A Forward Projection of the Cross-Country Income Distribution

Costas Azariadis

Department of Economics, University of California Los Angeles
405 Hilgard Avenue, Los Angeles, CA 90095-1477
azariadi@ucla.edu

John Stachurski

Center for Operations Research and Econometrics
Université catholique de Louvain
34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium
stachurski@core.ucl.ac.be

Abstract

This paper proposes and implements a method to predict the evolution of the cross-country income distribution using a stochastic parameterization of the Azariadis–Drazen (1990) nonconvex growth model. We estimate the dynamic structure of that model from data in the Penn World Tables, and define inductively all future distributions as a sequence in L_1. Elements of the sequence are calculated using Monte Carlo simulation. Our results suggest that nonlinearities in the growth process are responsible for the emerging bimodality in the distribution of income, but that bimodality will eventually peak and then decline. In the long run we predict convergence.

1 Introduction

The Bible says that the poor shall always be with us. Interpreted as a model of economic development, Solow's 1956 paper points in the opposite direction.

* We thank Omar Licandro, Kiminori Matsuyama and Kazuo Nishimura for helpful comments. The paper also benefitted from criticism at various seminars. We mention in particular the Conference on Irregular Growth, Université Paris 1 Panthéon-Sorbonne, June 2003. The second author thanks the Japan Society for Promotion of Science and the Economics Department and Economic Theory Center at the University of Melbourne for financial support.

Marginal returns to capital diminish and growth slows down; if production technology is intrinsically nonrival then the poor must catch up to the rich. Cass (1965), Koopmans (1965) and Brock and Mirman (1972) showed these conclusions also hold for rational consumers and uncertain production.

Prior to Solow, the development literature was less sanguine, arguing that complementarities, indivisibilities and external effects made growth self-reinforcing and poverty persistent (Young 1928, Rosenstein-Rodan 1943, Nurkse 1953, Nelson 1956). Many of these models have recently been formalized (Murphy, Shleifer and Vishny 1989, Azariadis and Drazen 1990, Matsuyama 1991, Durlauf 1993, Young 1993, Arthur 1994, Azariadis 1996, Acemoglu and Zilibotti 1997, etc.), sparking intense debate on the question of convergence and—implicitly—about the future shape of the cross-country income distribution.

By now the empirical literature on convergence is very large. Early contributions include De Long (1988), Barro (1991) and Mankiw et al. (1992). For a typical perspective from the linear econometric tradition see Sala-i-Martin (1996). More eclectic studies include Durlauf and Johnson (1995), Quah (1993, 1996), Bianchi (1997), Durlauf and Quah (1999), Graham and Temple (2001) and Masanjala and Papageorgiou (2004).

One trend in the new growth literature is greater acknowledgement of the role played by nonconvexities in the aggregate production set. Nonconvexities tend to generate nonlinearities in the associated dynamical system, which make interpretation of traditional econometric studies somewhat problematic. (Appealing to linear approximation in the neighborhood of a steady state is dubious, as convergence explicitly deals with the behavior of countries *at opposite ends* of the income spectrum.)

Monte Carlo methods provide researchers with various alternatives to linear extrapolation over unbounded phase spaces. If growth dynamics are known then they can be simulated. By appropriate resampling, the current state can be projected into the future. This paper implements such a scheme.

In the growth literature, explicit forward projection of the cross-country income distribution has previously been undertaken by Quah (1993), Jones (1997) and Kremer, Onatski and Stock (2001). Jones constructs steady state outcomes implied by neoclassical models. Quah and Kremer et al. give both finite horizon and steady state projections based on iteration of discrete state Markov chains, the matrices for which are calculated according to observed transition frequencies across states of a discretized income space. Finally, Quah (1996) suggests (but does not implement) a forward projection of continuous state chains.

This paper implements both short run and asymptotic projections using parametric one-sector growth models interpreted as Markov chains on the positive half-ray. We recover the dynamic structure by estimating the parameters, and then project forward from the current state.

One intended contribution of the paper is techniques for density projection. Algorithms for generating the sequence of cross-country income distributions are proposed and implemented. These algorithms are based on successive numerical computation of the integral operations associated with continuous state Markov chains.

The model we use to represent underlying dynamics is one with threshold externalities (Azariadis and Drazen 1990). Technological externalities are influential over a critical mass region of capital stock. This region is determined by the data, along with the size and direction of the external effect. Estimation of these parameters is by threshold autoregression (Chan and Tong 1986).

The nonconvexities induced by these external effects can potentially account for both the persistence of poverty and the transition to rapid growth. Overall, the cross-country income distribution may show either persistence and polarization or convergence, depending on the estimated dynamical structure (Azariadis and Drazen 1990, Galor 1996, Quah 1996, Stachurski 2003a).

The structure of the paper is as follows. Section 2 outlines the method of projection. Section 3 presents the model. Section 4 estimates parameters and implements the projections. Section 5 concludes. Details of the estimation procedure and proofs are contained in the appendices.

2 Methodology

In this section we outline the method of projection. Suppose we have sufficient knowledge of production relationships, market structures and the state-contingent policies of economic agents to estimate the law of motion for any economy. All economies have the same dynamics, represented by nonlinear autoregression

$$y_{t+1} = T(y_t, \varepsilon_t), \quad T\colon [0, \infty) \times [0, \infty) \to [0, \infty). \tag{1}$$

The state variable y is per capita income. The disturbance term ε_t is a country specific shock, independent across both time and space. The map T captures the dynamic structure of the economy. The distribution of the shock ε is also known to us. Let it be represented by density ψ.

It follows from (1) that for each fixed current level of income y, next period's income y' is a random variable, the distribution of which depends on y, the map T and the density ψ. In fact, if $\mathbb{1}_B$ is the indicator function for the set B (that is, $\mathbb{1}_B(x) = 1$ if $x \in B$ and zero otherwise), then

$$\mathrm{Prob}(y_{t+1} \in B \,|\, y_t = y) = \int \mathbb{1}_B[T(y, z)]\psi(z)dz. \tag{2}$$

Let us assume that for each y, this conditional distribution can also be represented by a density function, which is denoted $p(y, \cdot)$.[1] Here p is called the *stochastic kernel* associated with the law of motion (1), and provides a nonnegative real function over all state pairs (y, y'). By definition,

$$\int_B p(y, y')dy' = \int \mathbb{1}_B[T(y, z)]\psi(z)dz \tag{3}$$

for all B (more precisely, for all Borel sets B), so that p fully encodes the growth dynamics implied by both the deterministic structure in T and the stochastic structure in ψ.

Given that the model is stationary, from p—which gives the one-period transition probabilities—it is straightforward to construct the sequence $(p^t)_{t\geq 0}$ of t-period transition probabilities. In fact, if we define inductively $p^1 := p$ and

$$p^{t+1}(y, y') := \int p(s, y')p^t(y, s)\, ds, \quad t \in \mathbb{N}, \tag{4}$$

then it is known that $\int_B p^t(y, y')dy' = \text{Prob}(y_{s+t} \in B \mid y_s = y)$ as desired.[2]

Index the cross-section of economies by $n = 1, \ldots, N$. Suppose that the current cross country income distribution is given by vector $(y_0^n)_{n=1}^N$, where y_0^n is the income of country n when $t = 0$. We wish to calculate projections for future income distributions, which are probability distributions rather than realized observations. Clearly the time t distribution must have the interpretation of providing income probabilities for an arbitrary country selected at random from $\{1, \ldots, N\}$ at time t.

Consider first the case $t = 1$. The probability that a random country has income in set B at time 1 is the probability that a given country is selected (i.e., $1/N$), times the probability that country transitions into B given its current income, summed across all countries. That is,

$$\text{Prob}(y_1 \in B) = \sum_{n=1}^{N} \text{Prob}(y_1 \in B \mid y_0 = y_0^n)\frac{1}{N}.$$

Applying (2) and (3), then, $\text{Prob}(y_1 \in B) = \int_B \frac{1}{N}\sum_{n=1}^N p(y_0^n, y')dy'$. Since B is arbitrary and $\text{Prob}(y_1 \in B)$ is represented by an integral over B, the integrand is a density function for this distribution. Denote it (the density) by φ_1. That is, $\varphi_1(y') := \frac{1}{N}\sum_{n=1}^N p(y_0^n, y')$.

[1] Actually, the density representation is necessary only for almost every y, as p can be altered on a null set without affecting the projected income distributions (see the appendix).

[2] Here (4) is just the Chapman-Kolmogorov relation.

The same argument works for any t, using p^t instead of p. Thus

$$\text{Prob}(y_t \in B) = \int_B \frac{1}{N} \sum_{n=1}^{N} p^t(y_0^n, y')dy',$$

and if φ_t denotes the time t density, then

$$\varphi_t(y') := \frac{1}{N} \sum_{n=1}^{N} p^t(y_0^n, y'). \tag{5}$$

This is precisely what we wish to calculate. Once the initial state (y_0^n) and the stochastic kernel p are known, the calculation can proceed via (4).

There is another way of looking at the expression (5) that proves useful for implementation. From (4) and (5) we have

$$\begin{aligned}\varphi_{t+1}(y') = \frac{1}{N} \sum_{n=1}^{N} p^{t+1}(y_0^n, y') &= \frac{1}{N} \sum_{n=1}^{N} \int p(y, y')p^t(y_0^n, y)dy \\ &= \int p(y, y') \left[\frac{1}{N} \sum_{n=1}^{N} p^t(y_0^n, y) \right] dy.\end{aligned}$$

In other words,

$$\varphi_{t+1}(y') = \int p(y, y')\varphi_t(y)dy, \quad \forall t \in \mathbb{N}. \tag{6}$$

This property of being able to link successive marginal distributions recursively is characteristic of Markov models. It clearly demonstrates that under the current assumptions all of the relevant dynamical properties are encoded in kernel p. Note that from (6) the complete sequence of future cross-country income distributions is inductively defined.

2.1 Consistency of the projections

In reality the transition rule T and the density ψ that determine p are not initially known, but rather must be estimated, say from some parametric class (T_θ, ψ_θ). Each estimate $\theta \in \Theta \subset \mathbb{R}^m$ is a random function of the data, determining a stochastic kernel p_θ through (3), and hence the sequence of projections (5). It would be reassuring to know that these projections converge to the true distributions as the amount of data for estimating θ becomes large. The next theorem states that this is so whenever $\theta \mapsto p_\theta$ is continuous.

Theorem 1 (Consistency) *Let $(T_\theta, \psi_\theta)_{\theta \in \Theta}$ be a parametric class for the economy (1), where $\theta_0 \in \Theta \subset \mathbb{R}^m$ is the true parameter. Let θ_k be a consistent sequence of estimators for θ_0, in the sense that $\theta_k \to \theta_0$ $\mathbb{P}$-a.s. on some* *given probability space* $(\Omega, \mathscr{F}, \mathbb{P})$ *as* $k \to \infty$*. For* $\theta \in \Theta$*, let* p_θ *be the stochastic kernel corresponding to* θ*, as defined by* (3)*. Finally, let*

$$\varphi_t^k(y') := \frac{1}{N}\sum_{n=1}^{N} p_{\theta_k}^t(y_0^n, y') \quad \textit{and} \quad \varphi_t(y') := \frac{1}{N}\sum_{n=1}^{N} p_{\theta_0}^t(y_0^n, y')$$

be the respective projections from fixed initial $(y_0^n)_{n=1}^N$ *as defined by* (4) *and* (5)*. If* $\Theta \ni \theta \mapsto p_\theta(y, y') \in \mathbb{R}$ *is continuous for all* $(y, y') \in [0, \infty) \times [0, \infty)$*, then for any* $t \in \mathbb{N}$ *we have* $\varphi_t^k \to \varphi_t$ *in* L_1 *norm* $\mathbb{P}$*-a.s as* $k \to \infty$*.*

The proof of the theorem is rather long, and is deferred to the Appendix.

2.2 Long-run dynamics

There is some interest also in the long run behavior of the income distribution implied by the estimated dynamic model. In other words, does the sequence of density functions $(\varphi_t)_{t=0}^\infty$ defined by (5) converge to some limiting distribution? Are there many such limits depending on the starting point, or is the process history independent? In this section we provide one framework for addressing these questions. [3]

In the sequel, a dynamical system is a pair (X, S), where X is a topological space, and S is a continuous map from X to X. (Usually these are provided in the form of a difference equation $x_{t+1} = Sx_t$ defined on X.) For S and indeed for any other map, let the t-th power be t compositions of the map with itself. For any $x \in X$, the sequence $(S^t x)_{t=1}^\infty$ is called the *orbit*, or *trajectory* of x under S. A point x^* is called *invariant*, *stationary* or *fixed* for S if $S\,x^* = x^*$. Finally,

Definition 2 *The dynamical system* (X, S) *is called* asymptotically stable *if there exists a unique stationary point* x^* *and, for all* $x \in X$*, the trajectory of* x *under* S *converges to* x^* *as* $t \to \infty$*.*

Now take the stochastic kernel p from (3), let $\mathscr{D}$ be the set of all densities on the state space $[0, \infty)$, and consider an operator $\mathbf{P}$ from $\mathscr{D}$ into itself such that the image $\mathbf{P}\varphi$ of $\varphi \in \mathscr{D}$ is defined at the point y' by

$$(\mathbf{P}\varphi)(y') = \int p(y, y')\varphi(y)dy. \tag{7}$$

[3] Ergodic distributions, which are limits of the sequence $(\varphi_t)_{t=0}^\infty$, are calculated for the discrete case by Quah (1993) and Kremer et al. (2001) via iteration of a Markov matrix. In the continuous case Johnson (2000) computes an ergodic distribution for US state GDP data, where the transition probability function is estimated non-parametrically. Jones (1997) takes a different approach, calculating the steady state of individual countries via cross-country regression and then backing out implied steady states.

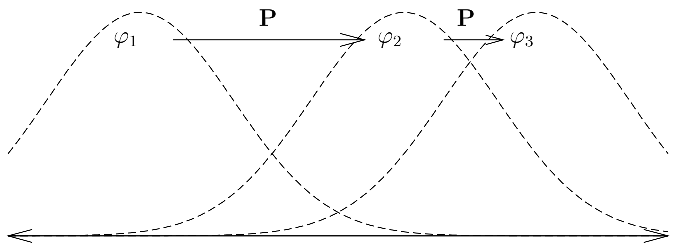

Fig. 1. Projection by the Markov operator.

Comparing (6) and (7), we see that the sequence of projected income distributions $(\varphi_t)_{t=1}^{\infty}$ defined by (4) and (5) obeys $\varphi_{t+1} = \mathbf{P}\varphi_t$ for all $t \geq 1$. This situation is illustrated in Figure 1. The operator $\mathbf{P}$ is called the *Markov* (or *stochastic*) *operator* associated with stochastic kernel p.

It is not difficult to show that $(\mathscr{D}, \mathbf{P})$ is a dynamical system in the sense defined above.[4] In Section 4 we prove that the estimated model $(\mathscr{D}, \mathbf{P})$ is asymptotically stable in the sense of Definition 2. (Asymptotic stability of $(\mathscr{D}, \mathbf{P})$ in this sense coincides with the traditional notion of global stability for Markovian economic models due to Brock and Mirman, 1972.) We calculate the limiting distribution, which is the unique solution to the functional equation $\mathbf{P}\varphi^* = \varphi^*$.

2.3 The convergence debate

On an abstract level the convergence debate can be summarized as follows. Assume again the hypotheses of the previous section. We have seen that the stochastic kernel p and initial distribution characterize completely the dynamics of the cross-country income distribution. Denote by $\mathscr{P}$ the class of all stochastic kernels on $[0, \infty)$.[5] Suppose as before that p can be recovered from the data. Before we can answer the question of whether convergence is occuring, we require a well-defined classification on $\mathscr{P}$ (a random function of the data) which associates to each $p \in \mathscr{P}$ one and only one of the labels "convergent" or "not convergent."

[4] Technically $\mathscr{D}$ is the class of nonnegative functions in $L_1[0, \infty)$ that integrate to unity, and $\mathbf{P}$ is a linear operator on L_1 (see the appendix for details). The operator is easily shown to be continuous, and, moreover, $\mathbf{P}$ sends $\mathscr{D}$ into itself (see, e.g., Lasota and Mackey, 1994, §§3.1).

[5] Precisely, $\mathscr{P}$ is the class of all $p\colon [0, \infty) \times [0, \infty) \to [0, \infty)$ with $\int_0^\infty p(y, y')dy' = 1$ for all y.

Needless to say, the existence of such a scheme upon which all growth theorists can agree is not assured (consider, for example, the problem of ranking inequality in income distributions).[6] In its absence, one suboptimal solution is the traditional approach, which has been to restrict attention to a linear parametric subclass of $\mathscr{P}$ for which classification is straightforward. For example, absolute β-convergence is defined (Sala-i-Martin, 1996) as a negative value for the coefficient β returned by the regression

$$\hat{y}_t = \text{const} + \beta \ln y_t + \varepsilon_t, \tag{8}$$

where $\hat{y} := \ln y_{t+1} - \ln y_t$ is the growth of per capita income over the period. In the same vein, conditional β-convergence is said to hold if the negative sign of the coefficient is realized for some collection of "growth-related" right hand side variables.

Aside from what Karl Popper would have had to say about the validity of conditional β-convergence as a falsifiable scientific hypothesis, a major shortcoming of these classifications is the assumption of linearity. Also, the influence of random factors is not incorporated into convergence dynamics (estimates of $\mathrm{Var}(\varepsilon)$ are calculated, but only to generate regression fit and other summary statistics). This practice involves significant loss of information, as the distribution of the shock is a key factor in determining convergence rates and other dynamic properties of Markov chains.

In an important paper, Quah (1996) suggested instead a way to estimate the stochastic kernel p from $\mathscr{P}$ using a purely nonparametric methodology which circumvents the need for any linear structure, and incorporates automatically the influence of random factors. Our work extends this line of research by using parametric estimation based on an economic model. The idea is to bring to bear on the estimation of p the restrictions implied by growth theory.

3 The Model

We assume that all economies obey an identical overlapping generations model with threshold type knowledge spillovers (Azariadis and Drazen 1990). In this economy, agents live for two periods, working in the first and living off savings in the second. Savings in the first period buys physical capital, which in the following period will be combined with the labor of a new generation of young agents to produce a single good under the private constant returns technology

$$y_t = A_t k_t^{\alpha} \ell_t^{1-\alpha} \varepsilon_{t-1}^{\sigma}, \quad 0 < \alpha < 1. \tag{9}$$

[6] One definition of convergence might require that the dynamical system $(\mathscr{D}, \mathbf{P})$ corresponding to $\mathbf{P}$ in (7) is asymptotically stable in the sense of Definition 2. In fact this definition is much *too weak*. See Stachurski (2003a).

Here y_t is income, k_t is capital and ℓ_t is the number of young agents, all of whom supply inelastically one homogeneous labor unit. Production is perturbed in each period by uncorrelated shocks ε_t, all having the same density ψ. The exponent $\sigma \geq 0$ is used to parameterize their variance.

Assumption 3.1 *The shock ε is lognormal. That is,* $\ln \varepsilon \sim N(0, 1)$.

This assumption is natural from a statistical perspective, but has considerable influence on the dynamics of the process. For example, countries that experience large positive shocks can potentially accumulate enough capital to overcome growth thresholds. In other words, the probability of growth miracles are nonzero. This accords well with the data (Parente and Prescott 1993). At the same time, the small tails of the distribution mean that large shocks are unlikely—poor countries will on average remain for a long time below the threshold required to sustain rapid growth. Small tails are also important in guaranteeing ergodicity of the process, restricting the probability of collapse or unbounded growth (Stachurski 2002, 2003a).

Young agents maximize $\ln c_t + \beta \mathbb{E} \ln c'_{t+1}$ subject to the budget constraints $s_t = (w_t - c_t)$ and $c'_{t+1} = s_t R_{t+1}$, where s is savings from wage income, c (resp. c') is consumption while young (resp. old), w is the wage rate and R is the gross rate of return on savings. Competitive markets imply that firms pay inputs their marginal factor product. Firms observe ε_{t-1} and then set

$$R_t(k_t, \varepsilon_{t-1}) = A_t \alpha k_t^{\alpha-1} \varepsilon_{t-1}^{\sigma}, \quad w_t(k_t, \varepsilon_{t-1}) = A_t (1-\alpha) k_t^{\alpha} \varepsilon_{t-1}^{\sigma},$$

where for convenience we are assuming that $\ell_t = 1$ for all t.

The scale parameter A_t is nonstationary due to the effect—external to firms—of capital aggregates on production technology. This effect has the form of a critical mass requirement. For now we assume only that $A_t = A(k_t)$, and that

Assumption 3.2 *The map $k \mapsto A(k)$ is strictly positive, nondecreasing and bounded.*

Thus at time t, households choose s_t to maximize

$$\ln(w_t(k_t, \varepsilon_{t-1}) - s_t) + \beta \int \ln[s_t \cdot R_{t+1}(k_{t+1}, z)] \psi(z) dz,$$

using their current belief for next period capital stock k_{t+1}. In a rational expectations equilibrium their beliefs are realized, with $k_{t+1} = s_t$. This gives the law of motion

$$k_{t+1} = \frac{\beta}{1+\beta} A(k_t)(1-\alpha) k_t^{\alpha} \varepsilon_{t-1}^{\sigma}. \tag{10}$$

Following Azariadis and Drazen (1990), we are interested in the case where the transition rule $k \mapsto A(k)$ increases mainly over a "critical mass" region of the state space. There are many possible causes of such nonlinear feedback in economic development, including capital market imperfections, patterns

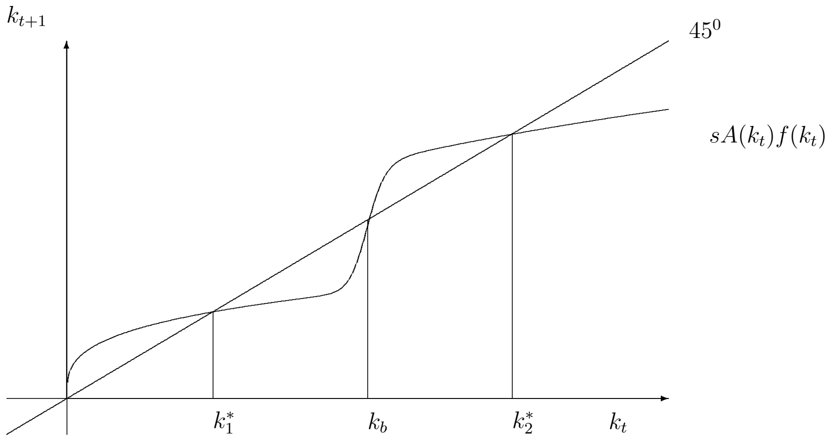

Fig. 2. Threshold externalities.

of human capital accumulation, endogenous fertility and other demographics, impatience traps and so on. See Azariadis (1996) for discussion and references. Figure 2 illustrates the deterministic part of (10) for one such specification of A.

As drawn, the model has multiple, locally stable steady states when $\sigma = 0$. If the noise level is positive, however, the situation is different. Indeed for $\sigma > 0$, it can be shown that the corresponding Markov operator $\mathbf{P}$ is strictly contracting over the density space $\mathscr{D}$.[7] That is,

$$\|\mathbf{P}\varphi - \mathbf{P}\varphi'\| < \|\varphi - \varphi'\|, \quad \text{all pairs } \varphi, \varphi' \in \mathscr{D} \text{ with } \varphi \neq \varphi'. \tag{11}$$

Such models can have at most one stationary point $\varphi^* \in \mathscr{D}$, because if φ_1^* and φ_2^* are both stationary and distinct, then $\|\mathbf{P}\varphi_1^* - \mathbf{P}\varphi_2^*\| = \|\varphi_1^* - \varphi_2^*\|$, contradicting (11). Nevertheless, the regions around k_1^* and k_2^* are "metastable," in the sense that nearby orbits are statistically likely to remain in the region (Stachurski 2003a, Proposition 5).

Next, let us convert the state variable to y rather than k in order to project income distributions. To begin, define $\tau(k) := A(k)k^\alpha$. Under Assumption 3.2, τ is strictly increasing and hence invertible. From (9) we get $y_t = \tau(k_t)\varepsilon_{t-1}^\sigma$ and so $k_t = \tau^{-1}(y_t/\varepsilon_{t-1}^\sigma)$ for all t. Combining this with (10) and applying τ to both sides gives $y_{t+1} = \tau(c_1 y_t)\varepsilon_t^\sigma = c_2 A(c_3 y_t) y_t^\alpha \varepsilon_t^\sigma$, where each c_i is a positive constant. But for A satisfying only Assumption 3.2, nothing changes if the constants are dropped, which gives the final form

$$y_{t+1} = A(y_t) y_t^\alpha \varepsilon_t^\sigma. \tag{12}$$

This is our parameterized version of the general system (1).

[7] See, for example, Stachurski (2002, Proposition 7.2). The distance metric is L_1 norm.

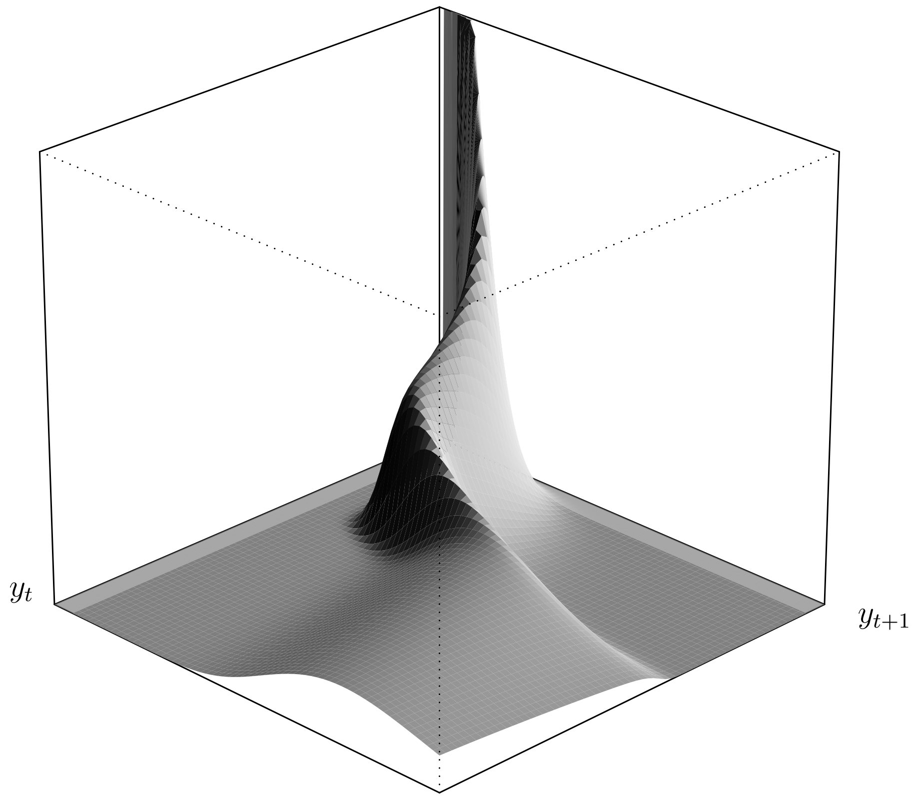

Fig. 3. The stochastic kernel

How might the stochastic kernel look for (12)? An elementary change of variable argument (e.g., Casella and Berger 1990, Theorem 2.1.2) shows that when the law of motion is (12), equation (3) is satisfied by

$$p(y, y') = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{y'} \exp\left[\frac{-(\ln y' - \ln A(y) - \alpha \ln y)^2}{2\sigma^2}\right]. \tag{13}$$

Here p is defined on $(0, \infty) \times (0, \infty)$.[8]

A plot of (13) for the critical mass rule of Figure 2 is shown in Figure 3. The origin is the furthest corner from the viewer. The y_t axis increases to the left, while the y_{t+1} axis increases to the right. For each point y on the y_t axis, the two dimensional section running parallel to the y_{t+1} axis is a density function which records the conditional distribution of next period income, given that current income is y.

[8] Alternatively, observe that $\ln y_{t+1}$ is normal for given y_t. Calculating its density and converting back to lognormal gives (13).

4 Estimation and Results

In this section we choose a parametric form for A, fit the parameters to the data using threshold autoregression, and calculate the projections.[9]

4.1 Specification

To begin, let $A(y) = A_0 \exp[h\Psi(y; \lambda, \gamma)]$, where A_0 is a positive constant, $h \in \mathbb{R}$ gives the size and direction of the external effect, and Ψ is the logistic transition function $\Psi(y; \lambda, \gamma) = (1 + (y/\gamma)^{-\frac{1}{\lambda}})^{-1}$. This function is standard in threshold autoregression. The first parameter λ determines the smoothness of the transition, while $\gamma > 0$ is the threshold point that locates it. The map $y \to \Psi(y)$ is nondecreasing and has limits $\lim_{y \downarrow 0} \Psi(y) = 0$, $\lim_{y \uparrow \infty} \Psi(y) = 1$.

Note that the standard convex model is included as the special case $h = 0$. Thus *positive external effects are not imposed a priori*; they may or may not be returned by the data during estimation.

4.2 Estimation procedure

It is convenient for the estimation to transfer the model to one residing on the real line. To this end, take logs of both sides of (12) to obtain

$$y_{t+1} = S(y_t) + \sigma \ln \varepsilon_t, \quad S(y) := a + h\Psi(e^y; \lambda, \gamma) + \alpha y, \tag{14}$$

where y is now log of income, and $a := \ln A_0$. For such a representation the unknown parameters can be solved by a smooth transition threshold autoregression scheme (or indeed any nonlinear regression procedure).

Cobb-Douglas technology provides the usual advantage that the single parameter α has the interpretation of being the capital share of income under perfect competition, which can be identified prior to estimation. In recent empirical work, this value found to be is temporally stable at around 0.3, relatively constant across countries and independent of the current income state (Gollin 2002). Taking the capital share as given, then, the remaining parameters a, h, λ, γ and σ are determined by regression.

The data is Penn World Tables, Version 6.1 (Heston, Summers and Aten 2002). In the PWT data, national income accounts are adjusted to measure purchasing power at international prices. Currently the series are valued in constant 1996 US dollars. For the two time points we select $t = 1970$ and $t + 1 = 2000$. The 30 year gap is perhaps on the long side for a generation in the OLG model, but helps to remove serial correlation from the shocks.

[9] All of the procedures are coded in R (Ihaka and Gentleman 1996).

There is a total of 104 countries with both data points available in this set. All were included in the sample. The law of motion (14) is then estimated in cross-section.

	a	h	γ	λ
estimate	4.590	3.090	3,478	0.702
st. error	0.444	0.873	716.4	0.305

Table 1

Estimation returned the parameters given in Table 1. The reported standard errors are asymptotic approximations for the nonlinear regression calculated by a standard algorithm. The values with immediate interpretation are h and γ. The former is associated with positive external effects, and has a 95% confidence interval of $[1.37, 4.81]$. The naive hypothesis that the model is convex ($h = 0$) is rejected at 99%. The second parameter γ is the threshold value, measured in 1996 US dollars. The standard error for γ is quite large, but the qualitative results of the projection are largely robust over the 95% confidence interval. [10]

Finally, the parameter σ introduced in (9) can be calculated from the sum of squared residuals in the regression (14). The estimation returned $\sigma = 0.538$.

One way to understand the implications of these parameters is to calculate the expected growth rate as a function of current income y, which is given by

$$y \mapsto \int \left[\frac{y' - y}{y} \right] p(y, y') dy'.$$

This function is plotted in Figure 4. The x-axis is current per capita income. The y-axis is the expected growth rate over the 30 year period, expressed in an annualized rate. The high growth rates at extremely low levels of income is mainly a facet of the model architecture—Cobb-Douglas technology must have an infinite derivate at zero. More interesting is the low levels of growth rates for the poorer countries, and the relatively high expected growth rate for middle income levels. As will become clear, this feature has a major influence on the projected evolution of the income distribution.

4.3 Projection

For the projection it is more convenient to work directly with logarithmic values via (14) rather than the original system (12). [11] Recall the method of

[10] As discussed below, the projections predict eventual convergence. Towards the top end of the 95% confidence interval for γ, the threshold is so large that eventual convergence fails to occur.

[11] These two ways are completely equivalent—see the appendix for details.

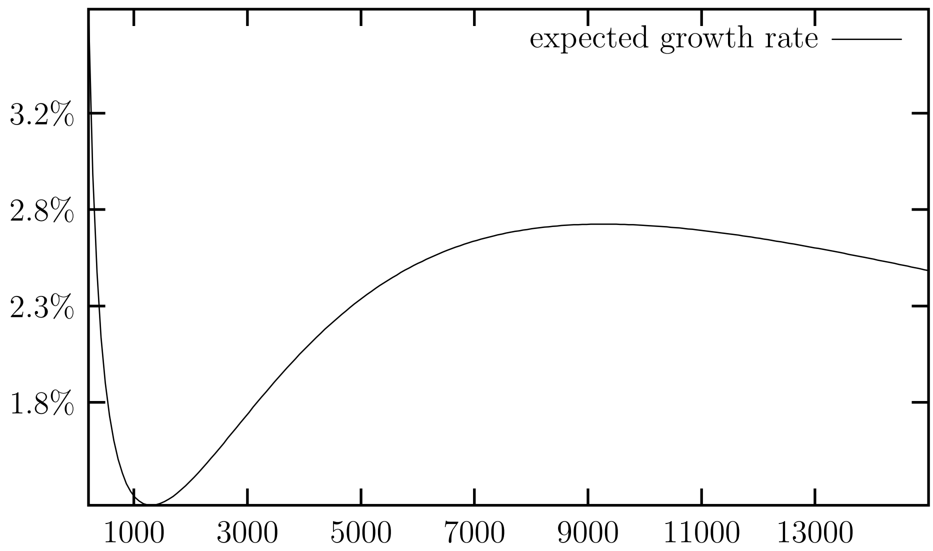

Fig. 4. Expected growth rate given current income

Section 2. Our objective is to take (14) as the estimated version of (1), and the 2000 cross section of income data as corresponding to the vector of current observations $(y_0^n)_{n=1}^N$. The projection can now take place using (4) and (5), substituting in the kernel associated with (14), which is readily seen to be

$$q(y, y') := \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(y' - S(y))^2}{2\sigma^2}\right), \tag{15}$$

a function on $\mathbb{R} \times \mathbb{R}$.[12]

The brute force approach is simply to calculate all of the integrals in Section 2 with a standard numerical routine. When t is large, however, calculation of the t-th stochastic kernel via (4) is computationally expensive. A faster and easier technique suggests itself through (6), which in the present case becomes

$$\varphi_{t+1}(y') = \int q(y, y')\varphi_t(y)dy. \tag{16}$$

Suppose that we can generate M random deviates $(y_m)_{m=1}^M$ from the density φ_t. In this case the law of large numbers gives

$$\frac{1}{M}\sum_{m=1}^{M} q(y_m, y') \overset{\text{a.s.}}{\to} \int q(y, y')\varphi_t(y)dy = \varphi_{t+1}(y') \quad \text{as} \quad M \to \infty.$$

Indeed to verify that Kolmogorov's strong law is applicable we need only check that the integral (16) is finite at every point y'. But clearly this is the case, as

[12] No change of variable is required to see this. That y' given y is normally distributed with mean $S(y)$ and variance σ^2 is immediate from (14).

q is bounded by a finite constant on $\mathbb{R} \times \mathbb{R}$.[13]

The only remaining question is how to generate deviates from φ_t. First note that by (5), $\varphi_t(y') = \frac{1}{N}\sum_{n=1}^{N} q^t(y_0^n, y')$. This suggests the following algorithm, which returns one such deviate.

Algorithm 1 *Setup: Provide the function S from (14).*

1. *Pick n from $1, \ldots, N$ with prob. $1/N$; set CurrentIncome $= y_0^n$.*
2. *Draw t independent observations $u_1, \ldots, u_t$ from $N(0, 1)$.*
3. *For i in 1 to t, let CurrentIncome $= S($CurrentIncome$) + \sigma u_i$.*
4. *Deliver CurrentIncome.*

The projection is shown in Figure 5. The quantities on the x-axis are 1996 US dollars of per capita GDP measured in logs. The main features are as follows. First, not surprisingly, probability mass continues to shift to the right. Expected per capita income, given at time t by $\int x\varphi_t(x)dx$, increases from \$3,288 in 1970 through \$4,954 in 2000 and \$9,377 in 2030 to \$11,114 in 2060 and \$13,480 in 2090.

The second major qualitative feature of our results is that no distribution puts a large amount of probability mass on the region between $8.5 \cong \$5,000$ and $9.5 \cong \$13,500$. The explanation for this behavior is contained in the dynamical structure of the fitted stochastic kernel. In essence, this is where the effects of increasing returns on growth are the strongest (recall Figure 4), and since countries in this region grow relatively quickly, statistically *they spend more time in other areas of the distribution.*

In the main our results support the conjectures of Quah (1996), who estimates the stochastic kernel p nonparametrically, and infers a tendency towards emergent bimodality from its shape. The modes of the distributions are concentrated on so-called metastable sets, which attract the state variable "on average." The degree of polarization depends on the estimated structure, which determines the strength of attraction, and the level of noise, which diminishes the importance of local effects.

4.4 Long run behavior

A third feature of our results deserves comment. Note from Figure 5 that in the long run probability mass is shifting to the right hand mode. As a result, the level of polarization in the distribution peaks and then begins to decline.

[13] The symbol $\xrightarrow{\text{a.s.}}$ means convergence almost surely. It has not been made clear to which probability space $(\Omega, \mathscr{F}, \mathbb{P})$ this notation refers. In the present case, we can take Ω to be $\mathbb{R}^{\mathbb{N}}$, $\mathscr{F} := \otimes_{t=1}^{\infty}\mathscr{B}$, where $\mathscr{B}$ is the Borel sets on $\mathbb{R}$, and $\mathbb{P} := \varphi_t \times \varphi_t \times \cdots$, the infinite dimensional product measure. Each random variable y_m is the natural projection π_m of $\mathbb{R}^{\mathbb{N}}$ onto $\mathbb{R}$.

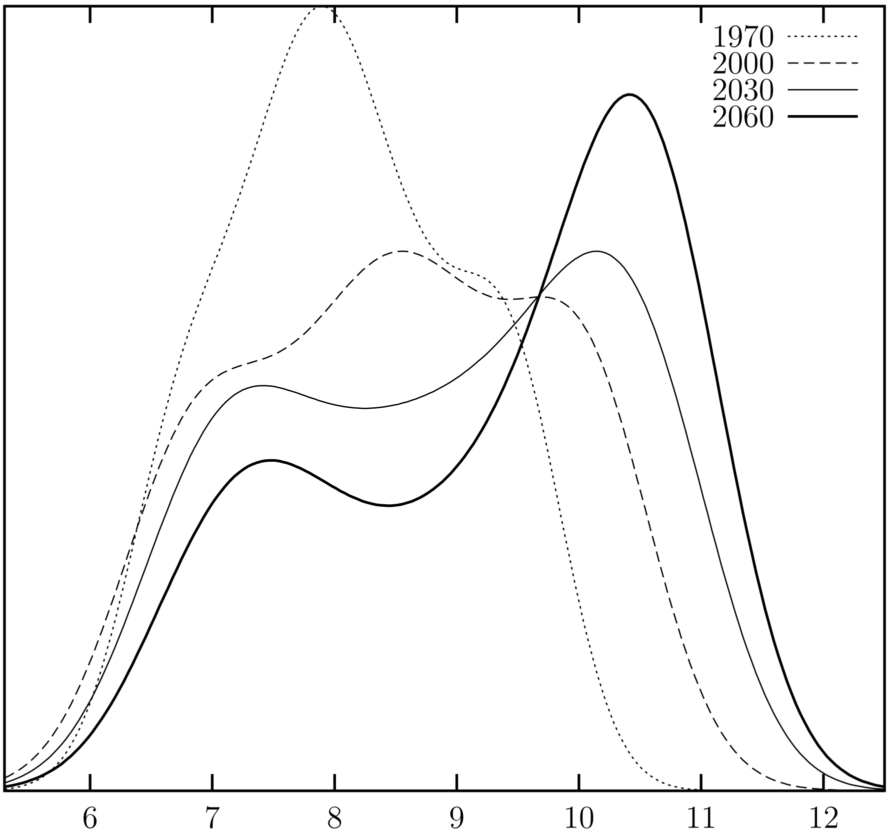

Fig. 5. Projected Cross-Country Income Distributions

That is, it forms an inverted U.[14] Evidently the estimated dynamical structure is such that countries will continue making the transition from poor to rich through the critical mass region, but few will move in the opposite direction. Growth miracles outweigh growth disasters.[15]

Next we investigate these longer-run features. There are, however, a number of good reasons to treat these results with more caution than the short-term projections—see below.

Following Kremer, Onatski and Stock (2001) one can estimate the Gini index of predicted distributions to gauge the evolution of inequality. (This is essentially the same as measuring so-called σ-convergence by estimating the standard deviation of each distribution.) Consider Figure 6. The x-axis is time,

[14] This could be called a "cross-country Kuznets hypothesis."

[15] This possibility has previously been discussed by Kremer, Onatski and Stock (2001), who provide the elegant interpretation of optimal experimentation—countries who find by experimentation a successful policy mix transit to the right hand mode, but are rarely foolish enough to persist with policies that prove measurably worse than the current rule.

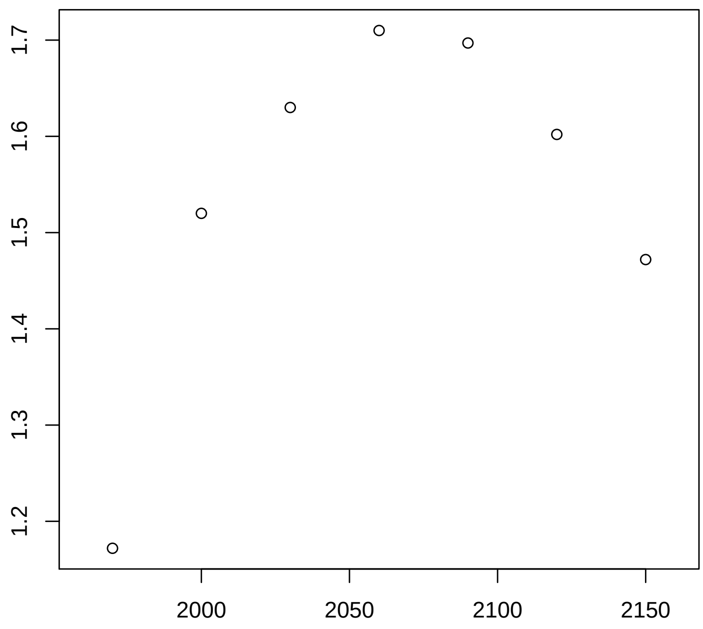

Fig. 6. Inequality in the Income Distributions.

and the y-axis is the Gini index, defined here for $\varphi \in \mathscr{D}$ as

$$g(\varphi) := \int \int |x - y| \varphi(x) \varphi(y) dx dy.$$

When finite, the value of g at φ returns the expected distance between two countries selected at random from that distribution. If we associate convergence with decreasing inequality as measured by the Gini index, then our results indicate that the current trend of divergence will peak around 2060, after which convergence will set in.

Consider finally the asymptotic behavior implied by the estimated kernel. Let $\mathbf{P}$ be the Markov operator associated with p in (13), let $\mathbf{Q}$ be the Markov operator associated with q in (15), let $\mathscr{D}$ be the density functions on $(0, \infty)$, and let $\mathscr{E}$ be the density functions on $\mathbb{R}$. Thus $(\mathscr{D}, \mathbf{P})$ is the dynamical system on density space corresponding to the fitted model (12), while $(\mathscr{E}, \mathbf{Q})$ is that corresponding to its transformation onto the real line (14). We wish to analyze and then estimate their long run behavior.

There are several probable conjectures. First, the two systems are related by a smooth transformation, and hence one would expect their dynamic behavior to be similar. Second, $(\mathscr{E}, \mathbf{Q})$ is derived from a smooth transition autoregression, which is known to be ergodic when α in (14) satisfies $|\alpha| < 1$ (Chan and Tong 1986, Proposition 2.1). Combining the first and second conjectures, then,

$(\mathscr{D}, \mathbf{P})$ should also be stable. Indeed,

Proposition 3 *The following statements are true.*

1. *For any stationary distribution $\varphi_p^* \in \mathscr{D}$ of $(\mathscr{D}, \mathbf{P})$, there exists a corresponding stationary distribution $\varphi_q^* \in \mathscr{E}$ for $(\mathscr{E}, \mathbf{Q})$ and vice versa. They are related by $\varphi_q^*(y) = \varphi_p^*(\exp y) \exp y$, $y \in [0, \infty)$.*
2. *Both $(\mathscr{D}, \mathbf{P})$ and $(\mathscr{E}, \mathbf{Q})$ have a unique stationary distribution and are asymptotically stable.*
3. *For both systems the L_1 distance between the current density and the limiting density declines monotonically in t.*

A proof is given in the appendix.[16]

With this result in hand we can calculate φ_q^* as the unique solution to

$$\mathbf{Q}\varphi(y') := \int q(y, y')\varphi(y)dy = \varphi(y'). \tag{17}$$

There are many known algorithms for solving the integral equation (17), such as collocation and the Galerkin projection. For our purposes, however, the most appropriate is the method of Glynn and Henderson (2001), who proposed the estimator

$$\frac{1}{T}\sum_{t=1}^{T} q(y_t, y'), \quad (y_t)_{t=1}^{T} \text{ a sequence generated by (14).} \tag{18}$$

Although this algorithm is new in the literature, it is easy to program, computationally efficient and converges to the true distribution in L_1 norm. The intuition simple: For t large, y_t is approximately distributed according to φ_q^* by Proposition 3, Part 2. In which case one might expect the sum in (18) to converge to $\int q(y, y')\varphi_q^*(y)dy = \mathbf{Q}\varphi_q^*(y') = \varphi_q^*(y')$ by the law of large numbers. Glynn and Henderson show that this is indeed the case.

The result of the estimation is given in Figure 7. The 1970 distribution is plotted as well for reference. Clearly our estimates imply that the asymptotic distribution is unimodal—the process of convergence suggested in Figure 5 continues until all probability mass shifts into the right hand mode. According to that figure, the convergence process may be in full swing by the end of the century.

5 Conclusion

Nonlinear regressions are used in this paper to estimate threshold externalities in a stochastic parameterization of the Azariadis-Drazen (1990) model

[16] See also Stachurski (2003a, Theorem 1).

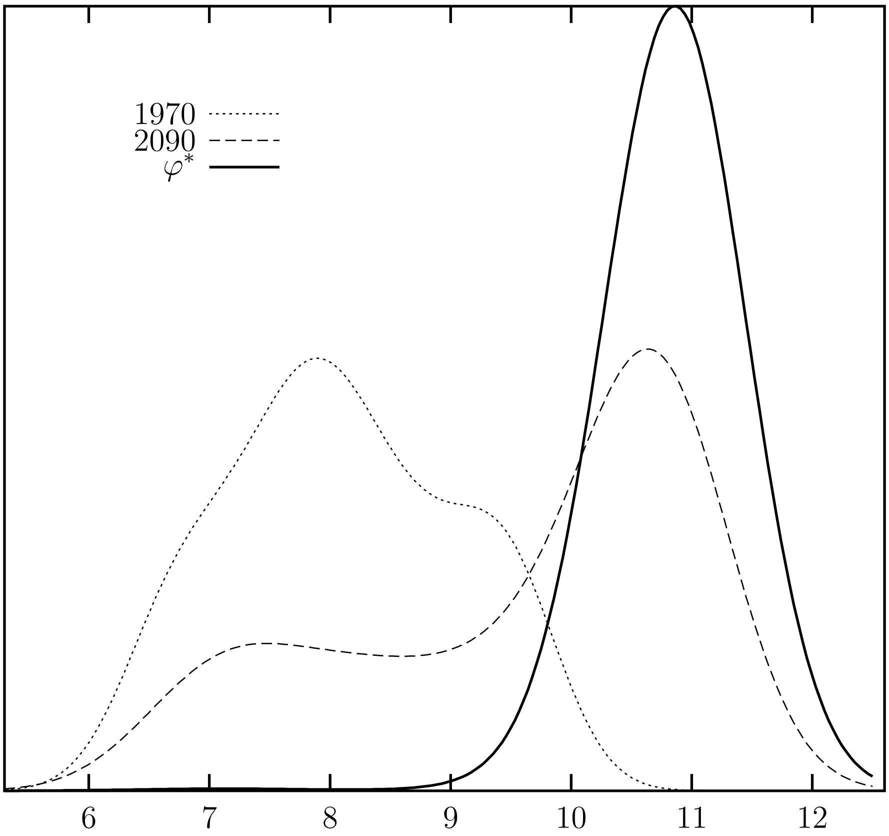

Fig. 7. Ergodic distribution

fitted to data from the Penn World Tables. In this model, dynamical behavior is determined by the interaction of productivity shocks and the influence of external effects. Parameter estimates are used to infer stochastic kernels generating transitions between income states, leading to a complete description of the evolution of the world distribution of per capita income over time.

The projections indicate that the world income distribution will show increasing inequality until about 2060, and decreasing inequality thereafter, as underlying growth and natural variation from output shocks allow more nations to attain the critical production scale necessary for rapid growth.

There are a number of important caveats to our findings. First, the longer term predictions are jeopardized by underlying nonstationarity of the model. Second, the estimation may suffer from sample bias. In particular, if countries that experience growth disasters drop out of the sample set, then the estimated structure will not reflect the true degree of persistence in the kernel. Finally, the degree of noise—the parameter σ—is estimated via the sum of squared errors from the regression. To the extent that our underlying model is biased, the regression fit is weakened and our estimate of σ increases. This will likely increase the rate of convergence. For example, local steady states with small basins of attraction will not hold the state variable over long periods when output is repeatedly perturbed by shocks.

6 Appendix A

Some definitions need to be made precise. Let X be any set, let $\mathscr{B}$ be a σ-field on X, and let μ be a σ-finite measure on $(X, \mathscr{B})$. Define $L_1 = L_1(X)$ to be the set of $\mathscr{B}$-measurable real functions on X with μ-finite integral. As usual, L_1 is given norm $\|g\| = \int |g| d\mu$ and pointwise ordering. (Functions equal almost everywhere are identified.) The space of densities $\mathscr{D} = \mathscr{D}(X)$ is all those elements of L_1 that are nonnegative and integrate to unity. Unless otherwise stated, the topology on L_1 is always the norm topology, and $\mathscr{D}$ inherets the same.

In general, by a stochastic kernel is meant a $\mathscr{B} \otimes \mathscr{B}$-measurable function on $X \times X$ with the property that $p(x, \cdot) \in \mathscr{D}$ for all $x \in X$. By the Markov operator associated with stochastic kernel p is meant the map $\mathbf{P} \colon L_1 \to L_1$ defined by

$$\mathbf{P}g(y') = \int p(y, y') g(x) \mu(dx). \tag{19}$$

It is well-known that $\mathbf{P}$ is a positive linear contraction on L_1 (Lasota and Mackey 1994). [17]

The first task is to prove Theorem 1. To do so we prove following result, which is a bit more general.

Theorem 4 *Let $(X, \mathscr{B}, \mu)$ be any σ-finite measure space, let (p_k) and p be stochastic kernels on $(X, \mathscr{B}, \mu)$, and let*

$$\varphi_t^k(y') := \frac{1}{N} \sum_{n=1}^{N} p_k^t(y_0^n, y') \quad \text{and} \quad \varphi_t(y') := \frac{1}{N} \sum_{n=1}^{N} p^t(y_0^n, y')$$

be the respective projections from fixed initial $(y_n)_{n=1}^N$ as defined by (4) and (5). If for some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we have $p_k(y, y') \to p(y, y')$ $\mathbb{P}$-a.s. as $k \to \infty$ for all (y, y') in $X \times X$, then

$$\|\varphi_t^k - \varphi_t\| \to 0 \;\; \mathbb{P}\text{-a.s as } k \to \infty \text{ for all } t \in \mathbb{N}.$$

The proof is broken down into several lemmata. In what follows, dx means $\mu(dx)$.

Lemma 5 *Let p_k and p be as in the statement of Theorem 4. Let $\mathbf{P}_k$ and $\mathbf{P}$ be their respective Markov operators. For all $g \in \mathscr{D}$, $\mathbf{P}_k g \to \mathbf{P}g$ as $k \to \infty$.*

[17] By positive is meant that $f \geq 0 \implies \mathbf{P}f \geq 0$. By contraction is meant that $\|\mathbf{P}f\| \leq \|f\|$, for all $f \in L_1$.

Proof We have

$$\begin{aligned}\|\mathbf{P}_k g - \mathbf{P} g\| &= \int \left| \int p_k(y,y')g(y)dy - \int p(y,y')g(y)dy \right| dy' \\ &\leq \int \int |p_k(y,y') - p(y,y')| g(y) dy dy' \\ &= \int \int |p_k(y,y') - p(y,y')| dy' g(y) dy.\end{aligned}$$

Define $h_k(y) := \int |p_k(y,y') - p(y,y')|dy'$. By assumption, $p_k(y,y') \to p(y,y')$ for all $(y,y') \in X \times X$.[18] Also, for fixed $y \in X$, $p_k(y,\cdot)$ and $p(y,\cdot)$ are densities, so that $p_k(y,\cdot) \to p(y,\cdot)$ in L_1 norm by Schéffe's Lemma. It follows that $h_k(y) \to 0$ in $\mathbb{R}$ for all $y \in X$. Thus to show that

$$\|\mathbf{P}_k g - \mathbf{P} g\| = \int h_k(y) g(y) dy \to 0 \text{ as } k \to \infty,$$

it is sufficient to verify that $h_k(y)$ is uniformly bounded in $\mathbb{R}$; as $g \in \mathscr{D}$ the result then follows from the Dominated Convergence Theorem. Since

$$h_k(y) = \int |p_k(y,y') - p(y,y')| dy' \leq \int p_k(y,y')dy' + \int p(y,y')dy' = 2$$

the result is confirmed.

We also need the following lemma.

Lemma 6 *Let $(\mathbf{P}_k)$ and $\mathbf{P}$ be Markov operators on L_1, and let $(g_k) \subset \mathscr{D}$, $g \in \mathscr{D}$. If $\mathbf{P}_k \to \mathbf{P}$ pointwise on $\mathscr{D}$ and $g_k \to g$, then $\mathbf{P}_k g_k \to \mathbf{P} g$.*

Proof Since all Markov operators are linear contractions on L_1,

$$\|\mathbf{P}_k g_k - \mathbf{P} g\| \leq \|\mathbf{P}_k g_k - \mathbf{P}_k g\| + \|\mathbf{P}_k g - \mathbf{P} g\| \leq \|g_k - g\| + \|\mathbf{P}_k g - \mathbf{P} g\|.$$

The result follows from the hypotheses.

In fact the result holds for any t.

Lemma 7 *Let $\mathbf{P}_k$, $\mathbf{P}$, g_k and g be as in Lemma* (6)*. For all $t \in \mathbb{N}$,*

$$\|\mathbf{P}_k^t g_k - \mathbf{P}^t g\| \to 0 \text{ as } k \to \infty.$$

Proof Since $\|\mathbf{P}_k^t g_k - \mathbf{P}^t g\| = \|\mathbf{P}_k(\mathbf{P}_k^{t-1} g_k) - \mathbf{P}(\mathbf{P}^{t-1} g)\|$ the result is true by induction and Lemma 6.

[18] Here and below we supress the notation $\mathbb{P}$-a.s. for simplicity.

Theorem 4 can now be proved. Let $\mathbf{P}_k$ and $\mathbf{P}$ be the Markov operators corresponding to p_k and p respectively. By (4) and (19),

$$p_k^t(y, y') = \int p(s, y') p^{t-1}(y, s) ds = \mathbf{P}_k p_k^{t-1}(y, y').$$

Repeating this argument t times gives $p_k^t(y, y') = \mathbf{P}_k^{t-1} p_k(y, y')$, and, similarly, $p^t(y, y') = \mathbf{P}^{t-1} p(y, y')$. We then have

$$\begin{aligned}
\|\varphi_t^k - \varphi_t\| &= \left\| \frac{1}{N} \sum_{n=1}^{N} p_k^t(y_n, \cdot) - \frac{1}{N} \sum_{n=1}^{N} p^t(y_n, \cdot) \right\| \\
&\leq \frac{1}{N} \sum_{n=1}^{N} \|p_k^t(y_n, \cdot) - p^t(y_n, \cdot)\| \\
&= \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{P}_k^{t-1} p_k(y_n, \cdot) - \mathbf{P}^{t-1} p(y_n, \cdot)\|.
\end{aligned}$$

As each $p_k(y_n, \cdot)$ and $p(y_n, \cdot)$ is a density, and the former converges to the latter in L_1 by Schéffe's Lemma, the desired convergence now follows from Lemma 7.

6.1 Stability

Next we give the following result, which proves stability of (12).

Lemma 8 *Let* $\mathbf{P}$ *be the Markov operator associated with stochastic kernel* (13). *Under the assumptions of Section 4,* $(\mathscr{D}, \mathbf{P})$ *is asymptotically stable in the sense of Definition 2.*

Proof Immediate from Stachurski (2003a, Theorem 1), for which all of the hypotheses are verifed.

Rather than prove the case of $(\mathscr{E}, \mathbf{Q})$ separately—consult, for example, Lasota and Mackey (1994, Chapter 10, especially p. 326)—we can establish it in the following way, which makes clear the relationship between the two systems and yields Part 1 of Proposition 3 along the way. Recall that two dynamical systems (X, S) and (Y, T) are called topologically conjugate if there exists a topological isomorphism τ from X to Y such that the two systems commute with τ, in the sense that $S = \tau^{-1} T \tau$ on X. [19] The next result is well-known.

Lemma 9 *Let* (X, S) *and* (Y, T) *be topologically conjugate under* $\tau \colon X \to Y$. *If* (X, S) *is asymptotically stable then so is* (Y, T). *If* x^* *is any stationary point for* S, *then* τx^* *is stationary for* T.

[19] By topological isomorphism is meant that τ is a continuous bijection with continuous inverse.

In the present case we seek a topological isomorphism $\tau \colon \mathscr{D} \to \mathscr{E}$ such that $\mathbf{P}$ and $\mathbf{Q}$ commute with τ.

Lemma 10 *Define* $\tau \colon \mathscr{D} \ni \varphi \mapsto \tau\varphi \in \mathscr{E}$ *by* $(\tau\varphi)(x) = \varphi(e^x)e^x$. *Then* τ *is a topological isomorphism and* $\mathbf{P} = \tau^{-1}\mathbf{Q}\tau$ *on* $\mathscr{D}$.

Proofs and further details are available in Stachurski (2003b). However, it is not difficult to confirm the result independently. For example, τ is not only continuous but an isometry from $L_1[0,\infty)$ to $L_1(\mathbb{R})$, because if $g \in L_1[0,\infty)$, then

$$\|\tau g\| = \int_{-\infty}^{\infty} |g(e^x)e^x| \, dx = \int_0^{\infty} |g(y)y| \frac{1}{y} \, dy = \int_0^{\infty} |g(y)| \, dy = \|g\|.$$

Parts 1 and 2 of Proposition 3 are now established: Combining Lemmas 8, 9 and 10, evidently $(\mathscr{E}, \mathbf{Q})$ is asymptotically stable, and, moreover, φ_q^*, the invariant distribution for $(\mathscr{E}, \mathbf{Q})$, satisfies $\varphi_q^*(x) = (\tau\varphi_p^*)(x) = \varphi_p^*(e^x)e^x$.

Incidentally, τ can be used for the finite t projections to move between those calculated using q and those calculated using p. In particular, it is easy to show via Lemma 10 that $\mathbf{P}^t = \tau^{-1}\mathbf{Q}^t\tau$ for any t. Hence a projection of one system can immediately be recovered from a projection of the other.

Regarding Part 3 of Proposition 3, this follows from the fact that both $\mathbf{P}$ and $\mathbf{Q}$ are nonexpansive.[20] In the case of $\mathbf{P}$, for example,

$$\|\mathbf{P}^t\varphi - \varphi_1^*\| = \|\mathbf{P}^t\varphi - \mathbf{P}^t\varphi_1^*\| \leq \|\mathbf{P}^{t-1}\varphi - \mathbf{P}^{t-1}\varphi_1^*\| = \|\mathbf{P}^{t-1}\varphi - \varphi_1^*\|.$$

References

[1] Acemoglu, D. and F. Zilibotti (1997): "Was Prometheus Unbound by Chance? Risk, Diversification and Growth," *Journal of Political Economy,* 105 (4), 709–751.

[2] Arthur, W. B. (1994): *Increasing Returns and Path Dependence in the Economy.* Ann Arbor: The University of Michigan Press.

[3] Azariadis, C. (1996): "The Economics of Poverty Traps," *Journal of Economic Growth,* 1, 449–486.

[4] Azariadis, C. and A. Drazen (1990): "Threshold Externalities in Economic Development," *Quarterly Journal of Economics,* 105, 501–526.

[5] Barro, R. (1991): "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics,* 106, 407–443.

[20] See Lasota and Mackey (1994, Proposition 3.1.1).

[6] Bianchi, M. (1997): "Testing for Convergence: Evidence from Non-Parametric Multimodality Tests," *Journal of Applied Econometrics,* 12, 393–409.

[7] Brock, W. A. and L. Mirman (1972): "Optimal Economic Growth and Uncertainty: The Discounted Case," *Journal of Economic Theory,* 4, 479–513.

[8] Casella, G. and R. L. Berger (1990): *Statistical Inference.* California: Duxbury Press.

[9] Cass, D. (1965): "Optimum Growth in an Aggregative Model of Capital Accumulation," *Review of Economic Studies,* 32, 233–240.

[10] Chan, K. S. and Tong, H. (1986): "On Estimating Thresholds in Autoregressive Models," *Journal of Time Series Analysis,* 7, 179–90.

[11] De Long, B. (1988): "Productivity Growth, Convergence and Welfare: Comment," *American Economic Review,* 78, 1138–54.

[12] Durlauf, S. N. (1993): "Nonergodic Economic Growth," *Review of Economic Studies,* 60, 349–366.

[13] Durlauf, S. N. and P. A. Johnson (1995): "Multiple Regimes and Cross-Country Growth Behavior", *Journal of Applied Econometrics,* 10 (4), 365-384.

[14] Durlauf, S. N. and D. Quah (1999): "The New Empirics of Economic Growth," in *Handbook of Macroeconomics,* ed. by J. B. Taylor and M. Woodford. North-Holland: Elsevier Science, 231–304.

[15] Galor, O. (1996): "Convergence? Inferences from Theoretical Models," *Economic Journal,* 106, 1056–1069.

[16] Glynn, P. W. and S. G. Henderson (2001): "Computing Densities for Markov Chains via Simulation," *Mathematics of Operations Research,* 26, 375–400.

[17] Gollin, D. (2002): "Getting Income Shares Right," *Journal of Political Economy,* 110 (2), 458–474.

[18] Graham, B. S. and J. Temple (2001): "Rich Nations Poor Nations: How Much Can Multiple Equilibria Explain?" *Center for International Development Working Paper,* 76, Harvard.

[19] Heston, A., R. Summers and B. Aten (2002): "Penn World Table Version 6.1." Center for International Comparisons, University of Pennsylvania.

[20] Ihaka, R. and R. Gentleman (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics,* 5 (3), 299–314.

[21] Johnson, P. A. (2000): "A Nonparametric Analysis of Income Convergence Across the US States," *Economics Letters,* 69, 219–223.

[22] Jones, C. I. (1997): "On the Evolution of the World Income Distribution," *Journal of Economic Perspectives,* 11, 3, 19–36.

[23] Koopmans, T. (1965): "On the Concept of Optimal Economic Growth," *Pontificae Academiae Scientiarum Scripta Varia,* 28, 225–300.

[24] Kremer, M., A. Onatski and J. Stock (2001): "Searching for Prosperity," *Carnegie Rochester Conference Series on Public Policy,* 55, 275–303.

[25] Lasota, A. and M. Mackey (1994): *Chaos, Fractals and Noise: Stochastic Aspects of Dynamics,* Applied Mathematical Sciences Vol. 97, New York: Springer-Verlag

[26] Mankiw, G., D. Romer and D. N. Weil (1992): "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics,* 107, 407-437.

[27] Masanjala, W and C. Papageorgiou (2004): "Parameter Heterogeneity and Nonlinearities in the Aggregate Production Function: Investigating the Solow Growth Model with CES Technology," *Journal of Applied Econometrics,* forthcoming.

[28] Matsuyama, K. (1991): "Increasing Returns, Industrialization and Indeterminacy of Equilibrium," *Quarterly Journal of Economics,* 106, 617–650.

[29] Murphy, K. M., A. Shleifer and R. W. Vishny (1989): "Industrialization and the Big Push," *Journal of Political Economy,* 97, 1003-1026.

[30] Nelson, R. R. (1956): "A Theory of the Low Level Equilibrium Trap", *American Economic Review,* 46, 894–908.

[31] Nurkse, R (1953): *Problems of Capital-Formation in Underdeveloped Countries,* 1962 edition, New York: Oxford Univeristy Press.

[32] Parente, S. L. and E. C. Prescott (1993): "Changes in the Wealth of Nations," *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring 1993, 3–14.

[33] Quah, D. T. (1993): "Empirical Cross-section Dynamics in Economic Growth," *European Economic Review,* 37, 426-434.

[34] Quah, D. (1996): "Convergence Empirics Across Economies with (some) Capital Mobility," *Journal of Economic Growth,* 1, 95–124.

[35] Rosenstein-Rodan, P. (1943): "The Problem of Industrialization of Eastern and South-Eastern Europe," *Economic Journal,* Vol. 53, p.202-11

[36] Sala-i-Martin, X. (1996): "The Classical Approach to Convergence Analysis," *Economic Journal,* 106, 437, 1019–1036.

[37] Solow, R. M. (1956): "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics,* 70, 65–94.

[38] Stachurski, J. (2002): "Stochastic Optimal Growth with Unbounded Shock," *Journal of Economic Theory,* 106, 45–60.

[39] Stachurski, J. (2003a): "Stochastic Growth with Increasing Returns: Stability and Path Dependence," *Studies in Nonlinear Dynamics and Econometrics,* in press.

[40] Stachurski, J. (2003b): "Log-linearization of Stochastic Economic Models," Kyoto Institute of Economic Research Discussion Paper No. 564.

[41] Young, A. A. (1928): "Increasing Returns and Economic Progress," *Economic Journal,* 28.

[42] Young, A. (1993): "Substitution and Complementarity in Endogenous Innovation," *Quarterly Journal of Economics,* 108 (3), 775–807.