

Occupational Choice and the Process of Development

Abhijit V. Banerjee

Harvard University

Andrew F. Newman

Northwestern University

This paper models economic development as a process of institutional transformation by focusing on the interplay between agents' occupational decisions and the distribution of wealth. Because of capital market imperfections, poor agents choose working for a wage over self-employment, and wealthy agents become entrepreneurs who monitor workers. Only with sufficient inequality, however, will there be employment contracts; otherwise, there is either subsistence or self-employment. Thus, in static equilibrium, the occupational structure depends on distribution. Since the latter is itself endogenous, we demonstrate the robustness of this result by extending the model dynamically and studying examples in which initial wealth distributions have long-run effects. In one case the economy develops either widespread cottage industry (self-employment) or factory production (employment contracts), depending on the initial distribution; in the other example, it develops into prosperity or stagnation.

I. Introduction

Why does one country remain populated by small proprietors, artisans, and peasants while another becomes a nation of entrepreneurs employing industrial workers in large factories? Why should two

We are grateful to Tim Besley, Allan Drazen, Drew Fudenberg, Maitreesh Ghatak, Rodi Manuelli, Kiminori Matsuyama, Joel Mokyr, Andrei Shleifer, Nancy Stokey, Aldo Rustichini, Peter Streufert, Chris Udry, an anonymous referee, and the editors of this *Journal* for helpful discussion and comments.

[*Journal of Political Economy*, 1993, vol. 101, no. 2]

© 1993 by The University of Chicago. All rights reserved. 0022-3808/93/0102-0007\$01.50

seemingly identical countries follow radically different development paths, one leading to prosperity, the other to stagnation? Questions like these are of central concern to both development economists and economic historians, who have been interested in the study of the evolution of institutional forms, particularly those under which production and exchange are organized. Yet most of these institutional questions have resisted formal treatment except in a static context (see Stiglitz [1988] for a review), whereas the dynamic issues that are peculiarly developmental have for the most part been restricted to the narrower questions of output growth or technical change. This paper takes a first step in the direction of providing a dynamic account of institutional change by focusing on the evolution of occupational patterns, the contractual forms through which people exchange labor services.¹

There are several ways in which the dynamics of occupational choice influence the process of development. Most obvious among them is the effect on the distribution of income and wealth. Insofar as distribution can affect saving, investment, risk bearing, fertility, and the composition of demand and production, there is a clear link with the economy's rate of growth and hence with development in its narrowest sense.

Just as important is the connection that arises when one considers development to mean institutional transformation as well as economic growth (Stiglitz 1988; Townsend 1988; Khan 1989). One of the most significant elements of the institutional structure of any economy is the dominant form of organization of production: it has "external" consequences considerably beyond the efficiency of current production. Some of these effects may be politico-economic, but there are also some that are purely economic. It has been argued, for example, that the introduction of the factory system in the early years of the Industrial Revolution left the technology unaffected and generated little efficiency gain initially. But it seems very likely that in the long run this new form of production organization helped to make possible the major innovations of the Industrial Revolution (see, e.g., Cohen 1981; Millward 1981; North 1981).

Conversely, the process of development also affects the structure of occupations. It alters the demand for and supply of different types of labor and, hence, the returns to and allocations of occupations. It transforms the nature of risks and the possibilities for innovations. And, of course, it changes the distribution of wealth. Since one's wealth typically affects one's incentives to enter different occupations,

¹ We use the term "occupation" to mean a contractual arrangement rather than a productive activity. A bricklayer and an accountant are in the same occupation if each is an independent contractor or if each works for a wage.

the effect on the wealth distribution generates a parallel effect on the occupational structure.

Our aim here is to build a model that focuses directly on this interplay between the pattern of occupational choice and the process of development. The basic structure of interaction is very simple. Because of capital market imperfections, people can borrow only limited amounts. As a result, occupations that require high levels of investment are beyond the reach of poor people, who choose instead to work for other, wealthier, employers; thus wage contracts are viewed primarily as substitutes for financial contracts. The wage rate and the pattern of occupational choice are then determined by the condition that the labor market must clear.² Depending on labor market conditions and on their wealth, other agents become self-employed in low-scale production or remain idle.

The pattern of occupational choice is therefore determined by the initial distribution of wealth, but the structure of occupational choice in turn determines how much people save and what risks they bear. These factors then give rise to the new distribution of wealth. We shall be concerned with the long-run behavior of this dynamic process.

Despite its simplicity, our model's structure is somewhat nonstandard. As a rule, the dynamics are nonlinear and the state space—the set of all wealth distributions—is very large, so that reasonably complicated behavior may be expected. While a complete mathematical analysis of the model is beyond the scope of this paper, we confine our attention to two special cases that admit considerable dimensional reduction. These examples afford complete study: they are simple enough to allow diagrammatic exposition in which we trace out entire paths of development, including institutional evolution, and with them we generate robust and natural instances of hysteresis or long-run dependence on initial conditions.

In one of our examples (Sec. *IVD*), the ultimate fate of the economy—prosperity or stagnation—depends in a crucial way on the initial distribution of wealth. If the economy initially has a high ratio of very poor people to very rich people, then the process of development runs out of steam and ends up in a situation of low employment and low wages (this may happen even when the initial per capita income is quite high, as long as the distribution is sufficiently skewed). By contrast, if the economy initially has few very poor people (the per capita income can still be quite low), it will “take off” and converge to a high-wage, high-employment steady state.

² This static model of occupational choice is a simplified version of the one in Newman (1991), which also discusses the advantages of the capital market imperfections approach over preference-based approaches such as that of Kihlstrom and Laffont (1979). See also the related work of Eswaran and Kotwal (1989).

That an economy's long-term prosperity may depend on initial conditions is a familiar idea in the development literature, and some recent papers capture different aspects of this phenomenon in a formal model (e.g., Romer 1986; Lucas 1988; Murphy, Shleifer, and Vishny 1989a, 1989b; Matsuyama 1991; Galor and Zeira, in press). Our paper differs from these in several respects. First, most of the papers study technological increasing returns, originating either in the production technology itself or in various kinds of productivity spillovers. We consider instead a kind of "pecuniary" increasing returns stemming from an imperfect capital market (Galor and Zeira also follow this tack). Second, distribution tends not to play a causal role in this literature. A notable exception is Murphy et al. (1989a), but there the mechanism is the structure of demand for produced commodities rather than the occupational choice mediated by the capital market: moreover, their model is static and therefore does not endogenize the distribution.

Third, and most important, none of these papers emphasizes the endogeneity of economic institutions as part of the process of development. This distinction is highlighted by the example we examine in Section IVC, in which there appears a different kind of dependence on initial conditions. We show that the economy might converge to a steady state in which there is (almost) only self-employment in small-scale production; alternatively, it may end up in a situation in which an active labor market and both large- and small-scale production prevail. Which of the two types of production organization eventually predominates once again depends on the initial distribution of wealth. Specifically, an economy that starts with a large number of relatively poor people is more likely to develop wage employment and large-scale production than an economy with few very poor people. This result provides a formalization of the classical view that despite the fact that capitalism is the more dynamic economic system, its initial emergence does depend on the existence of a population of dispossessed whose best choice is to work for a wage.

In Section II we set up the basic model. Section III examines single-period equilibrium. The main results on the dynamics of occupational choice and the process of development are in Section IV. We conclude in Section V with a brief discussion of some qualitative properties of this class of models.

II. The Model

A. Environment

There is a large population (a continuum) of agents with identical preferences; the population at time t is described by a distribution

function $G_t(w)$, which gives the measure of the population with wealth less than w .

At the beginning of life, agents receive their initial wealth in the form of a bequest from their parents. They also have an endowment of one unit of labor; the effort they actually exert, however, is not observable except under costly monitoring by another agent.

When agents become economically active, they may apply for a loan. Enforcement of loan contracts is imperfect, and agents immediately have an opportunity to renege; lenders will limit borrowing and require collateral in order to ensure that agents do not. The agents choose an occupation, which determines how they invest their labor and capital. They then learn investment outcomes and settle outside claims. Finally, they bequeath to their children, consume what remains, and pass from the scene.

Although the model is naturally recursive, we prefer to study dynamics in continuous time and to impose an overlapping demographic structure. These modifications permit us to avoid unrealistic jumps and overshooting, which can arise as artifacts of discrete time and simultaneous demographics. We therefore shall assume that all the economic activity other than inheritance—borrowing, investment, work, and bequests—takes place at the instant the agents reach maturity. The age of maturity in turn is distributed exponentially with parameter λ across the population and independently from wealth.³ The total population is stationary and is normalized to unity; that is, a cohort of size λ is active at each instant.

These assumptions, though artificial, greatly simplify the analysis. For instance, they imply that in an interval of time dt , a measure $\lambda G_t(w)dt$ of agents with wealth below w are active: the measure of active agents in a wealth interval is always proportional to the measure of the entire (immature) population in that interval. Thus differential changes in the wealth distribution at each instant will depend only on the current distribution. Moreover, the differential dynamics will be related to the recursive dynamics in a transparent manner so that it will be easy to switch attention from the (recursive) dynamics of a lineage to the (continuous) dynamics of the economy.

Agents are risk-neutral: preferences over commodities are represented by $c^\gamma b^{1-\gamma} - z$, where c is an agent's consumption of the sole physical good in the economy, b is the amount of this good left as a bequest to his offspring (the "warm glow" [Andreoni 1989] is much more tractable than other bequest motives), and z is the amount of

³ That is, an agent born at s is "immature" with probability $e^{\lambda(s-t)}$ at time $t > s$ ($1/\lambda$ is the average age of maturity of the population). These demographics resemble those in Blanchard (1985), although he does not assume instantaneous economic activity.

labor he supplies. Denote the income realization by y ; utility then takes the form $\delta y - z$, where $\delta \equiv \gamma^\gamma(1 - \gamma)^{1-\gamma}$.

B. *Production Technology and Occupations*

The economy's single good may be used for consumption or as capital. There are three ways to invest. First, there is a divisible, safe asset that requires no labor and yields a fixed gross return $\hat{r} < 1/(1 - \gamma)$.⁴ One may think of it as financial claims mediated by foreign banks that borrow and lend at the fixed international interest rate $\hat{r} - 1$.⁵ Agents may invest in this asset regardless of how they use their labor. Anyone who invests only in the safe asset is said to be idle or to be subsisting.

Second, there is a risky, indivisible investment project such as a farm or machine that requires no special skill to operate. To succeed, it must have an initial investment of I units of capital and one unit of labor; with any lower level of either input, it will not generate any returns. If the project succeeds, it generates a random return rI , where r is r_0 or r_1 with probabilities $1 - q$ and q , respectively ($0 < r_0 < r_1$), and has mean \bar{r} . Such a project may be operated efficiently by a self-employed agent insofar as it produces enough output to cover its labor cost: $I(\bar{r} - \hat{r}) - (1/\delta) \geq \max\{0, I(r_0 - \hat{r})\}$.

Finally, there is a monitoring technology that permits aggregated production. By putting in an effort of one, one entrepreneur can perfectly monitor the actions of $\mu > 1$ individuals; less effort yields no information. This activity is indivisible, and it is impossible to monitor another monitor.

Using this technology, an entrepreneur can hire μ workers, each at a competitive wage v . Workers undertake projects that require I' units of capital and one unit of labor and generate random returns $r'I'$; r' takes on the values r'_0 and r'_1 (also with $0 < r'_0 < r'_1$) with probabilities $1 - q'$ and q' . It is natural to imagine that the projects individual workers are running are similar to the projects being run by the self-employed. To facilitate this interpretation, we assume that $I' = I$ and that r' and r have the same mean (note that $q' \neq q$, however). The returns on each of the projects belonging to a single entrepreneur are perfectly correlated. Entrepreneurial production is feasible in the sense that at the lowest possible wage rate (which is $1/\delta$, since at a lower wage the worker is better off idle) it is more

⁴ The restriction on the safe return ensures that the long-run dynamics are reasonable in the sense that people's wealth levels do not grow without bound.

⁵ Of course, \hat{r} might instead represent the return to some physical subsistence activity that requires wealth but no effort; arbitrage considerations then dictate that this also be the return on loans.

profitable than self-employment: $\mu[I(\bar{r} - \hat{r}) - (1/\delta)] - (1/\delta) \geq \max\{I(\bar{r} - \hat{r}) - (1/\delta), \mu[I(r'_0 - \hat{r}) - (1/\delta)]\}$.

The main difference between the two types of production lies not so much in the technology but rather in the contracts under which output is distributed. In one, the worker runs a project for himself: he is the claimant on output and therefore needs no monitoring. In the other, the worker runs it for someone else, which entails the monitoring function of the entrepreneur.

To summarize, there are four occupational options: (1) subsistence, (2) working, (3) self-employment, and (4) entrepreneurship. There may be a question of how we rule out other possibilities. Entrepreneurs cannot control more than μ projects because one cannot monitor a monitor. Being a part-time entrepreneur (sharing with someone else) is ruled out by the indivisible monitoring technology and in any case would not be attractive because of risk neutrality. Raising capital through partnership is precluded by the same contract enforcement problems that exist between the bank and borrowers: one partner could as easily default on another partner as default on the bank (thus without loss of generality we need consider only debt and can ignore equity). The same arguments rule out combining self-employment with any other activity.

C. Markets

In the market for labor, demand comes from entrepreneurial production and supply from individuals' occupational choices. This market is competitive, with the wage moving to equate supply and demand. The goods market is competitive as well, but it is otherwise pretty trivial.

It remains to discuss the market for loans. We assume that lenders can enter freely; what distinguishes this market is the possibility that a borrower might renege on a debt. The story we have in mind is similar to that proposed by Kehoe and Levine (in press). To abstract from bankruptcy issues, assume that project returns are always high enough to ensure that borrowers can afford repayment. Suppose that an agent puts up all his wealth w (the maximum he can provide) as collateral and borrows an amount L . He may now attempt to avoid his obligations by fleeing from his village, albeit at the cost of lost collateral $w\hat{r}$; flight makes any income accruing to the borrower inaccessible to lenders. Fleeing does not diminish investment opportunities, however, and having L in hand permits the agent to achieve $V(L)$ in expected gross income net of effort (under our assumptions, his ensuing decisions and therefore $V(L)$ are independent of his choice whether to renege). At the end of the production period, he will have

succeeded in escaping the lender's attempts to find him with a large probability $1 - \pi$, in which case he avoids paying $L\hat{r}$. Should he be caught, though, he will have had ample time to dispose of his income, and therefore he can be subjected to only a nonmonetary punishment F (such as flogging or imprisonment), which enters additively into his utility. Reneging therefore yields a payoff of $V(L) - \pi F$, and repaying yields $V(L) + w\hat{r} - L\hat{r}$; the borrower will renege whenever $w\hat{r} + \pi F < L\hat{r}$. Knowing this, lenders will make only loans that satisfy $L \leq w + (\pi F/\hat{r})$. All loans made in equilibrium will satisfy this constraint, and the borrower will never renege.⁶

The only reason to borrow in this model is to finance self-employment or entrepreneurship. The target levels of capital are therefore I and μI (we assume that wages are paid at the end of the period so there is no need to finance them). Someone with a wealth level $w < I$ who wants to become self-employed therefore uses w as collateral and needs to borrow I .⁷ He will be able to borrow this amount if and only if $I \leq w + (\pi F/\hat{r})$. Thus the minimum wealth level w^* necessary to qualify for a loan large enough to finance self-employment is equal to $I - (\pi F/\hat{r})$ (the escape probability $1 - \pi$ is large enough that $w^* > 0$). The smallest wealth needed to borrow enough to be an entrepreneur, denoted w^{**} , is derived by a parallel argument and is equal to $\mu I - (\pi F/\hat{r})$. Since μ exceeds unity, w^{**} is greater than w^* ; moreover, neither of these values depends on the wage.

The model of the capital market we have chosen here yields a rather extreme version of increasing returns to wealth. In effect, it is not terribly different from the models of Sappington (1983) and Bernanke and Gertler (1989, 1990) or the numerous discussions of credit markets in the development literature (see Bell [1988] for a survey). Using such models would not alter the dependence of borrowing costs on wealth or of occupational structure on distribution. But as we shall see, the present model is simple enough in some cases to allow reduction to a dynamical system on the two-dimensional simplex, a procedure that would be impossible with a more elaborate specification.

III. Static Equilibrium

Recall that the distribution of wealth at time t is denoted by $G_t(w)$ and that because the age to maturity is exponentially distributed and

⁶ An alternative interpretation is that πF is equal to a moving cost incurred by the borrower when he flees, with no chance for the lender to catch him.

⁷ By using all his wealth as collateral, the borrower maximizes the size of the loan he can obtain.

independent of wealth, $\lambda G_t(w)$ represents the distribution of wealth for the cohort active at t . The (expected) returns to self-employment and subsistence are given exogenously by the model's parameters; the wage v determines the returns to the other two occupations. The returns and the borrowing constraints determine the occupational choice made at each level of wealth. Integrating these choices with respect to $\lambda G_t(w)$ gives us the demand for and the supply of labor. To find the instantaneous equilibrium, we need only find the wage that clears the labor market (we can assume that the goods market clears; as for the capital market, the interest rate has already been fixed at \hat{r}).

All agents who do not choose subsistence will have the incentive to expend full effort. Therefore, the payoffs to each occupation (for someone who can choose any of them) are subsistence, $\delta w \hat{r}$; worker, $\delta(w \hat{r} + v) - 1$; self-employed, $\delta[w \bar{r} + I(\bar{r} - \hat{r})] - 1$; and entrepreneur, $\delta[w \hat{r} + \mu I(\bar{r} - \hat{r}) - \mu v] - 1$. Since only entrepreneurs demand labor, these expressions imply that demand will be positive only if the wage does not exceed $\bar{v} \equiv [(\mu - 1)/\mu]I(\bar{r} - \hat{r})$. Moreover, since only agents with $w \geq w^{**}$ will be entrepreneurs, the labor demand correspondence is

$$\begin{aligned} 0 & \quad \text{if } v > \bar{v}, \\ [0, \mu\lambda[1 - G_t(w^{**})]] & \quad \text{if } v = \bar{v}, \\ \mu\lambda[1 - G_t(w^{**})] & \quad \text{if } v < \bar{v}. \end{aligned}$$

Similar reasoning tells us that the supply of labor is (denote the minimum wage $1/\delta$ by \underline{v})

$$\begin{aligned} 0 & \quad \text{if } v < \underline{v}, \\ [0, \lambda G_t(w^*)] & \quad \text{if } v = \underline{v}, \\ \lambda G_t(w^*) & \quad \text{if } \underline{v} < v < I(\bar{r} - \hat{r}), \\ [\lambda G_t(w^*), \lambda] & \quad \text{if } v = I(\bar{r} - \hat{r}), \\ \lambda & \quad \text{if } v > I(\bar{r} - \hat{r}). \end{aligned}$$

The equilibrium wage will be \underline{v} if $G_t(w^*) > \mu[1 - G_t(w^{**})]$ and \bar{v} if $G_t(w^*) < \mu[1 - G_t(w^{**})]$. The singular case in which $G_t(w^*) = \mu[1 - G_t(w^{**})]$ gives rise to an indeterminate wage in $[\underline{v}, \bar{v}]$. The facts that the wage generically assumes one of only two values, that it depends on no more information about the distribution $G_t(\cdot)$ than its value at w^* and w^{**} , and that w^* and w^{**} do not depend on any endogenous variables of the model are the keys to the dimensional reduction that so simplifies our analysis below.

To summarize, the pattern of occupational choice that is generated in equilibrium is as follows: (1) Anyone with initial wealth less than w^* will be a worker unless wages are exactly \underline{v} , in which case the labor market clears by having some of the potential workers remain idle. (2) Agents with initial wealth between w^* and w^{**} will become self-employed; although they could choose working, they would do so only if $v \geq I(\bar{r} - \hat{r})$, which cannot occur in equilibrium. (3) Anybody who starts with wealth at or above w^{**} will be an entrepreneur as long as $v < \bar{v}$. If $v = \bar{v}$, all the potential entrepreneurs are equally happy with self-employment, so $1 - [G_t(w^*)/\mu] - G_t(w^{**})$ of them opt for the latter, and the labor market clears.

Thus despite the fact that everybody has the same abilities and the same preferences, different people choose different occupations. What is more, the occupational choices made by individuals depend on the distribution of wealth. For example, if everyone is above w^* , everyone will be self-employed. Employment contracts emerge only if some people are below w^* and others are above w^{**} . With everyone below w^* , subsistence becomes the only option. Thus, as in Newman (1991), the institutional structure of the economy, represented by the pattern of occupations, depends on the distribution of wealth.⁸ The question, of course, is whether this dependence of institutional structure on distribution that obtains in the short run also obtains in the long run, when the distribution itself is endogenous.

IV. Dynamics

We have described how the equilibrium wage and occupational choices at time t are determined, given an initial wealth distribution. Knowledge of the realization of project returns then gives us each person's income and bequests, from which we can calculate the rate of change of this distribution.

A. Individual Dynamics

A person active at t leaves $1 - \gamma$ of his realized income as a bequest b_t . The intergenerational evolution of wealth is then represented as follows: (1) subsistence: $b_t = (1 - \gamma)w_t\hat{r}$; (2) working: $b_t = (1 - \gamma)(w_t\hat{r} + v)$; (3) self-employment: $b_t = (1 - \gamma)[w_t\hat{r} + I(r - \hat{r})]$, which is

⁸ So does static efficiency. In this model, a first-best Pareto optimum is achieved only when everyone is self-employed. Even though the employment contract is optimal from the point of view of the parties involved, an equilibrium with employment contracts cannot be first-best efficient (some resources are being spent on monitoring instead of direct production).

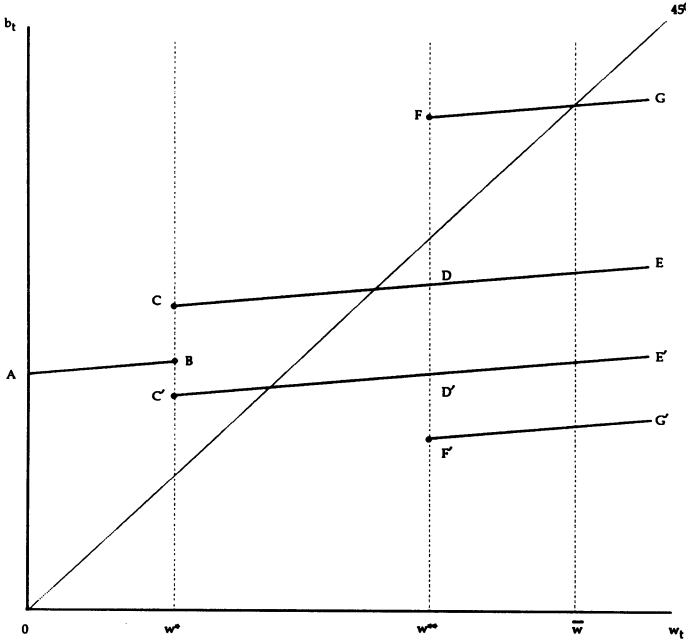


FIG. 1.—Individual recursion diagram for $v = \bar{v}$

random; and (4) entrepreneurship: $b_t = (1 - \gamma)\{w_t \hat{r} + \mu[I(r' - \hat{r}) - v]\}$, also random.

The transition diagram in figure 1 represents the dynamics of lineage wealth for the case $v = \bar{v}$. Everybody with wealth between zero and w^* will choose working, and their offspring's wealth as a function of their own wealth is given by the line segment AB . Agents between w^* and w^{**} will be self-employed, and their wealth dynamics are given by the two parallel lines CD and $C'D'$, each indicating one realization of the random variable r . Since the wage is \bar{v} , everyone above w^{**} will either be an entrepreneur or be self-employed; the two parallel lines DE and $D'E'$ represent the dynamics for a self-employed person and FG and $F'G'$ represent those for an entrepreneur.

A similar diagram can be constructed for the case in which $v = v$. The specific positions of the different lines in these diagrams depend, of course, on the parameters of the model.

B. The Dynamics of Distribution and Occupational Choice

From the point of view of an individual lineage, wealth follows a Markov process. If this process were stationary, we could go ahead

and use the standard techniques (see, e.g., Stokey and Lucas 1989) to establish existence and global stability of an ergodic measure on the wealth space and, since we are assuming a continuum of agents, reinterpret this to be the limiting wealth distribution for the economy. Under the stationarity assumption, one can study Markov processes by considering (deterministic) maps from the space of distributions to itself; such maps are well known to be linear.

In our model, however, the stationarity assumption is not justified. At the time a lineage is active, its transition rule depends on the prevailing wage. The wage in turn depends on the current distribution of wealth across all active agents in the economy (which, as we have said, is the same as that for the entire population); as the distribution changes over time, so does the wage, thereby destroying the stationarity of the process.

In short, the state space for our model is not simply the wealth interval, but the set of distributions on that interval: this is the smallest set that provides us with all the information we need to fully describe the economy and predict its path through time. We have already shown that given the current distribution of wealth, we can determine the equilibrium level of wages and the pattern of occupational choices. Then, using the transition equations, the current distribution of wealth $G_t(\cdot)$, and the fact that we have a large number of agents receiving independent project returns, we can in principle derive the (deterministic) change in the distribution of wealth at time t . We therefore have a well-defined, deterministic, dynamical system on the space of wealth distributions.

Ordinarily, the dynamical system so derived may be quite complex, and unlike a system induced by the familiar stationary Markov process, which is defined on the same space, it is nonlinear. The nonlinearity already tells us that uniqueness, global stability, and other nice, easy-to-verify properties of linear systems are unlikely to obtain. But we want to say more about our economy than to simply state abstractly that it might display hysteresis, nonuniqueness, cycles, or other nonlinear behavior.⁹

Fortunately, if we restrict attention to certain sets of parameter values, we can achieve a rather precise characterization of the economy's behavior using methods that are elementary. In the rest of this section we shall look at two examples that obtain when the individual transition diagrams like figure 1 have certain configurations; these

⁹ As this article was going to press, we became aware of the work of Conlisk (1976) on interactive Markov chains, to which our model is closely related. His results do not apply to our case, however.

cases are illustrative of interesting historical patterns of development and occupational structure.

C. *The Cottage versus the Factory*

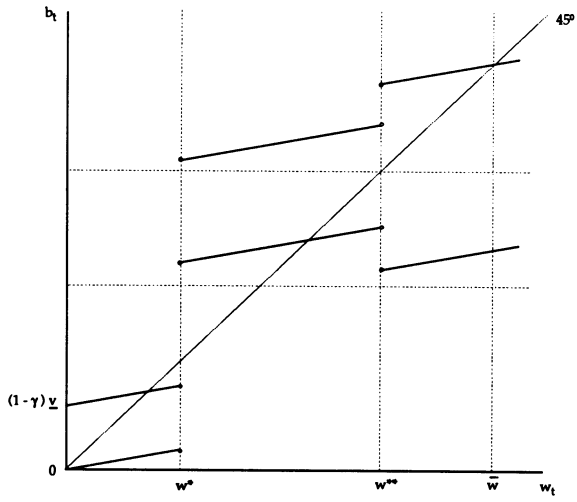
Consider the case in which the transition diagrams for $v = \underline{v}$ and $v = \bar{v}$ are given by figure 2a and b. The configuration represented in these diagrams will obtain when \bar{v} is relatively high, $1 - \gamma$ is relatively low, and the riskiness of production (given by $r_1 - r_0$ and $r'_1 - r'_0$) is quite large.

Look now at figure 2a. Define \bar{w} to be the fixed point of the intergenerational wealth transition map $b(w_t) = (1 - \gamma)\{w_t\hat{r} + \mu[I(r'_1 - \hat{r}) - v]\}$, and observe that this is the highest possible wealth level that can be sustained in the long run (any lineage with wealth greater than this value is sure to fall below it eventually). Without loss of generality then, we restrict all our attention to wealth distributions on the interval $[0, \bar{w}]$.

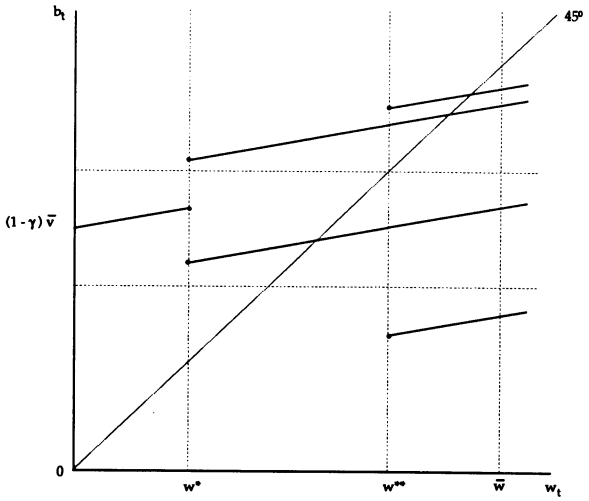
Observe now that in figure 2a, a lineage currently with wealth in $[0, w^*)$ remains in that range in the next period. Any lineage initially in $[w^*, w^{**})$ either goes to $[w^{**}, \bar{w}]$ (if the project return is high) or remains in $[w^*, w^{**})$ (if the project return is low). Finally, the offspring of an agent who is in $[w^{**}, \bar{w}]$ either remains there (if lucky) or goes to $[w^*, w^{**})$ (if unlucky). The important point is that these transitions depend only on what interval one is in and not on the precise wealth level within that interval. Similarly, inspection of figure 2b shows that when the prevailing wage is \bar{v} , the transitions between the same three intervals also depend only on those intervals and not on the wealth levels within them.

As we showed in Section III, the equilibrium wage and the occupational structure depend only on the ratio of the number of people in $[0, w^*)$ and the number of people in $[w^{**}, \bar{w}]$, and not on any other properties of the distribution. Identify the three intervals $[0, w^*)$, $[w^*, w^{**})$, and $[w^{**}, \bar{w}]$ with three "classes" L , M , and U (for lower, middle, and upper); wealth distributions (fractions of the population in the three classes) are then given by probability vectors $\mathbf{p} = (p_L, p_M, p_U)$, that is, points in Δ^2 , the two-dimensional unit simplex. The state space for our economy is then just this simplex: for our purposes, it contains all the information we need.¹⁰

¹⁰ Thus if $G(\cdot)$ is the current wealth distribution, then $p_L = G(w^*)$, $p_M = G(w^{**}) - G(w^*)$, and $p_U = 1 - G(w^{**})$. Of course, some information is lost by our dimensional reduction: if $H(\cdot)$ is another distribution with $H(w^*) = G(w^*)$ and $H(w^{**}) = G(w^{**})$, then it will be indistinguishable from $G(\cdot)$, even if the two distributions have different means. The limits to which they converge will generally differ as well but will be equal at w^* and w^{**} .



(a)



(b)

FIG. 2.—a, $v = \underline{v}$. b, $v = \bar{v}$

Now suppose that at some instant t , $\lambda p_L > \mu \lambda p_U$ so that there is excess supply in the labor market and $v = \underline{v}$. In an interval of time dt a measure $\lambda p_U dt$ of the current upper class is active. The people in this class are replaced by their children, of whom a fraction q' will have parents who are lucky with their investment and therefore remain in the upper class. Among the children in the currently active middle class, q have lucky parents and ascend into the upper class. The change in the upper-class population in this interval is therefore

$$dp_U = \lambda(qp_M dt + q'p_U dt - p_U dt).$$

The evolution of the entire wealth distribution can be represented by a dynamical system on Δ^2 , which may be written

$$\frac{d\mathbf{p}}{dt} = \mathbf{A}(\mathbf{p}(t))\mathbf{p}(t), \quad (1)$$

where $\mathbf{A}(\mathbf{p}(t))$ is a 3×3 matrix that depends on the current distribution $\mathbf{p}(t)$ in the sense that it takes two different forms depending on whether p_L is greater or less than μp_U . If $\lambda p_L > \mu \lambda p_U$, so that $v = \underline{v}$, then we have (for brevity, we set $\lambda = 1$ for the remainder of the paper)

$$\mathbf{A}(\mathbf{p}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -q & 1 - q' \\ 0 & q & q' - 1 \end{bmatrix}, \quad p_L > \mu p_U. \quad (2)$$

For the case $v = \bar{v}$, the situation is slightly more complicated since the individual transition probabilities for members of the class U depend on their occupation:

$$\mathbf{A}(\mathbf{p}) = \begin{bmatrix} -1 & 0 & (1 - q')p_L/\mu p_U \\ 1 & -q & (1 - q)[1 - (p_L/\mu p_U)] \\ 0 & q & q + (q' - q)(p_L/\mu p_U) - 1 \end{bmatrix}, \quad p_L < \mu p_U. \quad (3)$$

The third column of this matrix is derived by noting that $p_L/\mu p_U$ of the agents with wealth greater than w^{**} become entrepreneurs; of these, q' get the high return and remain above w^{**} , and $1 - q'$ fall below w^* ; the remaining agents in U become self-employed and enter L and U in the proportions $1 - q$ and q .

Now it will be convenient to study the dynamics of our economy by using a phase diagram; to do so we restrict our attention to the two variables p_L and p_U , since knowledge of them gives us p_M . This

procedure gives us a piecewise-linear system of differential equations:

$$\dot{p}_L = \begin{cases} 0, & p_L > \mu p_U \\ \left(\frac{1 - q'}{\mu} - 1 \right) p_L, & p_L < \mu p_U \end{cases} \quad (4)$$

and

$$\dot{p}_U = \begin{cases} q - qp_L + (q' - q - 1)p_U, & p_L > \mu p_U \\ q - \left(q + \frac{q}{\mu} - \frac{q'}{\mu} \right) p_L - p_U, & p_L < \mu p_U. \end{cases} \quad (5)$$

The phase diagram for this set of differential equations is given in figure 3a. The upper triangle represents distributions for which $v = \bar{v}$, and the lower triangle represents those for which $v = \underline{v}$. The heavy line is the “boundary” $p_L = \mu p_U$ between the two linear systems.¹¹

In the upper triangle the point C represents a stationary distribution that is locally stable. In the lower triangle there is a continuum of stationary distributions since the $\dot{p}_L = 0$ locus includes the whole lower triangle. This is a consequence of the fact that there is no way in or out of state L . Hysteresis of a degenerate sort is therefore built into this model.

Since our interest lies in hysteresis generated by the workings of the labor market, we feel that it is best to eliminate the degeneracy. This is legitimate since all we need to do to get rid of it is to perturb the dynamics slightly by allowing individuals very small probabilities of moving from state L to the other two states and from the other two states to L .¹² The phase diagram for one such perturbation is given in figure 3b. As expected, the $\dot{p}_U = 0$ loci in both triangles have moved only very slightly, as has the $\dot{p}_L = 0$ locus in the upper triangle. The most significant change is that now we have a $\dot{p}_L = 0$ locus in the lower triangle that intersects the $\dot{p}_U = 0$ locus in that triangle at the point F' .

Both F' and C' represent stationary distributions, and both are locally stable. But they represent very different social situations. Point F' is an economy in which there are three distinct classes with very little social mobility between the top two and the bottom one (all mobility in and out of L is due to the small random perturbations we used to eliminate the degeneracy). The principal reason behind the

¹¹ We have assumed that on the boundary the high-wage dynamics apply. The behavior at the boundary is, of course, affected by which wage prevails there. Making alternative assumptions will not significantly change our results.

¹² Think of these small probabilities as corresponding to winning the lottery and having a thunderbolt hit your house and factory.

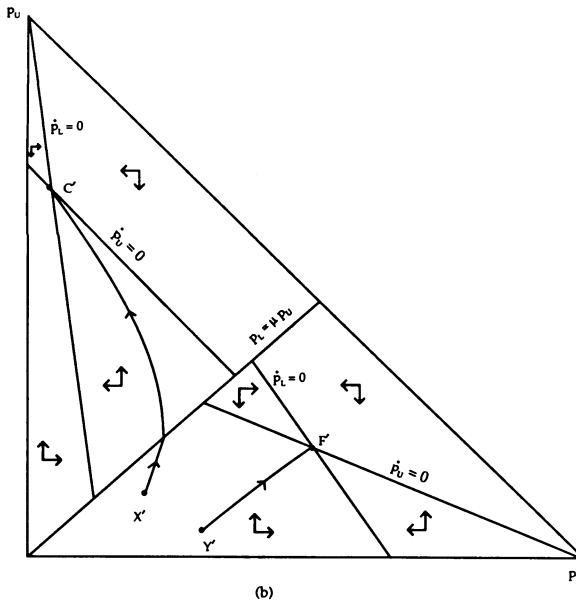
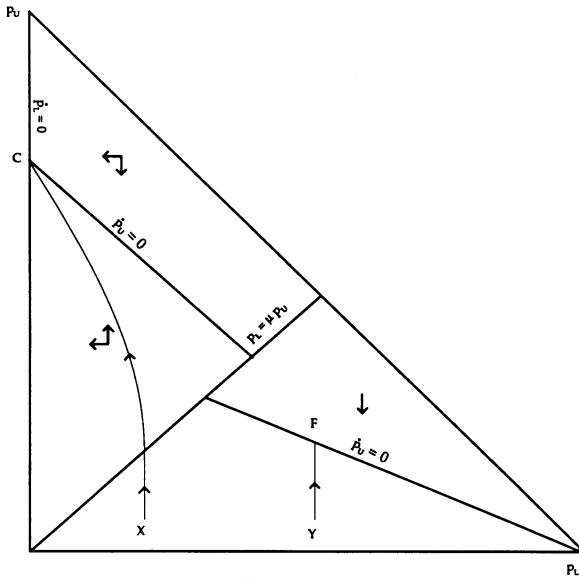


FIG. 3.—The cottage and the factory: *a*, original dynamics; *b*, perturbed dynamics

limited mobility is that the ratio of workers to entrepreneurs is high; the consequent low wage rate makes it virtually impossible, given the propensity to bequest, for workers to accumulate enough wealth to enter state M . At the same time, the wage rate is low enough and the project returns (in particular the low ones) are high enough to ensure the self-employed and entrepreneurs against going to L .

By contrast, C' is a situation in which there is really only one occupation in the economy: the overwhelming majority of the population (in the unperturbed version of the model, *everyone*) is self-employed. While there are a substantial number of people in class U who therefore are wealthy enough to be entrepreneurs, most of them are self-employed because they cannot find any workers. Since the low outcome for the self-employed is still high enough to keep the next generation in state M , the supply of people in state L remains small and the original configuration is able to reproduce itself.

The economy always converges to one of these stationary states. Which of the two will result depends on the initial conditions. With the aid of the phase diagram we see what types of economies converge to C' rather than to F' . Roughly speaking, economies with a small fraction of poor relative to middle- and upper-class people tend to converge to C' .

By looking at some trajectories, we can be more precise and better understand the dynamics. The points X' and Y' are two points close to each other in the lower triangle that both have a small upper class but have slightly different mixes of the classes. Consider the trajectory starting at X' , which has the relatively smaller lower class. Since the middle class is large and the upper class small, those moving up from M to U outnumber those who are moving the other way. The upper class grows. Because the size of the lower class changes very slowly, the ratio of the upper class to the lower class increases over time until μp_U becomes greater than p_L . At this point the wage increases to \bar{v} and the dynamics change. The workers start rising into the middle class, reducing the fraction of potential entrepreneurs who can find workers. The rest of the upper class now adopts self-employment and the transitions into the lower class decline (the self-employed remain in the middle class even when they are unlucky). The fraction of the lower class in the population thus continues to decline, and the economy converges to a distribution like C' .

The trajectory that starts at Y' also moves in the same direction at first, but since the initial fraction of the middle class was smaller, the rate of increase in the upper class will be smaller. For this reason, and also because the initial fraction of the lower class was larger, p_L remains larger than μp_U , wages do not rise, and employing people remains profitable. Instead of converging to C' , the economy ends

up at F' , which is a situation with both self-employment and entrepreneurial production.

If we identify self-employment with self-sufficient peasants and cottage industries and entrepreneurial production with large-scale capitalist agriculture and factory production, the dynamic patterns we describe above have historical parallels. The most famous of these might be the instance of England and France, which in terms of the level of development and technology were roughly comparable at the middle of the eighteenth century (O'Brien and Keyder 1978; Crafts 1985; Crouzet 1990) and yet went through radically different paths of development. England went on to develop and benefit hugely from the factory system and large-scale production, whereas France remained a nation of small farms and cottage industries for the next hundred years. In terms of our model, one possible explanation would be that England started at a point like Y' and France started at a point like X' .¹³

D. Prosperity and Stagnation

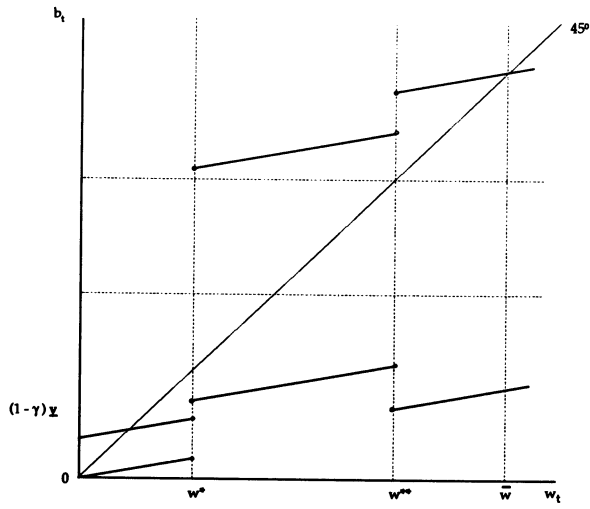
A somewhat different set of development paths can be generated with an alternative configuration of parameter values. Consider the case in which the transition map is as in figure 4a and b (corresponding once again to the cases $v = \underline{v}$ and $v = \bar{v}$). As before, the aggregate dynamic behavior can be reduced to a two-dimensional dynamical system in the simplex. Using the same definitions for the states as above, we follow a similar procedure to derive the dynamics of the wealth distribution. This process is described by the following system of piecewise-linear differential equations:

$$\dot{p}_L = \begin{cases} 1 - q - (1 - q)p_L + (q - q')p_U, & p_L > \mu p_U \\ 1 - q - \left(2 - q + \frac{q'}{\mu} - \frac{q}{\mu}\right)p_L, & p_L < \mu p_U \end{cases} \quad (6)$$

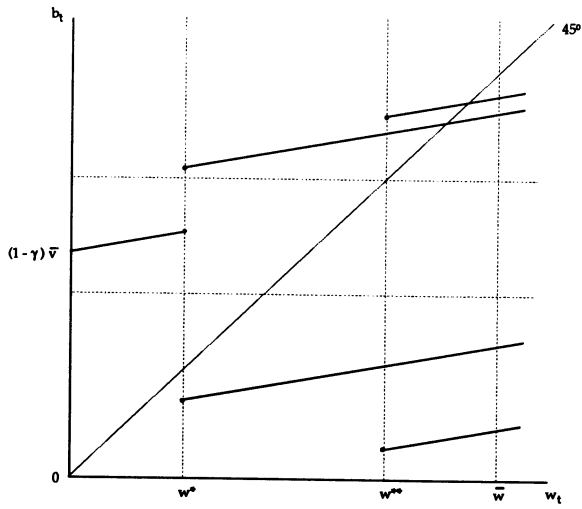
and

$$\dot{p}_U = \begin{cases} q - qp_L + (q' - q - 1)p_U, & p_L > \mu p_U \\ q - \left(q + \frac{q}{\mu} - \frac{q'}{\mu}\right)p_L - p_U, & p_L < \mu p_U. \end{cases} \quad (7)$$

¹³ A full study of the relevant data would be the subject of another paper, but there seems to be abundant evidence both for the poor performance of credit markets, at least in England (Deane 1965; Shapiro 1967; Ashton 1968), and for a more equal land distribution in France (especially after the Revolution) than in England (where the enclosure movement had generated a large population of landless poor). See Clapham (1936), Grantham (1975), and Soltow (1980).



(a)



(b)

FIG. 4.—a, $v = \underline{v}$. b, $v = \bar{v}$

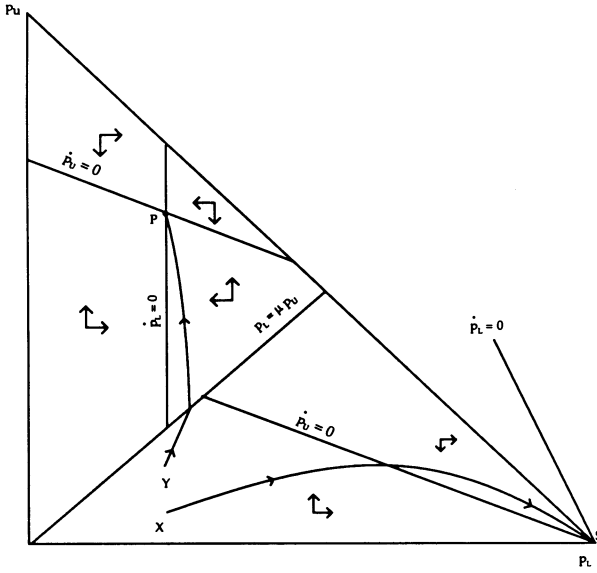


FIG. 5.—Prosperity and stagnation

The corresponding phase diagram appears in figure 5. There are two stationary distributions, labeled S and P , and both are locally stable, with large basins of attraction.¹⁴ Again, these stationary distributions are very different from each other. The distribution S is a state of economic collapse or stagnation: $p_L = 1$, so all agents have low wealth, which entails that they all remain in the subsistence sector. By contrast, P is a prosperous economy with both self-employment and an active labor market in which workers receive high wages; since the transition probabilities between the states are relatively high, there is also considerable social mobility. This contrasts with the case of factory production discussed above (point F' in fig. 3*b*) in which there is little mobility between L and the other two states.

As before, the long-run behavior of this economy depends on the initial conditions: economies in which the initial ratio of workers to entrepreneurs is low are more likely to be above the boundary line, where they will be subject to the high-wage dynamics, and are therefore more likely to converge to P . Where the initial ratio of poor to

¹⁴ Figure 5 is not the only possible phase diagram that can correspond to the configurations in fig. 4*a* and *b*. If q , q' , and μ satisfy $\mu q(1 - q) < 1 + q' + q(q - q')$, the stationary point of the high-wage dynamics will actually lie *below* the $p_L = \mu p_U$ boundary. Then there is a unique steady state since in converging to the high-wage stationary point, the economy crosses the boundary and the low-wage dynamics take over: the economy inevitably stagnates.

wealthy is high, the economy will be subject instead to the low-wage dynamics.

Of course, by examining figure 5, we can see that even if an economy initially has a high ratio of poor to wealthy, it is not necessarily doomed to stagnate, particularly if the middle class is sufficiently large (distributions with a large middle class are located near the origin). Consider the path starting at the point Y . Here most agents in the economy are self-employed, and the few workers that there are receive low wages because there are so few entrepreneurs demanding their labor (recall that some agents in state L must be idle). Over time, some of the self-employed become entrepreneurs and the rest fall into the lower wealth class. Along this particular path, the number of agents in U grows sufficiently fast that all agents in L are eventually hired as workers, and the economy is brought to the boundary. Now there is excess demand for labor and the high-wage dynamics take over, with the number of wealthy agents growing rapidly (the number of workers declines slightly along this part of the development path, from which we infer that the ranks of the self-employed must be growing). Thus even though this economy begins with a high ratio of poor to wealthy, it eventually achieves prosperity.

Notice, however, that if we start at the nearby point X instead of Y , the upper class grows slightly faster than the lower class, with both growing at the expense of the middle class of self-employed. The wage remains low, however, and eventually the lower class begins to dominate until the economy collapses to the stationary point S .

We can also check whether an economy might adhere to standard accounts of development such as the Kuznets hypothesis. The present example shows that the path to prosperity need not follow this pattern. Along the path emanating from Y , equality, measured by the relative size of the middle class, declines all the way to the prosperous steady state P . We can, however, easily generate versions of figure 5 in which some paths to prosperity are indeed of the Kuznets type. An example is shown in figure 6, which is obtained when the probability q' of high returns for entrepreneurs is fairly large. Beginning at Y , the middle class declines until point Z , after which it grows as the economy converges to P . Thus, as Kuznets suggested, while mean wealth rises along the entire development path, inequality first increases and then decreases.

V. Conclusion

In dynamic studies of income and wealth distribution, economists have tended to rely on what we have referred to as linear models, in which individual transitions are independent of aggregate variables

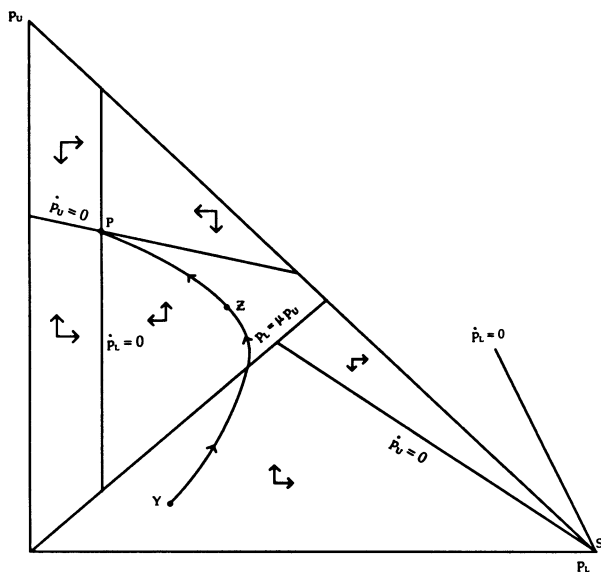


FIG. 6.—A development path that follows a Kuznets curve

(see Banerjee and Newman [1991] and the references therein). Our model of a developing economy, by contrast, is nonlinear because it violates this property of individual dynamics (see also Aghion and Bolton 1991). While it seems unlikely that other nonlinear models will admit the kind of dimensional reduction we have exploited, our examples do illustrate some of the fundamental differences between the two types of model.

For one thing, they may have distinct policy implications. Under the guidance of the linear model, which usually displays global stability, one is led to conclude that continual redistributive taxation, with the distortion it often entails, is required for achieving equity. The nonlinear model, by contrast, raises the possibility that one-time redistributions may have permanent effects, thereby alleviating the need for distortionary policy.

The nonlinear model also provides a way to capture the empirically appealing notion that the same individual characteristics (e.g., wealth levels) can be observed under different stationary distributions. For all practical purposes, the very richest people in India are as wealthy as the very richest in the United States, and the very poorest Americans are no wealthier than their Indian counterparts. Yet standard Markov process models (including deterministic representative agent models) that give rise to multiple steady states or hysteresis preclude this possibility: any state observed under one stationary distribution

cannot be observed under another, so that if India and the United States correspond to different equilibria of the same standard model, then no Indian can enjoy the same wealth as any American.

Our examples (particularly C' and F' in fig. 3*b*) underscore a related point. Individual lineages can travel all over the wealth space under two very different stationary distributions.¹⁵ Moreover, random perturbations to the individual-level dynamics will not significantly affect these distributions and cannot destroy the dependence of aggregate behavior on initial conditions. Contrary to the lessons of linear models, there need be no contradiction between individual mobility and aggregate hysteresis.

References

- Aghion, Philippe, and Bolton, Patrick. "A Trickle-down Theory of Growth and Development with Debt Overhang." Manuscript. Paris: DELTA, 1991.
- Andreoni, James. "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." *J.P.E.* 97 (December 1989): 1447–58.
- Ashton, Thomas S. *The Industrial Revolution, 1760–1830*. London: Oxford Univ. Press, 1968.
- Banerjee, Abhijit V., and Newman, Andrew F. "Risk-bearing and the Theory of Income Distribution." *Rev. Econ. Studies* 58 (April 1991): 211–35.
- Bell, Clive. "Credit Markets, Contracts and Interlinked Transactions." In *Handbook of Development Economics*, edited by Hollis Chenery and T. N. Srinivasan. New York: North-Holland, 1988.
- Bernanke, Ben, and Gertler, Mark. "Agency Costs, Net Worth, and Business Fluctuations." *A.E.R.* 79 (March 1989): 14–31.
- . "Financial Fragility and Economic Performance." *Q.J.E.* 105 (February 1990): 87–114.
- Blanchard, Olivier J. "Debt, Deficits, and Finite Horizons." *J.P.E.* 93 (April 1985): 223–47.
- Clapham, John H. *Economic Development of France and Germany, 1815–1914*. 4th ed. Cambridge: Cambridge Univ. Press, 1936.
- Cohen, Jon S. "Managers and Machinery: An Analysis of the Rise of Factory Production." *Australian Econ. Papers* 20 (June 1981): 24–41.
- Conlisk, John. "Interactive Markov Chains." *J. Math. Sociology* 4 (July 1976): 157–85.
- Crafts, N. F. R. "Industrial Revolution in England and France: Some Thoughts on the Question, 'Why Was England First?'" In *The Economics of the Industrial Revolution*, edited by Joel Mokyr. Totowa, N.J.: Rowman & Allanheld, 1985.
- Crouzet, François. *Britain Ascendant: Comparative Studies in Franco-British Economic History*. Cambridge: Cambridge Univ. Press, 1990.
- Deane, Phyllis. *The First Industrial Revolution*. Cambridge: Cambridge Univ. Press, 1965.

¹⁵ The idea that a stationary economy is one in which aggregate characteristics are fixed, but in which individuals may occupy different states over time, is already common in economics (examples are Loury [1981], Banerjee and Newman [1991], and Hopenhayn [1992]); it is one motivation for seeking ergodic distributions. What is new here is the presence of multiple ergodic distributions with common support.

- Eswaran, Mukesh, and Kotwal, Ashok. "Why Are Capitalists the Bosses?" *Econ. J.* 99 (March 1989): 162-76.
- Galor, Oded, and Zeira, Joseph. "Income Distribution and Macroeconomics." *Rev. Econ. Studies* (in press).
- Grantham, George W. "Scale and Organization in French Farming, 1840-1880." In *European Peasants and Their Markets: Essays in Agrarian Economic History*, edited by William N. Parker and Eric L. Jones. Princeton, N.J.: Princeton Univ. Press, 1975.
- Hopenhayn, Hugo Andreas. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica* 60 (September 1992): 1127-50.
- Kehoe, Timothy J., and Levine, David K. "Debt-constrained Asset Markets." *Rev. Econ. Studies* (in press).
- Khan, M. Ali. "In Praise of Development Economics." Manuscript. Baltimore: Johns Hopkins Univ., 1989.
- Kihlstrom, Richard E., and Laffont, Jean-Jacques. "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion." *J.P.E.* 87 (August 1979): 719-48.
- Loury, Glenn C. "Intergenerational Transfers and the Distribution of Earnings." *Econometrica* 49 (July 1981): 843-67.
- Lucas, Robert E., Jr. "On the Mechanics of Economic Development." *J. Monetary Econ.* 22 (July 1988): 3-42.
- Matsuyama, Kiminori. "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium." *Q.J.E.* 106 (May 1991): 617-50.
- Millward, R. "The Emergence of Wage Labor in Early Modern England." *Explorations Econ. Hist.* 18 (January 1981): 21-39.
- Murphy, Kevin M.; Shleifer, Andrei; and Vishny, Robert W. "Income Distribution, Market Size, and Industrialization." *Q.J.E.* 104 (August 1989): 537-64. (a)
- . "Industrialization and the Big Push." *J.P.E.* 97 (October 1989): 1003-26. (b)
- Newman, Andrew F. "The Capital Market, Inequality, and the Employment Relation." Manuscript. Evanston, Ill.: Northwestern Univ., 1991.
- North, Douglass C. *Structure and Change in Economic History*. New York: Norton, 1981.
- O'Brien, Patrick K., and Keyder, Caglar. *Economic Growth in Britain and France, 1780-1914: Two Paths to the Twentieth Century*. London: Allen & Unwin, 1978.
- Romer, Paul M. "Increasing Returns and Long-Run Growth." *J.P.E.* 94 (October 1986): 1002-37.
- Sappington, David. "Limited Liability Contracts between Principal and Agent." *J. Econ. Theory* 29 (February 1983): 1-21.
- Shapiro, Seymour. *Capital and the Cotton Industry in the Industrial Revolution*. Ithaca, N.Y.: Cornell Univ. Press, 1967.
- Soltow, Lee. "Long-Run Changes in British Income Inequality." In *Wealth, Income, and Inequality*, 2d ed., edited by Anthony B. Atkinson. Oxford: Oxford Univ. Press, 1980.
- Stiglitz, Joseph E. "Economic Organization, Information, and Development." In *Handbook of Development Economics*, edited by Hollis Chenery and T. N. Srinivasan. New York: North-Holland, 1988.
- Stokey, Nancy L., and Lucas, Robert E., Jr. *Recursive Methods in Economic Dynamics*. Cambridge, Mass.: Harvard Univ. Press, 1989.
- Townsend, Robert M. "Models as Economies." *Econ. J.* 98 (suppl., 1988): 1-24.