Persistent Inequality

Dilip Mookherjee
Boston University

Debraj Ray
New York University and Instituto de Análisis Económico (CSIC)

This version: September 2001

Abstract

Existing literature explains persistent inequality either by ongoing shocks to abilities or preferences, or by a combination of technological indivisibilities, capital market imperfections and ad hoc assumptions concerning savings behavior. We focus on the role of pecuniary externalities — driven by endogenous movements in relative prices — in explaining both the emergence and persistence of long-run inequality. With imperfect capital markets, it turns out that long-run inequality is inevitable, even if investments are divisible, agents maximize dynastic utility, and there are no random shocks. However, the divisibility of investment does matter in determining the multiplicity of steady states: with perfect divisibility such multiplicity typically disappears. We subsequently characterize efficient steady states, and study non-steady-state dynamics in a two occupation context.

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. Address correspondence to dilipm@bu.edu and debraj.ray@nyu.edu.

1 Introduction

A central prediction of the neoclassical growth model is that the market mechanism intrinsically 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] and Mulligan [1997]) therefore rely on the presence of random factors ("luck") in explaining the persistence of inequality, despite the overall tendency towards convergence.

It is well known that the convergence proposition relies on (strict) convexity of feasible sets at the individual level, implying diminishing returns to investment.[1] In part, this is a matter of technology. But in part it is a matter of relative prices, which inevitably matter once the neoclassical model is extended to incorporate multiple forms of capital. And such extensions are essential to incorporate issues central to income distribution theory, such as occupational choice and human capital investments. Whether investment continues to be characterized by diminishing returns is then no longer an assumption to be conveniently invoked; it becomes endogenous to the model.

In the presence of multiple forms of capital, questions of convergence which involve non-steady-state dynamics are intrinsically hard to analyze. One may therefore pose the following more limited question: even if all agents (or families) were to start equal initially and share the same utility function, must they all remain equal at every date in the future (absent random shocks)? If not, there would be a natural tendency towards divergence rather than convergence.

Clearly if (in equilibrium) individual feasible sets were to remain convex (and if preferences are convex as well) then *ex-ante* equality must imply *ex-post* equality. In fact, one can go further: if there are perfect capital markets, the observation above must be true even if the returns to education are nonconvex. This is because perfect capital markets effectively separate occupational choices from consumption choices: the former would be based on maximization of net present value, and the latter on intertemporal smoothing of consumption via borrowing and lending. People might select different occupations, but market prices will cause their net returns to be equalized.

Now suppose we introduce capital market imperfections. For expositional ease rule out any form of borrowing and lending. As always, occupational choices will be based on the comparison of costs and returns from different options, with the difference that the costs must be incurred up front. Notice that these costs and returns are endogenously determined in the market, creating a fundamental pecuniary externality: occupational choices of any individual must depend on the choices made by others. What are the implications of this externality? If the marginal returns to educational investment continue to diminish with the magnitude of the investment, *ex-ante* equality must continue to imply *ex-post* equality. Indeed, if this were the case an obvious extension of the convergence argument in the neoclassical model would yield long-run equality even if there is inequality to begin with.

But will educational investments be subject to diminishing returns? If not, a *pecuniary (non)convexity* would arise endogenously. It is important to reiterate that while an occupational choice is, by its very nature, nonconvex (if one cannot mix occupations), this is of no intrinsic interest: what matters instead is whether the relation between *financial* cost and returns of different occupations exhibits a nonconvexity. While this depends on the fineness of occupational structure, it also depends fundamentally on endogenous market prices.

Specifically, suppose the set of occupations is $\mathcal{H}$, with occupation h involving a training cost $x(h)$ and generating wage earnings $w(h)$ later. Define a function $W(x)$ with the property that $W(x(h)) = w(h)$, which represents the relation between cost and returns from different occupational choices available. If the occupational structure is coarse, there may not exist occupations with training cost corresponding to certain intermediate values of x: in those cases we may without loss of generality put $w(x)$ equal to the wage corresponding to the occupation with the highest training cost below x (with the notion that the excess is disposed of). The function $W(x)$ thus represents the relevant "investment frontier" available to any given agent "in equilibrium".

Clearly if there are only finitely many occupations $W(x)$ will be flat over intermediate ranges and jump upwards discontinuously at those values of x that correspond to the training costs of some occupation: hence it must necessarily be nonconvex. But the significance of such nonconvexities would shrink as the occupational structure became richer, providing agents with a large range of investment options. The relevant question is: even as the occupational structure becomes sufficiently fine that for every x there exists an occupation with training cost equal to x, will the $W(x)$ function be characterized by diminishing returns "in equilibrium"? If not, what are the implications of such pecuniary nonconvexities for long-run inequality? This question motivates the first theme of this paper.

We consider a model in which there are several dynasties of individuals, each dynasty composed of an infinite number of generations. In each generation, individuals acquire occupations, and consume a single final good. Capital markets are imperfect: parents cannot borrow against their children's earnings, and must bear the cost of training their children for an occupation. Training may involve material resources as well as teachers and other service-providers, and so may entail the hiring of people of different occupations (at market wages). Therefore training costs — and certainly wages — are, in principle, endogenous, and consequently so is the function $W(x)$.

We prove the following result: *every* steady state of the economy must be characterized by *ex-post* inequality and zero occupational mobility, under very general conditions (that there are at least two active occupations with distinct training costs). Hence no matter whether families start out equal or unequal, if the economy converges to a steady state, they must end up with persistently unequal consumption and utility.

The logic is very straightforward. Altruistic parents deciding to train their children for occupation h involving higher training cost than occupation h' incur a higher sacrifice of current consumption. They must be compensated for this by an assurance that their children will obtain a higher net utility as a result of entering occupation h rather than h'. Hence not only must the incomes earned by members of occupation h be higher than those of occupation h', the same must be true of their consumption levels as well (which equal their incomes, less the cost they incur for training their children in turn).

This result is both simple and robust. It applies irrespective of the nature of the credit market imperfection (all that is needed is that higher investments correspond to higher current sacrifice of consumption), of the nature of intergenerational altruism that motivates educational investment, or of the divisibility of possible levels of investment. Families "trapped" in low income and consumption occupations will not be able to "escape", no matter how much they care about the welfare of their descendants, and no matter how gradual and smooth possible paths of upward mobility may be. The market mechanism thus has an inherent tendency to create persistent inequality, a conclusion diametrically opposite to the predictions of the aggregative neoclassical model. In particular, market prices must necessarily create a pecuniary nonconvexity in the investment frontier facing different households. This result extends and generalizes similar arguments by Freeman [1996], Ljungqvist [1993] and Ray [1990].

Our second principal result concerns the uniqueness of steady states. Many papers in the literature on dynamics of inequality with capital market imperfections have pointed out the possibility of multiple steady states with varying levels of inequality and per capita income (e.g., Banerjee and Newman [1993], Galor and Zeira [1993], Ljungqvist [1993], Ray and Streufert [1993], followed by Bandopadhyay [1997], Freeman [1996], Mani (2001), and Piketty [1997]). This creates the history-dependence of eventual inequality and macroeconomic performance, and a new role for policy intervention (see, for instance, the survey in Hoff and Stiglitz [2001]). A policy need not affect the *set* of steady state outcomes, and in particular it need not be a continuing intervention. A judiciously chosen *temporary* policy can tip initial conditions into the basin of attraction for a new steady state, and therefore have a permanent effect.

We show that such multiplicity must rely on investment indivisibilities. If instead the occupational structure is sufficiently rich (in the sense that corresponding to every training cost x there exists an occupational choice which is selected by some families), the steady state must be unique (under some weak conditions). This uniqueness at the societal level coexists with chronic multiplicity at the individual level: the fortunes of a single dynasty are fundamentally history-dependent in this model. Societal equilibrium may require that there be occupants of various professional slots — winners and losers — in certain proportions, and these proportions be invariant in the aggregate. On the other hand, with a small number of occupations there may be history dependence at both micro and macro levels. Thus occupational discreteness — while irrelevant to the evolution and persistence of inequality — does matter for multiplicity.

A third theme of this paper concerns the *efficiency* of long-run outcomes. A key market — the credit market — is missing, so it stands to reason that steady states of our model will typically be inefficient. Somewhat surprisingly, this is not always the case. We provide a (nearly) complete characterization of those steady states which are efficient. Inefficiency turns out to involve either general underinvestment (the rate of return on education which exceeds the discount rate), or a misallocation of investment (where the rates of return on investing in different occupations are not equalized). As a corollary of this characterization, it turns out that for certain kinds of training technologies (where training involves only material resources, or is recursive in terms of human capital inputs needed) there is always an efficient steady state. Moreover, the unique steady state of the economy with perfectly divisible investment is also efficient.

On the other hand, economies with a small number of occupations typically possess a continuum of efficient steady states *and* a continuum of inefficient steady states. While the steady states are mutually non-comparable under the Pareto-ranking, the inefficient ones are associated with high inequality and the efficient with low inequality (since underinvestment tends to occur when there is high inequality). Here there is a potential for temporary policies to simultaneously reduce long-run inequality and raise per-capita income.

Finally, we address the difficult question of non-steady-state dynamics. This is the only point at which we need to simplify the model: we study an economy with only two occupations. We establish the uniqueness of competitive equilibrium (given initial conditions), and convergence to steady states from every initial condition. The comparative dynamics of redistributive policy are subsequently explored in this context. This model also helps provide an explicit and complete account of how *ex-post* inequality can emerge from a situation of perfect *ex-ante* equality.

We end this introduction by placing our model in the context of existing literature. We have already explained the fundamental differences from a view of persistent inequality as the outcome of ongoing random shocks. Indeed, we abstract from such shocks altogether, and emphasize instead the pecuniary nonconvexities that inevitably arise from the existence of occupational disparity. This is not motivated by considerations of "what's correct", for clearly randomness in ability and incomes is pervasive and an important source of inequality and mobility. Our purpose instead is to reconsider the question whether the market mechanism intrinsically tends to generate or dissolve inequality, a question most fruitfully addressed in a context *sans* uncertainty, as in the original formulation of the neoclassical model by Solow [1956]. Moreover, uncertainty may create the impression of mobility or ergodicity when — for all practical purposes — there is none.[2]

To be sure, competitive versions of the neoclassical model can explain the coexistence of multiple wealth distributions with a unique macroeconomic steady state outcome (see, e.g., Chatterjee [1994]).[3] While this coexistence bears some superficial resemblance to our result concerning steady state uniqueness with perfect divisibility of investments, there is a fundamental difference. The fates of individual dynasties may be highly path-dependent, but our model produces a unique occupational and consumption *distribution* in the aggregate, in contrast to the indeterminacy of consumption and wealth inequality in the competitive neoclassical model. Moreover, steady state inequality is inevitable in our model, whereas the competitive neoclassical model is perfectly consistent with perfect equality.

This last distinction is also true of much of the remaining literature on inequality with capital market imperfections. For instance, long-run inequality is a possibility in the models of Banerjee and Newman [1993], Galor and Zeira [1993], and Ray and Streufert [1993], but so is perfect equality. In addition, this literature assumes the existence of a small number of occupational or investment options (thus imposing technological nonconvexities), and many of the models impose *ad hoc* assumptions on intertemporal behavior.[4] Thus, an additional contribution of the present exercise is to relax these assumptions, in order to explain their role. Our analysis shows that neither assumption is crucial to the result concerning the existence of persistent inequality. For instance, inequality persists despite the attempts of poor parents to help their children escape poverty, no matter how much parents care about the welfare of their descendants. And it persists even if there are no indivisibilities in the investment options. Indeed, we provide stronger results concerning the inevitability of inequality in the long run, owing to the operation of pecuniary externalities. We show that indivisibilities matter instead for the multiplicity or efficiency of steady states: in their absence our results imply that there is no role for history dependence nor any scope for social policy that is not fundamentally redistributive.

In summary then, our model incorporates a richer specification of occupational structure, examines the inevitability of inequality, provides connections between investment divisibility and steady state uniqueness, characterizes steady states, and explores non-steady-state dynamics in special cases. The methods here led to a more comprehensive exploration of inequality on both positive and normative grounds.

In related work (Mookherjee and Ray [2000a]) we have examined analogous questions concerning inequality persistence and history dependence in the context of a contracting model. Similar to this paper, we show there that such phenomena can arise despite savings motivated by long term utility maximization, and a convex investment technology. That paper differs by endogenously modeling the capital market imperfection in terms of an underlying moral hazard problem, and abstracting from interactions across agents. Instead pecuniary nonconvexities in the returns to investment arise endogenously under particular allocations of bargaining power between contracting parties.

The paper is organized as follows. Section 2 introduces the model. Section 3 presents results concerning inequality and immobility characteristic of every steady state. Then

Section 4 discusses steady state multiplicity, Section 5 discusses efficiency, and then Section 6 studies non-steady-state dynamics in a two occupation context. Finally, Section 7 concludes, while the Appendix gathers all 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. Thus the index actually refers to a dynasty, with i at date t serving to label a member of generation t belonging to dynasty i. 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.[5] We take u to be 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 the real line. A *population distribution over professions* is simply a measure λ 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:

$$z \equiv (\lambda, c, \lambda'),$$

where λ 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 λ' 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,[6] that at least one profession requires no training,[7] and that owners of firms (in either producton or training sectors) seek to maximize profits.

2.3 Prices and Behavior

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, which we denote by the function $w(h)$ describing the wage earned by a member of profession h, or more compactly by w. Second, there are the training costs of *acquiring* professional skills for different occupations, denoted by the function $x(h)$ (more compactly by x). The latter represent costs incurred by investing parents, and revenues earned by educational institutions.

Given prices at any date t, the economy generates (input) demands for professions (λ_t), supply (c_t) of the final good and of trained professionals ($\lambda'_t = \lambda_{t+1}$) for the *next* generation at period $t+1$. Give wages w_t and training costs x_t at date t, profit maximization implies that $(\lambda_t, c_t, \lambda_{t+1})$ must solve

$$\max c + x_t \lambda' - w_t \lambda \tag{2}$$

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

Now turn to household responses. Given some sequence of prices $\{w_s, x_s\}_{s \geq t}$, 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

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

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 λ_0 on $\mathcal{H}$. Denote by $\{c_t(i), h_t(i)\}$ the consumption and professional choices made at every date.

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. For this reason most of our results extend to alternative formulations of capital market imperfections, or nature of intergenerational altruism, as explained further below.

2.4 Equilibrium

Given some initial distribution λ, an *equilibrium* is a collection $\{\lambda_t, c_t, w_t, x_t\}$ (with $\lambda_0 = \lambda$) such that:

[1] At each date t, $(\lambda_t, c_t, \lambda_{t+1})$ solves (2), given the price sequence $\{w_t, x_t\}$.

[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 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 (λ, c, w, x) is a steady state if there exists an equilibrium $\{\lambda_t, c_t, w_t, x_t\}$ with $(\lambda_t, c_t, w_t, x_t) = (\lambda, c, w, x)$ for all t.

2.5 Examples

The preceding model is general enough to incorporate several commonly studied models as special cases.

[1] The Neoclassical (Ramsey) Model. The set $\mathcal{H}$ can be reinterpreted as different possible levels of physical (rather than human) capital. The one good Ramsey model is obtained when $\mathcal{H}$ is an interval of the real line and the production set takes the following form: $\mathcal{T} = \{(\lambda, c, \lambda')|c = f(\int_{\mathcal{H}} h d\lambda) - \int_{\mathcal{H}} h d\lambda'\}$, where f is a concave smooth production function generating output which is divided between consumption and capital stock next period. Here $w(h)$ equals hf' and $x(h) = h$. This can be extended to heterogenous capital goods.[8]

[2] Models of Skill Acquisition without Interaction. In the simplest models that display a link between inequality and other features of economic development (e.g., the introductory model in Galor and Zeira [1993]), some exogenous setup cost has to be paid to acquire a skilled profession. It is easy to mimic this setup by assuming a

single final good, two professions: 1 (unskilled) and 2 (skilled), so that $\mathcal{H} = \{1, 2\}$, some constant cost x of acquiring the skill, and constant wage rates $w(1)$ and $w(2)$ to unskilled and skilled labor respectively.[9]

[3] Models of Skill Acquisition with Interaction. More sophisticated models (such as the extended version in Galor and Zeira [1993] as well as Banerjee and Newman [1993], Ray and Streufert [1993], Bandyopadhyay [1997], Maoz and Moav [1999] and others) display interaction across agents. One way of doing this is to suppose that the returns to skilled and unskilled labor in the previous example depend on the aggregate supplies of these two forms of labor.[10]

[4] Entrepreneurship. We need not interpret distinct professions represent different grades of *skill*, but rather as different occupations. As in Banerjee and Newman [1993] or Freeman [1996], we might conceive of one profession as standing for "worker", the other for "entrepreneur". Postulate some fixed investment I that must be made to set up a business: this is the relevant notion of 'training cost' for entrepreneurship. As in example 3, this model creates a natural form of interaction: the returns to workers and entrepreneurs respectively depend on the relative number of workers and entrpreneurs in the economy. With a production function $F(L)$ describing output produced by any given entrepreneur from employing L workers, the workers return is $F'(\frac{\lambda(2)}{\lambda(1)})$ and the entrepreneur's return is $F\left(\frac{\lambda(2)}{\lambda(1)}\right) - F'\left(\frac{\lambda(2)}{\lambda(1)}\right)$, where the number of workers per entrepreneur in the economy is $\frac{\lambda(2)}{\lambda(1)}$.

[5] Endogenous Training Cost. These models can be extended to accommodate endogenous training cost, as in the model of Ljungqvist [1993]. Let $\mathcal{H} = \{1, 2\}$, where 1 stands for unskilled worker and 2 stands for skilled worker. Skilled workers ($\lambda(2)$) are allocated between production ($p(2)$) and training sectors ($r(2)$), so $\lambda(2) = p(2) + r(2)$. There is a constant teacher-pupil ratio of $\alpha \in (0, 1)$, so $r(2) = \alpha\lambda'(2)$. Output of the consumption good depends on unskiled and skilled workers allocated to the production sector, as given by a production function $F(\lambda(1), \lambda(2) - \alpha\lambda'(2))$. Here $w(1) = F_1, w(2) = F_2, x(1) = 0, x(2) = \alpha w(2)$.

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 individuals constantly moving from one profession to another (as in the ergodic distribution of a Markov chain).

PROPOSITION **1 (Zero Mobility in Steady State)** *Let (λ, c, w, 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). In steady state, the present value utility of a generation currently occupying occupying occupation h and contemplating a permanent deviation to occupation g is given by $u\left(w(h) - x(g)\right) + \delta V(g)$ 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 families occupying the richest occupation (which must also entail 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 destiny of each family must be constant over time in any steady state, and $V(h) = \frac{u(w(h)-x(h))}{1-\delta}$ for all h. It leads directly to the conclusion concerning the necessity of inequality. Before this, we need the following definitions.

Let (λ, c, w, 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 would be the case where one of them requires more years of schooling than another — then irrespective of the precise set of prices, occupation h will involve a higher training cost than h'. More generally, with distinct training technologies for two professions, they will turn out to be distinct generically, though we do not pursue the exact conditions required to make this claim precise.

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 very simple. 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. Moreover, it must be the case that the earnings of occupation h net of training cost must also be higher: $w(h) - x(h) > w(h') - x(h')$. 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. This is a very weak requirement. For instance suppose that two professions are ordered in terms of input requirements (of every kind), 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 in the economy) must involve persistent inequality.

Endogenous market prices play an important role in generating and perpetuating this inequality. If several distinct professions are needed for economic activity, the behavior of prices must guarantee that each of those professions are actually chosen. Since parents pay for their children's education, profession requiring a greater training cost entail 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 consumption, not just in incomes.

The examples of Freeman [1996], Ljungqvist [1993] and Ray [1990] drive this point home. In each case, there are two professions (skilled and unskilled labor in Ljungqvist and Ray, managers and workers in Freeman). Consider the Ljungqvist-Ray scenario in which there are two skills, and both types of labor enter as inputs in a concave production function satisfying 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 labor 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 labor 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.

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 verified in Section 6 below) in which both occupations are occupied. By Proposition 2, such a steady state must display (utility and consumption) inequality. This inequality is a fundamental implication of the price mechanism and does not rely on stochastic shocks.[11]

Clearly, the result depends on the assumption that two distinct professions prevail in a steady state. As the discussion above makes clear, this assumption is really one about the potential variation in *relative* prices, which in turn relies on the imperfect substitutability of professions. The assumption may not apply in some situations, however. For instance, the Ramsey model with 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. However, once we allow a minimal degree of diversity of occupations, the Proposition does apply, and the convergence conclusions of the Ramsey model no longer hold.

The result extends to alternative formulations of capital market imperfections or intergenerational altruism. All that is needed is that the marginal cost of finance is higher 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. 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 inequality in every steady state.[12]

4 Multiplicity

4.1 Multiple Steady States and Policy

Typically, there may be several steady-states. A profusion of multiplicity results may be found in the literature (see, e.g., Banerjee and Newman [1993], Lundqvist [1993], Galor and Zeira [1993], Ray and Streufert [1993], Mani [2001] and Piketty [1997]). In these models, the same economic parameters are consistent with numerous steady state outcomes, with varying degrees of inequality, output, unemployment, and productive efficiency. Historical inequality can cause convergence to steady states with lower per capita income, and hence can be viewed as a "cause" of underdevelopment. Indeed, the multiplicity of long-run outcomes may simply reflect the possible multiplicity of initial conditions; given initial conditions there may be a unique equilibrium and a unique long-run outcome. It is also well known (see, for example, Hoff and Stiglitz [2001]) 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 steady states themselves. Thus an exploration of multiplicity is important, in the sense that it tells us what sort of shocks to policy interventions are likely to have lasting impact.

It will be useful to distinguish between two notions of multiplicity. *Individual or micro-* multiplicity refers to the case where initial endowments or perturbations *at the level of a household* significantly shapes the long-run outcome of that household. Contrast this with *societal or macro*-multiplicity, in which initial conditions significantly affect the final destiny of the economy as a whole. The references cited at the start of this section contain numerous instances of macro-multiplicity. These may or may not co-exist with micro-multiplicity. For instance, the Galor and Zeira framework is an example of both. In contrast, in the Piketty model, there is macro-multiplicity, but given a particular societal steady state, individual behavior does *not* depend on initial conditions: there is no micro-multiplicity. Finally, theories such as those in Aghion and Bolton [1997] and Loury [1981] are examples of situations in which there is no multiplicity of either kind: there is a single ergodic distribution, *and* the members of each dynasty experience (over time) all the outcomes in the support of that distribution.

In this section, we demonstrate that the extent of societal multiplicity depends on the richness of the set of professions. In particular, if there is a small number of professions, such multiplicity is endemic. If, on the other hand, there are numerous professions with no "gaps" in their training costs, societal multiplicity disappears entirely. At the same time the long-run outcomes for *individual* dynasties continue to be highly path-dependent, so that micro-multiplicity persists.

Perhaps this is the starkest display of inequality: at the individual level, economic destinies appear as whimsical outcomes which can be changed through one-time interventions. Yet, when the set of professions is rich, long-run outcomes cannot be simultaneously affected for a large group of people, by *any* temporary policy. Societal equilibrium may require that there be occupants of various professional slots — winners and losers — in certain proportions, and these proportions be invariant in the aggregate.

4.2 Exploring Multiplicity

4.2.1 Characterizing Steady States

Throughout this section, we shall suppose that every profession is occupied in steady state.[13] We call this the full-support postulate, and is satisfied if every occupation is essential for producing the consumption good.[14] An example of this is where the consumption good is produced by a Cobb-Douglas production function. 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.

The first necessary condition for a steady state (λ, c, w, x) is that (λ, c) must be related to (w, x) via profit maximization; that is,

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

Secondly, no individual must contemplate a "one-shot deviation" to another profession, where (by the zero-mobility result and the full-occupation postulate) 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',

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

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 characterize the set of steady states.

4.2.2 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 Ljungqvist [1993] or Ray [1990]). For unskilled labor take the training cost to be zero. For skilled labor assume that there is a 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 labor inputs. Abusing notation slightly, 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.[15] This yields the following simple characterization: a fraction λ of skilled people is compatible with a steady state if and only if

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

The left hand side of (10) represents the utility sacrifice of a skilled parent (hereafter denoted by $\kappa^s(\lambda)$) in educating its child, while the right hand side 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

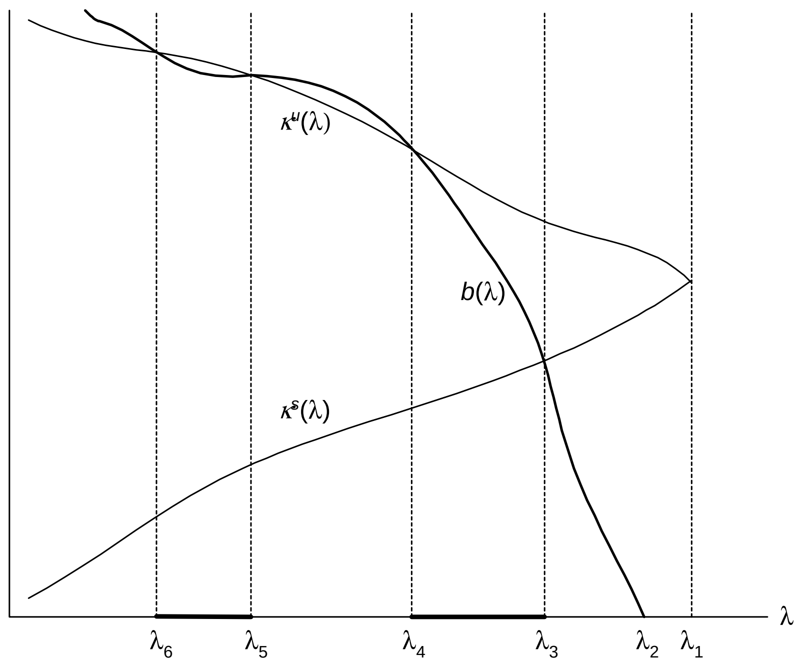

Figure 1: Education Costs and Benefits in Two-Profession Model

Proposition **3** *There is a continuum of steady states in the two-profession model with exogenous training costs, and total output net of training costs unambiguously rises 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. However, as we shall see, the extent of variation across steady states may vanish when the set of professions is 'rich'.

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 (λ_6, λ_5) and (λ_4, λ_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. At the same time, we must be careful not to confuse this finding with the Pareto-efficiency of a given steady state. It is true that there may be steady states "above" it that yield higher per-capita net output (and therefore higher per-capita utility) at every date. But this does not imply that the first steady state is Pareto-dominated. 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', whence there are no longer any indivisibilities in the set of investment options.

4.2.3 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 that there is a continuum of professions: $\mathcal{H} = [0, 1]$. It can be shown (see the working paper version of this paper, Mookherjee-Ray [2000b]) for a demonstration that the case of the continuum can indeed by viewed as the limit of a sequence of economies with progressively finer (but finite) occupational structures. So we simplify exposition by considering directly the continuum case.

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 $g(\mu^h, y^h, h)$, where μ^h is a measure on $[0, 1]$ denoting inputs from different occupations, and y^h the input of the final good, into the training of professionals in profession h. For the final good, the production function may simply be written as $f(\mu)$, describing net output of the consumption good from distribution μ over different inputs in the production sector. Hence $\mathcal{T}$ is generated by the collection of production functions $c + \int_{\mathcal{H}} y^h dh = f(\mu)$ and $\lambda'(h) = g(\mu^h, y^h, h)$, for $h \in \mathcal{H}$, subject to the aggregate resource constraint $\mu + \int_{\mathcal{H}} \mu^h dh \leq \lambda$. **[T.1]** implies the existence of a well-defined unit cost function for training profession h:

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

In a competitive equilibrium, $\psi(w, h)$ will equal the training cost function $x(h)$, given our assumption of constant returns to scale.

The next assumption we employ is

[T.2] *The unit cost function $\psi(w, h)$ 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.[16] The main use of this assumption is to ensure that the training cost function $x(h)$ 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.[17]

PROPOSITION **4** *Suppose that the space of professions is* $[0, 1]$*, that* **[T.1]** *and* **[T.2]** *apply, and that that the full-support postulate holds. Then, provided that some steady state exists with strictly positive wages for all occupations* h*, 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.*

One aspect of this proposition is very intuitive, so let us dispose of it first. Suppose, for the moment, that 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). Then there can only be one steady state wage function satisfying the full support property.

To see this, 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 be (almost) neglected, and all that matters 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 finding 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 with two occupations, interior steady states are characterized by incentive constraints that do not bind, which 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.[18]

So far, we assumed that x is exogenously given, and showed that there is a single w-function, *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.[19]

To see why the endogeneity of x does not jeopardize uniqueness, it may be best to look at a couple of 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

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

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

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

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.

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 cost minimization exercise generates the training cost function

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

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.

In both the examples, we have used what one might call a "recursive technology", in which the training of level-h individuals depend on indices labeled 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.

To complete the discussion of Proposition 4, notice that once the steady state wage function is pinned down, so is the unit cost function of acquiring a profession. Strict quasiconcavity of all production functions then implies that input demands in all sectors of the economy are uniquely determined, which determines the distribution across occupations, and hence the entire steady state.[20]

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 λ_0 over occupations, say that a allocation $\{c_t(i), \lambda_t\}$ is *efficient* if, first, it is feasible:

$$(\lambda_t, c_t, \lambda_{t+1}) \in \mathcal{T}$$

for all dates t, where $c_t \equiv \int_{[0,1]} c_t(i)di$, and if there is no other feasible allocation $\{c'_t(i), \lambda'_t\}$ (with $\lambda'_0 = \lambda_0$) such that for every date t:

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

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

Note that this notion of efficiency is actually a form of constrained Pareto efficiency, where the 'planner' is constrained from making intertemporal transfers in exactly the same way that market agents are.

Proposition **5** *Suppose that a steady state* (λ, x, 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 (assuming that $x(h) > x(h')$ and accordingly that $w(h) > w(h')$). 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 future generations will lose if this apparent overinvestment is eliminated. Note however that the condition also requires a balance between the extent of "overinvestment" in different occupations: 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 as well for steady states satisfying the full support property. The converse is not exact. We will assume that the number of professions is finite,[21] and that the technology set satisfies:

[T.3] $\mathcal{T}$ has a smooth boundary, in the sense that every weakly efficient[22] point of $\mathcal{T}$ has a unique supporting price vector of the form $(w, 1, x)$.

PROPOSITION **6** *Assume that* $\mathcal{H}$ *is finite and that* **[T.3]** *holds. Suppose that* (*18*) *fails at some steady state with all professions occupied. Then the 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*, whereby 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 somewhere 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 net output of 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 and $t+1$ 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 rate δ at which they weigh the utility of the next generation. 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.

While underinvestment or misallocation is therefore not consistent with Pareto efficiency, overinvestment (in the sense of a common rate of return on all educational investments which falls below the discount rate) is, as established by Proposition 5. The reason for this asymmetry is that increases in aggregate consumption in the case of such 'overinvestment' require a reduced scale of investment, which makes *future* generations worse off, and there is no way that current generations can compensate future generations for this change.[23] Such 'overinvestment' reflects the constraints on intertemporal transfers in this economy by means other than (nonnegative) human capital investments.

To apply the preceding characterization of efficient steady states, consider first the continuum case discussed in Section 4.2.3 satisfies the conditions of Proposition 5; therefore the unique steady state in that case is Pareto-efficient. Indeed, it is Pareto-efficient

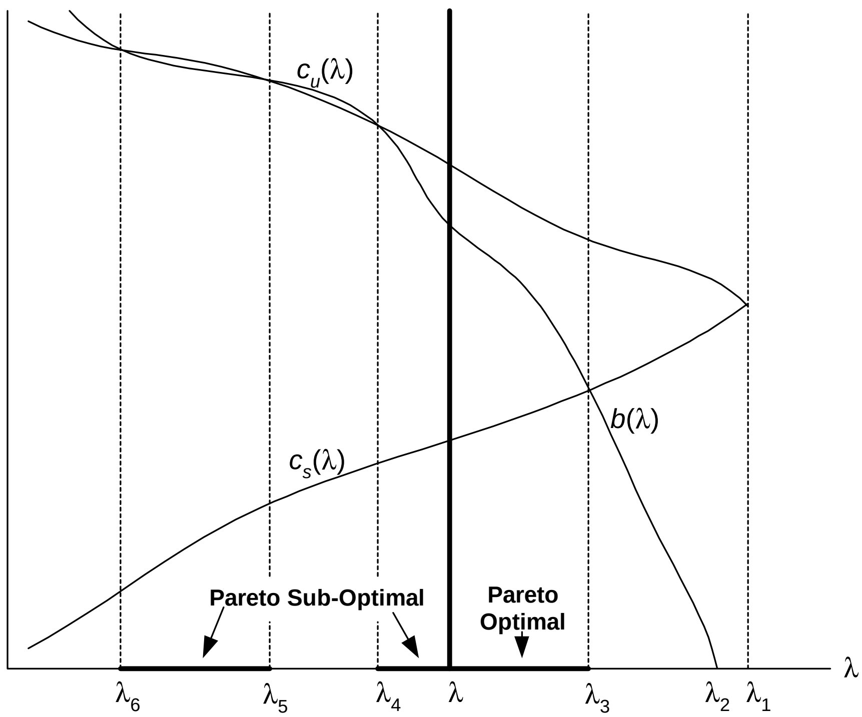

Figure 2: The Pareto-Efficiency Threshold with Two Professions.

in a first-best (i.e., unconstrained) sense as well, since the rate of return on investment is uniformly equal to the discount rate (i.e., it is not characterized by any overinvestment).

Next consider the two profession economy. It is easy to apply Propositions 5 and 6 to show that in the two-profession case, "high" inequality coexists with inefficiency. The reason is intuitive: high inequality is consistent with underinvestment in education given capital market imperfections. 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:

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

Define λ^* by the condition $w^s(\lambda) - w^u(\lambda) = x/\delta$. Notice that by Propositions 5 and 6, and 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 set of steady states. This is done by verifying that (19) is satisfied with *strict* inequality when $\lambda = \lambda^*$.[24]

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,[25] $w(0)$ 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.[26] 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 our uniqueness results of the preceding Section apply. But there may be scope for Pareto-improving policy in other contexts, e.g., where the occupational structure exhibits indivisibilities. Nevertheless, even in such case — and despite the missing credit market — an efficient steady state will exist in a large class of economies.

6 Dynamics

The discussion so far on the emergence of inequality (as opposed to its persistence) makes an important assumption. It is that starting from any initial configuration, an economy will indeed converge to a steady state. After all, Proposition 2 makes no claims regarding persistent inequality when the economy fails to converge to a steady state.

Moreover, a satisfactory theory of the long-run role of historical inequality should account for the dynamic process by which initial conditions determine long-run outcomes. For instance even if there are many possible steady states, it is conceivable that only a few of them are stable attractors, and others cannot be reached from a nontrivial set of initial conditions. In that case the steady state analysis overstates the multiplicity of long-run outcomes. And even if convergence can be established, the precise map between initial conditions and eventual steady state reached, and the transitory process is of interest in its own right (e.g., Does inequality tend to increase or decrease over time? How fast is the convergence? What are the transitory and long term effects of one-shot redistributions?)

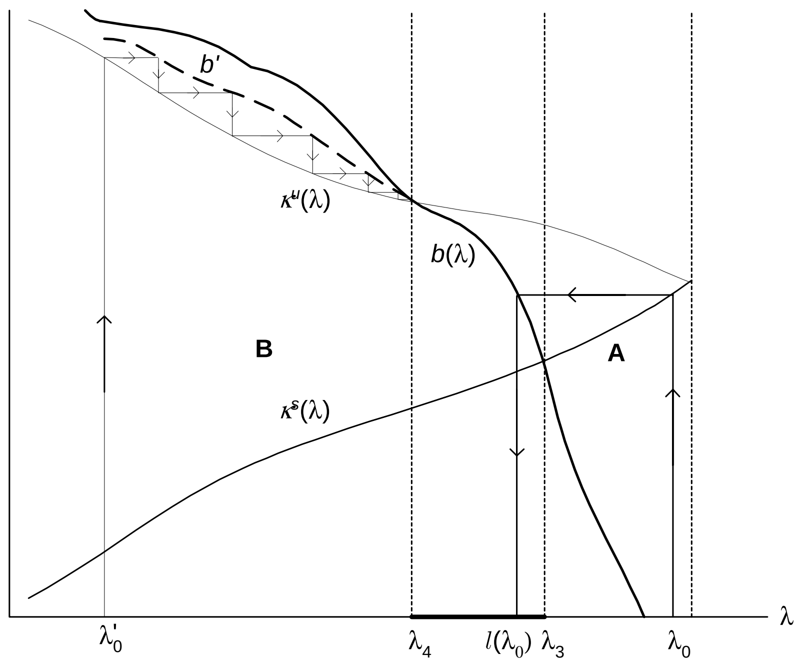

Figure 3: Dynamics in Two-Profession Model

To our knowledge, there is no general theorem that guarantees convergence in this class of models.[27] Competitive versions of the turnpike theorem are available (see Bewley [1982], Coles [1985] or Yano [1984]) but do not apply here, as those arguments rely on the equivalence between competitive equilibria and full Pareto-optimality. Such equivalence does not obtain in our setting because the credit market is missing.

The purpose of this section is to report on a special case for which we have been able to establish convergence, and characterize the dynamics completely.

We focus on the two-skill model with exogenous training cost, described in Section 4.2.2 above. To explain the nature of the dynamics, it will be necessary to consider two possible zones in which λ might lie, when λ is *not* a steady state. We divide the non-steady state space into two complementary parts: in the first subset (denoted by A), the steady state condition fails owing to insufficient incentive of skilled households to educate their children (the first inequality in (10) fails). Recall from Section 4.2.2 that λ_3 is the highest steady state value of λ, where skilled households are just indifferent between educating their children and not. Then A is the range of skill ratios that exceed λ_3 (see Figure 3).

In the second subset (denoted by B), the steady state condition fails because unskilled families strictly prefer not to educate their children. Equivalently, the second inequality in (10) fails. B is the union of all skill ratios lower than λ_3 that do not constitute steady states. In Figure 3 in which the set of steady states is simply the interval (λ_4, λ_3), the set B is the set of skill ratios $(0, \lambda_4)$. In general, it is clear that A and B are disjoint owing to the strict concavity of u. In what follows, we relate the dynamics of λ to membership of the initial skill ratio in one of the sets A and B.

PROPOSITION **7** *If $\lambda_0 \in A$, then there exists a unique competitive equilibrium from λ which goes to the steady state in one period: $\lambda = \lambda_0 > \lambda_1 = \lambda_t$ for all $t \geq 1$.*

If $\lambda_0 \in B$, then there exists a unique competitive equilibrium in which the proportion of skilled people increases strictly in every period, and converges to some steady state: $\lambda_t < \lambda_{t+1}$ for all $t \geq 0$.

If λ is a steady state, there is a unique competitive equilibrium from $\lambda_0 = \lambda$, given by $\lambda_t = \lambda$ for all t.

Hence from any initial condition, there is a unique competitive equilibrium which converges to a steady state. If the initial skill ratio is a steady state, the equilibrium involves that ratio for ever thereafter. If it is a high ratio (in the set A) then the skill ratio falls in just one generation to a steady state, and stays there forever after. This is depicted in Figure 3 for the initial skill ratio depicted λ_0. Since this ratio is very high, the skill premium is too low to motivate educational investments that are consistent with a steady state. Accordingly, at such a date, some skilled households will not educate their children, and every unskilled household will behave likewise. This lowers the skill ratio for the succeeding generation. The eventual steady state $l(\lambda_0)$ is pinned down by the requirement that generation 0 skilled households are just indifferent between educating and not educating their children. Since from generation 1 onwards the economy will be in a steady state, the present value benefit of educating children for generation 0 households is given by the steady state benefit function $b(\lambda)$, which must equal the sacrifice for generation 0 skilled households:

$$\kappa^s(\lambda_0) = b(l(\lambda_0)) \tag{20}$$

which determines the function $l(\lambda_0)$. The skill ratio $l(\lambda_0)$ must be a steady state because it is smaller than λ_0, so the sacrifice $\kappa^s(l(\lambda_0))$ for skilled households must be smaller than the benefit $b(l(\lambda_0))$ — using equation (20), while the sacrifice for unskilled families must be larger than $b(l(\lambda_0))$.

In contrast, if the initial skill ratio is too low to constitute a steady state (i.e., is in the set B), then convergence to the eventual steady state will occur gradually rather than in one step. In Figure 3 this is represented by the initial skill ratio λ_0'. Then the skill ratio will subsequently increase over time, with unskilled households progressively switching to the high skill status. The dynamics is determined by the condition that they

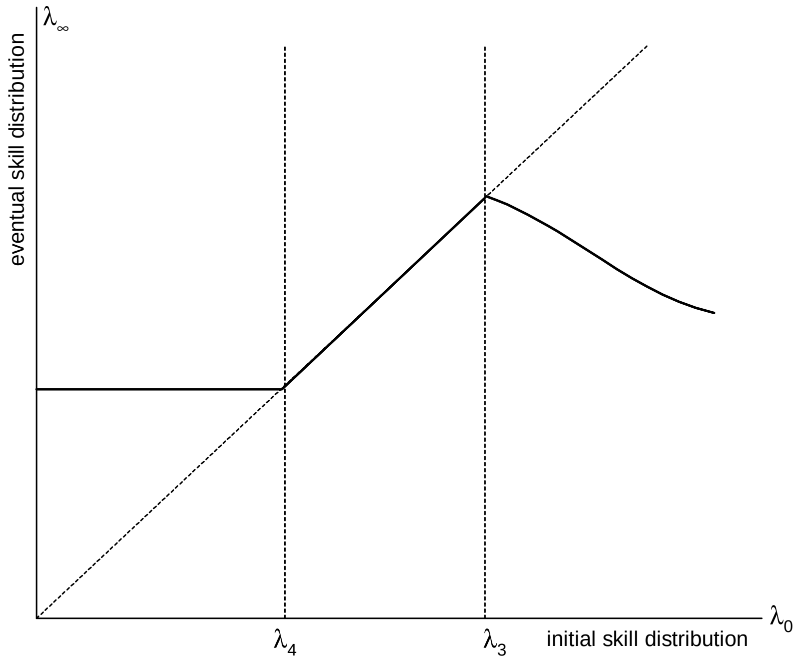

Figure 4: Map From Initial to Long-Run Skill Distribution

be indifferent between switching and not at every date. As more and more households become skilled in this fashion, the skill premium declines over time, reducing the benefit from switching. At the same time the cost of switching for unskilled households also falls, as the unskilled wage rises over time. Since the convergence does not occur in one step, the present value benefit of switching is not represented by the steady state benefit function $b(\lambda)$, but by the function depicted by b' in Figure 3 which is lower (reflecting the fact that the benefit of switching is falling over time). The dynamics is then pinned down by the equality of sacrifice for unskilled families and the present value benefit b', as depicted in Figure 3. It is evident from this that the eventual steady state will be the *smallest* steady state skill ratio lying above the initial skill ratio, in the case where the latter is in the low range B. In Figure 3 this is the skill ratio λ_4.

The corresponding map from initial to eventual skill ratios is depicted in Figure 4, and from the initial skill premium (a measure of inequality) to eventual long-run per capita income in Figure 5.[28] These help predict the effect of initial conditions of the economy to its long-run performance. These maps are nonmonotone, thus showing that orderings of countries by human capital, inequality and per capita income can get reversed over time. For instance countries with initial skill ratio higher than λ_3 will eventually end up with a lower skill ratio than countries starting with the skill ratio λ_3. Correspondingly

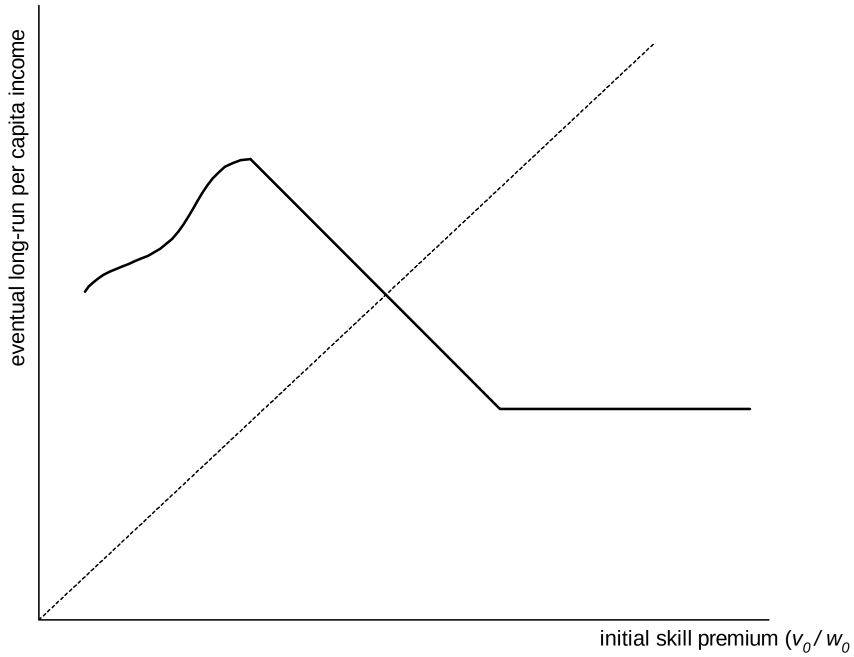

Figure 5: Map from Initial Skill Premium to Long-Run Per-Capita Income

countries that start with a high degree of equality (a low skill premium) end up more unequal and with a lower per capita income. Intuitively, if the economy starts with an excessively high proportion of skilled persons, the skill premium is low, reducing the earnings of first generation skilled families. In turn this raises the sacrifice these families must make to educate their children (since the education cost is fixed by assumption). In order to compensate for this larger sacrifice, the benefit their children receive from being skilled must rise. This requires that the eventual steady state must involve a higher skill premium, i.e., a smaller fraction of the economy must be skilled.

Such reversals can only occur at the high end of the spectrum of initial skill ratios (i.e., when starting in the set A). When initial skill ratios lie below (either in the steady state set or in B), initial conditions and eventual outcomes are ordered in the same way.

Particularly interesting is the case where the economy starts in B, i.e., with sufficient inequality. Then inequality falls over time, accompanied by a process of progressive increase in education and skill within the population, which serves to raise per capita income over time. However, the initial conditions do not (locally) affect the long-run outcome, which is invariably the nearest steady state (λ_4 in these figures). Hence one shot redistributions in this case only have a transitory impact, that speed up the skill upgrading process. In contrast when the economy starts in a steady state (e.g., in the interior of (λ_4, λ_3)), one shot redistributions do have an immediate and permanent impact. At the other extreme, when the economy starts with a very high skill ratio (in the set A), one shot redistributions have an immediate and permanent effect, but *which perversely causes a move in the opposite direction.*

7 Summary and Research Directions

We explored three themes in this paper. First, in contrast to a literature which views economic inequality as the outcome of ongoing stochastic shocks, we argued that there are fundamental reasons for the market to generate inequality, even in a world of perfect certainty and ex-ante identical agents. All that is required is that educational loan markets are imperfect. In particular, long run inequality is inevitable with multiple occupations, irrespective of the degree of foresight or intergenerational altruism of parents, or the divisibility of investment options.

Second, we show that while the fate of *individual* dynasties may be plagued by extreme and sensitive forms of 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 some conditions, there is a unique steady state, so that a one-time policy, while affecting some individual dynasties in a particular way, will have opposite and compensating effects on other dynasties.

Finally, we characterized efficient steady states in our model. Because the credit market is missing, it is of interest that some steady states may be efficient all the same. 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. Hence there is potential scope for temporary policies or historical shocks to raise long run per capita income while reducing inequality. Detailed dynamics of policy effects were subsequently explored in the two profession context.

We conclude by describing ideas for future research. First, we know little about non-steady-state dynamics outside the two profession case. Yet what happens outside steady state is important for understanding how inequality evolves over time. For example, the unique steady state in the case of perfectly divisible investments involves a linear (hence convex) investment frontier. So the nonconvexities that create inequality must then be entirely the effect of pecuniary externalities that appear endogenously *out of steady state*, and we do not yet fully understand how this happens. Moreover, the relevance of steady state analysis depends on whether the economy converges reasonably quickly to steady states from arbitrary initial conditions.

The second major research question concerns extension to the case where financial bequests can supplement investments parents undertake in their children's education. Then 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 availability of financial bequests may also modify non-steady-state dynamics significantly. Whether and how inequality evolves and persists in such contexts is an important and challenging open question.

Finally, we believe that the research program outlined in this paper can find several interesting applications to the interaction between trade and inequality. For instance, different levels of inequality across countries may constitute a source of comparative advantage in products involving differential forms or levels of human capital. Moreover, the model developed here may be used to predict the reciprocal effect of trade on inequality. Alternatively, one may reinterpret the results in this paper for a global economy with interaction. Suppose that the individuals of our model are countries, or more precisely the planning agencies of these countries. View 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 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 is sufficiently 'divisible' or 'rich'. Then our uniqueness results suggest that there will be high-tech exporters, but not too many of them. And not all developing countries need be primary commodity exporters, but there cannot be too few of them either. It may be hard to talk about economic policies that imitate a Korea or a Hong Kong in the world economy. Sequenced development that maintains global hierarchical compositions may be the rule rather than the exception.

References

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

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

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

Banerjee, A. and A. Newman (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 N. Tomes (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy* **87**, 1153-1189.

Bewley, T. (1982), "An Integration of Equilibrium Theory and Turnpike Theory," *Journal of Mathematical Economics* **10**, 233–267.

Champernowne, D. (1953), "A Model of Income Distribution," *Economic Journal* **63**, 318–351.

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.

Coles, J. (1985), "Equilibrium Turnpike Theory with Constant Returns to Scale and Possibly Heterogeneous Discount Factors," *International Economic Review* **26**, 671–679.

Dechert, R. and K. Nishimura (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 J. Zeira (1993), "Income Distribution and Macroeconomics,"

Hoff, K. and J. Stiglitz (2001), "Modern Economic Theory and Development," in G. Meier and J. Stiglitz (eds), *Frontiers of Development Economics*, World Bank and Oxford University Press, 389–459. *Review of Economic Studies* **60** (1), 35–52.

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 T. Mitra (1982), "Intertemporal Allocation with a Non-Convex Technology: The Aggregative Framework," *Journal of Economic Theory* **27**, 101–136.

Majumdar, M. and T. Mitra (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.

Matsuyama, K. (1999), "The Rise of Mass Consumption Societies," mimeo.

Maoz, Y.D. and O. Moav (1999), "Intergenerational Mobility and the Process of Development," *Economic Journal*, 109 (October), 677-697.

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

Mookherjee, D. and D. Ray (2000a), 'Contract Structure and Wealth Accumulation," Working Paper, Institute for Economic Development, Boston University.

——————— (2000b) 'Persistent Inequality," Working Paper, Institute for Economic Development, Boston University.

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.

Ray, D. and P. Streufert (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.

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

Yano, M. (1984), "The Turnpike of Dynamic General Equilibrium Paths and its Insensitivity to Initial Conditions," *Journal of Mathematical Economics* **13**, 235–254.

Notes

1. Chatterjee [1994] considers a version of the neoclassical model where individual feasible sets are defined by market prices and are therefore not strictly convex, while the aggregate production function for the economy as a whole is strictly convex. In this context inequality across individuals can persist, and at the same time the aggregate capital stock is uniquely determined in steady state. We discuss the relation of our approach to this version the neoclassical model below.

2. Notice that this feature, in a sense, is almost opposite in spirit to the first — inequality *creating* — aspect of uncertainty. For instance, Galor and Zeira [1993] show that when there are significant investment thresholds, then final outcomes for an individual may depend on initial conditions. This sort of path dependence is easy enough to knock out — at least formally — by introducing some uncertainty, however small, as long as the support of the uncertainty is quite large (for instance, if there is always some probability that any given person will win the State Lottery). The peculiar laws of stochastic processes running in infinite time then dictate that there is full ergodicity of the Galor-Zeira process. But this conveys a misleading impression that the Galor-Zeira model invariably tends to promote long-run equality. One sensible way of dealing with this problem is to use expected utility evaluations of all future streams, which would deal with low-probability events by giving them insignificant utility weight. This is exactly the approach we shall follow in evaluating welfares of households; consequently the results will not depend in any intrinsic manner on the existence of small amounts of randomness.

3. This arises from the linearity of the investment technology at the individual level, which coexists with strict concavity in the aggregate.

4. There is also an earlier literature on economic growth with nonconvex technologies which yields similar results (see, e.g., Clark [1971], Skiba [1978], Majumdar and Mitra [1982, 1983], Dechert and Nishimura [1983], and Mitra and Ray [1984]).

5. Much of the analysis extends to the many-consumption-good case. For details see the working paper version of this paper (Mookherjee-Ray [2000b]).

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

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

8. Then $\mathcal{H}$ is a set of vectors of capital stocks of various kinds, and the production function describes combinations of capital stocks for the next period and consumption good (λ', c) producible from current capital stocks λ. Using the consumption good as numeraire, $w(h)$ will equal the marginal product of capital goods in terms of the consumption good, and $x(h)$ will measure the marginal opportunity cost of producing the capital stock vector in terms of units of consumption god foregone.

9. Formally, derive the technology set from the following specifications. First, gross output $y = w(1)\mu(1) + w(2)\mu(2)$, where $\mu(h)$ is the input of labor of skill h. Next, a potential supply of skilled labor is created using y' units of the final good as input: $\sigma(2) = (1/x)y'$, while the potential supply of unskilled labor $\sigma(1)$ can be set to any nonnegative value (compactify this by setting some irrelevant upper bound $m > 1$). This creates the technology set $\mathcal{T} = \{(\mu, c, \sigma) \geq 0 | \; g = w(1)\mu(1) + w(2)\mu(2) - x\sigma(2)$ and $0 \leq \sigma(1) \leq m\}$.

10. Formally, we can modify the final output production function from a linear specification to any constant returns concave specification: $g = f(\mu(1), \mu(2))$. Then the returns $w(h)$ are obtained as the value of the partial derivatives of this function.

11. A model with "warm glow" bequests (see, e.g., Galor and Zeira [1993] or Maoz and Moav [1999]) 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.

12. For details, see the working paper version of this paper, Mookherjee and Ray [2000b].

13. 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.

14. There are two possible reasons why the full-support assumption might fail. First, certain professions may be inessential, either because the inputs they supply can be supplied more efficiently by some other profession, or because the input has a small enough marginal product at zero supply. For instance, suppose there are only two inputs, and a large number of possible professions, each of which supplies one of the two inputs. Then any profession which is not cost-effective in delivering its input relative to some other profession will be unoccupied. In this case there are effectively only two professions — those which deliver the respective inputs

cost-effectively — that will be occupied. 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 — and notwithstanding the fact that an unoccupied necessary occupation is infinitely lucrative — unoccupied professions may be prohibitively costly to acquire. We do not take this argument very seriously, as such 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.

15. 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 wage differentials that are incompatible with incentives for parents to educate their children.

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

17. 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 postulate 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.

18. 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] and Matsuyama [1999]) drive the story.

19. This is where we invoke the premise that a steady state exists with positive wages throughout. Given that $x(0) = 0$, it is possible that no steady state has $w(0) = 0$. Whether the proposition holds in that case remains an open question.

20. It is worthwhile to reiterate an important contrast with the competitive version of the one good Ramsey model which combines macro-determinacy with micro-

multiplicity (see, e.g., Chatterjee [1994]). There the aggregate capital stock is determined uniquely in steady state, while its distribution across different households is not. In particular, inequality is not tied down at all in that model. Proposition 4 in contrast provides conditions for the steady state distribution to be uniquely determined.

21. 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.

22. 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.

23. This suggests that 'overinvestment' will also be inconsistent with efficiency in the presence of alternative instruments by which current generations can leave bequests for their descendants, e.g., via financial assets. As explained in Section 7, consideration of such forms of bequests is beyond the scope of this paper.

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

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

26. Letting λ_h denote the number of people in occupation h, a steady state with positive consumption c requires 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 $(1 = a_0 + \sum_h w(h)a_h > a_0 + \sum_h x(h)a_h)$.

27. Even in the simplistic warm-glow formulations of intergenerational behavior, general convergence arguments are hard to come by (see, e.g., Banerjee and Newman [1993] for a discussion).

28. These figures correspond to the case depicted in Figure 3, where the set of steady states constitutes a single interval.

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 which enjoys positive measure under λ.

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 postive measure under λ.

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 function of starting at h under the going steady state. Then, because g' is feasible for h (after all, $x(g') < x(g)$),

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

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

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

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

$$u\left(w(h') - x(g')\right) - u\left(w(h) - x(g')\right) \geq u\left(w(h') - x(g)\right) - u\left(w(h) - x(g)\right). \tag{21}$$

However, given that $w(h) < w(h')$ and $x(g') < x(g)$, (21) 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

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

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 net output:

$$\max_{\lambda \geq 0} f(\lambda, 1-\lambda) - x\lambda. \tag{22}$$

This is a strictly concave problem in λ and attains a unique maximum when $f_1 - f_2 = x$. Recalling that $v = f_1$ while $w = f_2$, we conclude that this is the point λ such that $u(\lambda) - x = w(\lambda)$, which is precisely λ_2 in Figure 1. Because every steady state lies to the left of λ_2 and the maximization problem (22) is strictly concave, the result follows. ■

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

LEMMA **1** *The unit cost function $\psi(w, h)$ has the following properties:*

[1] *If two wage functions w and $\hat{w}$ satisfy $\hat{w}(h) \geq w(h)$ for every h, then $\psi(\hat{w}, h) \geq \psi(w, h)$ for every h.*

[2] *For every scalar $\alpha \geq 1$ and each h, $\psi(\alpha w, h) \leq \alpha\psi(w, h)$.*

[3] *For every scalar $\alpha \in [0, 1]$ and each h, $\psi(\alpha w, h) \geq \alpha\psi(w, h)$.*

[4] *In any steady state (λ, w, x, c), $x(h) = \psi(w, h)$ 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 postulate, there cannot be two steady states, with distinct wage functions $\hat{w}$ and 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 $\hat{\lambda}$ that produces the final good at level $\hat{c}$. By profit maximization and constant returns to scale in the production sector,

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

so that by the full-support postulate,

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

But (given constant returns to scale) this violates profit maximization at the steady state with wage function w. ■

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

If 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 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 postulate implies that every x in $[0, X]$ is chosen by some families.

This implies that W must be continuous. Otherwise some level of x in the neighborhood of a discontinuity will not be chosen, as it will be dominated by a neighboring 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,

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

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

$$u'\left(W(x)-[x+\epsilon]\right) \geq \frac{\delta}{1-\delta}u'\left(\theta(\epsilon)\right)\left[\frac{W(x+\epsilon)-W(x)}{\epsilon}-1\right] \geq u'\left(W(x+\epsilon)-x\right), \tag{23}$$

where $\theta(\epsilon)$ lies between $W(x) - x$ and $W(x + \epsilon) - (x + \epsilon)$. Now we may send ϵ to zero in (23) 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 $x = X$ (resp $x = 0$) to show the left-differentiability (resp. 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 w with strictly positive wages throughout. Then, by the claim,

$$w(h) = \frac{1}{\delta}x(h) + w(0). \tag{24}$$

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

$$\tilde{w}(h) = \frac{1}{\delta}\tilde{x}(h) + \tilde{w}(0). \tag{25}$$

for every h. Combining (24) and (25), we see that

$$\frac{\tilde{x}(h)}{x(h)} = \frac{\tilde{w}(h) - \tilde{w}(0)}{w(h) - w(0)} \tag{26}$$

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 w and $\tilde{w}$ 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

$$\frac{\tilde{x}(h^*)}{x(h^*)} = \frac{\tilde{w}(h^*) - \tilde{w}(0)}{w(h^*) - w(0)} > \alpha. \tag{27}$$

Define a new wage function 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],

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

while by Lemma 1, part [1],

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

Combining these two inequalities, we may conclude that

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

which contradicts (27).

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 (27), we see that

$$\frac{\tilde{x}(h_*)}{x(h_*)} = \frac{\tilde{w}(h_*) - \tilde{w}(0)}{w(h_*) - w(0)} < \beta. \tag{28}$$

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

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

while by Lemma 1, part [1],

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

Combining these two inequalities, we see that

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

which contradicts (28).

Thus, in both cases we have a contradiction, so that the first part of the proposition is established. The second part is obvious so does not need any proof. ■

Proof of Proposition 5.

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

$$\sum_{s=t}^{\infty} \delta^{s-t} u(c_s) \geq \sum_{s=t}^{\infty} \delta^{s-t} u(c), \tag{29}$$

we must have

$$\sum_{s=t}^{\infty} \delta^{s-t} c_s > \sum_{s=t}^{\infty} \delta^{s-t} c. \tag{30}$$

Proof. Suppose that there is a sequence of consumptions $\{c_s\}_{s=t}^{\infty}$, distinct from c at some $s \geq t$, such that (29) holds. By an elementary inequality involving strictly concave functions, we know that

$$u'(c)[c_s - c] \geq u(c_s) - u(c), \tag{31}$$

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

Combining (29) and (31), 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, on the contrary, that there is some Pareto-improving allocation $\{c_t(i), \lambda_t\}$ with

$$(c_t, \lambda_t, \lambda_{t+1}) \in \mathcal{T}$$

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 (30) at any date t and adding up over all agents, we see that

$$\sum_{s=t}^{\infty} \delta^{s-t} c_s \geq \sum_{s=t}^{\infty} \delta^{s-t} c, \tag{32}$$

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 (w, x), profits are maximized at the steady state allocation. Consequently, for each date s,

$$c + x\lambda - w\lambda \geq c_s + x\lambda_{s+1} - w\lambda_s \tag{33}$$

Taking discounted sums and invoking (32) from Lemma 3, we see that

$$\frac{x\lambda - w\lambda}{1 - \delta} \geq \sum_{s=t}^{\infty} \delta^{s-t} [x\lambda_{s+1} - w\lambda_s].$$

where strict inequality must hold at date 0. Using (18), it can be seen that

$$\frac{a-1}{1-\delta}w\lambda \geq \sum_{s=t}^{\infty} \delta^{s-t}[aw\lambda_{s+1} - w\lambda_s]. \tag{34}$$

[Recall once again that strict inequality must hold at date $t = 0$.]

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

$$a^{t-1}(a-\delta)\frac{a-1}{1-\delta}w\lambda \geq (a-\delta)\sum_{s=t}^{\infty} \delta^{s-t}[a^t w\lambda_{s+1} - a^{t-1}w\lambda_s]. \tag{35}$$

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

$$\begin{aligned}
-\frac{a-\delta}{1-\delta}w\lambda &\geq (a-\delta)\sum_{t=1}^{\infty}\sum_{s=t}^{\infty} \delta^{s-t}[a^t w\lambda_{s+1} - a^{t-1}w\lambda_s] \\
&= (a-\delta)\sum_{s=1}^{\infty}\sum_{t=1}^{s} \delta^{s-t}[a^t w\lambda_{s+1} - a^{t-1}w\lambda_s] \\
&= (a-\delta)\sum_{s=1}^{\infty}\left[\sum_{t=1}^{s} \delta^s \left(\frac{a}{\delta}\right)^t w\lambda_{s+1} - \sum_{t=1}^{s} \frac{\delta^s}{a}\left(\frac{a}{\delta}\right)^t w\lambda_s\right] \\
&= \sum_{s=1}^{\infty}[a(a^s-\delta^s)w\lambda_{s+1} - (a^s-\delta^s)w\lambda_s] \qquad (36)
\end{aligned}$$

Now add both sides of (36) to the corresponding sides of the inequality (34) for $t = 0$. Remembering that this latter inequality is strict, we see that

$$-w\lambda > \sum_{s=1}^{\infty}[a(a^s-\delta^s)w\lambda_{s+1} - (a^s-\delta^s)w\lambda_s] + \sum_{s=0}^{\infty}\delta^s[aw\lambda_{s+1} - w\lambda_s].$$

But careful inspection of the right-hand side of this inequality shows that it is also equal to $-w\lambda$, which is a contradiction. This completes the proof. ■

Proof of Proposition 6. The following standard lemma will be used.

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

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 (λ, c, w, 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

$$w(h_1) - w(h_2) > \frac{1}{\delta}[x(h_1) - x(h_2)]. \tag{37}$$

[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

$$\frac{w(h_2) - w(h_1)}{x(h_2) - x(h_1)} > \frac{w(h_4) - w(h_3)}{x(h_4) - x(h_3)}. \tag{38}$$

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

$$\frac{\nu w(h_1) - w(h_2)}{x(h_1) - x(h_2)} > \eta > \frac{1}{\delta}. \tag{39}$$

Fix these two numbers in what follows. For any $\epsilon > 0$ and small, define the distribution λ_ϵ by

$$\begin{aligned} \lambda_\epsilon(h_1) &\equiv \lambda(h_1) + \frac{\epsilon\nu}{x(h_1) - x(h_2)}, \text{ and} \\ \lambda_\epsilon(h_2) &\equiv \lambda(h_2) - \frac{\epsilon\nu}{x(h_1) - x(h_2)}, \end{aligned} \tag{40}$$

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

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

$$(\lambda, c - \epsilon, \lambda_\epsilon) \in \mathcal{T} \tag{41}$$

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

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

$$pz_\epsilon = (c - \epsilon) + x\lambda_\epsilon - w\lambda = (\nu - 1)\epsilon < 0,$$

where use has been made of (40) and the fact that $c + x\lambda - w\lambda = 0$. By Lemma 4, we know that for all $\alpha \in (0, 1)$ and sufficiently small, $(1 - \alpha)z + \alpha z_\epsilon \in \mathcal{T}$, where $z \equiv (\lambda, c, \lambda)$. Using (40), this is easily seen to be equivalent to (41) (for ϵ small enough), and the claim is established.

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

$$(\lambda_\epsilon, c + \eta\epsilon, \lambda) \in \mathcal{T} \tag{42}$$

for all $\epsilon \in (0, \epsilon_2)$, where η is defined in (39). To see this, calculate the "profit" generated by the allocation $z'_\epsilon \equiv (\lambda_\epsilon, c + \eta\epsilon, \lambda)$:

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

where the inequality in this string uses (39). So once again, by Lemma 4, we may conclude that for all $\alpha \in (0,1)$ and sufficiently small, $(1-\alpha)z + \alpha z'_\epsilon \in \mathcal{T}$. Using (40), this is easily seen to be equivalent to (42) (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: $(z_\epsilon, z'_\epsilon, z, z, \ldots)$, where $0 < \epsilon < \min\{\epsilon_1, \epsilon_2\}$. By (41) and (42), 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 the "transitional" differences equally among all agents. Notice that agents after period 1 are unaffected, while all agents at period 1 are strictly better off. It remains to check 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

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

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 (39)) 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 (38) in mind, choose ρ such that

$$\frac{w(h_2)-w(h_1)}{w(h_4)-w(h_3)} > \rho > \frac{x(h_2)-x(h_1)}{x(h_4)-x(h_3)}, \tag{43}$$

and then γ such that

$$0 < \gamma < \rho[x(h_4)-x(h_3)] - [x(h_2)-x(h_1)]. \tag{44}$$

Now adjust the steady state distribution λ as follows. For any $\epsilon > 0$ and small, define λ_ϵ by

$$\begin{aligned}
\lambda_\epsilon(h_1) &\equiv \lambda(h_1) - \epsilon, \\
\lambda_\epsilon(h_2) &\equiv \lambda(h_2) + \epsilon, \\
\lambda_\epsilon(h_3) &\equiv \lambda(h_3) + \rho\epsilon, \text{ and} \\
\lambda_\epsilon(h_4) &\equiv \lambda(h_4) - \rho\epsilon, \qquad (45) \\
& \qquad (46)
\end{aligned}$$

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

$$(\lambda, c + \gamma\epsilon, \lambda_\epsilon) \in \mathcal{T} \tag{47}$$

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

To establish this, observe that if $z_\epsilon \equiv (\lambda, c + \gamma\epsilon, \lambda_\epsilon)$ and $z \equiv (\lambda, c, \lambda)$, then

$$\begin{aligned} pz_\epsilon = pz_\epsilon - pz &= \gamma\epsilon - x(h_1)\epsilon + x(h_2)\epsilon - x(h_4)\rho\epsilon + x(h_3)\rho\epsilon \\ &= \gamma\epsilon + \{[x(h_2) - x(h_1)] - \rho[x(h_4) - x(h_3)]\}\epsilon \\ &< \gamma\epsilon - \gamma\epsilon = 0, \end{aligned}$$

where the last inequality uses (44). Applying Lemma 4 as before, we are done.

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

$$(\lambda_\epsilon, c, \lambda) \in \mathcal{T} \tag{48}$$

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

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

where the last inequality uses (43). 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 $(z_\epsilon, z'_\epsilon, z, z, \ldots)$, where $0 < \epsilon < \min\{\epsilon_3, \epsilon_4\}$. By (47) and (48), 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. ■

Proof of Proposition 7. We describe the maximization problem for each household in any competitive equilibrium of the two profession economy as follows. Define a sequence of *values* $\{\bar{V}_t, \underline{V}_t\}$ describing the infinite-horizon payoffs to each generation at each date, conditional on starting skilled or unskilled. That is, for each t,

$$\bar{V}_t = \max_{c_t, x_t} [u(c_t) + \delta V_{t+1}(x_t)]$$

subject to the conditions that

$$c_t + x_t = w^s_t,$$

and

$$\begin{aligned} V_{t+1}(x_t) &= \bar{V}_{t+1} \text{ if } x_t \geq x \\ &= \underline{V}_{t+1} \text{ if } x_t < x. \end{aligned}$$

Likewise, $\underline{V}_t$ denotes the maximum value of the above problem for a currently unskilled household with income w^u_t instead of w^s_t. For given initial skill distribution $\lambda_0 \in (0, 1)$, a competitive equilibrium is therefore a sequence of wages and subsequent skill distributions $\{w^s_t, w^u_t, \lambda_t\}_{t=0}^\infty$ such that

[i] Given λ_0, the path of subsequent skill distributions $\{\lambda_t\}$ is generated by the maximization problems just described,

[ii] For each t, $w_t^s = w^s(\lambda_t)$ and $w_t^u = w^u(\lambda_t)$ if $\lambda_t < \lambda_1$, and $w_t^s = w_t^u = w^s(\lambda_1) = w^u(\lambda_1)$ if $\lambda_t \geq \lambda_1$.

In order to avoid qualifying statements for the initial value of λ, we make the inessential assumption that $\lambda_0 \in (0, \lambda_1)$.

Next observe that in any competitive equilibrium, there cannot be any date at which an unskilled household decides to educate its children, while a skilled household does not. For if an unskilled household were to educate its children, then

$$u(w_t^u - x) + \delta \bar{V}_{t+1} \geq u(w_t^u) + \delta \underline{V}_{t+1},$$

or

$$u(w_t^u) - u(w_t^u - x) \leq \delta[\bar{V}_{t+1} - \underline{V}_{t+1}].$$

By strict concavity and the fact that $\lambda_t < \lambda_1$ for all t, we may conclude that

$$u(w_t^s) - u(w_t^s - x) < \delta[\bar{V}_{t+1} - \underline{V}_{t+1}].$$

But this means that a skilled household has a *strict* incentive to educate its children.

It follows from this observation that in any competitive equilibrium, if the proportion of skilled households increases from one generation t to the next, it must be the case that all skilled households at date t are educating their children, and some unskilled households as well. Moreover, some unskilled households must also be deciding to *not* educate their children — otherwise there would be no unskilled households at date $t+1$. It must then be the case that when some unskilled households switch professions, they must be exactly indifferent between switching and not switching professions. Conversely, if the proportion of skilled households goes down from one generation to the next, it must be the case that some skilled households change professions while being exactly indifferent between switching and not. Hence at every date, the lifetime utility of the skilled (and unskilled) must be equal to the utility they would have received were their descendants *never* to switch status: at every date t,

$$\bar{V}_t = \sum_{s=t}^{\infty} \delta^{s-t} u(w_s^s - x) \tag{49}$$

and

$$\underline{V}_t = \sum_{s=t}^{\infty} \delta^{s-t} u(w_s^u). \tag{50}$$

This implies that for a household that is skilled at date t:

$$\sum_{s=t}^{\infty} \delta^{s-t} u(w_s^s - x) \geq u(w_t^s) + \sum_{s=t+1}^{\infty} \delta^{s-t} u(w_s^u),$$

or equivalently,

$$u(w_t^s) - u(w_t^s - x) \leq \sum_{s=t+1}^{\infty} \delta^{s-t} [u(w_s^s - x) - u(w_s^u)], \tag{51}$$

with equality holding whenever a switch from "skilled" to "unskilled" occurs at date t. Likewise, for the currently unskilled, the sacrifice involved in educating their children exceeds or just equals the benefit of all their descendants switching from the unskilled to the skilled profession:

$$u(w_t^u) - u(w_t^u - x) \geq \sum_{s=t+1}^{\infty} \delta^{s-t}[u(w_s^s - x) - u(w_s^u)], \tag{52}$$

with equality holding whenever a switch from "unskilled" to "skilled" does occur along the equilibrium path.

Next, we establish some useful lemmas concerning dynamic properties of any competitive equilibrium.

Lemma **5** *If $\lambda_t > \lambda_{t+1}$, then $\lambda_t \in A$ and $\lambda_{t+1} = \lambda_{t+2}$.*
If $\lambda_t < \lambda_{t+1}$, then $\lambda_t \in B$ and $\lambda_{t+1} \leq \lambda_{t+2}$.

Warning. Note that the two statements in the lemma are *not* symmetric. The lack of symmetry will become even clearer later.

Proof of Lemma 5. *We begin by establishing the first part of the first statement.* Because $\lambda_t > \lambda_{t+1}$, (51) must hold with equality, and we have

$$u(w_t^s) - u(w_t^s - x) = \delta[u(w_{t+1}^s - x) - u(w_{t+1}^u)] + \delta^2 M \tag{53}$$

where $M \equiv \sum_{s=t+2}^{\infty} \delta^{s-(t+2)}[u(w_s^s - x) - u(w_s^u)]$. Using (51) for period $t+1$, we see that

$$u(w_{t+1}^s) - u(w_{t+1}^s - x) \leq \delta M. \tag{54}$$

Combining (53) and (54), we see that

$$u(w_t^s) - u(w_t^s - x) \geq \delta[u(w_{t+1}^s) - u(w_{t+1}^u)].$$

Because $\lambda_t > \lambda_{t+1}$, we see that $w_{t+1}^s > w_t^s$ and $w_{t+1}^u < w_t^u$. Therefore

$$u(w_t^s) - u(w_t^s - x) > \delta[u(w_t^s) - u(w_t^u)],$$

which shows that $\lambda_t \in A$.

The proof of the first part of the second statement is completely parallel, but because (as noted above) there is an asymmetry lurking here it will be useful to simply retrace these steps and convince ourselves that they indeed go through.

For this part, $\lambda_t < \lambda_{t+1}$, so that (52) must hold with equality, and we have

$$u(w_t^u) - u(w_t^u - x) = \delta[u(w_{t+1}^s - x) - u(w_{t+1}^u)] + \delta^2 M \tag{55}$$

where M is defined just as before. Using (52) for period $t+1$, we see that

$$u(w_{t+1}^u) - u(w_{t+1}^u - x) \geq \delta M. \tag{56}$$

Combining (55) and (56), we see that

$$u(w_t^u) - u(w_t^u - x) \leq \delta[u(w_{t+1}^s - x) - u(w_{t+1}^u - x)].$$

Because $\lambda_t < \lambda_{t+1}$, we see that $w_{t+1}^s < w_t^s$ and $w_{t+1}^u > w_t^u$. Therefore

$$u(w_t^u) - u(w_t^u - x) < \delta[u(w_t^s - x) - u(w_t^u - x)],$$

which shows that $\lambda_t \in B$.

Next, we establish the second part of the first statement: that $\lambda_{t+1} = \lambda_{t+2}$. Suppose this is false. Then there are two cases to consider.

Case 1: $\lambda_{t+1} < \lambda_{t+2}$. Then at date $t+1$, (52) must hold with equality, so that

$$u(w_{t+1}^u) - u(w_{t+1}^u - x) = \delta M. \tag{57}$$

Combining (53) and (57), we see that

$$u(w_t^s) - u(w_t^s - x) = \delta[u(w_{t+1}^s - x) - u(w_{t+1}^u - x)]. \tag{58}$$

Because $\lambda_t > \lambda_{t+1}$, we have $w_t^s > w_t^u > w_{t+1}^u$. Consequently, by the strict concavity of the utility function,

$$u(w_t^s) - u(w_t^s - x) < u(w_t^u) - u(w_t^u - x) < u(w_{t+1}^u) - u(w_{t+1}^u - x). \tag{59}$$

Combining (58) and (59), we may conclude that

$$u(w_{t+1}^u) - u(w_{t+1}^u - x) > \delta[u(w_{t+1}^s - x) - u(w_{t+1}^u - x)].$$

But this means that $\lambda_{t+1} \notin B$. On the other hand, we have $\lambda_{t+1} < \lambda_{t+2}$, and this contradicts the first part of the second statement of the lemma, which we have already proved.

Case 2: $\lambda_{t+1} > \lambda_{t+2}$. Then at date $t+1$, (51) must hold with equality, so that

$$u(w_{t+1}^s) - u(w_{t+1}^s - x) = \delta M. \tag{60}$$

Combining (53) and (60), we see that

$$u(w_t^s) - u(w_t^s - x) = \delta[u(w_{t+1}^s) - u(w_{t+1}^u)]. \tag{61}$$

Because $\lambda_t > \lambda_{t+1}$, we have $w_t^s < w_{t+1}^s$. Consequently, by the strict concavity of the utility function,

$$u(w_t^s) - u(w_t^s - x) > u(w_{t+1}^s) - u(w_{t+1}^s - x). \tag{62}$$

Combining (61) and (62), we may conclude that

$$u(w_{t+1}^s) - u(w_{t+1}^s - x) < \delta[u(w_{t+1}^s) - u(w_{t+1}^u)].$$

But this means that $\lambda_{t+1} \notin A$. On the other hand, we have $\lambda_{t+1} > \lambda_{t+2}$, and this contradicts the first part of the first statement of the lemma, which we have already proved.

Finally, we prove the second part of the second statement: that $\lambda_{t+1} \leq \lambda_{t+2}$. Suppose this is false. Then $\lambda_{t+1} > \lambda_{t+2}$. Thus at date $t+1$, (51) must hold with equality, so that

$$u(w_{t+1}^s) - u(w_{t+1}^s - x) = \delta M. \tag{63}$$

Combining (55) and (63), we see that

$$u(w_t^u) - u(w_t^u - x) = \delta[u(w_{t+1}^s) - u(w_{t+1}^u)]. \tag{64}$$

Because $\lambda_t < \lambda_{t+1}$, we have $w_t^u < w_{t+1}^u \leq w_{t+1}^s$. Consequently, by the strict concavity of the utility function,

$$u(w_t^u) - u(w_t^u - x) > u(w_{t+1}^u) - u(w_{t+1}^u - x) \geq u(w_{t+1}^s) - u(w_{t+1}^s - x). \tag{65}$$

Combining (64) and (65), we may conclude that

$$u(w_{t+1}^s) - u(w_{t+1}^s - x) < \delta[u(w_{t+1}^s) - u(w_{t+1}^u)].$$

But this means once again that λ_{t+1} satisfies the first inequality in (10), or equivalently, that $\lambda_{t+1} \notin A$. On the other hand, we have $\lambda_{t+1} > \lambda_{t+2}$, and this contradicts the first part of the first statement of the lemma, which we have already proved.

■

LEMMA **6** *If λ is a steady state, then there is a unique competitive equilibrium from $\lambda_0 = \lambda$, given by $\lambda_t = \lambda$ for all t.*

Proof. Immediate from Lemma 5. For if the competitive equilibrium is nonstationary, then it must be the case that either $\lambda \in A$ or $\lambda \in B$ (simply examine the first date that $\lambda_t \neq \lambda_{t+1}$ and apply Lemma 5). In either of these cases, λ cannot be a steady state. ■

A converse to this result is the subject of the next lemma.

LEMMA **7** *If at any date t along a competitive equilibrium we have $\lambda_t = \lambda_{t+1}$, then $\lambda \equiv \lambda_t = \lambda_{t+1}$ is a steady state, and in particular $\lambda_s = \lambda_t$ for all $s \geq t$.*

Proof. Suppose not. Then by Lemma 6, it must be the case that either $\lambda \in A$ or $\lambda \in B$.

CASE 1: $\lambda \in A$. In this case, renumbering time periods if necessary, we must have $\lambda_t = \lambda_{t+1} > \lambda_{t+2}$ (using Lemma 5). Thus (51) must hold with equality at date $t+1$, so that

$$u(w_{t+1}^s) - u(w_{t+1}^s - x) = \delta M, \tag{66}$$

while at date t

$$u(w_t^s) - u(w_t^s - x) \leq \delta[u(w_{t+1}^s - x) - u(w_{t+1}^u)] + \delta^2 M \tag{67}$$

Combining (66) and (67), we see that

$$\begin{aligned} u(w_t^s) - u(w_t^s - x) &\leq \delta[u(w_{t+1}^s) - u(w_{t+1}^u)] \\ &= \delta[u(w_t^s) - u(w_t^u)]. \end{aligned}$$

But this means that $\lambda \notin A$, which is a contradiction.

Case 2: $\lambda \in B$. In this case, renumbering time periods if necessary, we must have $\lambda_t = \lambda_{t+1} < \lambda_{t+2}$ (using Lemma 5). Thus (52) must hold with equality at date $t+1$, so that

$$u(w^u_{t+1}) - u(w^u_{t+1} - x) = \delta M, \tag{68}$$

while at date t

$$u(w^u_t) - u(w^u_t - x) \geq \delta[u(w^s_{t+1} - x) - u(w^u_{t+1})] + \delta^2 M \tag{69}$$

Combining (68) and (69), we see that

$$\begin{aligned} u(w^u_t) - u(w^u_t - x) &\geq \delta[u(w^s_{t+1} - x) - u(w^u_{t+1} - x)] \\ &= \delta[u(w^s_t - x) - u(w^u_t - x)]. \end{aligned}$$

But this means that $\lambda \notin B$, which is a contradiction.

So neither Case 1 nor Case 2 is possible. This means that λ is a steady state. Applying Lemma 6, we see that there is a unique stationary equilibrium, and we are done. ■

We now return to the proof of the proposition. To prove the first part of the proposition, note that *if* there is a competitive equilibrium, then by Lemmas 5 and 7, it must have the property discussed in the statement of the proposition. To check existence and uniqueness, define λ_1 by

$$u(w^s(\lambda)) - u(w^s(\lambda) - x) \equiv \delta(1-\delta)^{-1}[u(w^s(\lambda_1) - x) - u(w^u(\lambda_1))].$$

It is easy to see that λ_1 is well-defined and unique, and that $\lambda_1 < \lambda$. Now check that this gives us a competitive equilibrium, and that there is no other way of constructing an path that satisfies both (51) and (52).

To prove the second part of the proposition, we first need to strengthen the implication of Lemma 5 in this case. It will be enough to strengthen the second part of the statement of that lemma to: If $\lambda_t < \lambda_{t+1}$, then $\lambda_t \in B$ and $\lambda_{t+1} < \lambda_{t+2}$.

All of this is proved except for the stronger implication: $\lambda_{t+1} < \lambda_{t+2}$. To establish this, suppose that the assertion is false. Then, using Lemma 5, it must be the case that $\lambda_t < \lambda_{t+1} = \lambda_{t+2}$. By Lemma 7, we have $\lambda_{t+1} = \lambda_s$ for all $s \geq t+1$. Also, (52) must hold with equality at date t. Combining these two pieces of information, we see that

$$u(w^u_t) - u(w^u_t - x) = \delta(1-\delta)^{-1}[u(w^s_{t+1} - x) - u(w^u_{t+1})].$$

Now $\lambda_t < \lambda_{t+1}$, so that $w^u_t < w^u_{t+1}$. By the strict concavity of u and the equality above,

$$u(w^u_{t+1}) - u(w^u_{t+1} - x) < \delta(1-\delta)^{-1}[u(w^s_{t+1} - x) - u(w^u_{t+1})].$$

But this means that $\lambda_{t+1} \in B$ as well. But then by Lemma 7, it cannot be the case that $\lambda_{t+1} = \lambda_{t+2}$.

To prove existence and uniqueness from this initial condition, define recursively for each λ_t, the value of λ_{t+1} that solves the equation

$$u(w^u_t) - u(w^u_t - x) \equiv \delta[u(w^s_{t+1} - x) - u(w^u_{t+1} - x)], \tag{70}$$

where w^s_{t+1} and w^u_{t+1} are to be interpreted as the wages corresponding to λ_{t+1}.

To see that this is uniquely defined, note that

$$u(w_0^u) - u(w_0^u - x) < \delta[u(w_0^s - x) - u(w_0^u - x)],$$

because $\lambda_0 \in B$. So there is a unique λ_1 that solves (70) for $t = 0$. Note that λ_1 must exceed λ_0. And this will be so whenever $\lambda_t \in B$. So it only remains to show that if $\lambda_t \in B$, then $\lambda_{t+1} \in B$ as well. To see thus simply use the fact that $\lambda_{t+1} > \lambda_t$, which implies that $w_{t+1}^u > w_t^u$. Using this information in (70) along with the strict concavity of u, we are done.

Finally, part 3 of the proposition is already established. ■