Review of Economic Studies (2003) **70**, 369–393 0034-6527/03/00140369$02.00

Persistent Inequality

DILIP MOOKHERJEE
Boston University

and

DEBRAJ RAY
New York University and Instituto de Análisis Económico (CSIC)

First version received September 2000; *final version accepted October* 2002 (*Eds.*)

When human capital accumulation generates pecuniary externalities across professions, and capital markets are imperfect, persistent inequality *in utility and consumption* is inevitable in *any* steady state. This is true irrespective of the degree of divisibility in investments. However, divisibility (or fineness of occupational structure) has implications for both the multiplicity and Pareto-efficiency of steady states. Indivisibilities generate a continuum of inefficient and efficient steady states with varying *per capita* income. On the other hand, perfect divisibility typically implies the existence of a unique steady state distribution which is Pareto-efficient.

1. INTRODUCTION

A central prediction of the neoclassical growth model is that the market mechanism promotes the *convergence* of incomes of different agents, families or countries, so that historical inequality tends to vanish in the long run. Reformulations of this model in the context of intergenerational mobility (Becker and Tomes (1979), Loury (1981), Mulligan (1997)) therefore rely on the presence of random factors ("luck") in explaining the persistence of inequality, despite the overall tendency towards convergence. Of course, the key assumption that underlies these models is that the returns to investment are determined entirely by a convex technology.

A recent literature generates opposite predictions concerning the significance and persistence of long-run inequality. In this literature, there are steady states with inequality, where such inequality refers to variations in consumption and utility levels, and not just in gross incomes. Moreover, inequality persists across dynasties: poor families are unable to catch up with the rest of the population. This is the central contribution of Banerjee and Newman (1993), Galor and Zeira (1993), Ljungqvist (1993), Ray and Streufert (1993) and others.[1]

In all of these models, the economy displays both unequal and equal steady states, the convergence to one or another presumably depending on the initial conditions.[2] Hence history matters well into the future, at both the level of the household and the macroeconomy. As a corollary, one-shot interventions (such as a single redistribution) can have permanent effects.[3] Since these theories assign a significant role to historical inequality, and does so in a similar vein to classical theories of distribution and growth, we shall refer to them as the "new classical" theory.

1. See also Bandyopadhyay (1993, 1997), Freeman (1996), Aghion and Bolton (1997), Piketty (1997), Matsuyama (2000) and Mani (2001).
2. Different steady states typically correspond to different levels of *per capita* income.
3. Hoff and Stiglitz (2001) argue that such multiplicity creates a distinct role for policy (such as a one-time land reform). By changing initial conditions, the policy intervention may change the *particular* steady state that forms the attractor for the process and thereby generate permanent effects; there is no need to change the *set* of steady states.

Now, at a certain level, these two literatures are not as different as one might first think. The new classical literature emphasizes indivisibilities or nonconvexities in the relevant capital good (occupations, work capacities, skills), while the basic neoclassical growth model emphasizes convexities. Drop convexities in the neoclassical growth model and optimal programs may become history dependent.[4] Such history dependence can be readily translated into inequality, in line with the new classical models.

Indeed, one need not go as far as introducing nonconvexities. Simply consider the optimal growth model with a perfect capital market. Then the investment problem at the level of an individual is *linear*, and it is well known that a steady state is compatible with *any* distribution of wealth among the population, though the aggregate wealth level is fully pinned down.[5] Depending on the initial distribution of wealth, it is perfectly possible to converge to one of several final distributions of steady state wealth. For instance, as Chatterjee (1994) observes, "the influence of the initial distribution of wealth persists forever into the future", even in the competitive growth model with perfect capital markets.[6]

The purpose of this paper is to highlight the difference between the two approaches that we believe *is* critical. The postulate that we discuss is not present in all of the new classical models in equal degree, but in our opinion is central for understanding the persistence and evolution of inequality. This is the assumption that inputs supplied by different occupations are not perfect substitutes. The relative returns from different occupations then depends on the occupational distribution in the economy. The returns to a particular agent from investing (by selecting an occupation with higher human capital) therefore depend fundamentally on the *overall* distribution of investment choices of other agents in the economy—*i.e.* there are *pecuniary externalities* in investment. The purpose of this paper is to study the implications of this phenomenon when capital markets are imperfect.

To this end, we construct a dynastic model with missing capital markets.[7] Individuals have dynastic preferences (as in Barro, 1974), deriving utility from consumption and the utility of their offspring. They enjoy consumption, and bequeath "professions" to their offspring.[8] The return to these professions constitute the starting wealth of the descendants, and the process repeats itself *ad infinitum*. We abstract from uncertainty altogether.[9] The key feature of the model is the presence of pecuniary externalities across the investment decisions of different households.[10]

We first show that *every* steady state must involve no mobility at all across occupations with distinct wages: barring random shocks, each family is permanently locked into a particular level of earnings and consumption relative to the rest of the economy. And given a minimal extent of

4. See the literature on aggregative growth models with an increasing returns technology Clark (1971), Skiba (1978), Majumdar and Mitra (1982, 1983), Dechert and Nishimura (1983) and Mitra and Ray (1984).

5. At least, this is true for the case in which savings are based on long-run utility maximization, or a dynastic bequest motive. For instance, Becker and Tomes (1979) obtain convergence of household incomes despite linearity of investment frontiers, because they assume a paternalistic form of parental altruism. Note also that there is no obvious relationship between the imperfection of the capital market and the overall tendency to generate inequality: the creation of imperfections in the credit market restores strict convexity, and (combined with a dynastic bequest motive) removes all unequal steady states.

6. Chatterjee (1994) contains interesting results on the evolution of wealth inequalities over time (in the sense of comparing Lorenz curves) in the competitive version of the neoclassical growth model.

7. We follow Loury (1981) in assuming that they are entirely absent. However, the main results are qualitatively unaffected as long as the capital markets are imperfect. In Mookherjee and Ray (2002*a*), we explore the implications of endogenous capital market imperfections for asset accumulation strategies and the evolution of asset inequality.

8. The term "profession" is general and includes professions in the usual sense of the term as well as monetary bequests. We discuss these interpretations in more detail in what follows.

9. In part, we do so because we want to emphasize that there are inequality-creating tendencies in the system quite apart from exogenous stochastic shocks.

10. This model is closely related to those studied earlier by Ray (1990), Bandyopadhyay (1993, 1997), Ljungqvist (1993), Freeman (1996) and Matsuyama (2000).

occupational diversity (wherein there exist at least two active occupations with distinct wages or training costs), *every steady state must be associated with unequal consumption and utility across households in every generation*. Under these conditions, equal steady states cannot exist.

Note the contrast to the neoclassical framework and even several of the new classical models, in which perfect equality is never ruled out as a *possible* steady state (in the absence of uncertainty). Here, the inevitable nature of inequality is fundamentally a consequence of the pecuniary externalities that induce the population to sort into different occupations with unequal net incomes. With missing capital markets, earnings differentials in steady state must overcompensate for training cost margins, yielding higher levels of earnings net of training costs in occupations with higher gross earnings. These results are robust to the degree of divisibility of occupations or the exact form of the bequest motive.

While the divisibility of investment makes no difference to inequality, we show it has important implications for the history-dependence question. As mentioned above, the implications of divisibility have not been explored by the new classical literature. When there are just two occupations with differing training costs, there is a continuum of steady states that vary with respect to *per capita* income, occupational and earnings dispersion. This recaptures the results of the new classical theories under substantial occupational indivisibility. But when there is a continuum of occupations—reflecting perfect divisibility of occupational structure—the multiplicity disappears. We provide a broad set of conditions under which there is a unique steady state distribution. This suggests that while investment indivisibilities may be orthogonal to the study of inequality, they are an important ingredient of theories of history dependence at the macroeconomic level.[11]

Notice that this result is different from one which simply asserts the uniqueness of steady state *aggregates*. It is well known that the competitive version of the Ramsey model pins down the rate of return to investment in steady state, therefore—by the concavity of the production function—the aggregate capital stock, and consequently the wage rate, *per capita* income and so on. But there is, of course, an enormous multiplicity of steady state asset distributions (see, *e.g.* Chatterjee, 1994). The pecuniary externalities in our model precipitate the uniqueness of the asset *distribution*. More generally, our theory generates nontrivial restrictions on the distribution (even when investments are indivisible), whereas such restrictions are absent in the neoclassical model with perfectly divisible investments.[12]

The third set of results pertain to the efficiency of steady states. We provide an almost complete characterization of the (constrained) efficiency of steady states. The condition that characterizes efficient steady states is a common rate of return to all occupations, which does not exceed the discount rate. It follows that efficiency is compatible with optimal or even *overinvestment* in a first-best sense, but not with underinvestment. This characterization has implications for special cases. For instance, if there are only two occupations, there is a continuum of inefficient steady states involving underinvestment, and also a continuum of efficient steady states involving optimal or overinvestment. Moreover, there always exists a fully efficient steady state.

In the case of a continuum of occupations, the unique steady state turns out to be fully efficient. Once again, investment divisibilities play an important role. A continuum of occupations can act as a substitute for a missing capital market in restoring efficiency.

11. A nonconvexity is not the *only* pathway to history dependence, however. See, *e.g.* Moav (2002) and Mookherjee and Ray (2002*a*).

12. For instance, in Chatterjee's (1994) competitive version of the neoclassical model, the linearity of investment frontiers at the micro level implies an indeterminacy in steady state wealth for any given household, but strict concavity in the aggregate must pin down total wealth. It is true that *given* some starting wealth distribution, the wealth distribution thereafter evolves in predictable fashion to some limit steady state distribution, but this is quite different from asserting that there is only one such distribution.

It seems that efficiency-based arguments for interventionist policies must rely on the existence of investment indivisibilities, analogous to arguments concerning history dependence.

This paper is organized as follows. Section 2 introduces the model. Section 3 presents our results concerning inequality and immobility. Then Section 4 discusses steady state multiplicity, and Section 5 discusses efficiency. Finally, Section 6 concludes, while an Appendix gathers all the proofs.

2. MODEL

2.1. *Agents and professions*

There is a continuum of agents indexed by i on $[0, 1]$. Each agent lives for one period, and has one child who inherits the same index. Each infinite parent–child chain forms a dynasty. Dynasties are linked by fully altruistic preferences as in Barro (1974), so we may equivalently think of i as an infinitely lived individual.

Each individual enjoys the consumption of a single good c, with one-period utility u.[13] Assume u is increasing, smooth and strictly concave. Given altruistic preferences, if $\{c_s\}$ is an infinite sequence of consumptions, then generation t's payoff is given by the "tail sum"

$$\sum_{s=t}^{\infty} \delta^{s-t} u(c_s), \tag{1}$$

where $\delta \in (0, 1)$ is a discount factor, assumed common to all agents.

There is some set $\mathcal{H}$ of *professions* which individuals in each generation select from. Most cases of interest are accommodated by taking $\mathcal{H}$ to be some arbitrary compact subset of Euclidean space. One can therefore interpret a profession very widely, as ownership of a vector of different kinds of assets, or as an occupation described by a multidimensional attribute. A *population distribution over professions* is simply a measure $\boldsymbol{\lambda}$ on $\mathcal{H}$. We will be particularly interested in leading subcases in which $\mathcal{H}$ is finite or is an interval. This allows for arbitrary richness in the set of professions.

2.2. *Technology*

The technology combines a production sector with an educational or training sector. The consumption good is produced by workers of different professions, and inputs of the good itself. Trained professionals are produced by teachers and workers from different professions, besides material input of the consumption good. The technology is represented by means of a set $\mathcal{T}$, which contains various combinations of the form:

$$z \equiv (\boldsymbol{\lambda}, c, \boldsymbol{\lambda}'),$$

where $\boldsymbol{\lambda}$ represents the input vector (a measure on $\mathcal{H}$, the current population distribution), c is a real number representing net output of the consumption good, and $\boldsymbol{\lambda}'$ is a measure on $\mathcal{H}$ which denotes the supply of trained professionals (which forms the next period's population distribution).

Throughout the paper, we assume that $\mathcal{T}$ is a closed convex cone,[14] that at least one profession requires no training,[15] and that owners of firms (in either production or training sectors) seek to maximize profits.

13. Much of the analysis extends to the many-consumption-good case without difficulty, but we avoid this generality for expositional simplicity.

14. Closedness is relative to the (product) weak topology on population measures over the set of professions and the usual topology on c.

15. This captures the notion that each family has the option of not investing at all in their children's education.

2.3. *Prices and behaviour*

Normalize the price of the consumption good to unity. Then two sets of prices are relevant at each date. First, there are the returns to professions: let $\mathbf{w} = \{w(h)\}$ be the wage function summarizing these returns. Second, there are the training costs for different professions: call this function $\mathbf{x} = \{x(h)\}$. Note that $\mathbf{x}$ represents both the costs incurred by investing parents and the revenues earned by training institutions.

Given prices at any date t, the economy generates (input) demands for professions ($\boldsymbol{\lambda}_t$), a supply (c_t) of the final good and supplies of trained professionals ($\boldsymbol{\lambda}'_t = \boldsymbol{\lambda}_{t+1}$) for the *next* generation at period $t+1$. Given wages $\mathbf{w}_t$ and training costs $\mathbf{x}_t$ at date t, profit maximization implies that $(\boldsymbol{\lambda}_t, c_t, \boldsymbol{\lambda}_{t+1})$ must solve

$$\max c + \mathbf{x}_t\boldsymbol{\lambda}' - \mathbf{w}_t\boldsymbol{\lambda} \tag{2}$$

subject to $(\boldsymbol{\lambda}, c, \boldsymbol{\lambda}') \in \mathcal{T}$.

Now turn to household responses. Given some sequence of prices $\{\mathbf{w}_s, \mathbf{x}_s\}_{s\geq t}$, a generation t household i with current profession $h(i)$ will choose a sequence $\{h_s, c_s\}_{s\geq t}$ to solve

$$\max \sum_{s=t}^{\infty} \delta^s u(c_s) \tag{3}$$

subject to the constraints

$$h_t = h(i) \tag{4}$$

and

$$w_s(h_s) = c_s + x_s(h_{s+1}) \qquad \text{for all } s \geq t. \tag{5}$$

Because preferences are perfectly altruistic, there is no time inconsistency across generations, so we may as well restrict ourselves to the choices made by generation 0, with initial "endowment" of professions given by $\{h_0(i)\}_{i\in[0,1]}$, or equivalently, by the population distribution $\boldsymbol{\lambda}_0$ on $\mathcal{H}$. Denote by $\{c_t(i), h_t(i)\}$ the consumption and professional choices made at every date by dynasty i.

Observe that the optimization problem (3) formulated for an individual (or dynasty) incorporates the simplest description of a missing market for the accumulation of human capital. Generation $t+1$'s human capital must be paid for by generation t; no loans are possible. If preferences are strictly convex, this means that self-finance has different implications for people depending on their current economic status. Specifically, the poor have a higher marginal cost of finance.

2.4. *Equilibrium*

Given some initial distribution $\boldsymbol{\lambda}$, an *equilibrium* is a collection $\{\boldsymbol{\lambda}_t, c_t, \mathbf{w}_t, \mathbf{x}_t\}$ (with $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}$) such that:

[1] At each date t, $(\boldsymbol{\lambda}_t, c_t, \boldsymbol{\lambda}_{t+1})$ solves (2), given the price sequence $\{\mathbf{w}_t, \mathbf{x}_t\}$.

[2] There exists $\{h_t(i), c_t(i)\}$ (for $i \in [0, 1]$ and $t = 0, 1, 2, \ldots$) such that for all individuals i, $\{h_t(i), c_t(i)\}_{t=0}^{\infty}$ solves (3) starting from $h_0(i)$, and such that markets clear at any date:

$$c_t = \int_{[0,1]} c_t(i)di \tag{6}$$

and

$$\lambda_t(B) = \text{Measure}\{i : h_t(i) \in B\} \tag{7}$$

for every Borel subset of $\mathcal{H}$.

A particular type of equilibrium is a *steady state*, one in which all prices and aggregate quantities remain the same over time. Formally, a collection $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$ is a steady state if there exists an equilibrium $\{\boldsymbol{\lambda}_t, c_t, \mathbf{w}_t, \mathbf{x}_t\}$ with $(\boldsymbol{\lambda}_t, c_t, \mathbf{w}_t, \mathbf{x}_t) = (\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$ for all t.

The model is general enough to encompass several commonly studied models as special cases. The neoclassical (Ramsey) model corresponds to the case where $\mathcal{H}$ is interpreted as different possible levels of physical (rather than human) capital, and forms an interval of the real line.[16] Models of skill acquisition, such as Ray (1990), Bandyopadhyay (1993, 1997), Galor and Zeira (1993) or Ljungqvist (1993), correspond to the case of two professions 1 (unskilled) and 2 (skilled), with a constant cost x of acquiring the skill. In similar vein, models of entrepreneurship or occupational choice (as in Banerjee and Newman (1993) or Freeman (1996)) also correspond to a two-profession version. The relative wages of different occupations may be exogenous or endogenously determined. However, it appears that most existing literature takes training costs to be exogenous (with the exception of Ljungqvist, 1993).

3. PERSISTENT INEQUALITY

3.1. *Inequality at steady states*

Our first result states that even though a steady state is defined in terms of the stationarity of aggregates (such as the population distribution over professions, or the total production of the consumption good), it also involves stationarity at the individual level. Notice that this result does not automatically follow from the definition of a steady state. There is no reason why a steady state cannot involve a constant fraction of the population in each profession, while at the same time there are dynasties constantly moving from one profession to another (as in the ergodic distribution of a Markov chain).

Let $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$ be a steady state. Say that two professions h and h' are *distinct* (relative to this steady state) if they involve different training costs: $x(h) \neq x(h')$. Note a simple sufficient condition for two professions to be distinct in any equilibrium: if training someone for occupation h requires more of every material good and every kind of teacher than training someone for occupation h'—as is typically the case when h requires more years of schooling than h'—then irrespective of the precise set of prices, occupation h will involve a higher training cost than h'. More generally, two professions with distinct training technologies will turn out to be generically distinct, though we do not pursue the exact conditions required to make this claim precise.

Proposition 1 (Zero Mobility in Steady State). *Let $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$ be a steady state. Then no positive measure of individuals will switch across distinct professions.*

This "zero-mobility" result is based on a single-crossing property that stems from the convexity of preferences and the absence of credit markets (*i.e.* the fact that parents must pay for their children's education). Note first that if h and h' are distinct professions with $x(h) > x(h')$, it must be the case that $w(h) > w(h')$ for any family to be induced to choose occupation h. Hence in order to attain a higher income for their children, parents have to invest more in education. In steady state, the present value utility of a generation currently occupying occupation h and contemplating a permanent deviation to occupation g is given by $u(w(h) - x(g)) + \delta V(g)$ where $V(g)$ is the present (utility) value to the parent of moving the child to profession g. The strict concavity of u implies that richer families must endure a smaller utility sacrifice in educating their children, hence must be willing to invest more in education. Accordingly the children of

16. Further, the production set takes the following form: $\mathcal{T} = \{(\boldsymbol{\lambda}, c, \boldsymbol{\lambda}') | c = f(\int_{\mathcal{H}} hd\boldsymbol{\lambda}) - \int_{\mathcal{H}} hd\boldsymbol{\lambda}'\}$, where f is a smooth, production function with output divided between current consumption and capital stock next period.

families occupying the richest occupation (which entails the highest training costs) must be trained for the same occupation. Otherwise this occupation would not be filled at subsequent dates, contradicting the steady state assumption. When there are a finite number of professions, the same argument applies then to the next richest occupation, and so on down the line.

The no-switching property implies that the earnings and consumption of each family must be constant over time in any steady state, and $V(h) = [u(w(h) - x(h))]/(1 - \delta)$ for all h. It leads directly to the conclusion concerning the necessity of inequality:

Proposition 2 (Inequality in Steady State). *Suppose that two dynasties inhabit two distinct professions in some steady state. Then they must enjoy different levels of consumption* (*and utility*) *at every date.*

The reasoning is simple. If h involves a higher training cost than h', not only must it generate a higher level of earnings ($w(h) > w(h')$), but also a higher level of earnings *net of training cost*: $w(h) - x(h) > w(h') - x(h')$. Otherwise the parent selecting occupation h for its child would be better off reducing the educational investment from $x(h)$ to $x(h')$, and letting all its descendants move to occupation h' instead of h.

Proposition 2 states that inequality is an endemic feature of every steady state satisfying a minimal "diversity" criterion: two or more distinct professions should be inhabited. In contexts where the term "profession" corresponds to different occupations with distinct forms of human capital, this is really a very weak requirement. For instance if there are two occupations ordered in terms of input requirements (of every kind) in their training, and are both essential in the production of the consumption good (in the sense that without them the consumption good cannot be produced), then every steady state (with positive consumption) must involve persistent inequality.[17] On the other hand, if "profession" includes the inheritance of financial as well as human capital, the diversity condition is much more subtle, and requires careful examination (see Mookherjee and Ray, 2002*b* for details).

Endogenous market prices play a crucial role in generating and perpetuating this inequality. If several distinct professions are needed for economic activity, the behaviour of prices must guarantee that each of those professions are actually chosen. Since parents pay for their children's education, a profession requiring a greater training cost entails a greater sacrifice for parents. So to induce them to undertake this sacrifice it must be the case that their children are rendered better off in utility terms. Hence there must be inequality in utility and net consumption, not just in incomes.

The examples of Ray (1990), Ljungqvist (1993) and Freeman (1996) go further in explaining how the market can endogenously create inequality, starting from a position of equality. In each of these models, there are two professions (skilled and unskilled labour in Ljungqvist and Ray, managers and workers in Freeman). Consider the Ljungqvist–Ray scenario in which there are two skills, and both types of labour enter as inputs in a concave production function satisfying the Inada conditions. Now suppose all individuals in a particular generation have equal wealth. Is it possible for all of them to make the same *choices*? The answer is no. If all of them choose to leave their descendants unskilled, then the return to skilled labour will become enormously high, encouraging some fraction of the population to educate their children. Similarly, it is not possible for all parents to educate their children, if unskilled labour is also necessary in production. Even if all agents were identical to start with, they must sort into distinct occupations, owing to the interdependence of decisions of different families.

17. Notice that the Ramsey model with strictly concave investment technology at the level of each individual household exhibits convergence to a unique steady state for each household. Then every economy-wide steady state must involve the same "profession" for every household, and Proposition 2 does not apply.

To be sure, at this stage there are no implications for inequality. There is inequality of (earned) *incomes*, but no utility differences as far as the original generation is concerned. But utility differences do arise from the descendants onward. Suppose the economy converges to a steady state (as is verified in Ray (1990), Mookherjee and Ray (2000)) in which both occupations are occupied. By Proposition 2, such a steady state must display (utility and consumption) inequality. Hence the pecuniary externalities inherent in the market mechanism cause the fortunes of *ex ante* equal dynasties to diverge, in complete contrast to both neoclassical and new classical models.[18]

The result extends to alternative formulations of capital market imperfections or intergenerational altruism. All that is needed is a higher marginal cost of finance for poorer households, which almost any reasonable model of imperfect capital markets will satisfy. Or parents may have a "warm glow" bequest motive, where they care only about the size of their bequests (or educational investments), rather than their implication for the well-being of their descendants. Or they may care about child wealth. Irrespective of these details, the crucial "single-crossing" property that underlies Propositions 1 and 2 will obtain: richer households will have a greater willingness to invest in their children's education, implying both zero mobility and positive inequality in every steady state.

4. MULTIPLICITY

In this section, we explore the multiplicity of steady states, and relate it to occupational diversity.

4.1. *Characterizing steady states*

Throughout most of this section, we make the following "full-support" assumption: every profession is occupied in steady state.[19] A sufficient condition for full support is that every occupation is essential for producing the consumption good.[20] Alternatively, even if some inputs may not be necessary in production of the consumption good, they will be essential if they are necessary to train other occupations that are essential in production of the consumption good. Later we shall explain how the full-support assumption can be substantially weakened without affecting the results.

The first necessary condition for a steady state $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$ is that $(\boldsymbol{\lambda}, c)$ must be related to $(\mathbf{w}, \mathbf{x})$ via profit maximization; that is,

$$(\boldsymbol{\lambda}, c, \boldsymbol{\lambda}) \in \arg\max c + \mathbf{x} \cdot \boldsymbol{\lambda}' - \mathbf{w}\hat{\boldsymbol{\lambda}}, \qquad \text{subject to } (\hat{\boldsymbol{\lambda}}, c, \boldsymbol{\lambda}') \in \mathcal{T}. \tag{8}$$

18. A variant with "warm glow" bequests will exhibit similar properties. Typically, the optimal bequest will increase in wealth, so that the single-crossing property is once again satisfied: children of wealthier parents are more willing to invest in training. Hence in a steady state there can be no occupational mobility, parallel to Proposition 1. And Proposition 2 extends too, since lifetime utility must be strictly increasing in inheritance. For examples of the warm-glow model, see, *e.g.* Banerjee and Newman (1993), Galor and Zeira (1993) or Maoz and Moav (1999).

19. In particular, if the set of professions is an interval, and the steady state population distribution over this set admits a density, then we require that density to be positive throughout.

20. The reasons why the full-support assumption might fail include the following. First, certain professions may be inessential, because the inputs they supply can be supplied more efficiently by some other profession. In this case we may simply redefine the set of occupations to exclude those that are dominated by others. Secondly, even if all professions are necessary, there could be trivial equilibria with zero output simply because unoccupied professions may be prohibitively costly to acquire owing to a lack of teachers to train that profession. These steady states *literally* rely on the assumption of a totally missing capital market and a closed economy. With a slight perturbation of these assumptions—allowing teachers to be imported and/or borrowing at a higher rate than the lending rate—such steady states would no longer survive. And third, different professions may compete in supplying the same input, and some of them may be shut down depending on the precise pattern of prices. We explain in what follows how the results can be extended in such cases.

Secondly, no individual must contemplate a "one-shot deviation" to another profession, where (by the zero-mobility result and the full-support assumption) it may safely be conjectured that the new profession will be adhered to by all descendants. That is, for every individual at some occupation h and for every alternative occupation h',

$$u(w(h) - x(h)) \geq (1 - \delta)u(w(h) - x(h')) + \delta u(w(h') - x(h')). \tag{9}$$

Indeed, by the one-shot deviation principle (for discounted optimization problems) and the zero-mobility result, conditions (8) and (9) are necessary as well as sufficient to describe the set of steady states.

4.2. *Two professions*

First study (8) and (9) for the case of two professions with exogenous training cost. Call the professions "skilled" and "unskilled" (as in Ray (1990) or Ljungqvist (1993)). For unskilled labour take the training cost to be zero. For skilled labour assume that there is an exogenous training cost x, which is just the number of units of the consumption good used as input into the training process. This implicitly assumes that training does not require any labour inputs. Let λ denote the fraction of the population at any date that is skilled. If some well-behaved production function f (satisfying the usual curvature and Inada end-point conditions) determines the wage to skill categories, the skilled wage at that date will be given by $w^s(\lambda) \equiv f_1(\lambda, 1 - \lambda)$, while the unskilled wage will be given by $w^u(\lambda) \equiv f_2(\lambda, 1 - \lambda)$, where subscripts denote appropriate partial derivatives.[21] This yields the following simple characterization: a fraction λ of skilled people is compatible with a steady state if and only if

$$\begin{aligned} u(w^s(\lambda)) - u(w^s(\lambda) - x) &\leq \frac{\delta}{1-\delta}[u(w^s(\lambda) - x) - u(w^u(\lambda))] \\ &\leq u(w^u(\lambda)) - u(w^u(\lambda) - x). \end{aligned} \tag{10}$$

The L.H.S. of (10) represents the utility sacrifice of a skilled parent (hereafter denoted by $\kappa^s(\lambda)$) in educating its child, while the R.H.S. is the corresponding sacrifice for an unskilled parent (denoted by $\kappa^u(\lambda)$). The term in the middle is the present value benefit of all successive descendants being skilled rather than unskilled (which we shall denote by $b(\lambda)$).

These benefit and sacrifice functions are illustrated in Figure 1. $\lambda_1 \in (0, 1)$ denotes the skill intensity of the population at which the skill premium just disappears and the wages of the skilled and unskilled are equal. So κ^s and κ^u intersect there. Likewise, λ_2 is the point at which the wages of the skilled *net* of training equal those of the unskilled. So b drops to zero there. These observations can be used in conjunction with (10) to establish the following:

Proposition 3. *There is a continuum of steady states in the two-profession model with exogenous training costs, and both per capita income and consumption rise as the skill proportion in steady state increases.*

Proposition 3 tells us that multiplicity—in the sense of a continuum of steady states—is endemic for a small number of professions. While stated only for the two-profession case, it is easy enough to extend the argument to any finite number of distinct professions.

21. This applies only in the unrealistic event that skilled workers cannot perform unskilled tasks. More generally, if skilled workers can perform unskilled tasks, then the skilled wage cannot ever fall below the unskilled wage. So when the skill intensity λ is large enough that $f_1 < f_2$, wages will not be given by f_1 and f_2, but will be equalized (as a result of skilled workers filling unskilled positions whenever the latter pay higher wages). We omit this minor complication here because a competitive equilibrium with a positive fraction of skilled workers will never give rise to such wage configurations in any case.

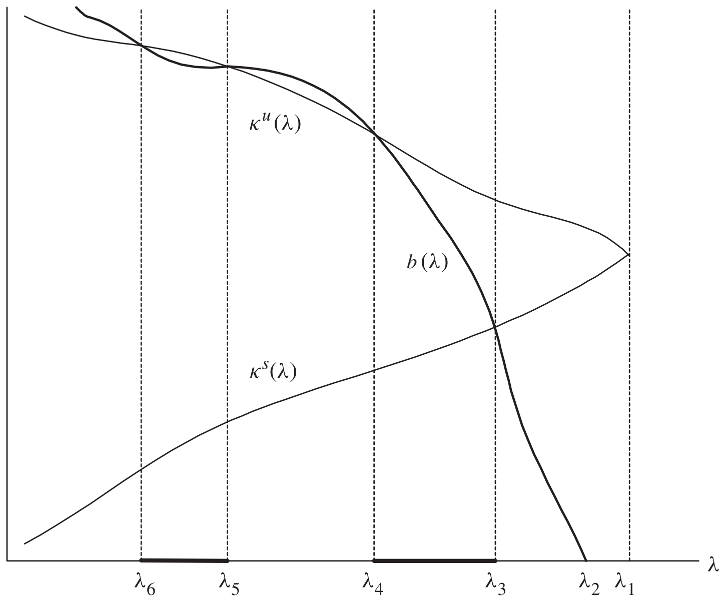

FIGURE 1

Education costs and benefits in two-profession model

Notice that the *structure* of the set of steady states may be complicated. In particular, the set need not be connected. For instance, in Figure 1, the set of steady states is the union of the two intervals $[\lambda_6, \lambda_5]$ and $[\lambda_4, \lambda_3]$.

The proposition also states that steady states are ordered not only in terms of skill premium but also *per capita* income: a steady state with a higher λ and lower skill premium corresponds to higher *per capita* income net of training costs. This does not, however, imply that these steady states are Pareto-ordered. We defer further discussion of efficiency to Section 5.

The societal multiplicity described in Proposition 3 is very much in line with existing literature. We now turn to the question of how this multiplicity is modified when the space of professions is "rich", so that there are no "gaps" in the set of investment options.

4.3. *A continuum of professions*

One way to conceptualize the notion of "richness" in a set of professions is by introducing some notion of continuity in the *cost* of creating professional slots. To this end, assume (to start with) that there is a one-dimensional continuum of professions: $\mathcal{H} = [0, 1]$.[22] We impose the following restriction on the nature of the technology: there is a well-defined unit cost function for each category of professional to be trained. This requires the following assumption.

[T.1] *The set $\mathcal{T}$ is generated from a collection of individual production functions, one for the consumption good, and one each for the training of a professional in every profession h.*

Thus for each professional category h, there is a well-defined production function described by $g(\boldsymbol{\mu}^h, y^h, h)$, where $\boldsymbol{\mu}^h$ is a measure on $[0, 1]$ denoting inputs from different occupations, and y^h the input of the consumption good, into the training of professionals in profession h.

22. Though we do not go into details here, the case of the continuum can indeed be viewed as the limit of a sequence of economies with progressively finer (but finite) occupational structures.

For the consumption good, the production function may simply be written as $f(\boldsymbol{\mu})$, describing net output of the consumption good from distribution $\boldsymbol{\mu}$ over different inputs in the production sector. Hence $\mathcal{T}$ is generated by the collection of production functions $c + \int_{\mathcal{H}} y^h dh = f(\boldsymbol{\mu})$ and $\lambda'(h) = g(\boldsymbol{\mu}^h, y^h, h)$, for $h \in \mathcal{H}$, subject to the aggregate resource constraint $\boldsymbol{\mu} + \int_{\mathcal{H}} \boldsymbol{\mu}^h dh \leq \boldsymbol{\lambda}$. **[T.1]** implies the existence of a well-defined unit cost function for training profession h:

$$\psi(\mathbf{w}, h) \equiv \inf_{\boldsymbol{\mu}, y'} \left\{ y' + \int_{\mathcal{H}} w(h') d\boldsymbol{\mu}(h') \right\}, \qquad \text{subject to } g(\boldsymbol{\mu}, y', h) \geq 1. \tag{11}$$

In a competitive equilibrium, $\psi(\mathbf{w}, h)$ will equal the training cost function $x(h)$, given our assumption of constant returns to scale. Our next assumption is as follows.

[T.2] *The unit cost function $\psi(\mathbf{w}, h)$ is continuous in h for every measurable* **w**.

[T.2] is typically satisfied when the technology is such that the required inputs to train a professional in occupation h can be represented by a *density* function over various professional inputs, which varies continuously in h.[23] The main use of this assumption is to ensure that the training cost function **x** in any steady state is continuous in occupations, thereby implying that every steady state must involve a perfectly "connected" range of investment options, in terms of financial cost and returns. One could just as easily replace this assumption by the weaker requirement that the range of possible training costs is an interval, so that the set of investment options is perfectly divisible.[24]

Proposition 4. *Suppose that the space of professions is* $[0, 1]$, *that* **[T.1]** *and* **[T.2]** *apply, and that the full-support assumption holds. Then, provided that some steady state exists with strictly positive wages for all occupations, there is no other steady state wage function. If, in addition, every production function* (*for the consumption good, as well as for training in each profession*) *is strictly quasiconcave, then there is no other steady state.*

Observe that the proposition pertains to the uniqueness of the entire occupational and earnings distribution. Because this distribution will generally involve inequality, the fortunes of individual families can be highly history dependent. In that sense the steady state is not unique. But since the population proportions are uniquely determined, different steady states can only amount to different permutations of families across occupation, income and consumption categories. The contrast with the Ramsey model (as in Chatterjee, 1994) is stark: the population proportions across different categories are uniquely determined here, whereas in the Ramsey model *any* distribution with the same (uniquely determined) mean constitutes a steady state.

Since the last part of Proposition 4 is straightforward, the part that needs explaining is the uniqueness of the steady state wage function. Focus initially on the case where the cost of acquiring a profession is *exogenously* given by some continuous function $x(h)$ on $[0, 1]$. This is the case where no human capital input of any sort is required in training.

Observe that the steady state condition (9) holds for every occupation h, by the full-support assumption. Imagine testing this condition by moving a tiny amount "up" or "down" in "profession space". For such movements, the curvature of the utility function can (almost) be neglected (since the consumption of each family is constant over time, the marginal disutility of the parent in decreasing its consumption is exactly counterbalanced by the marginal utility of an increase in

23. For instance, **[T.2]** rules out a technology in which profession h is the sole input in the production of professional capacity h.

24. If **[T.2]** is dropped, we can prove the following version of the result. Say that a steady state is *divisible* if the range of $x(h)$ is an interval. Then if **[T.1]** and the full-support assumption holds, and there exists a divisible steady state with a positive and bounded wage function $w(h)$, there cannot exist any other divisible steady state.

its child's future consumption). All that matters then is whether the discounted marginal return is greater or less than the marginal cost of this move. In fact, to make sure that every point *is* a steady state choice (which is required by the full-support postulate), the discounted marginal return must be exactly *equal* to the marginal cost. This proves that for a tiny change $\Delta(h)$,

$$w(h+\Delta h) - w(h) \simeq \frac{1}{\delta}[x(h+\Delta h) - x(h)].$$

By piecing this over all professions, and recalling that $x(0)$ must be zero, we conclude that

$$w(h) = \frac{1}{\delta}x(h) + w(0), \tag{12}$$

where $w(0)$ is just the wage for occupation 0 which does not require any training. Intuitively, there is no room for constructing local variations in the wage structure, owing to the divisibility of the occupational "space" that causes relevant local incentive constraints to bind (*i.e.* across adjacent occupations). When this divisibility is absent, as in the two-occupation case studied above, interior steady states are characterized by incentive constraints that may not bind. This leaves room for local variations in the wage structure that do not disturb the incentive constraints.

To complete the argument of uniqueness (given **x**), note that there cannot be two different values of $w(0)$ that satisfy the steady state condition (8). For if there were, the wage function associated with one must lie completely *above* the wage function associated with the other. Moreover, by profit maximization, *both* these wage functions must be compatible with some nontrivial profit-maximizing choice. But that cannot be, given constant returns to scale and the fact that the price of the consumption good is always normalized to unity.[25]

So far, we assumed that **x** is exogenously given, and showed that there is a single steady state **w**, *given* **x**. The less intuitive part of the proposition is that *there is only one* **w**-*function even when* **x** *is endogenously determined*. This part of the argument makes fundamental use of constant returns to scale, and the reader is invited to study the formal proof for details.[26]

To get a feel for why the endogeneity of **x** does not jeopardize uniqueness, consider the following examples. Recall that the endogeneity of this function arises from the possibility that it takes professionals to train professionals, so that **x** depends on **w**. One elementary formulation is a fixed-coefficients "recursive" training technology: workers proceed incrementally over successive training levels, and to increase one's level of training from $h - dh$ to h requires a fixed proportion $\alpha(h) > 0$ of teachers with training level h: this costs $\alpha(h)w(h)$. This corresponds to the cost function

$$x(h) = \psi(\mathbf{w}, h) = \int_0^h \alpha(h')w(h')dh' \tag{13}$$

which is obviously continuous in h for every measurable **w**, so that **[T.2]** is satisfied. Combining (13) with (12), we see that the wage profile in any limit steady state must belong to the family

$$w(h) = w(0)\exp\left[\int_0^h \frac{\alpha(h')}{\delta}dh'\right]. \tag{14}$$

Smooth steady states are thus pinned down entirely, except for their level, which correspond to the wage $w(0)$ of workers with no training at all. Note, however, that the initial condition $w(0)$ maps out a family of wage functions which is pointwise ordered. By an argument given earlier, it follows that only one value of $w(0)$ is consistent with profit maximization.

25. The argument—that in a "monotonic" family of wage functions there can be at most one member that is consistent with profit maximization—may need to be qualified when there are several consumption goods. In particular, the multiplicity question needs further examination when demand-side compositional effects (as in Baland and Ray (1991), Mani (2001), Matsuyama (2002)) drive the story.

26. This is where we invoke the premise that a steady state exists with positive wages throughout.

Or suppose, alternatively, that the training technology is Cobb–Douglas, with level-h training technology described by the function

$$\log s(h) = \int_0^h \alpha(h') \log t(h') dh' \tag{15}$$

where $s(h)$ is the number of type h students turned out by a process that uses $t(h')$ teachers of type $h' \in [0, h]$. Here, training an h-type requires teachers of all levels up to level h, but there is scope for substitutability among teachers of different levels. Higher level teachers may be more effective, but also more expensive. Hence cost-effective training requires educational institutions to select an optimal teacher mix of different levels given their wage profile, to minimize the cost of turning out each student. This minimization exercise generates the training cost function

$$x(h) = \psi(\mathbf{w}, h) = \exp\left[\int_0^h \alpha(h') \log \frac{w(h')}{\alpha(h')} dh'\right] \tag{16}$$

which once again satisfies **[T.2]**. Combining this with (12), we see that a limit steady state wage profile must satisfy the differential equation

$$w'(h) = \frac{1}{\delta}\alpha(h) \log \frac{w(h)}{\alpha(h)} \exp\left[\int_0^h \alpha(h') \log \frac{w(h')}{\alpha(h')} dh'\right]. \tag{17}$$

Once again, it is evident that the family of wage functions (determined up to a constant of integration) is pointwise ordered, so only one of them is consistent with profit maximization in the final goods sector.

Both examples involve a "recursive" technology, in which the training of level-h individuals depend on indices labelled h or below. This suggests that the set of professions may need to be ordered in some way for the result to work. However, the proof of Proposition 4 is very general and does not rely at all on a recursive technology.

We conclude this section by indicating how the preceding result can be extended when the full-support assumption does not hold. It is not essential that every potential occupation be necessarily chosen by some agents in the economy. What really matters is that the set of effective financial options available is perfectly divisible, *i.e.* the range of training costs and returns from different occupations that are selected in every steady state forms a continuum. For simplicity we consider the case of exogenous training costs.

Proposition 5. *Suppose that:* (a) *the space of professions* $\mathcal{H}$ *is a compact, connected subset of a metric space;* (b) *there is an exogenous training cost function* $x(h)$ *defined on* $\mathcal{H}$ *which is continuous, such that the minimum training cost across all professions is* 0*;* (c) *the set of professions active in any steady state is a connected subset of* $\mathcal{H}$*, which contains both a minimum training cost and a maximum training cost occupation. Then there is a unique steady state wage function. If, in addition, the production function is strictly quasiconcave, there is a unique steady state.*

We skip the proof of this since it is closely related to that of the previous proposition. Here it is not necessary that every occupation is active, only that the set of active occupations has no "holes" and spans the entire range of training costs. With a continuous (exogenous) training cost function, this ensures that the range of observed training costs in any steady state is an interval with a lower end-point equal to 0. All steady states must therefore involve the same span of training costs. The same argument concerning local indifference can now be applied with regard to selection of the level of training cost by each family, to pin down the steady state wage function.

5. EFFICIENCY

Are steady states efficient in the sense of Pareto-optimality? A crucial market is missing, so it would be no surprise if they failed to be efficient. It turns out, however, that the answer is somewhat more complex, and is once again related to the richness of the set of professions.

The concept of efficiency itself requires some discussion. We lay emphasis on the fact that a "continuation value" from any date t is not just the tail utility for generation 0, but *is* the utility of the generation born at date t. Therefore the universe of agents to whom the Pareto criterion should be applied may be described by the collection of all pairs (i, t), where i indexes the dynasty and t the particular member of that dynasty born at date t. Consequently, given some initial distribution $\boldsymbol{\lambda}_0$ over occupations, say that an allocation $\{c_t(i), \boldsymbol{\lambda}_t\}$ is *efficient* if, first, it is feasible

$$(\boldsymbol{\lambda}_t, c_t, \boldsymbol{\lambda}_{t+1}) \in \mathcal{T}$$

for all dates t, where $c_t \equiv \int_{[0,1]} c_t(i)di$, and if there is no other feasible allocation $\{c'_t(i), \boldsymbol{\lambda}'_t\}$ (with $\boldsymbol{\lambda}'_0 = \boldsymbol{\lambda}_0$) such that for every date t

$$\sum\nolimits_{s=t}^{\infty} \delta^{s-t} u(c'_s(i)) \geq \sum\nolimits_{s=t}^{\infty} \delta^{s-t} u(c_s(i)),$$

with strict inequality holding over a set of agents of positive measure at some date.

Proposition 6. *Suppose that a steady state* $(\boldsymbol{\lambda}, c, \mathbf{x}, \mathbf{w})$ *has the property that*

$$x(h) - x(h') = a[w(h) - w(h')] \tag{18}$$

for some $a \geq \delta$*, and for all occupations* h *and* h'*. Then such a steady state is Pareto-efficient.*

To interpret the proposition, note that $|x(h) - x(h')|$ is just the marginal cost of moving up to a "better" profession. The familiar Pareto-optimality condition states that the discounted returns from doing so should *equal* this cost; that is

$$x(h) - x(h') = \delta[w(h) - w(h')].$$

This condition is included in (18), but the latter is weaker. The incremental costs are permitted to *exceed* the incremental wages without threat to Pareto-optimality. In this sense "overinvestment" in human capital is not a source of Pareto-inefficiency. As elaborated below, the reason is that some future generation *must* lose if this apparent overinvestment is eliminated: essentially, the proposition states that there is no way that a current generation can compensate all future generations for the reduction in investment.

Note, however, that the condition also requires a balance between the extent of "overinvestment" in different occupations: that the incremental costs for *every* pair of professions be in excess of the discounted returns by exactly the same ratio (that is, the a in (18) is independent of professions). This balance ensures the absence of Pareto-improving reallocations across professions.

Next we provide a converse to the preceding result which shows that condition (18) is necessary for Pareto-efficiency for steady states satisfying the full-support property. The converse is not exact. We will assume that the number of professions is finite,[27] and that the technology set satisfies the following condition:

27. We make this assumption for technical reasons, and not to suggest that the proposition will fail if the number of professions is infinite. There are some technical conditions involving the appropriate negation of (18) which we would rather avoid.

[T.3] *$\mathcal{T}$ has a smooth boundary, in the sense that every weakly efficient*[28] *point of $\mathcal{T}$ has a unique supporting price vector of the form* $(\mathbf{w}, 1, \mathbf{x})$.

Proposition 7. *Assume that $\mathcal{H}$ is finite and that* **[T.3]** *holds. Suppose that* (18) *fails at some steady state with full support. Then that steady state cannot be Pareto-efficient.*

The proof of this proposition (in the Appendix) provides some understanding for the role of condition (18). This condition could either be violated by a general *underinvestment*, in which the rate of return is equalized across all occupations but this common rate of return exceeds the discount rate δ. Or there may be a misallocation in investment, with rates of return not equalized across occupations. In the former case, the planner can construct a Pareto improvement by investing more at some location in the occupation distribution (redistributing weight towards some occupation h_1 away from another h_2 involving a lower training cost) for some generation t and returning to the previous steady state from the following generation onwards. The deviation is constructed so as to raise consumption for generation t, while reducing it for the previous generation $t-1$, and leaving all generations from $t+1$ onwards unaffected. The changes in consumption for generations $t-1$ and t are distributed equally across all families. Hence those in generation t will be better off, and all those in succeeding generations are not affected at all by the variation. Finally, generation $t-1$ must be better off since the rate of return on education exceeds the factor δ by which the utility of the next generation is discounted. A similar variation can be constructed in the case of a misallocation: educational investments can be reallocated across occupations for some generation t so as to yield a higher aggregate consumption for that generation, while leaving aggregate consumption for future generations unchanged.

To apply the preceding characterization of efficient steady states, consider, first, the continuum case discussed in Section 4.3. The steady state wage function (12) satisfies the conditions of Proposition 6; therefore the unique steady state in that case is Pareto-efficient. Indeed, since the rate of return on investment is uniformly *equal* to the discount rate, there is no overinvestment in this steady state.

Next, consider the two-profession economy. It is easy to apply Propositions 6 and 7 to show that in this case, "high" inequality coexists with inefficiency. The reason is intuitive: given the missing capital market, high inequality is correlated with underinvestment in education. More precisely, we will show that the set of steady states, indexed by the proportion of individuals in the skilled profession, is always partitioned by a threshold proportion—call it λ^*—which itself must belong to the interior of the set of steady states. Steady states in which $\lambda < \lambda^*$ must be inefficient, while steady states with $\lambda \geq \lambda^*$ must be efficient (see Figure 2). This implies that a continuum of efficient and inefficient steady states coexist in the case of two professions.

To see this, simply recall the condition (10) that characterizes a steady state in the two-profession case

$$u(w^s(\lambda)) - u(w^s(\lambda) - x) \leq \frac{\delta}{1-\delta}[u(w^s(\lambda) - x) - u(w^u(\lambda))] \leq u(w^u(\lambda)) - u(w^u(\lambda) - x). \quad (19)$$

Define λ^* by the condition $w^s(\lambda) - w^u(\lambda) = x/\delta$. Notice that by Propositions 6 and 7, and by the particular properties of the functions $w^s(\lambda)$ and $w^u(\lambda)$, a steady state proportion λ is Pareto-efficient if and only if $\lambda \geq \lambda^*$. So it only remains to show that λ^* belongs to the interior of the

28. We look at weakly efficient points because professions that take no resources to produce can be created in unlimited quantities. Of course, the supporting price for such professional capacities (that is, $x(h)$ for profession h) will be zero.

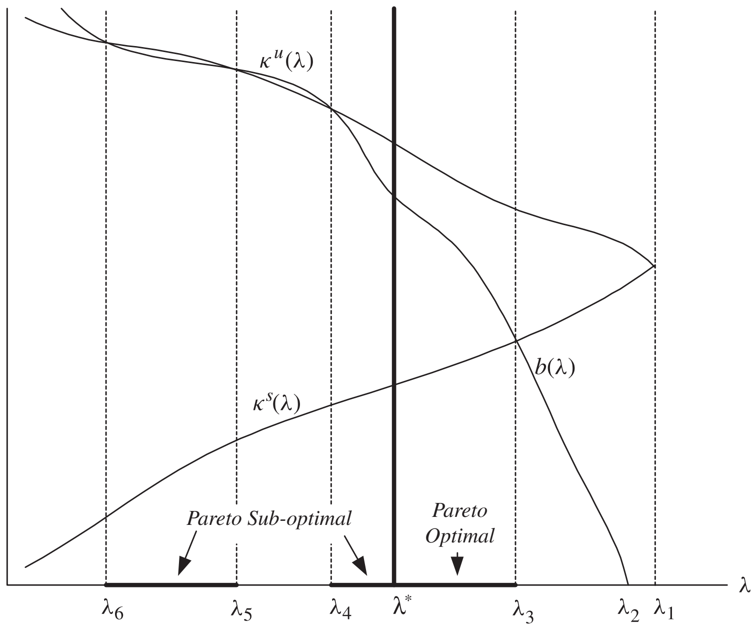

FIGURE 2

The Pareto-efficiency threshold with two professions

set of steady states. This is done by verifying that (19) is satisfied with *strict* inequality when $\lambda = \lambda^*$.[29]

Indeed, this observation for the case of two professions extends in several directions, though considerations of space preclude a full treatment here. For instance, with exogenous training costs (or equivalently, for the case in which training requires only material inputs) there always exists an efficient steady state. To see this, consider the following class of wage functions: $w(h) = \frac{1}{\delta}x(h) + w(0)$, and treat $w(0)$—the wage of the profession that requires no training—as a parameter for the moment. Under weak conditions on the technology,[30] $w(0)$ can be chosen to ensure zero maximal profits in the sector producing the consumption good. Once this is done, a steady state is easy to construct.[31] And exactly the same argument as in the two-profession case guarantees that the intertemporal utility maximization conditions are met. A similar argument extends the result to the case of a "recursive" training technology, where professions can be ordered in a way that the cost of training for any occupation h depends only on wages of occupations ordered below h.

Summarizing, there is no scope for Pareto-improving policies in the case of a continuum of professions where the uniqueness results of Section 4.3 apply. But there may be scope for Pareto-improving policy in other contexts, *e.g.* those in which the occupational structure exhibits

29. Exploit the strict concavity of u to see that $u(w^s(\lambda^*)) - u(w^s(\lambda^*) - x) < u'(w^s(\lambda^*) - x)x = u'(w^s(\lambda^*) - x)\frac{\delta}{1-\delta}[w^s(\lambda^*) - x - w^u(\lambda^*)] < \frac{\delta}{1-\delta}[u(w^s(\lambda^*) - x) - u(w^u(\lambda^*))] < u'(w^u(\lambda^*))\frac{\delta}{1-\delta}[w^s(\lambda^*) - x - w^u(\lambda^*)] = u'(w^u(\lambda^*))x < u(w^u(\lambda^*)) - u(w^u(\lambda^*) - x)$.

30. Essentially, these are Inada conditions on any subset of inputs needed to produce the final good.

31. Letting λ_h denote the number of people in occupation h, a steady state with positive consumption c requires the existence of a gross output λ_0 of the final good such that $\lambda_0 = a_0\lambda_0 + \sum_h x(h)\lambda_h + c$, where $\lambda_h = a_h\lambda_0$ for each h and a_0, a_h denote cost-minimizing input coefficients at the given wages. Such a λ_0 exists for any given c if $1 - a_0 - \sum_h a_h > 0$, which is guaranteed by the zero-profit condition in the final good sector ($1 = a_0 + \sum_h w(h)a_h > a_0 + \sum_h x(h)a_h$).

indivisibilities. Nevertheless, even in such cases—and despite the missing credit market—an efficient steady state will exist in a large class of economies.

6. SUMMARY AND RESEARCH DIRECTIONS

We explored three themes in this paper. First, in contrast with literature which views economic inequality as the outcome of ongoing stochastic shocks, we argue that there are fundamental non-stochastic reasons for persistent inequality when capital markets are imperfect. In particular, long-run inequality (in consumption and utility, not just gross incomes) is inevitable in any steady state with some occupational diversity, irrespective of the degree of foresight or intergenerational altruism of parents, or the divisibility of investment options.[32]

Second, we show that while the fate of *individual* dynasties may be plagued by path dependence, the same may not be true of an economy in which the set of professions is “rich enough” to eliminate indivisibilities in the set of investment options. Under a broad set of conditions there is a unique steady state. Hence if a one-time policy has a permanent effect on some families in a particular way, it will have opposite and compensating effects on other families. With indivisibilities the familiar multiplicity result is recovered, and there is room for one-time policy to have permanent effects.

Finally, we characterize efficient steady states in our model. Because the credit market is missing, it is of interest that some steady states may be efficient. At the same time if there are significant indivisibilities in occupational choice—such as the case of only two professions with an exogenous training cost—there are two continua of efficient and inefficient steady states. The inefficient steady states involve underinvestment and greater inequality than every efficient steady state. In such circumstances there is potential scope for temporary policies or historical shocks to raise long-run *per capita* income while reducing inequality.

We conclude by describing topics for subsequent research. First, there is the question of non-steady state dynamics. This is important for understanding how inequality evolves over time. We have been able to resolve this question fully in the two-occupation case, where uniqueness of competitive equilibrium from arbitrary initial conditions and global convergence to a steady state obtains (for details see Ray (1990), Mookherjee and Ray (2000)). It would be important to see whether these results extend to a more general occupational structure.

The second major research question concerns an extension to the case in which financial bequests can supplement human capital investments. While this is formally a special case of our model, the appropriate interpretation of this case implies that the occupational diversity condition (needed for persistent inequality) can be quite strong. Families can effectively lend while being restricted in their borrowing. Conceivably there might then be steady states without inequality, where less skilled dynasties compensate for their lower human capital by holding and bequeathing more financial wealth. The burden then falls more squarely on an examination of non-steady state dynamics. Some preliminary observations on the interplay between financial bequests and human capital investments are reported in Mookherjee and Ray (2002*b*).

Finally, we believe that the model developed here can be fruitful for the analysis of dynamic effects of trade, technical change, financial sector reforms, and redistributive policies. One may also use the model to understand the effects of economic integration on inequality across countries, by interpreting each family to represent a different country. For instance, one could view “occupational” setup costs as infrastructural investments made by the planners to facilitate a particular mix of economic activities in each country (*e.g.* a country may decide to subsidize

32. When financial bequests are available, the occupational diversity condition may be strong; see the remarks given in what follows.

agriculture, promote exports, or invest in high-technology production capabilities). Then—in the absence of a perfect international capital market to finance these investments—global inequality must emerge, with historical events determining the subsequent fate of individual countries. Nevertheless, while individual fates can be altered, the world economy must exhibit a certain compositional balance, if the investment technology allows a sufficiently diverse set of options.

APPENDIX

Proof of Proposition 1. Say that an occupation h is *dominated* if there is a distinct occupation g such that $x(g) \leq x(h)$ and $w(g) \geq w(h)$, with at least one of these inequalities strict. It should be obvious that there is no set of dominated occupations with positive measure under $\boldsymbol{\lambda}$.

Now suppose that the proposition is false, and there is a set of individuals of positive measure such that for each individual in this set, a switch (to a distinct profession) takes place at some date. Then—because there are only a countable infinity of dates—there is some *common* date at which a professional switch takes place for a positive measure of individuals.

Claim. *There exist undominated professions h, h', g and g' such that a person with occupation h moves to g, one with h' moves to g' and the following property is satisfied: $x(h) < x(h')$ and $x(g) > x(g')$.*

To prove this claim, note that if a positive measure of people switch professions (say "up" from h to g or "down" from h' to g'), then to maintain the steady state distribution there must be flows in the opposite direction. Moreover, all these professions must be undominated, because no set of dominated professions exhibits positive measure under $\boldsymbol{\lambda}$.

The claim implies that there exist initial professions h and h' such that $w(h) < w(h')$, but with the property that the optimal choice of professions (g and g' respectively) satisfies $x(g) > x(g')$. Let $V(h)$ denote the value of starting at h under the going steady state. Then, because g' is feasible for h (after all, $x(g') < x(g)$),

$$u(w(h) - x(g)) + \delta V(g) \geq u(w(h) - x(g')) + \delta V(g'),$$

while because g is feasible under $w(h')$ (as g is feasible under $w(h)$ and $w(h) < w(h')$),

$$u(w(h') - x(g')) + \delta V(g') \geq u(w(h') - x(g)) + \delta V(g).$$

Combining these two inequalities and cancelling common terms, we see that

$$u(w(h') - x(g')) - u(w(h) - x(g')) \geq u(w(h') - x(g)) - u(w(h) - x(g)). \tag{A.1}$$

However, given that $w(h) < w(h')$ and $x(g') < x(g)$, (A.1) contradicts the strict concavity of u. ||

Proof of Proposition 2. Let h and h' be distinct professions with $x(h) > x(h')$. Then (because dominated professions cannot be inhabited), $w(h) > w(h')$. Now we know by Proposition 1 that for a person at h, choosing h represents the best continuation. It follows that

$$\begin{aligned}\frac{u(w(h) - x(h))}{1-\delta} &\geq u(w(h) - x(h')) + \frac{\delta u(w(h') - x(h'))}{1-\delta}\\ &> u(w(h') - x(h')) + \frac{\delta u(w(h') - x(h'))}{1-\delta}\\ &= \frac{u(w(h') - x(h'))}{1-\delta},\end{aligned}$$

which shows that a person at h has higher lifetime utility than a person at h'. Because no person switches professions at a steady state (Proposition 1), the person at h must have a higher utility at every date compared to the person at h'. ||

Proof of Proposition 3. By the Inada conditions, there exists λ_3 such that $b(\lambda)$ and $\kappa^s(\lambda)$ are equalized. Notice that λ_3 must be strictly less than λ_2, which in turn is less than λ_1. So, using the strict concavity of the utility function, it must be the case that $\kappa^u(\lambda_3) > \kappa^s(\lambda_3) = b(\lambda_3)$. Thus (10) is satisfied at λ_3 and we have a steady state.

Now use the slopes of these curves to argue that for all $\lambda < \lambda_3$ but sufficiently close to it,

$$\kappa^u(\lambda) \geq b(\lambda_3) \geq \kappa^s(\lambda_3),$$

which establishes that there must be a continuum of steady states.

To see that the steady states are ordered in terms of net output, consider the following maximization problem for the net output:

$$\max_{\lambda \geq 0} f(\lambda, 1-\lambda) - x\lambda. \tag{A.2}$$

This is a strictly concave problem in λ and attains a unique maximum when $f_1 - f_2 = x$. Recalling that $w^s = f_1$ while $w^u = f_2$, we conclude that this is the point λ such that $w^s(\lambda) - x = w^u(\lambda)$, which is precisely λ_2 in Figure 1. Because every steady state lies to the left of λ_2 and the maximization problem (A.2) is strictly concave, the result follows. ‖

Proof of Proposition 4. The following elementary lemmas will be used.

Lemma 1. *The unit cost function $\psi(\mathbf{w}, h)$ has the following properties:*

[1] *If two wage functions $\mathbf{w}$ and $\hat{\mathbf{w}}$ satisfy $\hat{w}(h) \geq w(h)$ for every h, then $\psi(\hat{\mathbf{w}}, h) \geq \psi(\mathbf{w}, h)$ for every h.*
[2] *For every scalar $\alpha \geq 1$ and each h, $\psi(\alpha\mathbf{w}, h) \leq \alpha\psi(\mathbf{w}, h)$.*
[3] *For every scalar $\alpha \in [0, 1]$ and each h, $\psi(\alpha\mathbf{w}, h) \geq \alpha\psi(\mathbf{w}, h)$.*
[4] *In any steady state $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$, $x(h) = \psi(\mathbf{w}, h)$ for all h.*

The proofs are obvious and therefore omitted. The verification of [2] and [3] uses constant returns to scale, coupled with the fact that the price of the final good (which may be an input in the production of some h) is normalized to unity.

Lemma 2. *Under the full-support assumption, there cannot be two steady states with distinct wage functions $\hat{\mathbf{w}}$ and $\mathbf{w}$ such that $\hat{w}(h) \geq w(h)$ for all h.*

Proof. Suppose the lemma is false. Then not only is $\hat{w}(h) \geq w(h)$ for all h, strict inequality holds on a set of positive measure. Consider some steady state input distribution $\hat{\boldsymbol{\lambda}}$ that produces the final good at level $\hat{c}$. By profit maximization and constant returns to scale in the production sector,

$$\hat{c} - \hat{\mathbf{w}}\hat{\boldsymbol{\lambda}} = 0,$$

so that by the full-support postulate,

$$\hat{c} - \mathbf{w}\hat{\boldsymbol{\lambda}} > 0.$$

But (given constant returns to scale) this violates profit maximization at the steady state with wage function $\mathbf{w}$. ‖

For the main proof, we retrace the steps of the informal discussion. Fix some steady state $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$. We first prove the following claim: (12) holds for all h.

If $\mathbf{x}$ is zero throughout (12) follows trivially, as wages must be constant for all h. And if some training costs are positive, Lemma 1 (part [4]) and the continuity of ψ implies that $x(h)$ must be continuous in h, so the range of x is an interval of the form $[0, X]$ for some $X > 0$. Obviously, there is a function W defined on $[0, X]$ such that for every h with $x(h) > 0$, $w(h) = W(x(h))$. The full-support assumption implies that every x in $[0, X]$ is chosen by some family.

This implies that W must be continuous. Otherwise some level of x in the neighbourhood of a discontinuity will not be chosen, as it will be dominated by a neighbouring x' associated with a substantially higher wage.

Next, consider any x in the interior of $[0, X]$. Then, invoking (9) and using the same argument leading up to (10), we see that for every $\epsilon > 0$ and sufficiently small,

$$\begin{aligned} u(W(x) - x) - u(W(x) - (x+\epsilon)) &\geq \frac{\delta}{1-\delta}[u(W(x) - (x+\epsilon)) - u(W(x) - x)] \\ &\geq u(W(x+\epsilon) - x) - u(W(x+\epsilon) - (x+\epsilon)). \end{aligned}$$

Dividing these terms throughout by ϵ, applying the concavity of the utility function to the two side terms, and the mean value theorem to the central term, we see that

$$u'(W(x) - [x+\epsilon]) \geq \frac{\delta}{1-\delta} u'(\theta(\epsilon)) \left[\frac{W(x+\epsilon) - W(x)}{\epsilon} - 1\right] \geq u'(W(x+\epsilon) - x), \tag{A.3}$$

where $\theta(\epsilon)$ lies between $W(x) - x$ and $W(x+\epsilon) - (x+\epsilon)$. Now we may send ϵ to zero in (A.3) and use the continuous differentiability of u to conclude that

$$\lim_{\epsilon \downarrow 0} \frac{W(x+\epsilon) - W(x)}{\epsilon} \text{ exists, and equals } \frac{1}{\delta}.$$

Exactly the same argument applies when $x = X$ (respectively $x = 0$) to show the left-differentiability (respectively right-differentiability) of W at that point. We may therefore conclude that for all $x \in [0, X]$:

$$W(x) = \frac{1}{\delta}x + w(0).$$

This establishes our claim that every steady state must satisfy (12) for all h.

With this claim in hand, we can complete the proof. Suppose that there is a steady state wage function $\mathbf{w}$ with strictly positive wages throughout. Then, by the claim,

$$w(h) = \frac{1}{\delta}x(h) + w(0). \tag{A.4}$$

Suppose, in contrast to the proposition, that there is another steady state $(\tilde{\boldsymbol{\lambda}}, \tilde{c}, \tilde{\mathbf{w}}, \tilde{\mathbf{x}})$ with a distinct wage function. Applying the claim again, we know that

$$\tilde{w}(h) = \frac{1}{\delta}\tilde{x}(h) + \tilde{w}(0) \tag{A.5}$$

for every h. Combining (A.4) and (A.5), we see that

$$\frac{\tilde{x}(h)}{x(h)} = \frac{\tilde{w}(h) - \tilde{w}(0)}{w(h) - w(0)} \tag{A.6}$$

for all h such that both $x(h)$ and $\tilde{x}(h)$ are not simultaneously zero, interpreting this ratio to be ∞ in case $x(h) = 0$.

Now define $\alpha \equiv \max \frac{\tilde{w}(h)}{w(h)}$ and $\beta \equiv \min \frac{\tilde{w}(h)}{w(h)}$. Because $\mathbf{w}$ and $\tilde{\mathbf{w}}$ are continuous functions and $w(h) > 0$ everywhere, these terms are well defined. Notice, moreover, that $\alpha > 1$ and $\beta < 1$ if the two wage functions are distinct (by virtue of Lemma 2).

Case 1. $\frac{\tilde{w}(0)}{w(0)} < \alpha$. Let $h^* > 0$ be some value of h such that α is attained. Then it is easy to see that

$$\frac{\tilde{x}(h^*)}{x(h^*)} = \frac{\tilde{w}(h^*) - \tilde{w}(0)}{w(h^*) - w(0)} > \alpha. \tag{A.7}$$

Define a new wage function $\mathbf{w}''$ such that $w''(h) \equiv \alpha w(h)$ for all h. Then, using the fact that $\alpha > 1$ and invoking Lemma 1, part [2],

$$\psi(\mathbf{w}'', h^*) \leq \alpha\psi(\mathbf{w}, h^*) = \alpha x(h^*),$$

while by Lemma 1, part [1],

$$\tilde{x}(h^*) = \psi(\tilde{\mathbf{w}}, h^*) \leq \psi(\mathbf{w}'', h^*).$$

Combining these two inequalities, we may conclude that

$$\tilde{x}(h^*) \leq \alpha x(h^*),$$

which contradicts (A.7).

Case 2. $\frac{\tilde{w}(0)}{w(0)} = \alpha$. Let $h_* > 0$ be some value of h such that β is attained. Then, parallel to (A.7), we see that

$$\frac{\tilde{x}(h_*)}{x(h_*)} = \frac{\tilde{w}(h_*) - \tilde{w}(0)}{w(h_*) - w(0)} < \beta. \tag{A.8}$$

Continuing the parallel argument, define a function $\mathbf{w}'''$ such that $w'''(h) \equiv \beta w(h)$ for all h. Then, using the fact that $\beta < 1$ and Lemma 1, part [3],

$$\psi(\mathbf{w}''', h_*) \geq \beta\psi(\mathbf{w}, h_*) = \beta x(h_*),$$

while by Lemma 1, part [1],

$$\tilde{x}(h_*) = \psi(\tilde{\mathbf{w}}, h_*) \geq \psi(\mathbf{w}''', h_*).$$

Combining these two inequalities, we see that

$$\tilde{x}(h_*) \geq \beta x(h_*),$$

which contradicts (A.8).

Thus, in both cases we have a contradiction, so that the first part of the proposition is established. The second part is obvious and needs no proof. ||

Proof of Proposition 6. We rely on the following result.

Lemma 3. *Fix $c \geq 0$, and suppose that $\{c_s\}$ is a non-negative sequence starting from date t, not identical to c at every $s \geq t$. Then, provided that*

$$\sum_{s=t}^{\infty} \delta^{s-t} u(c_s) \geq \sum_{s=t}^{\infty} \delta^{s-t} u(c), \tag{A.9}$$

we must have

$$\sum_{s=t}^{\infty} \delta^{s-t} c_s > \sum_{s=t}^{\infty} \delta^{s-t} c. \tag{A.10}$$

Proof. Suppose that there is a sequence of consumptions $\{c_s\}_{s=t}^{\infty}$, distinct from c at some $s \geq t$, such that (A.9) holds. By an elementary inequality involving strictly concave functions, we know that

$$u'(c)[c_s - c] \geq u(c_s) - u(c), \tag{A.11}$$

with strict inequality holding whenever $c_s \neq c$.

Combining (A.9) and (A.11), we see that

$$u'(c) \sum_{s=t}^{\infty} \delta^{s-t}[c_s - c] > \sum_{s=t}^{\infty} \delta^{s-t}[u(c_s) - u(c)] \geq 0,$$

and this completes the proof. ‖

Now return to the proof of the theorem. Suppose, in contrast, that there is some Pareto-improving allocation $\{c_t(i), \boldsymbol{\lambda}_t\}$ with

$$(c_t, \boldsymbol{\lambda}_t, \boldsymbol{\lambda}_{t+1}) \in \mathcal{T}$$

for all dates t, where $c_t \equiv \int_{[0,1]} c_t(i)di$, and such that initial conditions are respected.

Then two things must happen. First the new allocation must be distinct for a positive measure of individuals (at some date) from the old, and second, no individual at *any* date can be worse off. Consequently, using (A.10) at any date t and adding up over all agents, we see that

$$\sum_{s=t}^{\infty} \delta^{s-t} c_s \geq \sum_{s=t}^{\infty} \delta^{s-t} c, \tag{A.12}$$

where c is aggregate steady state consumption. Moreover, strict inequality must hold for *some* date t.

Now, we know that at the steady state prices $(\mathbf{w}, \mathbf{x})$, profits are maximized at the steady state allocation. Consequently, for each date s,

$$c + \mathbf{x} \cdot \boldsymbol{\lambda} - \mathbf{w} \cdot \boldsymbol{\lambda} \geq c_s + \mathbf{x} \cdot \boldsymbol{\lambda}_{s+1} - \mathbf{w} \cdot \boldsymbol{\lambda}_s. \tag{A.13}$$

Taking discounted sums and invoking (A.12) from Lemma 3, we see that

$$\frac{\mathbf{x} \cdot \boldsymbol{\lambda} - \mathbf{w} \cdot \boldsymbol{\lambda}}{1-\delta} \geq \sum_{s=t}^{\infty} \delta^{s-t}[\mathbf{x} \cdot \boldsymbol{\lambda}_{s+1} - \mathbf{w} \cdot \boldsymbol{\lambda}_s]$$

for every t, with strict inequality at date 0. Using (18), it can be seen that

$$\frac{a-1}{1-\delta} \mathbf{w} \cdot \boldsymbol{\lambda} \geq \sum_{s=t}^{\infty} \delta^{s-t}[a\mathbf{w} \cdot \boldsymbol{\lambda}_{s+1} - \mathbf{w} \cdot \boldsymbol{\lambda}_s] \tag{A.14}$$

for every t, recalling once again that strict inequality must hold at date 0.

Leave the inequality at $t = 0$ undisturbed, but for $t \geq 1$ multiply the corresponding inequality on both sides by $(a-\delta)a^{t-1}$. Then for any $t \geq 1$, we have

$$a^{t-1}(a-\delta)\frac{a-1}{1-\delta} \mathbf{w} \cdot \boldsymbol{\lambda} \geq (a-\delta) \sum_{s=t}^{\infty} \delta^{s-t}[a^t \mathbf{w} \cdot \boldsymbol{\lambda}_{s+1} - a^{t-1}\mathbf{w} \cdot \boldsymbol{\lambda}_s]. \tag{A.15}$$

Add these inequalities over all $t \geq 1$. Notice that $a < 1$, otherwise we cannot have a steady state competitive equilibrium. Therefore

$$\begin{aligned}
-\frac{a-\delta}{1-\delta} \mathbf{w} \cdot \boldsymbol{\lambda} &\geq (a-\delta) \sum_{t=1}^{\infty} \sum_{s=t}^{\infty} \delta^{s-t}[a^t \mathbf{w} \cdot \boldsymbol{\lambda}_{s+1} - a^{t-1}\mathbf{w} \cdot \boldsymbol{\lambda}_s] \\
&= (a-\delta) \sum_{s=1}^{\infty} \sum_{t=1}^{s} \delta^{s-t}[a^t \mathbf{w} \cdot \boldsymbol{\lambda}_{s+1} - a^{t-1}\mathbf{w} \cdot \boldsymbol{\lambda}_s] \\
&= (a-\delta) \sum_{s=1}^{\infty} \left[\sum_{t=1}^{s} \delta^s \left(\frac{a}{\delta}\right)^t \mathbf{w} \cdot \boldsymbol{\lambda}_{s+1} - \sum_{t=1}^{s} \frac{\delta^s}{a} \left(\frac{a}{\delta}\right)^t \mathbf{w} \cdot \boldsymbol{\lambda}_s\right] \\
&= \sum_{s=1}^{\infty} [a(a^s - \delta^s)\mathbf{w} \cdot \boldsymbol{\lambda}_{s+1} - (a^s - \delta^s)\mathbf{w} \cdot \boldsymbol{\lambda}_s].
\end{aligned} \tag{A.16}$$

Now add both sides of (A.16) to the corresponding sides of the inequality (A.14) for $t = 0$. Remembering that this latter inequality is strict, we see that

$$-\mathbf{w}\cdot\boldsymbol{\lambda} > \sum_{s=1}^{\infty}[a(a^s-\delta^s)\mathbf{w}\cdot\boldsymbol{\lambda}_{s+1}-(a^s-\delta^s)\mathbf{w}\cdot\boldsymbol{\lambda}_s]+\sum_{s=0}^{\infty}\delta^s[a\mathbf{w}\cdot\boldsymbol{\lambda}_{s+1}-\mathbf{w}\cdot\boldsymbol{\lambda}_s].$$

But careful inspection of the R.H.S. of this inequality shows that it is also equal to $-\mathbf{w}\cdot\boldsymbol{\lambda}$, which is a contradiction. This completes the proof. ‖

Proof of Proposition 7. The following standard lemma will be used.

Lemma 4. *Suppose that a boundary point* $\mathbf{z} = (\boldsymbol{\mu}, c, \boldsymbol{\sigma})$ *of* $\mathcal{T}$ *has a unique supporting price* $\mathbf{p} = (\mathbf{w}, 1, \mathbf{x})$. *Suppose further that for some alternative allocation* $\mathbf{z}'$ *(not necessarily feasible),* $\mathbf{p}\cdot\mathbf{z}' < 0$. *Then for all* $\alpha \in (0, 1)$ *and sufficiently close to zero,* $(1-\alpha)\mathbf{z}+\alpha\mathbf{z}' \in \mathcal{T}$.

Proof. Standard. See, *e.g.* Rockafellar (1979, Theorem 2). ‖

We now turn to the proof of the proposition. Suppose that (18) is false at some steady state $(\boldsymbol{\lambda}, c, \mathbf{w}, \mathbf{x})$. Then one of the following must be true.

[I] There are professions h_1 and h_2 with $x(h_1) > x(h_2)$ such that

$$w(h_1) - w(h_2) > \frac{1}{\delta}[x(h_1) - x(h_2)]. \tag{A.17}$$

[II] There are four professions h_1, h_2, h_3 and h_4 (not necessarily all distinct) with $x(h_1) < x(h_2)$ and $x(h_3) < x(h_4)$ such that

$$\frac{w(h_2) - w(h_1)}{x(h_2) - x(h_1)} > \frac{w(h_4) - w(h_3)}{x(h_4) - x(h_3)}. \tag{A.18}$$

Accordingly, we divide the analysis into two cases.

Case 1. [I] is true. Then there is $\nu \in (0, 1)$ and $\eta > 0$ such that

$$\frac{\nu w(h_1) - w(h_2)}{x(h_1) - x(h_2)} > \eta > \frac{1}{\delta}. \tag{A.19}$$

Fix these two numbers in what follows. For any $\epsilon > 0$ and small, define the distribution $\boldsymbol{\lambda}_\epsilon$ by

$$\lambda_\epsilon(h_1) \equiv \lambda(h_1) + \frac{\epsilon\nu}{x(h_1) - x(h_2)},$$

and

$$\lambda_\epsilon(h_2) \equiv \lambda(h_2) - \frac{\epsilon\nu}{x(h_1) - x(h_2)}, \tag{A.20}$$

while $\lambda_\epsilon(h) = \lambda(h)$ otherwise (where $\boldsymbol{\lambda}$ is the original steady state distribution).

We first claim that there exists $\epsilon_1 > 0$ such that

$$(\boldsymbol{\lambda}, c - \epsilon, \boldsymbol{\lambda}_\epsilon) \in \mathcal{T} \tag{A.21}$$

for all $\epsilon \in (0, \epsilon_1)$.

To establish this claim, calculate the "profit" generated by the allocation $\mathbf{z}_\epsilon \equiv (\boldsymbol{\lambda}, c - \epsilon, \boldsymbol{\lambda}_\epsilon)$ (not necessarily feasible) at the steady state price vector $\mathbf{p} = (\mathbf{w}, 1, \mathbf{x})$. We see that

$$\mathbf{p}\cdot\mathbf{z}_\epsilon = (c - \epsilon) + \mathbf{x}\cdot\boldsymbol{\lambda}_\epsilon - \mathbf{w}\cdot\boldsymbol{\lambda} = (\nu - 1)\epsilon < 0,$$

where use has been made of (A.20) and the fact that $c+\mathbf{x}\cdot\boldsymbol{\lambda}-\mathbf{w}\cdot\boldsymbol{\lambda} = 0$. By Lemma 4, we know that for all $\alpha \in (0, 1)$ and sufficiently small, $(1-\alpha)\mathbf{z}+\alpha\mathbf{z}_\epsilon \in \mathcal{T}$, where $\mathbf{z} \equiv (\boldsymbol{\lambda}, c, \boldsymbol{\lambda})$. Using (A.20), this is easily seen to be equivalent to (A.21) (for ϵ small enough), and the claim is established.

Next, we claim that there exists $\epsilon_2 > 0$ such that

$$(\boldsymbol{\lambda}_\epsilon, c + \eta\epsilon, \boldsymbol{\lambda}) \in \mathcal{T} \tag{A.22}$$

for all $\epsilon \in (0, \epsilon_2)$, where η is defined in (A.19). To see this, calculate the "profit" generated by the allocation $\mathbf{z}'_\epsilon \equiv (\boldsymbol{\lambda}_\epsilon, c + \eta\epsilon, \boldsymbol{\lambda})$:

$$\begin{aligned}\mathbf{p} \cdot \mathbf{z}'_\epsilon &= (c + \eta\epsilon) + \mathbf{x} \cdot \boldsymbol{\lambda} - \mathbf{w} \cdot \boldsymbol{\lambda}_\epsilon \\ &= c + \eta\epsilon + \mathbf{x} \cdot \boldsymbol{\lambda} - \mathbf{w} \cdot \boldsymbol{\lambda} - \epsilon\nu \frac{w(h_1) - w(h_2)}{x(h_1) - x(h_2)} \\ &< c + \eta\epsilon + \mathbf{x} \cdot \boldsymbol{\lambda} - \mathbf{w} \cdot \boldsymbol{\lambda} - \eta\epsilon = c + \mathbf{x} \cdot \boldsymbol{\lambda} - \mathbf{w} \cdot \boldsymbol{\lambda} = 0,\end{aligned}$$

where the inequality in this string uses (A.19). So once again, by Lemma 4, we may conclude that for all $\alpha \in (0, 1)$ and sufficiently small, $(1 - \alpha)\mathbf{z} + \alpha\mathbf{z}'_\epsilon \in \mathcal{T}$. Using (A.20), this is easily seen to be equivalent to (A.22) (for ϵ small enough).

We use these constructions to create a path that Pareto-dominates the steady state. Consider the following sequence of production plans: $(\mathbf{z}_\epsilon, \mathbf{z}'_\epsilon, \mathbf{z}, \mathbf{z}, \ldots)$, where $0 < \epsilon < \min\{\epsilon_1, \epsilon_2\}$. By (A.21) and (A.22), such a path is (technologically) feasible.

Relative to the steady state, this path displays an aggregate consumption shortfall of ϵ in period 0, an aggregate consumption excess of $\eta\epsilon$ in period 1, and no difference thereafter. Divide these transitory differences equally among all the agents. Notice that agents after period 1 are unaffected, while all agents at period 1 are strictly better off. It remains to check agents at period 0. The gain in utility for each person i at date 0 is just $\Delta(i) \equiv [u(c(i) - \epsilon) + \delta u(c(i) + \eta\epsilon)] - [u(c(i)) + \delta u(c(i))]$. Notice that

$$\Delta(i) \geq \delta u'(c(i) + \eta\epsilon)\eta\epsilon - u'(c(i) - \epsilon)\epsilon = \frac{\epsilon}{u'(c(i) - \epsilon)}\left[\delta\eta \frac{u'(c(i) + \eta\epsilon)}{u'(c(i) - \epsilon)} - 1\right].$$

Now, there are only a finite number of possible values which $c(i)$ can assume, and all of them are strictly positive. Use this information together with the smoothness of u, and the fact that $\delta\eta > 1$ (from (A.19)) to conclude that for ϵ small enough,

$$\Delta(i) > 0$$

for every agent i. This completes the proof in Case 1.

Case 2. [II] is true. With (A.18) in mind, choose ρ such that

$$\frac{w(h_2) - w(h_1)}{w(h_4) - w(h_3)} > \rho > \frac{x(h_2) - x(h_1)}{x(h_4) - x(h_3)}, \tag{A.23}$$

and then γ such that

$$0 < \gamma < \rho[x(h_4) - x(h_3)] - [x(h_2) - x(h_1)]. \tag{A.24}$$

Now adjust the steady state distribution $\boldsymbol{\lambda}$ as follows. For any $\epsilon > 0$ and small, define $\boldsymbol{\lambda}_\epsilon$ by

$$\begin{aligned}\lambda_\epsilon(h_1) &\equiv \lambda(h_1) - \epsilon, \\ \lambda_\epsilon(h_2) &\equiv \lambda(h_2) + \epsilon, \\ \lambda_\epsilon(h_3) &\equiv \lambda(h_3) + \rho\epsilon,\end{aligned}$$

and

$$\lambda_\epsilon(h_4) \equiv \lambda(h_4) - \rho\epsilon, \tag{A.25}$$

while $\lambda_\epsilon(h) = \lambda(h)$ otherwise. We claim that there exists $\epsilon_3 > 0$ such that

$$(\boldsymbol{\lambda}, c + \gamma\epsilon, \boldsymbol{\lambda}_\epsilon) \in \mathcal{T} \tag{A.26}$$

for all $\epsilon \in (0, \epsilon_3)$.

To establish this, observe that if $\mathbf{z}_\epsilon \equiv (\boldsymbol{\lambda}, c + \gamma\epsilon, \boldsymbol{\lambda}_\epsilon)$ and $\mathbf{z} \equiv (\boldsymbol{\lambda}, c, \boldsymbol{\lambda})$, then

$$\begin{aligned}\mathbf{p} \cdot \mathbf{z}_\epsilon = \mathbf{p} \cdot \mathbf{z}_\epsilon - \mathbf{p} \cdot \mathbf{z} &= \gamma\epsilon - x(h_1)\epsilon + x(h_2)\epsilon - x(h_4)\rho\epsilon + x(h_3)\rho\epsilon \\ &= \gamma\epsilon + \{[x(h_2) - x(h_1)] - \rho[x(h_4) - x(h_3)]\}\epsilon \\ &< \gamma\epsilon - \gamma\epsilon = 0,\end{aligned}$$

where the last inequality uses (A.24). Applying Lemma 4 as before, we are finished.

Next, we claim that there exists $\epsilon_4 > 0$ such that

$$(\boldsymbol{\lambda}_\epsilon, c, \boldsymbol{\lambda}) \in \mathcal{T} \tag{A.27}$$

for all $\epsilon \in (0, \epsilon_4)$. To prove this, define $\mathbf{z}'_\epsilon \equiv (\boldsymbol{\lambda}_\epsilon, c, \boldsymbol{\lambda})$ and note that

$$\begin{aligned}\mathbf{p} \cdot \mathbf{z}'_\epsilon = \mathbf{p} \cdot \mathbf{z}'_\epsilon - \mathbf{p} \cdot \mathbf{z} &= -w(h_2)\epsilon + w(h_1)\epsilon - w(h_3)\rho\epsilon + w(h_4)\rho\epsilon \\ &= \epsilon\{[w(h_4) - w(h_3)]\rho - [w(h_2) - w(h_1)]\} \\ &< 0,\end{aligned}$$

where the last inequality uses (A.23). The claim then follows from a final application of Lemma 4.

Just as in Case 1, we may now construct a Pareto-dominating path. Consider the sequence of production plans $(\mathbf{z}_\epsilon, \mathbf{z}'_\epsilon, \mathbf{z}, \mathbf{z}, \ldots)$, where $0 < \epsilon < \min\{\epsilon_3, \epsilon_4\}$. By (A.26) and (A.27), such a path is (technologically) feasible. Relative to the steady state, this path displays an aggregate consumption surplus of $\gamma\epsilon$ in period 0 and no difference thereafter. Divide this surplus equally among all date-0 agents. Clearly, a Pareto-improvement has taken place, and the proof is complete. ‖

Acknowledgements. We thank Abhijit Banerjee, Michael Kremer, Glenn Loury, Andy Newman and Rajiv Vohra for helpful discussions, and seminar participants at Columbia, Harvard and the July 2000 NBER Income Distribution workshop for their comments. An Editor and two anonymous referees provided helpful comments on an earlier draft. This research was supported by the National Science Foundation (NSF grant No. SBS-9709254), and Mookherjee was additionally funded by the MacArthur Foundation.

REFERENCES

AGHION, P. and BOLTON, P. (1997), "A Theory of Trickle-Down Growth and Development", *Review of Economic Studies*, **64**, 151–172.

BALAND, J.-M. and RAY, D. (1991), "Why Does Asset Inequality Affect Unemployment? A Study of the Demand Composition Problem", *Journal of Development Economics*, **35**, 69–92.

BANDYOPADHYAY, D. (1993), "Distribution of Human Capital, Income Inequality and Rate of Growth" (Ph.D. Thesis, University of Minnesota).

BANDYOPADHYAY, D. (1997), "Distribution of Human Capital and Economic Growth" (Working Paper No. 173, Department of Economics, Auckland Business School).

BANERJEE, A. and NEWMAN, A. (1993), "Occupational Choice and the Process of Development", *Journal of Political Economy*, **101**, 274–298.

BARRO, R. (1974), "Are Government Bonds Net Wealth?", *Journal of Political Economy*, **82**, 1095–1117.

BECKER, G. and TOMES, N. (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility", *Journal of Political Economy*, **87**, 1153–1189.

CHATTERJEE, S. (1994), "Transitional Dynamics and the Distribution of Wealth in a Neoclassical Growth Model", *Journal of Public Economics*, **54**, 97–119.

CLARK, C. (1971), "Economically Optimal Policies for the Utilization of Biologically Renewable Resources", *Mathematical Biosciences*, **12**, 245–260.

DECHERT, R. and NISHIMURA, K. (1983), "A Complete Characterization of Optimal Growth Paths in an Aggregated Model With a Non-Concave Production Function", *Journal of Economic Theory*, **31**, 332–354.

FREEMAN, S. (1996), "Equilibrium Income Inequality Among Identical Agents", *Journal of Political Economy*, **104** (5), 1047–1064.

GALOR, O. and ZEIRA, J. (1993), "Income Distribution and Macroeconomics", *Review of Economic Studies*, **60** (1), 35–52.

HOFF, K. and STIGLITZ, J. (2001), "Modern Economic Theory and Development", in G. Meier and J. Stiglitz (eds.) *Frontiers of Development Economics* (New York: World Bank and Oxford University Press) 389–459.

LJUNGQVIST, L. (1993), "Economic Underdevelopment: The Case of Missing Market for Human Capital", *Journal of Development Economics*, **40**, 219–239.

LOURY, G. C. (1981), "Intergenerational Transfers and the Distribution of Earnings", *Econometrica*, **49**, 843–867.

MAJUMDAR, M. and MITRA, T. (1982), "Intertemporal Allocation With a Non-Convex Technology: The Aggregative Framework", *Journal of Economic Theory*, **27**, 101–136.

MAJUMDAR, M. and MITRA, T. (1983), "Dynamic Optimization With a Non-Convex Technology: The Case of a Linear Objective Function", *Review of Economic Studies*, **50**, 143–151.

MANI, A. (2001), "Income Distribution and the Demand Constraint", *Journal of Economic Growth*, **6**, 107–133.

MAOZ, Y. D. and MOAV, O. (1999), "Intergenerational Mobility and the Process of Development", *Economic Journal*, **109**, 677–697.

MATSUYAMA, K. (2000), "Endogenous Inequality", *Review of Economic Studies*, **67**, 743–759; "Constraint", *Journal of Economic Growth*, **6**, 107–133.

MATSUYAMA, K. (2002), "The Rise of Mass Consumption Societies", *Journal of Political Economy*, **110**, 1035–1070.

MITRA, T. and RAY, D. (1984), "Dynamic Optimization on a Non-Convex Feasible Set: Some General Results for Non-Smooth Technologies", *Zeitschrift für Nationalökonomie*, **44**, 151–175.

MOAV, O. (2002), "Income Distribution and Macroeconomics: The Persistence of Inequality in a Convex Technology Framework", *Economics Letters*, **75**, 187–192.

MOOKHERJEE, D. and RAY, D. (2000), “Persistent Inequality” (Working Paper, Institute for Economic Development, Boston University).

MOOKHERJEE, D. and RAY, D. (2002*a*), “Contractual Structure and Wealth Accumulation”, *American Economic Review*, **92** (4), 818–849.

MOOKHERJEE, D. and RAY, D. (2002*b*), “Is Equality Stable?”, *American Economic Review*, **92**, 253–259.

MULLIGAN, C. (1997) *Parental Priorities and Economic Inequality* (Chicago, IL: University of Chicago Press).

PIKETTY, T. (1997), “The Dynamics of the Wealth Distribution and the Interest Rate With Credit Rationing”, *Review of Economic Studies*, **64**, 173–189.

RAY, D. (1990), “Income Distribution and Macroeconomic Behavior” (Mimeo, New York University; `http://www.econ.nyu.edu/user/debraj/DevEcon/Notes/incdist.pdf`).

RAY, D. and STREUFERT, P. (1993), “Dynamic Equilibria with Unemployment due to Undernourishment”, *Economic Theory*, **3**, 61–85.

ROCKAFELLAR, R. (1979), “Clarke’s Tangent Cones and the Boundaries of Closed Sets in $\mathbb{R}^n$”, *Nonlinear Analysis*, **3**, 145–154.

SKIBA, A. (1978), “Optimal Growth with a Convex–Concave Production Function”, *Econometrica*, **46**, 527–540.