the calculation:

$$\begin{aligned} G^{(k)}(t) - G^{(k+1)}(t) &= \Pr\left(\sum_{m=1}^{k} \tau_m \leq t \right) - \Pr\left(\sum_{m=1}^{k+1} \tau_m \leq t \right) \\ &= \Pr\left(\sum_{m=1}^{k} \tau_m \leq t \right) - \Pr\left(\left\{ \sum_{m=1}^{k} \tau_m \leq t \right\} \bigcap \left\{ \sum_{m=1}^{k+1} \tau_m \leq t \right\} \right) \\ &= \Pr\left(\left\{ \sum_{m=1}^{k} \tau_m \leq t \right\} \bigcap \left\{ \sum_{m=1}^{k+1} \tau_m > t \right\} \right) \end{aligned}$$

where the second equality follows from $\tau \geq 0$ a.s., and the third from the relation $\Pr(A \cap B^c) = \Pr(A) - \Pr(A \cap B)$.) Thus, (5.8) describes aggregate transition behavior when mobility possibilities in a particular state are imagined to change systematically over time as in the "aging" in (5.7), but where, upon a transition out of that state, the mobility possibilities evolve as though the past has been forgotten.

For tractability, it will again be convenient to parametrize G as a gamma (α, β) distribution: since G is then stable under convolution, the quantities $G^{(k)}(t)$ above are easily calculated. Further, when α is set to 1, the waiting time distribution G becomes exponential (β), and consequently (5.8) collapses to (5.2). The intuition for this follows from observing that when the waiting time τ is distributed exponential, the renewal property holds regardless of whether or not a transition actually occurs. Thus, a model like (5.8), under $\alpha = 1$, ought to behave exactly like (5.2). This intuition is made rigorous by noting that under $\alpha = 1$,

$$\forall k \geq 1: \qquad G^{(k)}(t) = \Gamma(k)^{-1}\beta^k \int_0^t u^{k-1}e^{-\beta u}\,du = \Gamma(k)^{-1}\gamma(k, \beta t),$$

where γ is that part of the incomplete gamma function given by:

$$\gamma(a, x) = \int_0^x u^{a-1}e^{-u}\,du.$$

Using the series expansion:

$$\gamma(a, x) = e^{-x}x^a \sum_{m=0}^{\infty} \frac{\Gamma(a)}{\Gamma(a+1+m)} x^m,$$

it follows that:

$$G^{(k)}(t) = e^{-\beta t}(\beta t)^k \sum_{m=0}^{\infty} \frac{1}{\Gamma(k+1+m)}(\beta t)^m = e^{-\beta t} \sum_{m=k}^{\infty} \frac{1}{\Gamma(1+m)}(\beta t)^m,$$

and thus:

$$G^{(k)}(t) - G^{(k+1)}(t) = e^{-\beta t}\frac{(\beta t)^k}{\Gamma(k+1)} = e^{-\beta t}\frac{(\beta t)^k}{k!}.$$

The case for $k = 0$ is trivial. Thus, we have shown that restricting α to equal 1 when G is gamma (α, β) implies that (5.8) is just (5.2) with $\lambda = \beta$. In general, however, the model (5.8) strictly contains (5.2).

6. Empirical Results

This section applies the ideas of the previous sections to re-interpret the transition probabilities reported above. Table 2 imposes the different structures on transition probabilities for the 3-state case; Table 3 does the same for the 4-state case. In all cases, we estimate the underlying structural parameters—the Q matrix and the different auxiliary parameters describing waiting times between move exposures—by minimizing the sum of squared residuals between the implied observed transition probability matrix and the unrestricted versions calculated in Section 3. The minimized value of this criterion also implies a natural measure of goodness of fit analogous to R^2 in standard regression analysis. Each panel in the tables begins with the sum of squared deviations. The second number in the panel gives an R^2-like statistic: one minus the ratio of the sum of squared deviations to the maximum possible such number, given the estimates of the unrestricted transition probability matrices. Then each panel shows the matrix Q which describes transitions conditional on a move exposure. Next, each panel displays the auxiliary coefficients associated with each of the parametrizations (5.2), (5.5), (5.6), (5.7) (for both gamma and Weibull hazards), and (5.8). The final row in each panel shows the ergodic steady state distribution, if one exists, that is implied by the parameter estimates.[5] Note that a steady state distribution need not exist for models (5.5), (5.7), and (5.8). We found no convergence to such a distribution for those models. By contrast, (5.2) and (5.6) are restricted to admit a long run steady state distribution. For the three-state model, both of those show rapid convergence to an unconditional distribution with approximately one-third of the cross-section in each state. That convergence to a steady-state distribution is also rapid for the four-state model where the unconditional distribution now shows approximately one-fourth of the cross-section in each state.

Within each of the 3- and 4-state models, the estimated Q matrices are quite similar across different structures. All show substantial movement from the low-

[5] In Quah [1993]—which imposes much simpler parametrizations that those used here—ergodic distributions always exist.

growth state to high-growth and vice versa: the top right-hand and bottom left-hand entries are large in all cases; for the 4-state model, those entries are the largest in their respective rows. In the 4-state model, there is a tendency for the middle-growth economies to transit into other middle-growth states; at the extremes, on the other hand, the movement appears to be fairly dramatic across growth states.

For model (5.2), the exit rate is estimated to have Poisson parameter 0.97 in the 3-state model and 0.90 in the 4-state model. In words, the average rate of movement out of any state is found to be a little less than once a year. When heterogeneity in the form of a distribution of exit rates is allowed (heterogeneity varying with country), the average exit rate (α/β) is estimated to be somewhat higher, and there appears to be significant heterogeneity in the cross-section: the standard deviation $\sqrt{\alpha}/\beta$ is about the same magnitude as the mean α/β.

The third panel gives estimates for (5.6) which allows for state-dependence in exit rates. In the very lowest growth state, exit appears to be quite rapid: for the 3-state model, the Poisson rate $\lambda_1 = 1.29$ implies an nine-month average time to exit. This exit parameter declines as one moves into the high-growth states: roughly speaking, there tends to be an increasing persistence as an economy moves from low to high growth.

The Poisson and Weibull unconditional aging models (5.7) tell the same story: the hazard rate is declining towards (close to) zero in both cases. In this time-heterogeneity sense then, there does appear to be evidence against the convergence hypothesis: the longer an economy spends within a certain growth state, the lower is the resulting exit probability. Within the 3- and 4-state models, both fit and Q matrices are approximately the same for both Poisson and Weibull hazards (not surprising given the estimated parameter values).

Model (5.8) adds a renewal-type property to these state transitions: however, the fit is no better here than in (5.7), and in the 4-state case is actually worse.

Recall that models (5.7) and (5.8) allow time-heterogeneity in mobility behavior whereas models (5.5) and (5.6) allow cross-section heterogeneity. Goodness of fit does not differ significantly across the broad classes of models, although the time-heterogeneity models marginally better describe the data.

In summary, it is important to emphasize that all the models fit surprisingly well—there is little to distinguish among them. Improvement in fit from the simplest (5.2) to any of the others is on the order of only about 1%; thus, strong conclusions shouldn't be taken from any single one of the models. All the results taken together suggest, however, a picture of significant cross-section mobility in growth rates. Should a steady-state distribution of growth rates exist, it appears likely to be uniform across different states. Such a statement, of course, depends on the coarseness of the discretization used in the empirical work: given the goodness of fit of the current models, it seems unlikely that sharp contradictory conclusions could be possible.

Since significant cross-section mobility is observed over both short- and long-horizons, the convergence hypothesis in growth rates appears to accurately describe the data.

7. Conclusion

This paper has analyzed the transition and mobility patterns of per capita income growth across the cross-section of world economies. I have presented modeling and econometric tools that allow empirical study of the convergence hypothesis. These methods bring together ideas from stochastic process theory and from empirical work in income distribution, sociology, and labor economics. In reviewing the theoretical discussion, it is worthwhile emphasizing that in the current work, we both observe and are interested in an evolving conditional distribution. By contrast, in typical stochastic process analysis, it is assumed that the realizations of a process are observed but the conditional distributions of the process are themselves unknown. We next used ideas that were first exploited in income distribution, sociology, and labor economics to impose structure on the evolving distributions—structure that allowed us to get at a number of questions of interest in modern growth theory.

The empirical evidence suggests significant cross-section mobility over both long- and short-horizons in income growth rates across the different economies in the world. In this sense then the convergence hypothesis appears to accurately describe the data. There is some positive, although weak, evidence that phenomena like persistence and path dependence are actually present in the data. However, given that relatively simple empirical models that don't allow for these effects already fit the data quite well, such phenomena may be unimportant for explaining growth.

In the current work, I have not considered explicitly separating growth dynamics into elements that persist from elements that don't (such a partition is, however, implicit in the discussion of long run steady state distributions). It would be interesting to attempt permanent and transitory decomposition exercises in this large cross-section application, such as have been performed for the time series case in Blanchard and Quah [1989], Quah [1992c], and elsewhere. That kind of analysis would necessarily involve quite different techniques from that used in time-series work; such modeling might, however, provide useful insights beyond those from the analysis of this paper.

TABLE 1
Annual Growth Rates
Unrestricted Empirical Transition Probabilities
Selected Sub-periods

3-states	**4-states**
$T_1 = \begin{pmatrix} 0.43 & 0.24 & 0.33 \\ 0.37 & 0.43 & 0.20 \\ 0.31 & 0.25 & 0.44 \end{pmatrix}$	$T_1 = \begin{pmatrix} 0.35 & 0.18 & 0.21 & 0.26 \\ 0.25 & 0.44 & 0.25 & 0.06 \\ 0.05 & 0.33 & 0.43 & 0.19 \\ 0.23 & 0.18 & 0.14 & 0.46 \end{pmatrix}$
$T_{13} = \begin{pmatrix} 0.21 & 0.21 & 0.57 \\ 0.17 & 0.26 & 0.57 \\ 0.22 & 0.22 & 0.56 \end{pmatrix}$	$T_{13} = \begin{pmatrix} 0.15 & 0.26 & 0.18 & 0.41 \\ 0.09 & 0.16 & 0.28 & 0.47 \\ 0.14 & 0.14 & 0.19 & 0.52 \\ 0.23 & 0.09 & 0.14 & 0.54 \end{pmatrix}$
$T_{24} = \begin{pmatrix} 0.48 & 0.36 & 0.17 \\ 0.43 & 0.43 & 0.14 \\ 0.34 & 0.47 & 0.19 \end{pmatrix}$	$T_{24} = \begin{pmatrix} 0.38 & 0.32 & 0.15 & 0.15 \\ 0.28 & 0.31 & 0.34 & 0.06 \\ 0.10 & 0.38 & 0.43 & 0.10 \\ 0.32 & 0.46 & 0.14 & 0.09 \end{pmatrix}$

NOTES:

1. Only three transition matrices are displayed here for the sake of brevity.
2. Row sums need not equal one exactly due to rounding.
3. The transition probabilities describe the growth experiences from 1960/61–1961/62 (T_1) through 1960/61–1984/85 (T_{24}) for 109 Summers-Heston market economies (excluding the major oil-exporters, Oman, UAE, Iran, Iraq, Saudi Arabia, Kuwait, Yemen, and Venezuela, and also Afghanistan and Indonesia because of data incompleteness).

4. Incomes are measured as disparities relative to world per capita income. The distributions do not necessarily have mean zero since it is only a population-weighted average that equals the world per capita income.

TABLE 2

Structured Transition Probabilities

3 states

Model (5.2) $SSR = 2.135$ $R^2 = 0.9671$

$$Q = \begin{pmatrix} 0.00 & 0.29 & 0.71 \\ 0.67 & 0.00 & 0.33 \\ 0.30 & 0.70 & 0.00 \end{pmatrix}$$

$$\lambda = 0.97; \qquad SS = \{0.33,\ 0.33,\ 0.34\}$$

Model (5.5) $SSR = 1.972$ $R^2 = 0.9696$

$$Q = \begin{pmatrix} 0.00 & 0.12 & 0.88 \\ 0.84 & 0.00 & 0.16 \\ 0.13 & 0.87 & 0.00 \end{pmatrix}$$

$$\alpha = 0.90,\ \beta = 0.60 \implies \frac{\alpha}{\beta} = 1.48,\ \sqrt{\frac{\alpha}{\beta^2}} = 1.57$$

Model (5.6) $SSR = 2.112$ $R^2 = 0.9674$

$$Q = \begin{pmatrix} 0.00 & 0.27 & 0.73 \\ 0.97 & 0.00 & 0.03 \\ 0.56 & 0.44 & 0.00 \end{pmatrix}$$

$$\lambda_1 = 1.29,\ \lambda_2 = 0.75,\ \lambda_3 = 0.92; \qquad SS = \{0.33,\ 0.33,\ 0.34\}$$

Model (5.7) [Gamma] $SSR = 1.872$ $R^2 = 0.9711$

$$Q = \begin{pmatrix} 0.00 & 0.46 & 0.54 \\ 0.52 & 0.00 & 0.48 \\ 0.44 & 0.56 & 0.00 \end{pmatrix}$$

$$\alpha = 0.05, \ \beta = 0.001$$

[Weibull] $SSR = 1.872$ $R^2 = 0.9712$

$$Q = \begin{pmatrix} 0.00 & 0.46 & 0.54 \\ 0.52 & 0.00 & 0.48 \\ 0.44 & 0.56 & 0.00 \end{pmatrix}$$

$$\lambda = 5.50, \ \kappa = 0.13$$

Model (5.8) $SSR = 1.871$ $R^2 = 0.9712$

$$Q = \begin{pmatrix} 0.00 & 0.36 & 0.64 \\ 0.61 & 0.00 & 0.39 \\ 0.36 & 0.64 & 0.00 \end{pmatrix}$$

$$\alpha = 0.05, \ \beta = 0.005$$

TABLE 3

Structured Transition Probabilities

4 states

Model (5.2) $SSR = 4.070$ $R^2 = 0.9590$

$$Q = \begin{pmatrix} 0.00 & 0.00 & 0.22 & 0.78 \\ 0.24 & 0.00 & 0.76 & 0.00 \\ 0.00 & 0.70 & 0.00 & 0.30 \\ 0.69 & 0.25 & 0.06 & 0.00 \end{pmatrix}$$

$$\lambda = 0.90 \qquad SS = \{0.24,\ 0.24,\ 0.26,\ 0.26\}$$

Model (5.5) $SSR = 3.694$ $R^2 = 0.9628$

$$Q = \begin{pmatrix} 0.00 & 0.41 & 0.00 & 0.59 \\ 0.26 & 0.00 & 0.74 & 0.00 \\ 0.00 & 0.52 & 0.00 & 0.48 \\ 0.65 & 0.06 & 0.29 & 0.00 \end{pmatrix}$$

$$\alpha = 0.89,\ \beta = 0.59 \implies \frac{\alpha}{\beta} = 1.52,\ \sqrt{\frac{\alpha}{\beta^2}} = 1.61$$

Model (5.6) $SSR = 3.965$ $R^2 = 0.9601$

$$Q = \begin{pmatrix} 0.00 & 0.30 & 0.05 & 0.65 \\ 0.42 & 0.00 & 0.51 & 0.07 \\ 0.00 & 0.88 & 0.00 & 0.12 \\ 0.79 & 0.21 & 0.00 & 0.00 \end{pmatrix}$$

$\lambda_1 = 1.08,\ \lambda_2 = 0.98,\ \lambda_3 = 0.52,\ \lambda = 0.76$ $SS = \{0.24,\ 0.24,\ 0.26,\ 0.26\}$

Model (5.7) [Gamma] $SSR = 3.484$ $R^2 = 0.9649$

$$Q = \begin{pmatrix} 0.00 & 0.44 & 0.10 & 0.46 \\ 0.44 & 0.00 & 0.56 & 0.00 \\ 0.00 & 0.39 & 0.00 & 0.61 \\ 0.48 & 0.16 & 0.36 & 0.00 \end{pmatrix}$$

$\alpha = 0.09,\ \beta = 0.01$

[Weibull] $SSR = 3.489$ $R^2 = 0.9648$

$$Q = \begin{pmatrix} 0.00 & 0.48 & 0.03 & 0.49 \\ 0.42 & 0.00 & 0.58 & 0.00 \\ 0.00 & 0.41 & 0.00 & 0.59 \\ 0.49 & 0.11 & 0.40 & 0.00 \end{pmatrix}$$

$\lambda = 2.06,\ \kappa = 0.23$

Model (5.8) $SSR = 3.561$ $R^2 = 0.9641$

$$Q = \begin{pmatrix} 0.00 & 0.48 & 0.00 & 0.52 \\ 0.31 & 0.00 & 0.69 & 0.00 \\ 0.00 & 0.46 & 0.00 & 0.54 \\ 0.61 & 0.05 & 0.34 & 0.00 \end{pmatrix}$$

$\alpha = 0.10,\ \beta = 0.02$

References

[1] Barro, Robert J. (1991), "Economic Growth in a Cross-Section of Countries", *Quarterly Journal of Economics*, 106(2):407–443, May.

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

[3] Baumol, William J. (1986), "Productivity Growth, Convergence, and Welfare", *American Economic Review*, 76(5):1072–85, December.

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

[5] Beveridge, Stephen and Nelson, Charles R. (1981), "A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the 'Business Cycle' ", *Journal of Monetary Economics*, 7:151–174, March.

[6] Blanchard, Olivier Jean and Quah, Danny. (1989), "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, 79(4):655–673, September.

[7] Campbell, John Y. and Mankiw, N. Gregory. (1987), "Are Output Fluctuations Transitory?", *Quarterly Journal of Economics*, 102:857–880, November.

[8] Cochrane, John H. (1988), "How Big is the Random Walk Component in GNP?", *Journal of Political Economy*, 96(5):893–920, October.

[9] Cox, D. R. and Lewis, P. A. W. (1966), *The Statistical Analysis of Series of Events*, Chapman and Hall, New York NY.

[10] DeLong, J. Bradford. (1988), "Productivity Growth, Convergence, and Welfare: A Comment", *American Economic Review*, 78(5):1138–55, December.

[11] Futia, Carl. (1982), "Invariant Distributions and the Limiting Behavior of Markovian Economic Models", *Econometrica*, 50(1):377–408, January.

[12] Geweke, John, Marshall, Robert C., and Zarkin, Gary A. (1986), "Mobility Indices in Continuous Time Markov Chains", *Econometrica*, 54(6):1407–1423, November.

[13] Heckman, James J. (1981), "Statistical Models for Discrete Panel Data", in Manski, C. F. and McFadden, D., (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, chapter 3, pages 114–178. MIT Press, Cambridge MA.

[14] King, Mervyn A. and Robson, Mark H. (1989), "Endogenous Growth and The Role of History", Discussion Paper 63, LSE Financial Markets Group, London, August.

[15] Krugman, Paul. (1991), "History Versus Expectations", *Quarterly Journal of Economics*, 106(2):651–667, May.

[16] Lucas Jr., Robert E. (1988), "On The Mechanics of Economic Development", *Journal of Monetary Economics*, 22(3):3–42, June.

[17] Nelson, Charles R. and Plosser, Charles I. (1982), "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 10:129–162, September.

[18] Quah, Danny. (1992), "International Patterns of Growth: I. Persistence in Cross-Country Disparities", Working paper, LSE, London, October.

[19] Quah, Danny. (1992), "The Relative Importance of Permanent and Transitory Components: Identification and Some Theoretical Bounds", *Econometrica*, 60(1):107–118, January.

[20] Quah, Danny. (1993), "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review*, 37(2/3):426–434, April.

[21] Romer, Paul M. (1986), "Increasing Returns and Long-run Growth", *Journal of Political Economy*, 94(5):1002–1037, October.

[22] Singer, Burton and Spilerman, Seymour. (1976), "Some Methodological Issues in the Analysis of Longtitudinal Surveys", *Annals of Economic and Social Measurement*, 5:447–474.

[23] Spilerman, Seymour. (1972), "Extensions of the Mover-Stayer Model", *American Journal of Sociology*, 78(3):599–626, November.

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

[25] Summers, Robert and Heston, Alan. (1988), "A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950-1985", *Review of Income and Wealth*, 34(1):1–25, March.

[26] Tamura, Robert. (1989), "Dual Convergence: An Equilibrium Model of the World Distribution of Income", mimeo, University of Iowa, Iowa City, IO, October.

[27] Watson, Mark W. (1986), "Univariate Detrending Methods with Stochastic Trends", *Journal of Monetary Economics*, 18:49–75, July.