American Economic Journal: Microeconomics 2 (November 2010): 38–76
http://www.aeaweb.org/articles.php?doi=10.1257/mic.2.4.38

Inequality and Markets: Some Implications of Occupational Diversity†

By Dilip Mookherjee and Debraj Ray*

This paper studies income distribution in an economy with borrowing constraints. Parents leave both financial and educational bequests; these determine the occupational choices of children. Occupational returns are determined by market conditions. If the span of occupational investments is large, long-run wealth distributions display persistent inequality. With a "rich" set of occupations, so that training costs form an interval, the distribution is unique and the average return to education must rise with educational investment. This finding contrasts with the usual presumption of diminishing returns to human capital. It is the central testable proposition of this paper. (*JEL* D14, D31, J24)

This paper studies the nature of long-run income distribution in a competitive economy with borrowing constraints. Parents can invest in their children's education and leave financial bequests to them. All financial bequests must be nonnegative. In other words, parents cannot borrow from their children's future earnings. There is a continuum of occupations with continuously varying entry (education or setup) costs. Owing to this lack of indivisibility in investment opportunities, the steady-state income distribution is unique (in contrast to a large literature on occupational choice with indivisibilities). Different occupations are imperfect substitutes for one another in the production process (in contrast to theories of Glenn C. Loury 1981, and Gary S. Becker and Nigel Tomes 1979, 1986). Hence, occupational returns are endogenously determined by a combination of supply-side and demand-side factors.

We show that if the span of occupational investments is wide enough, the wealth distribution is nondegenerate and long-run inequality arises. In this case, the average return to education must rise with the level of educational investment—the return to human capital is endogenously nonconcave. This finding (see Propositions 1 and 2), which contrasts with the usual presumption that the private return to human capital is decreasing, constitutes the central empirically testable proposition of this paper.

* Mookherjee: Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215 (e-mail: dilipm@bu.edu); Ray: Department of Economics, New York University, 19 West 4th Street, New York, NY 10012 (e-mail: debraj.ray@nyu.edu). Mookherjee's research was supported by the MacArthur Foundation and by National Science Foundation Grant No. 0617874, and Ray's research was supported by National Science Foundation Grant Nos. 0241070 and 0617827. We thank Joan Esteban, Kiminori Matsuyama, and Andrew Newman for their helpful comments. Some of the results in this paper were reported informally in Mookherjee and Ray (2002).

† To comment on this article in the online discussion forum, or to view additional materials, visit the articles page at http://www.aeaweb.org/articles.php?doi=10.1257/mic.2.4.38.

Consider two approaches that dominate the existing literature on long-run income distribution in the presence of borrowing constraints. In the "neoclassical" Becker-Tomes-Loury models, the return to human capital is determined entirely by an exogenous technology, which is *assumed* to be concave in investment.[1] All movements in relative wages across different occupations are avoided by presuming that human capital is reducible to "efficiency units." As Becker and Tomes (1986) express it:

> Although human capital takes many forms, including skills and abilities, personality, appearance, reputation and appropriate credentials, we further simplify by assuming that it is homogeneous and the same "stuff" in different families. (Becker and Tomes 1986, S6)

Our model differs markedly from this approach in that we explicitly permit relative wages to vary, and therefore we *derive* a particular pattern of returns rather than assume one. Moreover, the particular pattern we derive contrasts with a typical assumption made in this literature—that the return to human capital investment is declining.

A distinct approach[2] emphasizes the role of indivisibilities in occupational choice. Such indivisibilities, by their very nature, induce a nonconcavity in the return to occupational investment. (Think of training costs to acquire "skills" or upfront costs to start a business.) Once again, a particular pattern of returns is imposed by assumption, though in much of this literature the relative returns to occupations is allowed to vary. We deliberately remove all indivisibilities by allowing for a diverse occupational structure, with training costs running smoothly from zero to some upper bound. In this way the shape of the human capital returns function is *fully* contingent on the equilibrium rates of return to a large multiplicity of occupations; whether or not it is concave is a question to be settled rather than an assumption.

Actually, the main focus of this literature is not the shape of occupational returns, but, instead, the multiplicity of steady states, i.e., questions of long-run history dependence. It turns out (Proposition 5) that such multiplicity disappears when there is a continuum of occupations with continuously varying training costs.[3] In this particular sense—but only in this sense—we are more in line with the "neoclassical" approach described above.

The assumptions concerning richness and imperfect substitutability of the set of occupations are key in our model, so a few remarks are in order. The assumptions on technology underlying the two principal approaches described above seem to us rather extreme and unrealistic. The presumption (in the second literature) of a sparse set of occupations with large gaps in training or setup costs is at best a simplifying device. It is hard to argue against the statement that most economies are

[1] In the endogenous growth models studied by Paul M. Romer (1986), Robert E. Lucas Jr. (1988), and others, increasing returns are permitted by assumption, but only via nonpecuniary externalities at the aggregate level.

[2] See, e.g., Ray (1990, 2006), Abhijit V. Banerjee and Andrew F. Newman (1993), Oded Galor and Joseph Zeira (1993), Lars Ljungqvist (1993), Debasis Bandyopadhyay (1997), Maitreesh Ghatak and Neville Nien-Huei Jiang (2002), and Kiminori Matsuyama (2000, 2006).

[3] Mookherjee and Ray (2003), which fundamentally emphasizes the question of multiplicity, also establish uniqueness with a continuum of occupations. However, in that model, there are no financial bequests at all. The extension of these results to the case of financial bequests is both substantive and nontrivial, as a perusal of the arguments used in this paper will reveal. See also the discussion following Proposition 5.

characterized by a broad spectrum of occupations with finely varying training costs.[4] At the same time, this is not an excuse for adopting an approach based on efficiency units (as in the first literature). The returns to occupational choices, earnings distributions, in particular, are sensitive to the supply of agents in different occupations. Considerable empirical evidence to this effect is available: e.g., Lawrence F. Katz and Kevin M. Murphy (1992) and a large subsequent literature documenting how skill premia in wages decline as the relative supply of skilled labor expands.

We also remark on another distinctive feature of the model. Parents and children are linked by intergenerational altruism. We adopt a formulation that permits any combination of *nonpaternalistic* (or "dynastic") preferences (in which parents care about the utilities of their children) and *paternalistic* preferences (in which parents get direct utility from the wealth of their children).[5] Our main approach depends on the existence of some degree of paternalistic altruism, though it could be vanishingly small. The literature typically assumes one extreme case or the other,[6] and serious attention is not paid to their different implications. An important merit of our approach is that it allows us to draw out the essential similarities and differences.

Sections I and II set up the general model that we use. In Sections III and IV, we describe the main results of the paper, summarized here:

- Proposition 1 provides a full characterization of steady-state returns to different occupations. The pattern of earnings involves two "phases," or two ranges of occupations, separated by a threshold level of training (or entry) cost. For the range of occupations with training costs below the threshold, the rate of return on educational investments is constant and equal to the rate of return on financial bequests. For occupations with higher training costs, the average rate of return on education is higher than the rate of return on financial bequests.

- Proposition 2 establishes that in the second phase, the average return to human capital is increasing with the level of occupational investment. Hence, if that phase is nondegenerate (whether and when that happens to be the case is described below), the returns to investment *cannot* be concave. It is to be emphasized that this is a derived result and not an ex ante presumption based on efficiency units, as in the neoclassical models, or indivisibilities, as in the models of history-dependence.

Propositions 1 and 2 together constitute the central testable findings of the paper.

[4] While there are large differences in training costs between unskilled occupations (such as farm workers or manual jobs) and skilled occupations (such as engineers, doctors, and lawyers), there is a large variety of semi-skilled occupations (technicians, nurses, and clerks) with intermediate training costs and wages. Besides, there are large differences in the quality of education within any given occupation, which translate into corresponding differences in education costs and wages.

[5] Paternalistic preferences are similar to, but not the same as, "warm-glow" preferences, in which parents get utility from the bequest itself. Reformulating our model with warm-glow preferences makes absolutely no difference to the results.

[6] For instance, Banerjee and Newman (1993), Galor and Zeira (1993), and Ghatak and Jiang (2002) all adopt a warm-glow formulation in which parental utility depends on the bequest itself, while Ljungqvist (1993) and Mookherjee and Ray (2003) use dynastic preferences.

- Apart from the determination of baseline wages for unskilled labor, the shape of the wage function depends only on preferences and is independent of technology. See additional discussion in Section VIIA. (To be sure, the quantities employed in each occupation do depend on the technology; all long-run adjustments to technology take place through quantities.) We provide a first-order differential equation for the wage function which can be solved in closed form for specific utility functions.

- Proposition 5 establishes that with a continuum of occupations, the steady state is unique. This proposition extends the corresponding result in Mookherjee and Ray (2003) to incorporate the presence of financial bequests.

- Whether or not this unique steady state exhibits inequality is related to whether the second phase of the wage function is nondegenerate. Proposition 6 provides a necessary and sufficient condition for that second phase to be nonempty. The condition combines three parameters: occupational "span" or the overall variation in training costs, total factor productivity (TFP), and the bequest motive. It requires occupational span to be suitably large relative to TFP and the strength of the bequest motive. Intuitively, a wide occupational span implies that a constant rate of return to education (equal to the rate of return on financial bequests) at a certain baseline wage (given by overall TFP) provides insufficient incentive to parents to train children for highly skilled occupations. When this "widespan condition" is met, high-end occupations must earn higher average returns, leading to the shape in Proposition 2, and the necessary persistence of inequality.

The widespan condition has interesting and novel implications for effects of various parameters on long-run inequality, some of which are briefly discussed in Section VIIE.

It is important to note that one of our key results, specifically Proposition 2 on the increasing average return to human capital, does depend on the existence of *some* degree of paternalistic altruism in the bequest motive. This point is explored in detail in Section II and especially in Section VI, where we study the purely dynastic version of our model, and compare the findings to our formulation. In the dynastic formulation, the return to human capital is a constant, but all of our other results go through if we are willing to entertain a vanishingly small degree of paternalism.

Our results have both substantive empirical content and broader implications concerning the role of markets in generating inequality. Chiara Binelli's recent findings for Latin America (see Binelli 2008) are consistent with our predictions, though additional empirical evidence would be necessary for sharper tests of our predictions.[7]

[7] Her paper finds evidence from Mexican micro-data of convexity of wage functions, with returns to higher education exceeding the returns to intermediate or primary education. She also finds evidence of the importance of supply-side factors; the convexity of the wage function became intensified as a result of a drop in returns to intermediate levels of education relative to high and low levels, owing to changes in supply patterns into intermediate education levels.

I. Model

A. *Occupations and Training*

A compact space $\mathcal{H}$ of *occupations* is used (along with physical capital) in the production of a single final good. There is an exogenous *training cost* $x(h)$ for occupation $h \in \mathcal{H}$, denominated in units of final output.[8] For reasons explained at the outset of this paper, we assume the occupational structure is *rich* in the following sense:

[R] The set of all possible training costs is a compact interval of the form $[0, X]$.

B. *Production*

A single output y is produced by physical capital[9] and population distribution $\boldsymbol{\lambda}$ over occupations in $\mathcal{H}$. The production function $y = f(k, \boldsymbol{\lambda})$ is assumed to be continuous,[10] strictly quasiconcave, and homogeneous of degree one. We often interpret different occupations as corresponding to different kinds of human capital. In Section VII, we explain how to extend this interpretation to the ownership of closely held firms of differing scales that produce different intermediate goods.

The following assumption ensures that the support of *observed* training costs must be connected.

[E] For every subset $C \subseteq [0, X]$ of positive Lebesgue measure, if the occupational distribution has zero value over every occupation h with $x(h) \in C$, then no output can be produced.

This is the key assumption that captures occupational "richness." For (almost) every training cost in $[0, X]$, there is a continuum of occupations with neighboring training costs that must be chosen by a positive measure of households in every generation. Observe that conditions [R] and [E] really go together as a pair; without some restriction like [E], [R] can always be trivially met by simply inventing useless occupations. Together, [R] and [E] imply that whenever positive output is produced, the chosen range of "equilibrium training costs" is always $[0, X]$.

C. *Prices and Firms*

Firms maximize profits at given prices. Normalize the price of final output to 1. We assume that the rate of interest r is exogenously given and time stationary. One simple interpretation is that capital is internationally mobile, and that our economy is a price taker on the world market. See Section VIIB.

Let $\mathbf{w} \equiv \{w(h)\}$ denote the wage function, and $c(\mathbf{w}, r)$ the unit cost function.

[8] As in Mookherjee and Ray (2003), this may be generalized to allow training costs to depend on the pattern of wages. We conjecture that the principal qualitative results of this paper will continue to hold in that setup.

[9] As long as capital goods are alienable, and shares in them can be divisibly held, having several capital goods makes no difference to the analysis.

[10] Endow the space of all nonnegative finite Borel measures on $\mathcal{H}$ with the topology of weak convergence. We ask that output be continuous with respect to the product of this topology and the usual topology on k.

By constant returns to scale, profit maximization at positive output is possible if and only if $c(\mathbf{w}, r) = 1$; in that case, call $\mathbf{w}$ a *supporting price*.

D. *Families*

There is a continuum of families indexed by $i \in [0,1]$. All families are ex ante identical. Each family i has a single representative at each date or generation.

Each agent receives utility from her own consumption c. As for intergenerational altruism, we adopt a general perspective. We assume that a parent derives utility from the overall utility (the "value function") of her descendant, as well as directly from descendant wealth y'. The first component may be viewed as *nonpaternalistic* altruism, and the second component as *paternalistic* altruism. We write overall utility, then, as

$$U(c) + \delta[\theta V + (1 - \theta)W(y')], \tag{1}$$

where V is the anticipated utility of the child (i.e., the value function), and W is some exogenous function defined on child wealth, $\delta \in (0,1)$ and $\theta \in [0,1]$.

We assume that both U and W are smooth, increasing and strictly concave, and that U has unbounded steepness at zero consumption. Given the assumption that parents cannot borrow from their children, the strict concavity of U plays a key role in the analysis, ensuring that the marginal cost of investment is higher for poorer parents.

The constant $\delta \in (0,1)$ in (1) is to be interpreted as a discount factor, while $\theta \in [0,1]$ reflects the strength of nonpaternalistic versus paternalistic altruism. When $\theta = 1$, we have the well-known formulation with value functions (as in Loury 1981 and Ljungqvist 1993), and when $\theta = 0$ all altruism is paternalistic (as in Galor and Zeira 1993 and Banerjee and Newman 1993).

Our hybrid specification is motivated not by greater generality for its own sake, though the model may have some descriptive value. It enables us to understand the role of different forms of altruism, an issue discussed in greater detail below.[11]

E. *Bequests*

Consider a member of generation t. She begins adult life with a financial bequest b and an occupation h, both selected by her parent. Her overall wealth is then $y \equiv b(1 + r) + w_t(h)$, where $w_t(h)$ is the going wage for occupation h at date t.

The agent anticipates factor prices $(\mathbf{w}_{t+1}, r)$, as well as the value function V_{t+1} for the next generation $t + 1$, and selects her own financial and educational bequests (b', h') to maximize

$$U(y - x(h') - b') + \delta[\theta V_{t+1}(y') + (1 - \theta)W(y')], \tag{2}$$

[11] As for generality alone, all of what we do below extends, with no additional insights, to utility indicators of the form $U(c) + \Psi(V, y)$, provided mild restrictions are adopted on Ψ (including a slope assumption to mimic discounting). We adopt the simpler specification in equation (1) for expositional ease.

where $y' = (1 + r)b' + w_{t+1}(h')$, and the no-intergenerational-debt constraint $b' \geq 0$ is respected.

Now b' and h' become the financial and educational inheritance of her child—generation $t + 1$—and the entire process repeats itself ad infinitum.

The condition $b' \geq 0$ is a fundamental restriction stating that children cannot be held responsible for debts incurred by their parents. The capital market is active in all other senses; households can make financial bequests at the going rate r, and firms can freely hire in capital at the very same rate.

An alternative interpretation of (2) is that the parent makes a fully financial bequest, while the *child* uses that bequest to make occupational choices under borrowing constraints. Because the child will always seek to maximize lifetime wealth, this leads to a fully equivalent formulation. In similar fashion, our model accommodates any situation in which some aspects of the occupational choice decision are delegated to the parent and the remainder to the child. None of this makes any difference to the formalization that we adopt.

F. *Equilibrium*

Fix an initial distribution of financial wealth and occupational choices. A *competitive equilibrium* is a sequence of wage functions $\mathbf{w}_t$, occupational and financial bequests $\{h_t(i), b_t(i)\}$ for each family i, value functions V_t defined on overall starting wealth, as well as occupational distributions $\boldsymbol{\lambda}_t$ such that for each t and each family i:

- person (t, i) chooses $(b_{t+1}(i), h_{t+1}(i))$ to maximize the utility function in (2), given that her own starting wealth equals $(1 + r)b_t(i) + w_t(h_t(i))$;

- these decisions aggregate to $\boldsymbol{\lambda}_t$:

$$\lambda_t(H) = \text{Measure}\{i \in [0,1] | h_t(i) \in H\},$$

for every (measurable) subset of occupations;

- $\mathbf{w}_t$ is a supporting price, and $(\boldsymbol{\lambda}_t, k_t)$ (for some k_t) is a profit-maximizing input combination at that price, and

- for each starting wealth y, $V_t(y)$ is precisely the maximized value of (2).

Observe that equilibrium conditions place no restrictions on k_t. Because there is international capital mobility, financial holdings by households need bear no relation to capital used in production.[12]

A *steady state* is a competitive equilibrium for which all time subscripts can be dropped—$(\mathbf{w}_t, \boldsymbol{\lambda}_t, V_t) = (\mathbf{w}, \boldsymbol{\lambda}, V)$ for all t—and which exhibits positive output as

[12] When there is no international capital mobility, k_t must equal the aggregate of financial holdings, and r must adjust to assure this equalization in equilibrium.

well as positive wages for all occupations.[13] A *steady-state wage function* is a wage function that is part of some steady state.

This paper restricts attention to steady states. The question of convergence to steady states from non-steady state initial conditions is beyond the scope of this paper, and represents an important issue for future research.

In what follows, an *equal steady state* will refer to a steady state with a degenerate wealth distribution. All other steady states will be called *unequal*. These are therefore distinguished with respect to history dependence at the level of individual households. History dependence at the macro level pertains to multiplicity of steady states.

II. A Benchmark

A. *A Textbook Exercise with No Occupational Choice*

A special case of this model is the following elementary and entirely abstract exercise: *only* financial bequests are possible (earning interest $r \geq -1$), and everyone earns a fixed positive wage w. If an individual makes a bequest of b, the child's lifetime wealth is

$$y' = w + (1 + r)b. \tag{3}$$

A parent with wealth y selects $b \in [0, y]$ to maximize

$$U(y - b) + \delta[\theta V(y') + (1 - \theta)W(y')],$$

subject to (3), where V is the (time-invariant) value function. Optimally chosen descendant wealth y' is some function $\phi(y; w, r)$. Obviously $y' \geq w$. If strict inequality holds, y' is fully described by the first-order conditions

$$U'\left(y - \frac{y' - w}{1 + r}\right) = \delta(1 + r)[\theta V'(y') + (1 - \theta)W'(y')]. \tag{4}$$

B. *Limit Wealth in the Benchmark Exercise*

A single-crossing argument shows that $\phi(y; w, r)$ is nondecreasing in y. So an iteration of ϕ from any initial condition $y > 0$ will yield long-run wealth starting from y. Call this long-run wealth Ω_y. It could be infinite, if the bequest motive is strong and the rate of return high. It could just be the baseline wage in the opposite case.

[13] Our definition of a steady state implies that the associated total wealth of each family is finite. Notice that our definition excludes the trivial stationary outcome in which produced output is zero.

Following Becker and Tomes (1979), it is standard in the literature on intergenerational wealth transmission to impose the "limited persistence" restriction $\partial\phi/\partial y \in (0,1)$: any increase in parental wealth translates into a smaller increase in child's wealth. This implies that the wealth of all families will converge to a common limit *independent of initial wealth*; Ω_y will not vary with y.

To be sure, such a condition is a joint restriction on tastes *and* the rate of return on capital, and is typically justified on empirical grounds.[14] We impose a weaker restriction on tastes alone: [UN] *Uniqueness*. For every $w > 0$ and $r > -1$, Ω_y has only one value (possibly infinite), permitting us to drop the subscript y on Ω.

While Ω is thereby made independent of initial y, it certainly depends on (w, r), so we generally write it as $\Omega(w, r)$. Only later do we additionally assume that $\Omega(w, r)$ is finite (see Condition F in Section IVA), but it may make for easier reading to presume throughout that $\Omega(w, r)$ is finite as well.

Condition [UN] is easy to check. By passing to the limit in (4), and using the envelope theorem to eliminate V', it is easy to see that any finite level of long-run wealth $\Omega(w, r)$ must be fully characterized by the solution in Ω to

$$(5) \qquad [1 - \delta(1 + r)\theta]U'\left(\frac{r\Omega + w}{1 + r}\right) \geq \delta(1 + r)(1 - \theta)W'(\Omega),$$

with equality if $w < \Omega(w, r) < \infty$ and the opposite inequality holding strictly throughout if $\Omega(w, r) = \infty$. (See Lemma 4 for a precise statement and proof.)

It follows that if $\delta(1 + r)\theta > 1$ (or even if $\delta(1 + r)\theta = 1$, but $\theta < 1$), $\Omega(w, r) = \infty$ and [UN] is automatically met.

If $\delta(1 + r)\theta < 1$, then [UN] requires a mild restriction on the relative movements of U' and W', one that is easy enough to verify. For instance:

OBSERVATION 1: *Suppose that U and W are the same member of the HARA family, so that*

$$-U''(c)/U'(c) = -W''(c)/W'(c) = 1/(\alpha + \beta c),$$

with $(\alpha, \beta) \geq 0$ *and nonzero.*

Then the uniqueness property in [UN] is satisfied, except when $\delta(1 + r) = \theta = 1$.

Provided that the paternalistic components of preferences (U and W) have the same functional form, Observation 1 states that [UN] holds for utility functions that are iso-elastic or exponential, or belong to the HARA class which nests these as special cases.

But there is one significant and important qualification to [UN]. It *never* holds when both $\delta(1 + r)$ and θ equal 1. This is a leading special case; with purely nonpaternalistic or "dynastic" preferences, $\theta = 1$, and even though r is exogenous, the

[14] A substantial empirical literature measures this slope and finds it to lie typically between 0 and 0.6, depending on the precise measure of income or wealth. See Samuel Bowles and Herbert Gintis (2001) or Casey B. Mulligan (1997) for a review of various empirical estimates.

case in which $\delta(1 + r) = 1$ is often invoked from general equilibrium arguments. Under this configuration, *every* initial stock is a stationary optimum, so there are many limit wealths.

This case requires separate analysis. We provide it in Section VI after we describe the results for which [UN] is assumed.

III. A Characterization of Steady States

Fix a steady state. Say that an occupation (or training cost) is *chosen* if some family chooses that occupation (or incurs that training cost). By conditions [R] and [E], we know that every steady state, which has positive output by definition, must exhibit a full measure of chosen training costs. Moreover, any two actively chosen occupations with the same training costs must earn the same wage. Hence, a steady state can be represented equivalently as a wage function $w(x)$ defined on the interval $[0, X]$ of training costs, rather than the set of occupations. In this representation, we can go back and forth (with some minor abuse of notation) between $\mathbf{w}$ defined on occupations or on training costs. It can also be shown that such a function must be continuous in training costs. We note this below as:

OBSERVATION 2: *Assume* [R] *and* [E]. *Then every steady state wage function has an equivalent representation continuous in training costs, and it is the unique representation with this property.*

From this point on we will refer (often without qualification) to the continuous equivalent representation.

A. *Steady-State Wage Functions*

Our first main result describes steady-state wage functions:

PROPOSITION 1: *Assume* [R], [E], *and* [UN] *hold.*

The continuous equivalent representation of a steady-state wage function is fully described as follows:

(i) *There exists $w > 0$—the "baseline wage"—and a threshold $z \in [0, X]$ such that for all training costs $x \le z$,*

$$w(x) = w + (1 + r)x. \tag{6}$$

(ii) *If $z < X$, then for all $x > z$,*

$$w'(x) = \frac{U'(w(x) - x)}{\delta[\theta U'(w(x) - x) + (1 - \theta)W'(w(x))]}, \tag{7}$$

with the endpoint constraint $w(z) = w + (1 + r)z$.

(iii) $\Omega(w, r) \geq w(z)$, *with equality if* $x < z$.

Any steady-state wage function must have a wage for unskilled labor; this is precisely w in the proposition. The proposition states that the wage function must be linear (with slope $1 + r$) at every training cost x with $w + (1 + r)x \leq \Omega(w,r)$, where $\Omega(w,r)$ is the limit wealth in the benchmark model with parameters (w,r). This is the joint content of parts (i) and (iii).

It is not hard to see why financial and occupational bequests must generate the same rate of return in this region. Every family has the same options as in the benchmark model with parameters (w,r), and more (they have access to occupational bequests as well). Therefore they must accumulate at least as much wealth as in the benchmark model. At the same time, there is a variety of low-level occupations that must be chosen in order for positive production. To induce individuals to choose such occupations, the occupational rate of return must equal the financial rate of return in this range; hence the linearity of the wage function.

Investments that generate wealth beyond $\Omega(w,r)$ are another matter altogether. Individuals must be given the incentive to make those investments, which necessitates the occupational rate of return to rise above the financial rate of return. The occupational return (at the margin) is now described by the differential equation (7).

Figure 1 describes a typical steady-state wage function.

If there are no such occupations, the steady-state wage function is linear throughout. This is the case described by $z = X$. Note, however, that w and z are endogenous. We postpone a fuller description of how they are determined.

B. *The Shape of the Wage Function*

The production technology plays only an indirect and limited role in determining the shape of the wage function. The richness conditions are technological, and they permit us to use a differential equation to describe wage functions over the zone $[z,X]$. It is also the case that the baseline wage w is endogenous and the production technology will allow us to pin it down. But apart from these two lines of influence, the production technology is unimportant in determining the shape of the steady state wage function. That depends entirely on preferences; see equation (7).

At the same time, we can say quite a bit about this equilibrium shape without making any further assumptions on preferences, except for those implicit in [UN]. Recall that the wage function has two "phases." In the first phase, which begins at $x = 0$ and ends at $x = z$ (the value at which wages exactly equal benchmark limit wealth evaluated at (w,r)), wages are linear in training costs with slope $(1 + r)$. If $z < X$, which is the maximum training cost, more occupations need to be populated, but a return of r will not suffice. It is therefore no surprise that the marginal return to occupational investment must exceed the financial rate of return in this second phase.

How this marginal return itself changes with the level of occupational investment is unclear, and will require more structure to pin down (see, e.g., Observation 3). Put another way, we cannot tell at this level of generality whether a steady-state wage

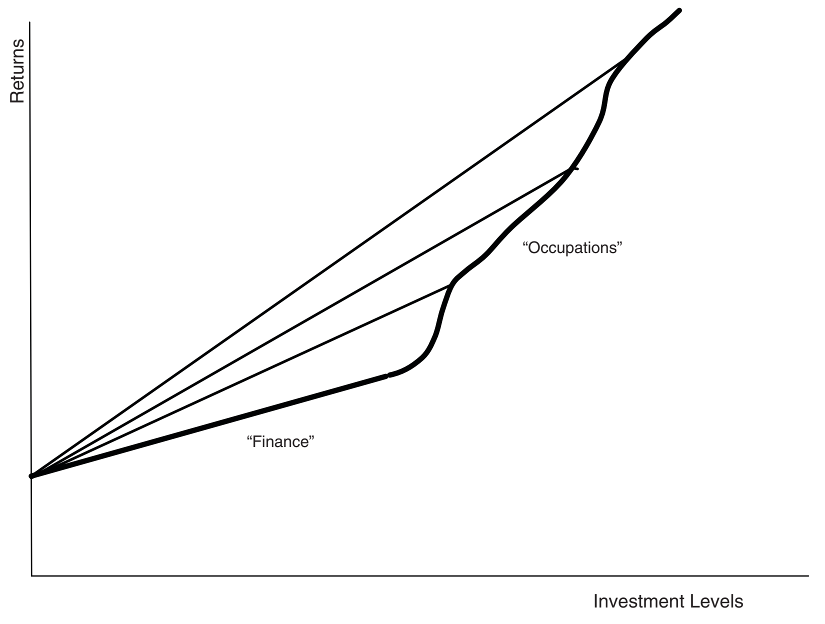

FIGURE 1. A STEADY STATE WAGE FUNCTION

function must be convex everywhere or whether it admits some (local) concavities. In general, both are possible. It *cannot* be concave everywhere; this much we already know from the fact that the marginal return exceeds r.

But more can be said. Define the *average return* to occupational investment x by

$$\rho(x) \equiv \frac{w(x) - w}{x}.$$

PROPOSITION 2: *The average return to occupational investment is strictly increasing in x on $[z, X]$.*

Figure 1 illustrates the proposition. It is a stronger statement than the mere absence of global concavity. It is also the central testable proposition of the paper because it stands the traditional theory on its head. That theory presumes—usually by assumption—that the rates of return to human capital must be *declining* in training cost (see, for instance, Loury 1981 and Becker and Tomes 1986). Therefore the poorer families make all the human capital investment, and once families are rich enough so that the marginal return on human capital falls to the constant rate assumed for financial capital, all other bequests are financial.

In contrast, we allow for flexible relative prices, a situation for which there is considerable empirical evidence (e.g., Katz and Murphy 1992). Proposition 2 then asserts that the theory *endogenously* generates rates of return that run counter to the assumptions made in the literature. Financial bequests are made at the low end, while "occupational bequests" carry a higher average rate of return, which increases with the level of such bequests.[15]

With sharper restrictions on preferences, the wage function exhibits a "global" nonconvexity, in the sense that the *marginal* rate of return rises monotonically with training costs beyond z. For instance:

OBSERVATION 3: *Consider the purely paternalistic model* $(\theta = 0)$, *in which both U and W have the same constant elasticity.*[16] *Then the marginal rate of return on occupations monotonically increases with training cost beyond the boundary z described in Proposition 1, though it is uniformly bounded over all possible training costs.*

It is also possible to provide an example with CARA utility in which the marginal rate of return increases with educational investment. But it is worth reiterating that the results on marginal returns are not robust to more general specifications. In contrast, the behavior of average rates of return highlighted in Proposition 2 is robust and general.

IV. Existence and Uniqueness of Steady States

A. *Existence*

Our definition of a steady state includes the requirement that output and wages are positive, so that existence is typically nontrivial. Proposition 1 tells us that a steady state must assume a particular form. *Given* some baseline wage w for unskilled labor, that proposition fully pins down the wage function. The only scope for variation lies in w. It comes as no surprise, then, that the existence of a (nondegenerate) steady state depends on the economy being productive enough to sustain positive profit at one of these conceivable wage functions.[17]

To this end, we construct the family of all *two-phase functions*. For any baseline wage $w > 0$, set $w(x) = w + (1 + r)x$ for all training costs no greater than

$$\zeta \equiv \frac{\Omega(w,r) - w}{1 + r}. \tag{8}$$

For $x > \zeta$, the wage function is set to satisfy the differential equation (7):

$$w'(x) = \frac{U'(w(x) - x)}{\delta[\theta U'(w(x) - x) + (1 - \theta)W'(w(x))]}, \tag{9}$$

[15] Some empirical support for these predictions is to be found in Binelli (2008), as explained in footnote 7.
[16] That is, $U(c) = (c^{1-\sigma} - 1)/(1 - \sigma)$ and $W(y) = M(c^{1-\sigma} - 1)/(1 - \sigma)$, for $\sigma > 0$ and $M > 0$.
[17] There is a family resemblance here to the existence of steady states in the multisectoral optimal growth model, which requires a "productivity condition" (see M. Ali Khan and Tapan Mitra 1986).

with the endpoint constraint $w(z) = w + (1 + r)z = \Omega(w, r)$.

We have defined a two-phase function for every training cost. To obtain a bonafide wage function, we simply truncate at the maximal training training cost X; call this a *two-phase wage function*. This procedure generates a unique wage function corresponding to any baseline w. Indeed, Proposition 1 tells that a steady-state wage function must be two-phase, with $z = \min\{\zeta, X\}$.

Our definition of a steady state presumes that all quantities are finite, so it is time to formally impose the condition:

$[\mathrm{F}]$ The value of r is such that $\Omega(w, r) < \infty$ for every $w > 0$.

This is tantamount to an upper bound on the interest rate. Consider, for the sake of an illustration, iso-elastic utility and purely paternalistic altruism. Thus, set $U(c) = (c^{1-\sigma} - 1)/(1 - \sigma)$ with $\sigma > 0$, and $W \equiv \delta U$. Define $\rho \equiv [\delta(1 + r)]^{1/\sigma}$. Intergenerational wealth movements in the benchmark model with parameters (w, r) then take the form

$$y' = \frac{(1 + r)\rho}{1 + \rho + r} y + \frac{\rho}{1 + \rho + r} w$$

if $y \geq w/\rho$, and $y' = w$ otherwise. This allows us to calculate limit wealth:

$$\Omega(w, r) = \begin{cases} w & \text{if } \rho \leq 1, \\ \dfrac{\rho}{1 - r(\rho - 1)} w & \text{if } \rho \in \left(1, 1 + \frac{1}{r}\right), \\ \infty & \text{if } \rho \geq 1 + \frac{1}{r}. \end{cases} \tag{10}$$

In line with Observation 1, $\Omega(w, r)$ is always well-defined. It is easy to see that $[\mathrm{F}]$ holds if and only if $\rho < 1 + (1/r)$.

PROPOSITION 3: *Under* $[\mathrm{R}]$, $[\mathrm{E}]$, $[\mathrm{UN}]$, *and* $[\mathrm{F}]$, *a steady state with positive production exists if and only if the following condition holds:*

$[P]$ *Unit cost* $c(\mathbf{w}, r) < 1$ *for some two-phase wage function.*

Given Proposition 1, necessity is immediate. Sufficiency is more delicate. It requires us to verify that a two-phase wage function indeed satisfies all the properties of a steady state.

Condition P in Proposition 3 is easy to check. As an example, suppose that each training cost x corresponds to a unique occupation (so name it x as well), and that the production function takes the Cobb-Douglas form

$$\ln y = (1 - \alpha) \ln k + \int_0^X \alpha(x) \ln(\lambda(x)) dx + \ln A,$$

where A is a productivity parameter, $\alpha(x) \geq 0$ and $\int \alpha(x)dx = \alpha \in (0,1)$. Then, it is easy to see that for any wage function $\mathbf{w}$,

$$\ln c(\mathbf{w}, r) = (1 - \alpha)[\ln r - \ln(1 - \alpha)] + \int_0^X \alpha(x)[\ln(w(x)) - \ln(\alpha(x))]\, dx - \ln A.$$

The verification of $[\mathrm{P}]$ therefore simply entails the choice of a wage function that minimizes $\int \alpha(x) w(x)$, and then checking whether the resulting expression above is nonpositive.

In particular, it is trivial to check that $[\mathrm{P}]$ is true in this example whenever A is large enough. This is more generally true:

COROLLARY 1: *Suppose that the aggregate production function is parameterized by a parameter A, so that $f(\boldsymbol{\lambda}, k) = A f^*(\boldsymbol{\lambda}, k)$. Under $[R]$, $[E]$, and $[UN]$, there exists $A^* > 0$ such that a steady state exists if and only if $A > A^*$.*

B. *Uniqueness*

We now examine the issue of macro-multiplicity.

PROPOSITION 4: *Assume* $[\mathrm{E}]$, $[\mathrm{R}]$, *and* $[\mathrm{UN}]$. *Then*, *apart from equivalent representations which change no observed outcome*, *there is at most one steady state.*

This proposition is a substantial extension of the uniqueness theorem in Mookherjee and Ray (2003) to a context in which financial capital coexists with human capital. Indeed, given the simplified context of our model,[18] the uniqueness result of Mookherjee and Ray (2003) can be seen very easily and intuitively.

Imagine reworking Proposition 1 by imposing the additional constraint that no financial bequests are permitted. One would reasonably suppose, then, that the first phase of the two-phase function would disappear, and that any steady-state wage function must be governed by the differential equation (7) *throughout*. It is easy to see why there can be only one such wage function. If we begin at two different initial conditions and apply (7) thereafter, the two wage trajectories cannot cross—a well-known property for this class of differential equations. In short, if there are two steady-state wage functions, *one must lie entirely above the other*. But now we have a contradiction, for two wage functions ordered in this way cannot *both* serve as *bona fide* supporting prices for profit maximization. We obtain uniqueness when there are no financial bequests.

While this serves as some intuition for the result at hand, different considerations emerge when financial bequests are permitted. Now crossings of the two putative steady-state wage functions must be ruled out by entirely different arguments. After

[18] Apart from the central difference of financial bequests, there are two differences between our model and that of Mookherjee and Ray (2003). First, they use a nonpaternalistic bequest motive. Second, training costs are endogenously determined in their model. However, these differences are minor and can be readily accommodated.

all, the behavior of the wage functions is *not* governed throughout by the differential equation (7); a nontrivial "first phase" makes an appearance. Our formal proof relies on revealed-preference arguments based on household optimization to rule out such crossings.

It should be noted that our uniqueness proposition does depend on the free international mobility of working capital at some fixed rate of return. If all physical capital must be domestically produced, then multiple steady states *may* be possible, and additional assumptions would be required to restore uniqueness.

V. Inequality

A. *Wealth Distributions*

The following proposition describes steady-state wealth for different families.

PROPOSITION 5: *Let* $\mathbf{w}$ *be a steady-state wage function with baseline wage* w *and threshold* z *as described in Proposition* 1.

(i) *Consider two families that choose occupations with training costs less than* z. *Their occupational income will generally be different. But their overall wealth must be the same*, *and equal to* $\Omega(w, r)$.

(ii) *A family that chooses an occupation with cost* $x > z$ *makes only occupational bequests*, *and has steady-state wealth* $w(x)$. *In particular*, *families in this region exhibit persistent wealth differences.*

We have already discussed why part (i) must hold. To see part (ii), consider any family inhabiting a training cost in excess of z. By Proposition 2, that family has access to rates of return that exceed r. So it cannot use financial bequests, at least up to the maximal occupation. To understand why no financial bequests are present beyond this point, observe that the slope of the wage function at the maximal training cost X is *just* sufficient to induce families to settle there. That slope is no less than $1 + r$. It follows from [UN] that no family can be lured into maintaining steady-state wealth beyond X by simply using the rate of return r.

Proposition 5 implies, in particular, that there is persistent inequality in steady state if and only if the threshold z is smaller than X, the maximum training cost. In other words, the second phase of the two-phase wage function must be nonempty. Under what conditions is this the case?

B. *Conditions for Persistent Inequality*

A simple preliminary exercise lays the groundwork for a complete solution to this question. This exercise concerns the production technology alone and has nothing to do with preferences.

Consider the class of all linear wage functions of the form $w(x) = w + (1 + r)x$ defined on *all* of $[0, X]$, parameterized by $w \geq 0$.

OBSERVATION 4: *Assume* $[P]$. *Then there is a unique value of* w*—call it* a*—and a corresponding linear wage function* $\mathbf{w}^*$ *with* $w^*(x) = a + (1 + r)x$ *for all* x*—such that* $c(\mathbf{w}^*, r) = 1$.

Now, a is not an explicit parameter of our model. But for all intents and purposes it is an exogenous primitive. To compute a all one needs is a knowledge of the production function.

As an example, recall the Cobb-Douglas case studied in Section IVA, in which each training cost corresponds to a single occupation: Cobb-Douglas form

$$\ln y = (1 - \alpha) \ln k + \int_0^X \alpha(x) \ln(\lambda(x))\, dx.$$

Using the same logic as in that case, it is easy to see that a must solve the equation

$$(1 - \alpha)[\ln r - \ln(1 - \alpha)] + \int_0^X \alpha(x)[\ln(a + [1 + r]x) - \ln(\alpha(x))]\, dx = 0,$$

provided that condition $[P]$ holds.

We are now in a position to state our central result concerning persistent inequality.

PROPOSITION 6: *Under* [R], [E], [UN], [F], *and* [P], *the unique steady state is unequal if and only if*

$$X > \frac{\Omega(a,r) - a}{1 + r}. \tag{11}$$

We shall refer to (11) as the *widespan condition*. It is made up of three parts: two of them have to do with technology, and one has to do with preferences. First, there is overall productivity in the final goods sector, which is proxied by the parameter a. (The higher the productivity, the higher the intercept of our linear wage function in Observation 4 must be, so as to meet the zero profit condition.) Then there is the range of occupations proxied here by X, the span of occupational costs. (The span X also reflects—inversely—productivity in the "educational sector.") Finally, we have the limit wealth function Ω, which is entirely a feature of preferences. The widespan condition states that limit wealth—commencing from a—is *not* enough to "span" the entire range of occupations. Net of a and discounted by the interest rate, it is smaller than the span X. Proposition 6 declares that in all such cases, the steady state must exhibit persistent inequality.

One might equivalently declare this to be a "low TFP" condition. For the term $\Omega(a,r) - a$ is typically increasing in a. See the discussion in Section VIIE.

In principle, (11) may not prescribe a unique threshold for the span X. Both a and X depend on the training cost technology. On the other hand, consider economies that are identical in all respects except for their training cost functions, which are drawn from an ordered family (all starting at 0 for some occupation). Such an ordered family may be parameterized by the highest training cost X.

In this class, it is easy to see that a depends negatively on X. Therefore, provided again that $\Omega(a, r) - a$ is increasing in a, we do generate a single-threshold restriction on span; "there is X^* such that widespan holds if and only if $X > X^*$."[19]

Uniqueness plus widespan tells us that there is just one steady state, but it must treat individuals unequally. The discussion around the statement of Proposition 4 continues to be relevant here. There is no history dependence "in the large," as the steady state is unique. But just where an individual family will end up in that distribution is significantly affected by the distant history of that family.

VI. Dynastic Preferences

In this section, we take up the special case of "dynastic preferences," or pure non-paternalistic altruism, under the additional restriction that $\delta(1 + r) = 1$. We will refer to this case as the *dynastic model*.

Recall that condition [UN] was used throughout the main analysis. That is, in the benchmark case of no occupational choice, there is at most one steady state wealth for any household. That property is natural when there is some degree of paternalistic altruism, but it is violated in the dynastic model. It is well known that *every* initial wealth is also a stationary wealth, so that there is a continuum of limit wealths. This indeterminacy of financial bequests in the long run has consequences for steady-state equilibrium. In describing them, as we shall do now, we also see the role played by a paternalistic component to altruism.

A. *Linearity of the Steady-State Wage Function*

As in the model with [UN], there is a unique steady-state wage function, but it *must* be affine, with slope $1/\delta$. We omit the formalities of this argument,[20] but there is an easy intuition for it. Because the Euler equation that maintains stationary outcomes holds at any level of wealth, any departure of the slope of the wage function from $1/\delta$ will result in some occupations not being chosen. Therefore, any steady-state wage function must be affine, with common slope equal to $(1/\delta) = 1 + r$.

But there is constant returns to scale in production, and so (just as in Observation 4), there is at most one such wage function that just supports profit maximization. Indeed that wage function must be precisely the function $\mathbf{w}^*$ described in Observation 4.

The linearity of the wage function is a key difference between the dynastic model and the framework we analyze, in which [UN] is maintained. In the latter case, an agent in the benchmark world would move toward a particular "target" wealth level. To induce her to maintain wealth above that "target" requires wage functions that offer more than the rate of interest on physical capital, which leads to the increas-

[19] More generally, though, widespan must hold for all training costs large enough (even though (11) may not imply a single threshold for X).

[20] It is straightforward to adapt—to the case of simultaneous financial and occupation bequests—the arguments leading up to equation (12) in Mookherjee and Ray (2003).

ing average returns property captured in Proposition 2. In contrast, under dynastic preferences (with $\delta(1 + r) = 1$), *any* wealth level serves as a suitable target, and limited persistence fails. This is why linear wage functions suffice to maintain incentives. Whether we accept Proposition 2 or are content with linearity must rest on how reasonable we consider [UN] to be.[21]

B. *Multiplicity of Wealth Distributions*

Another feature of the dynastic model is that it admits multiple steady-state distributions of total wealth.

Fix the unique steady-state wage function $\mathbf{w}^*$. Under the assumed conditions on the production technology, there is a unique associated occupational distribution λ^* that firms will demand in order to maximize their profits.

Assign population to occupations using the measure λ^*, and for any individual located at occupation x, assign *any* wealth no less than $w^*(x)$. That individual will gladly pass on the same wealth to her descendant, and she will be happy to bequeath $w^*(x)$ of it as an occupational bequest, so that the same dynasty inhabits an unchanged occupation over generations. Therefore, for any such wealth and occupational assignment, we have a steady state.

In particular, there are always equal steady states, and there are always unequal steady states. For the former, consider the highest wage along our affine wage function, $w^*(X)$, and give every agent a common wealth level of $\hat{w}$, where $\hat{w} \geq w^*(X)$. Using the logic of the previous paragraph, it is easy to see that there is a steady state with every dynasty's total (physical and human) wealth equal to $\hat{w}$. As for the latter, allocate all agents to just human wealth ($w^*(x)$ to a dynasty located at x), and no financial wealth at all.[22]

C. *A Refinement*

At the same time, an infinitesimal degree of paternalistic altruism serves to refine the class of steady-state distributions.

Recall condition (5) that characterizes limit wealth Ω in the benchmark case:

$$[1 - \delta(1 + r)\theta]U'\left(\frac{r\Omega + w}{1 + r}\right) \geq \delta(1 + r)(1 - \theta)W'(\Omega),$$

with equality if $w < \Omega(w,r) < \infty$ and the opposite inequality holding strictly throughout if $\Omega(w,r) = \infty$. Taking θ to 1, we see that the corresponding solutions for Ω must converge to $\Omega^*(w)$, defined as the solution to the following conditions:

$$W'(w) \leq U'(w) \text{ implies } \Omega^*(w) = w;$$

[21] To be more precise, it isn't exactly [UN] that we need for the shape predicted in Proposition 2; multiple but isolated steady states would also generate a similar result.

[22] These constructions do depend on the assumption that there is full mobility of capital. In a closed economy, the total amount of wealth held as capital held must equal the total amount of capital used in production. This may or may not be enough to allow for an equal steady state, though unequal steady states must continue to exist.

$$W'(\Omega) = U'\left(\frac{r\Omega + w}{1 + r}\right) \text{for some } \Omega \in (0,\infty) \text{ implies } \Omega = \Omega^*(w);$$

$$W'(z) > U'\left(\frac{rz + w}{1 + r}\right) \text{for all } z \text{ implies } \Omega^*(w) = \infty.$$

It is easy to check that if [UN] is imposed for all $\theta \in (0,1)$, then these conditions define $\Omega^*(w)$ uniquely.

Of course, $\Omega^*(w)$ is a steady-state wealth in the dynastic model. But it is the only steady-state wealth which is robust to small perturbations in θ around 1. Using this criterion to select from multiple steady-state wealths at $\theta = 1$, our theory extends in a straightforward manner. The widespan condition can be restated as $\Omega^*(a) < a + (1 + r)X$, where a (as before) is the intercept of the function $\mathbf{w}^*$. Even though the wage function is linear, and the returns to investment are the same as in the world without occupational bequests, financial bequests will amount to $\Omega^*(a)$, and some households in the economy must select occupations generating higher earnings than this, implying the existence of long-run inequality. This is the only steady state which is robust to (arbitrarily) small doses of paternalism.

Under this refinement, then, all our results extend without change, with the exception of Proposition 2.

VII. Extensions

We now discuss the implications of extending our model or dropping some of the central assumptions.

A. *Transitional Dynamics and the Wage Function*

Starting from any initial condition, does the economy converge to some steady state? This is not a mere technicality; it is conceptually important and (partially) justifies our focus on steady states.

But quite apart from the usual reasons for exploring transitional dynamics, there is another reason that is specific to our framework. Proposition 1 tells us that in steady state, *the form of the wage function only depends on preferences*, at least up to an intercept term, which is the baseline wage. The reason that the wage function is nevertheless compatible with profit maximization is that the baseline wage (the "intercept") can be adjusted.

What happens, then, if there is some technological shock in a particular region of the production function? The answer is that in the new steady state, all of that shock is distributed over all the input space, leaving the contours of the new wage function once again impervious to technology (the baseline wage will adjust, of course). What *will* change are quantities of different occupational inputs, and therefore the skill composition in society.

That long-run finding contrasts with what we should expect to see just following a shock. Wages in the occupations that are in greater demand will surely be bid up, so that the short-run wage function *will* indeed reflect the technology. That reflection will go away as the dynamics proceed.

This intertemporal transformation of the wage function from one that is shaped by technology to one that is mediated by preferences is an essential feature of the transitional dynamics. It cannot be seen in steady state.

Ray (2006) studies transitional dynamics in a two-occupation model with fully forward-looking agents (versions that feature more myopic agents are to be found in Banerjee and Newman 1993; Galor and Zeira 1993; and Ghatak and Jiang 2002). It remains to be seen whether these results can be extended to the considerably more challenging framework described here.[23]

B. *Endogenous Returns to Physical Capital*

Under autarky, the interest rate is endogenously determined by the condition that the capital market must clear. The propositions in the paper still apply, conditional on the interest rate. In addition, the supply of physical capital comes from financial bequests, and this must equal the demand for capital from firms at the given schedule of factor prices. Now, both the baseline wage w and interest rate r are determined by the joint condition of profit maximization and the clearing of the capital market. The wage function itself will still be two-phase.

It is entirely possible that there are multiple steady states associated with different interest rates, though these must generically be isolated. It is therefore conceivable that long-run, cross-country income differences can be explained by initial conditions that led to different interest rates. This requires more research to understand, though related forms of steady-state multiplicity have been discussed in Thomas Piketty (1997) and Banerjee and Esther Duflo (2005). Presumably, countries with a higher interest rate will involve lower unskilled wages and a higher skill premium, while comparisons of skilled wages are ambiguous.[24]

This point of view also suggests that integration of capital markets across countries will tend to promote convergence across countries.[25]

C. *Wealth Distribution and Financial Bequests*

A seemingly strange prediction of the particular model we use is that financial bequests are made at the lower end of the income distribution, but not at the upper end. All bequests there are occupational.

[23] The techniques in Ray (2006) fully exploit the assumption that there are only two occupations, and it is unclear how to extend them. Moreover, that paper assumes that there are no financial bequests.

[24] It is possible that wages of all occupations are lower, if the interest rate is higher; this is consistent with profit-maximization.

[25] This is in contrast to Matsuyama (2004) and Claustre Bajona and Timothy J. Kehoe (2006), where the opposite can happen. These papers assume there are just two factors of production: capital and labor, and a conventional neoclassical production function. Under autarky, poorer countries with less capital obtain a higher rate of return to their investments, and thus grow faster. With capital market integration (or factor price equalization owing to product market integration), the rate of return becomes equal across all countries. Interest rates decline in poor countries and rise in rich countries, thus shutting down the key force towards convergence. These results do not apply in our model owing to the endogenous nonconvexities in returns to investment, associated with the heterogeneity of different forms of human capital.

Are our predictions counterfactual? On the face of it, the answer is in the affirmative, but there are at least two straightforward extensions of the model that bring the predictions more into line with the facts.

First, we've assumed that there is just a single rate of return on financial capital. A natural extension of the model would be to incorporate multiple rates of return on financial bequests. For instance, suppose that financial capital earns r up to some threshold, and a higher return r' once past the threshold. (This would happen in situations which call for large investments, as in a hedge fund, which cannot be broken up into smaller holdings.) Or it is possible that the bequest of property has a higher rate of return associated with the asset-specific utility gains of ownership. Then a steady-state wage function will have financial bequests both at the lower end of the wealth distribution, as we have here, as well as in its upper reaches (perhaps in the form of property transfers).

Second, by "occupations," we mean not just human capital, but every productive activity that is inalienable. This includes human capital, but it is certainly not restricted to it. In particular, it is possible to view large financial bequests observed at the top end of the distribution as a form of occupational investment by parents, in the form of transfer of ownership or control of (partly inalienable) business activities. This interpretation is pursued a bit further in the next subsection.[26]

D. *Firms as Occupations*

So far, the term "occupation" has been loosely used to describe a particular form of human capital. (This interpretation is reinforced by the term "wage function.") It is possible to entertain an alternative interpretation, in which "occupations" are firms of different sizes, the ownership of which cannot be fully diversified in a stock market. Firms rely on self-finance or an imperfect capital market to fund their set-up costs.[27]

As before, an individual starts life with wealth inherited from her parents, but now starts up one of several firms with varying setup costs (interpret "labor" as starting a firm with zero fixed costs). There are borrowing constraints that depend on starting wealth. Suppose that at each size (or setup) level, intermediate goods are produced that are essential in final production. This will guarantee rich diversity in firm size, and it implies our occupational richness conditions [R] and [E].

The rest is largely, but not entirely, reinterpretation. The production function for the final consumption good is given by $f(\boldsymbol{\lambda})$, where $\boldsymbol{\lambda}$ is a firm size distribution over the set of intermediate sectors $\mathcal{H}$. The setup cost for sector $h \in \mathcal{H}$ is just $x(h)$. Perfect competition in the final good sector determines prices $\mathbf{p}$ for intermediate goods.

[26] Two other points are to be noted. First, a large proportion of financial bequests may occur because death is imperfectly predicted. For instance, Jagadeesh Gokhale et. al. (2001) argue that most financial bequests in the US economy are unintentional, the result of premature death and imperfect annuitization. For a model of unintended bequests arising from uncertain life span, see, e.g., Luisa Fuster (2000). Second, if one compares earnings and income from assets, earnings inequality accounts for most of overall income inequality. For instance, Gary S. Fields (2004) summarizes observations from several studies, writing that "labor income inequality is as important or more important than *all other income sources combined* in explaining total income inequality."

[27] Huw Lloyd-Ellis and Dan Bernhardt (2000) and Matsuyama (2000, 2006) are related to this broad view.

It remains to specify the returns to intermediate-good production, the "wages" $w(h)$. These are the profits (not counting setup costs) in sector h. They will depend on the production function in that sector. If we take the simplest specification that labor alone is needed to produce intermediates, then this profit will depend on the baseline wage w. It is also easy to incorporate imperfect (rather than entirely missing) capital markets in the firm's setup decision.

Barring minor modifications, our previous analysis applies largely unchanged. In addition, we obtain a more natural interpretation of inequality at the top end of the distribution, compared to the case of pure human capital. All bequests are "financial" in this world, and the largest financial bequests are observed at the top rather than bottom end of the distribution. The right notion of "occupation" therefore seems to involve more than just human capital or the acquisition of skill. It also embraces those sectors of the economy with large start-up costs, and which, for one reason or another, cannot be fully incorporated.

E. *Some Informal Applications of Widespan*

To illustrate the implications of the span condition in Proposition 6, we return to the example of iso-elastic utility function and purely paternalistic altruism.

Recall, in particular, equation (10) and the discussion following it. If $\delta \leq 1/(1 + r)$, (11) is always satisfied and an equal steady state never exists. There are effectively no financial bequests in the limit, so the model reduces to the disequalization model in which financial bequests are not allowed. If, on the other hand, $\delta \geq 1 + (1/r)$, then $\Omega(a, r) = \infty$ and (11) fails. Financial bequests overwhelm any inequality arising from the need to provide occupational choice incentives, and an unequal steady state cannot exist.

In the intermediate case in which δ is neither too large nor too small, (10) tells us that

$$\Omega(a, r) = \frac{\rho}{1 - r(\rho - 1)} a, \tag{12}$$

where $\rho \equiv [\delta(1 + r)]^{1/\sigma}$, so that the widespan condition (11) reduces to

$$X > a \frac{\rho - 1}{1 - r(\rho - 1)}.$$

We now describe effects of varying parameters of the model, which are relevant to explaining cross-country differences, or effects of technological change.

Differences in TFP Levels.—Suppose we compare two countries that differ only in their levels of total factor productivity. Then for any common value of X, the poorer country has a lower value of a, implying that it is more prone to disequalization. Intuitively, the lower level of wages reduces the intensity of the parental bequest motive; they are less willing to undertake the educational investments for high-end occupations. The resulting shortage of people in high-end occupations causes a rise in the skill premium. This motivates some households to enter these high-end occupations, but makes wealth inequality more acute in the process. Technologically backward countries are therefore more prone to disequalization.

Of course, this argument is based partly on the assumption that the range of training costs X is unaffected by wages. However, it is easy to incorporate this extension under the plausible assumption that both human and physical inputs enter into production. Then X is lower in the unproductive country, but not by the same factor as a. The argument is obviously reinforced if poorer countries also possess a less productive educational technology.

Differences in TFP Growth Rates.—While TFP-related differences in poverty are positively associated with disequalization, higher *growth* may be positively related to it as well. For instance, if growth (from Hicks-neutral technical progress) causes all wages and costs to grow at a uniform rate, then—all other things being equal—the level of desired bequests will be dulled, raising the likelihood of disequalization.[28]

To the extent that poorer countries grow faster owing to a "catch-up" phenomenon in technology, the widespan condition is therefore more likely to hold on two counts: higher poverty *and* higher growth. Of course, the net result is ambiguous if subsequent growth isn't positively correlated with initial poverty.

Changes in Interest Rates.—A change in the rate of return to capital has subtle effects. When r rises, ρ also goes up. Both these effects work against the widespan condition, by raising the rate of return to financial bequests. So a first cut at this issue would suggest that an increase in the global rate of return to physical capital tends to be equality-enhancing. However, there is the possibility that a may be lowered by the increase in r. This effect runs in the opposite direction, and a full analysis is yet to be conducted.

Reliance on Physical Capital.—Now, let us compare economies with differing degrees of mechanization, i.e., reliance on physical capital vis-à-vis human capital in production. One simple way to do this is to suppose that final output is produced via a nested function

$$y = Ak^{\alpha}m^{1-\alpha},$$

where m is a composite of the occupational inputs: e.g., an intermediate good "produced" by workers. Then greater mechanization corresponds to a rise in α. "Optimizing out" capital by setting its marginal product equal to r, we see that the indirect "reduced-form" production function is linear in m:

$$y = Bm,$$

[28] We omit a formal demonstration of this assertion, which proceeds by deriving an equivalent of the widespan condition (12) in the presence of neutral technical progress.

where

$$B = A\left(\frac{\alpha A}{r}\right)^{\frac{\alpha}{1-\alpha}}.$$

Notice that B essentially prices the composite in terms of the final output. If B goes down for some reason, then the intercept a will decline. So a reduction in B, other things being equal, will contribute to a greater likelihood of disequalization. Whether B goes up or down with α depends on the ratio of A (TFP) to r.[29] In relatively "unproductive" economies in which A is small, an increase in physical capital intensity lowers B, making inequality more likely. The opposite is the case in "productive" economies in which A is large. We thus obtain an interesting answer to a classic question in the theory of distribution, the impact of greater mechanization in production on long-run inequality.

Wider Product Variety.—Wider occupational spans may be the outcome of introduction of new goods and services, owing to technological change. The production of new goods and services such as information and communication technology creates an entirely new set of occupations. Such occupations are likely to require high levels of education and training, which may be thought of as an increase in the span of occupations and associated training costs. Unlike the parameterization used in Proposition 6, such changes involve an increase both in X and in the productivity of the technology. In terms of (12), both X and a tend to rise and the net effect depends on the ratio of these two variables.

VIII. Conclusion

We have studied a model of intergenerational bequests which allows for both financial bequests as well the choice of a rich variety of occupations. Occupational inputs are imperfect substitutes, so that relative factor prices are market-determined. At the level of an individual household, occupational investments may be fine-tuned to an arbitrary degree. But the returns to those occupations are endogenous, so that *markets*, not *technology*, determine whether households face a convex or nonconvex investment frontier.

We fully characterize the household frontier in steady state. It must have a two-phase property. Initially, returns are linear in investment, with the rate of return on occupational investment exactly equal to the rate on financial bequests. Thereafter, the payoff frontier follows a differential equation which we can fully describe using the primitives of the model.

In this second phase, the average rate of return on occupational investment exceeds the financial rate of return, *and it must be strictly increasing in the size of that investment*. This is the central proposition of the paper, one that distinguishes it empirically from classical models of rich occupational choice. Those models typically assume that

[29] The intuitive reason why this is the appropriate comparison is that A becomes the productivity of capital in the limiting case when $\alpha = 1$.

all occupations and skills can be reduced to "efficiency units of human capital," and that the return to human capital is *declining*, in contrast to what we obtain here.

We also address an old question on inequality: are persistent economic differences across individuals the outcome of "luck," or must markets act to necessarily create such differences? The answer to this question is equivalent to the existence of a nonempty second phase in the household investment frontier described above. In this phase, the rate of return to occupational investment rises above the financial return, and markets must act to create persistent wealth inequalities, even in the absence of any uncertainty. We show that a certain "widespan condition" is necessary and sufficient for the existence of this second phase, and we examine how this condition relates to underlying parameters of preferences and technology.

Our theory generates a number of empirically testable predictions, concerning the returns to different occupations. Of particular interest are nonlinearities with respect to education or training costs. The Loury-Becker-Tomes theory predicts a linear or concave pattern of returns, whereas our theory predicts returns to higher end occupations will be higher than to lower end occupations. To take our model to the data will however necessitate extending it to incorporate shocks to abilities or income luck, as well as uncertain lifetimes (with corresponding implications for unintended financial bequests, as distinct from the intended financial bequests in the current model). Adequate empirical tests will also require a suitably broad definition of occupations (which include entrepreneurship, family firms and service sector firms in law, medicine, real estate etc. which require a combination of high-end human capital allied with high setup costs).

Another interesting avenue for extension are the implications for intergenerational mobility. The current model exhibits no mobility. It needs to be enriched with shocks to incomes or abilities in order to generate steady state mobility. We hope such extensions will generate interesting new insights as well as empirically testable predictions.

Appendix

In what follows, and where the context is clear, we shall freely switch between references to wage functions defined on training costs and wage functions defined on occupations. Proofs of the first four lemmas below are straightforward and suppressed for the sake of brevity. Lemma 1 follows from a standard single-crossing argument based on the concavity of U. The remaining three Lemmas follow from the optimization problem faced by parents.

LEMMA 1: *Given any steady-state wage function that is increasing in training cost x, and any initial wealth for a family, the aggregate of its human and financial wealth must be monotonic in time. In particular, the overall wealth of every family is stationary in any steady state.*

LEMMA 2: *For any steady-state wage function:*

(i) *If h is chosen, $x(h) = x(h')$ implies $w(h) \geq w(h')$.*

(ii) *If h is chosen, then $x(h) > x(h')$ implies $w(h) - w(h') \geq (1+r)[x(h) - x(h')]$.*

LEMMA 3: *In the benchmark case,*

$$V(y) = \max_{0 \leq b \leq y} U(y-b) + \delta[\theta V(y') + (1-\theta)W(y')],$$

where $y' = w + (1+r)b$.

Moreover, V is concave and differentiable with $V'(y) = U'(y-b)$ for all $y > 0$, where b is the optimal choice of bequest at y.

LEMMA 4: *In the benchmark case, let $y' = \phi(y; w, r)$ be the optimal choice at y. If $\max\{y', y\} > w$, $y' - y$ has exactly the same sign (including equality) as*

$$\delta(1+r)(1-\theta)W'(y) - [1 - \delta(1+r)\theta]U'\left(\frac{ry+w}{1+r}\right). \tag{13}$$

In particular, at any limit wealth Ω with $w < \Omega < \infty$,

$$\delta(1+r)(1-\theta)W'(\Omega) - [1 - \delta(1+r)\theta]U'\left(\frac{r\Omega+w}{1+r}\right) = 0, \tag{14}$$

and if w is a limit wealth, then

$$\delta(1+r)(1-\theta)W'(w) - [1 - \delta(1+r)\theta]U'(w) \leq 0. \tag{15}$$

Finally, if (13) is positive for all y, then the only steady-state wealth is $\Omega = \infty$.

PROOF OF OBSERVATION 1:

If $\delta(1+r)\theta > 1$, (13) is positive for all y. Hence, Lemma 4 implies steady-state wealth in the benchmark case is unique and equals ∞. The same is true if $\delta(1+r)\theta = 1$ and $\theta < 1$. If $\delta(1+r)\theta < 1$, uniqueness of steady-state wealth follows from Lemma 4 if

$$U'\left(\frac{ry+w}{1+r}\right) / W'(y) \text{ is increasing in } y, \tag{16}$$

for all $y > 0$, $w > 0$, and $r > -1$. By differentiating with respect to y, we see that it is sufficient to prove that

$$W'(y)\frac{r}{1+r}U''\left(\frac{ry+w}{1+r}\right) - U'\left(\frac{ry+w}{1+r}\right)W''(y) > 0,$$

which is equivalent to

$$-\frac{U''\left(\frac{ry+w}{1+r}\right)}{U'\left(\frac{ry+w}{1+r}\right)}\left(\frac{r}{1+r}\right) < -\frac{W''(y)}{W'(y)}.$$

This reduces to the condition that

$$\frac{r}{1+r} < \frac{\alpha + \beta\left[\frac{ry+w}{1+r}\right]}{\alpha + \beta y}, \tag{17}$$

i.e., $\beta w + \alpha > 0$.

PROOF OF OBSERVATION 2:

Let $\hat{\mathbf{w}}$ be a steady-state wage function. Let T be the set of all chosen training costs. Because a steady state must have positive output by definition, it follows from $[\mathrm{R}]$ and $[\mathrm{E}]$ that T must be of full measure. Moreover, $\hat{\mathbf{w}}$ (viewed as a function of x) must be continuous on T. For if not, we can select training costs x and x' in T that are arbitrarily close, but such that their wage difference is bounded away from zero. In that case no parent would select the (almost identical) training cost with a lower wage.

Therefore, we can find a unique continuous extension of $\hat{\mathbf{w}}$ to all of $[0,X]$; call it $\mathbf{w}$. Continuing the slight notational abuse, let $w(h) = w(x(h))$. We claim that $w(h) \geq \hat{w}(h)$ for all h that are uninhabited. For if this were false for some h, we can find a chosen occupation h' arbitrarily close to h, but with wages bounded below that of $\hat{w}(h)$, which means that all occupiers of h' would prefer h, a contradiction.[30]

By this claim, if we replace $\hat{\mathbf{w}}$ by $\mathbf{w}$, no firm will wish to change its desired input mix (unused inputs have not become any cheaper). To complete the equivalence, observe that no family occupying h' finds it strictly profitable to switch to an uninhabited occupation h once its wage has been replaced by $w(h)$. For if this were true, then by the definition of continuous extension we can find a third *chosen* occupation h'' such that the family must therefore also find it profitable to switch from h' to h''. But this is a contradiction, since that option is already available in the going steady state.

Observation 2 tells us that a continuous equivalent wage representation $\mathbf{w}$ exists for any steady-state wage function. In what follows, we focus on this representation. Define $w \equiv w(0)$.

[30] This argument, as well as its counterpart for (a), can be made entirely precise by showing that the preference for h over h' can be made uniform over all families, irrespective of their wealth.

To prepare for the proofs of the remaining propositions, we record several lemmas, and we presume (often implicitly) that [R], [E], and [UN] apply where needed.

If a wage function satisfies $w(x) - w(x') = (1 + r)(x - x')$ for all x and x' in some interval, say that it is *r-linear* over that interval.

LEMMA 5: *Suppose that a family in steady state chooses to make both an occupational bequest of x and an additional financial bequest. That is, it possesses (and bequeaths) total wealth W, where $W > w(x)$. Then $\mathbf{w}$ is r-linear over all $x' \geq x$ with $w(x) + (1 + r)(x' - x) \leq W$:*

$$w(x') = w(x) + (1 + r)(x' - x). \tag{18}$$

PROOF:

Pick any $x' > x$ with $w(x) + (1 + r)(x' - x) \leq W$. Our family is making a financial bequest of at least $x' - x$. If (18) were to fail at x', then by Lemma 2, part (ii), and the fact that $\mathbf{w}$ is a continuous equivalent representation, we must have

$$w(x') > w(x) + (1 + r)(x' - x),$$

which means that our family would certainly be strictly better off choosing an occupational bequest of x' combined with a zero financial bequest, a contradiction.

LEMMA 6: *Let Y be any lower bound on stationary wealth across all families.*

(i) *If for any occupation with training cost x, we have $w + (1 + r)x \leq Y$, then $w(x) = w + (1 + r)x$.*

(ii) *The stationary wealth of any family selecting an occupation whose training cost satisfies $w + (1 + r)x \leq Y$, must be $\Omega(w, r)$.*

(iii) *The stationary wealth of every family must be at least $\Omega(w, r)$.*

PROOF:

Part (i): Since there exist occupations with training costs arbitrarily close to 0, given any occupation with training cost $x > 0$ and $w + (1 + r)x \leq Y$, there exist occupations with training cost $x' < x$ which are chosen. Apply Lemma 5.

Part (ii): For each family selecting occupation h with $w + (1 + r)x(h) \leq Y$, we have $w(h) = w + (1 + r)x(h)$. Therefore, the realized rate of return to *all* the choices of such a family, financial and educational, is exactly r. Moreover, by Lemma 2, part (ii), and the fact that $\mathbf{w}$ is a continuous equivalent representation, we also know that wages yield *no less* a return than r for all educational investments. Yet, these families find it optimal (by part (i)) not to utilize such regions of educational investment. They must therefore be behaving in exactly the same way as in a benchmark world with parameters (w, r). We must conclude that their stationary wealth equals $\Omega(w, r)$.

Part (iii): Part (ii) and Lemma 5 together tell us that $w(h) = w + (1 + r)x(h)$ for all h with $w(h) \leq \Omega(w,r)$. It follows from [UN] that no family can have a wealth smaller than $\Omega(w,r)$.

By Lemma 6, then, a steady-state wage function $\mathbf{w}$ starting from w must be r-linear up to $\Omega(w,r)$. Define

$$z = \min\left\{\frac{\Omega(w,r) - w}{1 + r}, X\right\}. \tag{19}$$

We are now interested in the shape of $\mathbf{w}$ in the region $U \equiv [z,X]$, provided that $z < X$.

LEMMA 7: *Let I be some subinterval of U such that no financial bequests are made by any family that chooses some occupation with training cost in I. Then,* $\mathbf{w}$ *satisfies* (7) *on I.*

PROOF:

Let $\mathbf{w}$ be a steady-state wage function and V be the associated value function that goes with it. Fix any $x \in I$, with $x < \sup I$. For $\epsilon > 0$, but small enough, $x + \epsilon \in I$ as well. Assume provisionally that both x and $x + \epsilon$ are chosen. Then family wealth at x (resp. $x + \epsilon$) is merely $w(x)$ (resp. $w(x + \epsilon)$). Using the two optimality conditions, one for families with wealth $w(x)$ and the other for families with wealth $w(x + \epsilon)$, we see that

$$\begin{aligned} U(w(x) - x) - U(w(x) - (x + \epsilon)) &\geq \delta[\theta\{V(w(x + \epsilon)) - V(w(x))\} \\ &\quad + (1-\theta)\{W(w(x + \epsilon)) - W(w(x))\}] \\ &\geq U(w(x + \epsilon) - x) - U(w(x + \epsilon) - (x + \epsilon)). \end{aligned}$$

Now, using the fact that $\mathbf{w}$ is a continuous equivalent representation, and invoking [R] and [E], we can see that the inequality above actually applies to *all* x and $x + \epsilon$ in I, not just those that are chosen.[31]

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

$$\begin{aligned} U'(w(x) - (x + \epsilon)) &\geq \delta[\theta\Delta V(x,\epsilon) + (1 - \theta)W'(y)\Delta w(x,\epsilon)] \\ &\geq U'(w(x + \epsilon) - x), \end{aligned} \tag{20}$$

where y is a suitable intermediate value in $[w(x), w(x + \epsilon)]$, $\Delta V(x,\epsilon) \equiv [V(w(x + \epsilon)) - V(w(x))]/\epsilon$, and $\Delta w(x,\epsilon) \equiv [w(x + \epsilon) - w(x)]/\epsilon$.

[31] Given [R] and [E], we can approach both x and $x + \epsilon$ by a sequence of chosen training cost pairs in I. For each such pair the inequality holds. Note, moreover, that V is continuous.

Now, observe that for every x' in I,

$$V(w(x')) = U(w(x') - x') + \delta[\theta V(w(x')) + (1 - \theta)W(w(x'))],$$

so that

$$\Delta V(x,\epsilon) = \frac{U'(c)[\Delta w(x,\epsilon) - 1] + \delta(1 - \theta)W'(y')\Delta w(x,\epsilon)}{1 - \delta\theta}, \tag{21}$$

where the mean value theorem has been used again, c and y' are suitable intermediate values.

Now, combine (20) and (21). Send ϵ to 0, and use the continuous differentiability of U and Ψ to conclude that the right-hand derivative of w with respect to x—call it $w^+(x)$—exists, and

$$U'(w(x) - x) = \delta\theta \frac{U'(w(x) - x)[w^+(x) - 1] + \delta(1 - \theta)W'(w(x))w^+(x)}{1 - \delta\theta} + \delta(1 - \theta)W'(w(x))w^+(x).$$

Transposing terms and simplifying, we conclude that

$$w^+(x) = \frac{U'(w(x) - x)}{\delta[\theta U'(w(x) - x) + (1 - \theta)W'(w(x))]}.$$

By exactly the same argument applied to x (greater than $\inf I$) and $x - \epsilon$, we may conclude the same of the left-hand derivative, which verifies (7).

The next lemma summarizes what we know so far.

LEMMA 8: *The continuous equivalent representation of any steady-state wage function is r-linear up to z, followed by combinations of intervals over which either the differential equation* (7) *is obeyed, or r-linearity holds.*

PROOF:

Combine Lemmas 5 and 7.

We now discuss some properties of two-phase functions. These are defined in Section A; see (8) and (9). (The corresponding two-phase *wage* functions are found by restricting the domain to $[0,X]$.)

LEMMA 9: *Any two-phase function with positive baseline wage has* $w'(x) > 1 + r$ *for almost all* $x > \zeta \equiv (\Omega(w,r) - w)/(1 + r)$.

PROOF:

The continuous differentiability of U and W imply that $\mathbf{w}$ is continuously differentiable in its second phase, where it follows (7). Also, it is easy to see that

$w'(\zeta) = 1 + r$. Therefore, if the assertion is false, there is an interval $[x_1, x_2]$, with $x_1 \geq \zeta$, such that (a) $w'(x) \geq 1 + r$ for all $x \leq x_1$, (b) $w'(x_1) = 1 + r$, and (c) $w'(x) \leq 1 + r$ for all $x \in [x_1, x_2]$. Applying (9) at x_1 and using (b), we see that

$$\delta(1 + r)(1 - \theta)W'(y_1) - [1 - \delta(1 + r)\theta]U'(y_1 - x_1) = 0, \tag{22}$$

where we've defined $y_1 \equiv w(x_1)$.

Define $\hat{w} = y_1 - (1 + r)x_1$. By (a), $\hat{w} \geq w > 0$. It is also easy to see that $(ry_1 + \hat{w})/(1 + r) = y_1 - x_1$, so that (22) reduces to

$$\delta(1 + r)(1 - \theta)W'(y_1) - [1 - \delta(1 + r)\theta]U'\left(\frac{ry_1 + \hat{w}}{1 + r}\right) = 0. \tag{23}$$

By Lemma 4, we must conclude that y_1 is a limit wealth in the benchmark model with parameters $(\hat{w}, r)$.

At the same time, applying (9) at x_2, defining $y_2 \equiv w(x_2)$, and using (c), we have

$$\delta(1 + r)(1 - \theta)W'(y_2) - [1 - \delta(1 + r)\theta]U'(y_2 - x_2) \geq 0. \tag{24}$$

Item (c) tells us that $\hat{w} \geq y_2 - (1 + r)x_2$, so that $(ry_2 + \hat{w})/(1 + r) \geq y_2 - x_2$. Using this information together with the concavity of U in (24), (24) reduces to

$$\delta(1 + r)(1 - \theta)W'(y_2) - [1 - \delta(1 + r)\theta]U'\left(\frac{ry_2 + \hat{w}}{1 + r}\right) \geq 0, \tag{25}$$

By Lemma 4, again, we must conclude that there is some $y \geq y_2$ which is a limit wealth in the benchmark model with parameters $(\hat{w}, r)$. Because y_1 is already a limit wealth, this contradicts UN.

LEMMA 10: *Let* $\mathbf{w}$ *be a two-phase function with baseline wage* $w > 0$, *and associated* ζ *defined by* (8). *Under* $\mathbf{w}$*:*

(i) *A family with starting wealth* $y \geq w(\zeta)$ *will bequeath a total wealth of precisely* y.

(ii) *A family with starting wealth* $y < w(\zeta)$ *will monotonically accumulate wealth, with limit wealth* $w(\zeta) = \Omega(w, r)$.

PROOF:

Denote by V the value function under $\mathbf{w}$. First pick a family located at occupation with training cost $x \geq z$. Because $\mathbf{w}$ has a slope of at least $1 + r$, this family has no need to make financial bequests. Let $M(x, x') \equiv U(w(x) - x') + \delta[\theta V(w(x')) + (1 - \theta)W(w(x'))]$ be this family's expected payoff from leaving an educational bequest x', and let $N(x) \equiv M(x,x)$. Then, by Lemma 3, N is differentiable and it is easy to see that

$$N'(x) \geq U'(w(x) - x)\, w'(x) \text{ for all } x, \text{ with equality if } x \geq \zeta. \tag{26}$$

For any $x' \geq x \geq \zeta$, then, using the equality in (26),

$$\begin{aligned} M(x,x) &= M(x',x') - \int_x^{x'} U'(w(z) - z) w'(z)\, dz \\ &\geq M(x',x') - \int_x^{x'} U'(w(z) - x') w'(z)\, dz \\ &= M(x',x') + U(w(x) - x') - U(w(x') - x') \\ &= M(x,x'). \end{aligned}$$

Similarly, for $x' \leq x$, using the inequality in (26),

$$\begin{aligned} M(x,x) &= M(x',x') + \int_{x'}^{x} N'(z)\, dz \\ &\geq M(x',x') + \int_{x'}^{x} U'(w(z) - z) w'(z)\, dz \\ &\geq M(x',x') + \int_{x'}^{x} U'(w(z) - x') w'(z)\, dz \\ &= M(x',x') + U(w(x) - x') - U(w(x') - x') \\ &= M(x,x'). \end{aligned}$$

Therefore, $M(x,x) \geq M(x,x')$ for all x', so that the family with starting wealth $w(x) \geq w(\zeta)$ behaves optimally by bequeathing x. This proves (i).

Now, consider a family with $y < w(\zeta)$. By a standard single-crossing argument and part (i), that family will never bequeath more than ζ. Therefore, it must behave just as in a benchmark world with prices (w, r), which then proves part (ii).

LEMMA 11: *Let* $\mathbf{w}$ *be a two-phase function with associated* ζ *given by* (8). *Let* $\mathbf{w}^*$ *be any continuous function such that:* (*a*) $\mathbf{w}^*$ *coincides with* $\mathbf{w}$ *at some point* $x_1 \geq \zeta$, (*b*) $\mathbf{w}^*$ *is r-linear on* $[x_1, x_2)$ *for some* $x_2 > x_1$ (*possibly infinite*), (*c*) $w^*(x) - w^*(x') \geq (1 + r)[x - x']$ *for all* $x > x'$, *and* (*d*) *descendant wealth starting from some wealth* y *in* $[w^*(x_1), w^*(x_2))$ *remains forever in* $[w^*(x_1), w^*(x_2)]$. *Then limit wealth from that starting point* $y \in [w^*(x_1), w^*(x_2))$ *cannot exceed* $w^*(x_1)$.

PROOF:

Define $y_1 \equiv w^*(x_1)$. Note that by condition (a), $y_1 = w(x_1)$. Use (7) to conclude that

$$\frac{U'(y_1 - x_1)}{\theta U'(y_1 - x_1) + (1 - \theta)W(y_1)} = \delta w'(x_1) \geq \delta(1 + r),$$

so that

$$\delta(1 + r)(1 - \theta)W'(y_1) - [1 - \delta(1 + r)\theta]U'(y_1 - x_1) \leq 0.$$

Define $\hat{w} \equiv y_1 - (1 + r)x_1 > 0$. Then it is easy to see that $y_1 - x_1 = (ry_1 + \hat{w})/(1 + r)$, and so

$$\delta(1 + r)(1 - \theta)W'(y_1) - [1 - \delta(1 + r)\theta]U'\left(\frac{ry_1 + \hat{w}}{1 + r}\right) \leq 0.$$

This shows that there exists $y^* \leq y_1$ such that y^* is a limit wealth in the benchmark model with parameters $(\hat{w}, r)$.

Consider any initial wealth $y \in [w^*(x_1), w^*(x_2))$ such that condition (d) holds. Then descendant wealth lies in the same interval for all dates. Consequently, if limit wealth exceeds y_1, then it must lie in $(w^*(x_1), w^*(x_2)]$. But then, by conditions (b) and (c), it is also a limit wealth in the benchmark model with parameters $(\hat{w}, r)$. Together with UN, this contradicts the fact that there is another limit wealth $y^* \leq y_1$ in the same benchmark model.

LEMMA 12: *In an unequal steady state, the equivalent-representation wage function must be two-phase.*

PROOF:

Let $\mathbf{w}^*$ be a (continuous) steady state wage function, starting from $w > 0$. Denote by $\mathbf{w}$ the two-phase wage function starting from the same point. Lemma 6 tells us that the two functions coincide at least up to ζ (defined for $\mathbf{w}$). If $\zeta \geq X$ we are done. Otherwise, $\zeta < X$. Suppose, contrary to the assertion, that $w^*(x) \neq w(x)$ for some $x \in (\zeta, X]$. Let x_1 be the (first) point at which the two functions depart from each other. At the point of departure, by Lemma 8, $\mathbf{w}^*$ must be r-linear over some interval of the form $[x_1, x_2)$. Pick a family that inhabits $x \in (x_1, x_2)$ in steady state. Then, under $\mathbf{w}^*$, a family at occupation x has limit wealth that is at least $w^*(x)$, which strictly exceeds $w^*(x_1)$.

Now observe that conditions (a), (b) and (c) of Lemma 11 are satisfied, while the conclusion of that lemma fails for initial wealth $y = w^*(x)$. Therefore condition (d) must fail for y, and a family with occupation x must have limit wealth that strictly exceeds $w^*(x_2)$. However, $w^*(x') - w^*(x)$ must *strictly exceed* $(1 + r)(x' - x)$ for every $x' > x_2$.[32] But this implies that the family must make occupational bequests

[32] For no $x' > x_2$ is the wage function w^* r-linear on $[x_1, x']$.

that strictly exceed x_2 (after a finite number of periods), which contradicts the fact that it occupies x.

We have therefore shown that no training cost $x \in [x_1, x_2)$ can be occupied in steady state, which contradicts the richness conditions [R] and [E].

PROOF OF PROPOSITION 1:

Observation 2 establishes that there is a continuous equivalent representation to the wage function in every steady state. Lemmas 5 and 6 together show that the wage function must be r-linear over some initial stretch of training costs; to be exact, over all x such that $w(x) \leq \Omega(w, r)$. Lemma 7 proves that over all remaining values of x (if any), $\mathbf{w}$ must follow (7) *provided* no financial bequests are made in this zone. We subsequently show that indeed, no financial bequests will be made in this zone, which permits us to assert (Lemma 12) that a steady state wage function must indeed be two-phase.

PROOF OF PROPOSITION 2:

Suppose that the assertion is false. Then there exists x_1 and x_2 with $x_2 > x_1 \geq z$, such that

(a) $\rho(x_2) \leq \rho(x_1)$, and

(b) the line from $(0, w)$ to $(x_1, w(x_1))$ is a local tangent to the wage function at the latter point; i.e., if we define $\hat{r} \equiv w'(x_1)$, then

(b.1) $w\,(x_1) = w + (1 + \hat{r})x_1$; while

(b.2) $w\,(x_2) \leq w + (1 + \hat{r})x_2$.

Define $y_i \equiv w\,(x_i)$ for $i = 1, 2$. Condition (b.1) tells us that

$$y_1 - x_1 = \frac{\hat{r}y_1 + w}{1 + \hat{r}},$$

and using this in (7), we must conclude that

$$\frac{U'\left(\dfrac{\hat{r}y_1 + w}{1 + \hat{r}}\right)}{\theta U'\left(\dfrac{\hat{r}y_1 + w}{1 + \hat{r}}\right) + (1 - \theta)W'(y_1)} = \delta w'(x_1) = \delta(1 + \hat{r}), \tag{27}$$

so that y_1 is a limit wealth in the benchmark model under the parameters $(w, \hat{r})$.

On the other hand, condition (b.2) tells us that

$$y_2 - x_2 \leq \frac{\hat{r}y_2 + w}{1 + \hat{r}}, \tag{28}$$

while (7) informs us that

$$\frac{U'(y_2 - x_2)}{\theta U'(y_2 - x_2) + (1 - \theta)W'(y_2)} = \delta w'(x_2) \leq \delta(1 + \hat{r}). \tag{29}$$

Combining (28) and (29) and using the concavity of U, we see that

$$\frac{U'\left(\dfrac{\hat{r}y_2 + w}{1 + \hat{r}}\right)}{\theta U'\left(\dfrac{\hat{r}y_2 + w}{1 + \hat{r}}\right) + (1 - \theta)W'(y_2)} \leq \delta(1 + \hat{r}).$$

By Lemma 4, we must conclude that there is another limit wealth in the benchmark model (possibly infinity) that weakly exceeds y_2. Along with UN, this contradicts our earlier assertion (following (27) that y_1 is a limit wealth in the same benchmark model.

PROOF OF OBSERVATION 3:

Applying (7) to the constant elasticity case with $\theta = 0$, we see that for all $x \geq z$,

$$w'(x) = \frac{1}{\delta}\left[\frac{w(x)}{w(x) - x}\right]^{\sigma}.$$

Differentiation of this equality shows us that

$$w''(x) = \sigma\left[\frac{w(x)}{w(x) - x}\right]^{\sigma-1} \frac{w(x) - xw'(x)}{[w(x) - x]^2},$$

so that $w''(x)$ is continuous and has precisely the same sign as $w(x)/x - w'(x)$. Notice that

$$\frac{w(x)}{x} > w'(x)$$

at $x = z$. So $w'(x)$ increases just to the right of z, while—using (30)—$w(x)/x$ monotonically falls. But it must be the case throughout that $w(x)/x$ continues to exceed $w'(x)$, otherwise the very changes described in this paragraph cannot occur to begin with. Therefore, $w'(x)$ rises throughout, establishing strict convexity to the right of z.

However, w' cannot go to ∞, as another perusal of (30) will readily reveal. Indeed, w' converges to a finite limit, which is computed by setting both $w'(x)$ and $w(x)/x$ equal to the same value in (30).

PROOF OF PROPOSITION 3:

The necessity of [P] is obvious, given the characterization in Proposition 1, so we establish sufficiency.

Index each two-phase wage function $\mathbf{w}$ by its starting wage w, and for notational ease define $c^*(w) \equiv c(r, \mathbf{w})$. Condition P assures us that $c^*(w_1) < 1$ for some $w_1 \geq 0$. We claim that $c^*(w_2) > 1$ for some $w_2 > 0$. Suppose not; then $c^*(w) \leq 1$ for all w. Send $w \uparrow \infty$, then to maintain $c^*(w) \leq 1$ it must be that the associated cost-minimizing $\boldsymbol{\lambda}$—call it $\boldsymbol{\lambda}(w)$—converges weakly to 0. Fix any $k > 0$. Then, for w large enough, [E] implies that

$$\frac{f(k, \boldsymbol{\lambda}(w))}{k} < r.$$

For all such w, concavity of f in k tells us that the associated cost-minimizing capital input $k(w)$ must be bounded. But now the continuity of f (together with [E]) tells us that output goes to zero as $w \to \infty$, which contradicts unit cost minimization. This proves the claim.

Because c^* is continuous,[33] there exists $w^* > 0$ between w_1 and w_2 such that $c^*(w^*) = 1$.

We prove that the two-phase wage function $\mathbf{w}$ emanating from w^* satisfies all the conditions for a steady-state wage function. To this end, we specify a steady-state wealth and bequest distribution, and occupational choice.

First, let $\boldsymbol{\lambda}^*$ be the input mix associated with the supporting wage function $\mathbf{w}$. Arrange the population over occupations according to $\boldsymbol{\lambda}^*$. Let $z = z(w^*)$.

If a family i is assigned to occupation h with $x(h) \leq z$, set that family's wealth equal to $\Omega(w^*, r)$, its educational bequest equal to $x(h)$, and its financial bequest equal to $[\Omega(w^*, r) - x(h)]/(1 + r)$. By [F], $\Omega(w^*, r)$ is well-defined and finite.

Otherwise, if occupational assignment h has $x(h) > z$, set that family's wealth equal to $w^*(x(h))$, its educational bequest equal to $x(h)$, and its financial bequest equal to 0.

Now, invoke Lemma 10, which applies to all two-phase functions. The results of that lemma must apply a fortiori to any two-phase *wage* function, which is obtained by truncating the corresponding two-phase function at X. (Use Lemma 9.) It follows that families find their steady-state occupational and wealth assignments optimal.

PROOF OF COROLLARY 1:

Let $c^*(\mathbf{w}, r)$ denote the unit cost function under the function f^*. Then it is easy to see that under the production function f,

$$c(\mathbf{w}, r) = \frac{c^*(\mathbf{w}, r)}{A}.$$

The assertion in the proposition is now a trivial consequence of this equality.

[33] This is a consequence of the maximum theorem and the assumption that production is continuous in the weak topology over occupational distributions on $\mathcal{H}$.

PROOF OF PROPOSITION 4:

Suppose, on the contrary, that there are two steady -state wage functions (modulo equivalent representations). Denote these by $\mathbf{w}$ and $\mathbf{w}^*$, and observe from Proposition 1 that each of them must come from the two-phase family. Let w and w^* be the two baseline wages, with associated values of z and z^* as given by (19). Without loss of generality suppose that $z \leq z^*$.

These two wage functions *must* cross, otherwise the profit-maximization (support) condition cannot be satisfied for both. Beyond z^* both wage functions satisfy the same differential equation (7), which rules out a crossing in this region. The functions also cannot cross below z since both wage functions are r-linear in this region. So, $z < z^*$, and the functions cross at some $x_1 \in (z, z^*)$.

Define $x_2 \equiv z^*$ and notice that $\mathbf{w}$ and $\mathbf{w}^*$ satisfy conditions (a)–(c) of Lemma 11. However, the assertion in that lemma is violated for every $y \in [w^*(x_1), w^*(x_2))$, because every family under $\mathbf{w}^*$ accumulates a limit wealth of at least $\Omega(w^*, r) \geq w^*(x_2) = w^*(z^*) > w^*(x_1)$. Therefore condition (d) in that lemma is violated for all such y, and each family with initial $y \in [w^*(x_1), w^*(x_2))$ must accumulate *strictly* more than $w^*(x_2)$ in the limit. But this means that no training cost below x_2 can be occupied under the steady state $\mathbf{w}^*$,[34] which contradicts the richness conditions [R] and [E].

PROOF OF OBSERVATION 4:

Condition P tells us that for some two-phase wage function $\mathbf{w}$, $c(r, \mathbf{w}) < 1$. Define a new wage function $\hat{\mathbf{w}}$ that is r-linear from the same baseline wage as that for $\mathbf{w}$; then by Lemma 9, $\hat{w}(x) \leq w(x)$ for all x. It follows that $c(r, \hat{\mathbf{w}}) < 1$ as well. The existence of the required baseline wage a now follows from the same argument used in the proof of Proposition 3.

PROOF OF PROPOSITION 6:

First assume that (11) fails. Using the same technique as in the proof of Proposition 3, it is easy to see that the r-linear wage function starting at a is an equal steady state. Given Proposition 4, this completes the proof.

Indeed, by the characterization result of Proposition 1, an equal steady-state wage function *must* be the wage function that starts at a. If, therefore, (11) holds, that proposition assures us that an equal steady state cannot exist.

REFERENCES

Bajona, Claustre, and Timothy J. Kehoe. 2006. "Trade, Growth, and Convergence in a Dynamic Heckscher-Ohlin Model." National Bureau of Economic Research Working Paper 12567.

Bandyopadhyay, Debasis. 1997. "Distribution of Human Capital and Economic Growth." University of Auckland Economics Department Working Paper 173.

Banerjee, Abhijit V., and Esther Duflo. 2005. "Growth Theory through the Lens of Development Economics." In *Handbook of Economic Growth*, Vol. 1A, Handbooks in Economics 22, ed. Philippe Aghion and Steven N. Durlauf, 474–553. Amsterdam: Elsevier.

[34] Because $\mathbf{W}^*$ is *not* r-linear beyond $z^* = x_2$, the very fact that families enter this zone means that steady-state occupational bequests cannot lie below x_2.

Banerjee, Abhijit V., and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy*, 101(2): 274–98.

Becker, Gary S., and Nigel Tomes. 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy*, 87(6): 1153–89.

Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*, 4(3): S1–39.

Binelli, Chiara. 2008. "Returns to Education and Increasing Wage Inequality in Latin America." http://www.nip-lac.org/uploads/Chiara_Binelli.pdf.

Bowles, Samuel, and Herbert Gintis. 2001. "The Inheritance of Economic Status: Education, Class and Genetics." In *International Encyclopedia of the Social and Behavioral Sciences*, ed. Neil J. Smelser and Paul B. Baltes, 4132–41. Oxford, UK: Elsevier. http://www.umass.edu/preferen/gintis/feldman.pdf.

Fields, Gary S. 2004. "A Guide to Multisector Labor Market Models." Unpublished.

Freeman, Scott. 1996. "Equilibrium Income Inequality among Identical Agents." *Journal of Political Economy*, 104(5): 1047–64.

Fuster, Luisa. 2000. "Capital Accumulation in an Economy with Dynasties and Uncertain Lifetimes." *Review of Economic Dynamics*, 3(4): 650–74.

Galor, Oded, and Joseph Zeira. 1993. "Income Distribution and Macroeconomics." *Review of Economic Studies*, 60(1): 35–52.

Ghatak, Maitreesh, and Neville Nien-Huei Jiang. 2002. "A Simple Model of Inequality, Occupational Choice, and Development." *Journal of Development Economics*, 69(1): 205–26.

Gokhale, Jagadeesh, Laurence J. Kotlikoff, James Sefton, and Martin Weale. 2001. "Simulating the Transmission of Wealth Inequality via Bequests." *Journal of Public Economics*, 79(1): 93–128.

Katz, Lawrence F., and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963–1987: Supply and Demand Factors." *Quarterly Journal of Economics*, 107(1): 35–78.

Khan, M. Ali, and Tapan Mitra. 1986. "On the Existence of a Stationary Optimal Stock for a Multi-Sector Economy: A Primal Approach." *Journal of Economic Theory*, 40(2): 319–28.

Ljungqvist, Lars. 1993. "Economic Underdevelopment: The Case of Missing Market for Human Capital." *Journal of Development Economics*, 40(2): 219–39.

Lloyd-Ellis, Huw, and Dan Bernhardt. 2000. "Enterprise, Inequality and Economic Development." *Review of Economic Studies*, 67(1): 147–68.

Loury, Glenn C. 1981. "Intergenerational Transfers and the Distribution of Earnings." *Econometrica*, 49(4): 843–67.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, 22(1): 3–42.

Matsuyama, Kiminori. 2000. "Endogenous Inequality." *Review of Economic Studies*, 67(4): 743–59.

Matsuyama, Kiminori. 2004. "Financial Market Globalization, Symmetry-Breaking and Endogenous Inequality of Nations." *Econometrica*, 72(3): 853–84.

Matsuyama, Kiminori. 2006. "The 2005 Lawrence R. Klein Lecture: Emergent Class Structure." *International Economic Review*, 47(2): 327–60.

Mookherjee, Dilip, and Debraj Ray. 2002. "Is Equality Stable?" *American Economic Review*, 92(2): 253–59.

Mookherjee, Dilip, and Debraj Ray. 2003. "Persistent Inequality." *Review of Economic Studies*, 70(2): 369–93.

Mulligan, Casey B. 1997. *Parental Priorities and Economic Inequality*. Chicago, IL: University of Chicago Press.

Piketty, Thomas. 1997. "The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing." *Review of Economic Studies*, 64(2): 173–89.

Ray, Debraj. 1990. "Income Distribution and Macroeconomic Behavior." http://www.econ.nyu.edu/user/debraj/DevEcon/Notes/incdist.pdf.

Ray, Debraj. 2006. "On the Dynamics of Inequality." *Economic Theory*, 29(2): 291–306.

Romer, Paul M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, 94(5): 1002–37.

This article has been cited by:

1. Kiminori Matsuyama. 2010. Imperfect Credit Markets, Household Wealth Distribution, and Development. *Annual Review of Economics* **3**:1, 110301095653089. [CrossRef]