Knowledge Diffusion, Growth, and Inequality

Erzo G.J. Luttmer*
University of Minnesota

April 16, 2014

Abstract

Consider an economy in which knowledge transfer and individual experimentation are two ways of accumulating human capital. Knowledge transfer requires the simultaneous inputs of time on the part of both the seller (teacher) and the buyer (student) of knowledge. It happens after an uncertain amount of time, and then the student becomes like the teacher. The optimal assignment is for fast learners to study with the teachers who have the most productive knowledge. For example, consider managers who can produce consumption using fixed and variable labor. Only sufficiently productive managers will be active—employ workers and produce consumption. If production enables managers to learn more quickly, then the most productive active managers teach the least productive active managers, while managers with intermediate productivities teach inactive managers to become active again. Because managers also experiment on their own, the overall distribution of managerial productivities steadily improves over time. Although in general the long-run growth rate depends on the initial distribution of productivity, a unique long-run growth rate is implied by initial distributions with bounded support. Higher rates of experimentation lead to more inequality among managers and higher learning rates reduce inequality. A reduction in the amount of fixed labor needed for managers to remain active lowers the factor share of workers and leads to more earnings inequality.

*Work in progress; discussion of related literature, quantitative results, and conclusions incomplete. I thank seminar participants at the Toulouse School of Economics, Stanford University, and the Philadelphia Fed, as well as Fernando Alvarez, Ben Moll and Robert E. Lucas, Jr. for helpful comments. Fernando Alvarez suggested a calculation along the lines of the one reported in Section 2.1.

1. Introduction

Growth and inequality are two sides of the same coin. In any initially homogeneous group of individuals, human capital tends to diverge because not everyone learns at the same rate. Ability must be an important factor accounting for differences in learning, but the role of choice and serendipity is also likely to be significant. Individuals choose and are exposed to different life and career experiences, and some of those experiences provide greater learning opportunities than others. To the extent that such randomness gives rise to different levels of human capital, there are incentives for those who have learned more to teach those who have lagged behind. This paper describes a model of those incentives and characterizes the relation between long-run growth and heterogeneity in human capital.

Consider an economy in which consumption is produced by teams of managers, production workers, and workers who cover overhead. Management requires useful knowledge. Managers with more of it lead larger teams of workers that produce more consumption and generate higher profits for themselves. Learning-by-doing and an ever changing environment cause the useful knowledge of a manager in charge of a team of workers to evolve stochastically over time. Individuals begin life without useful knowledge and differ in their ability to learn. Relatively slow learners choose to become workers and fast learners try to become managers. Initially, these fast learners are "inactive managers" who supply labor just like workers. At the same time they try to learn from "active managers" who are actually managing teams of workers. These active managers cannot supply labor on the side, but they do have time to either teach other managers, or to try to learn themselves from more knowledgeable active managers. The context provided by their status as active managers allows them to learn more quickly than inactive managers. Knowledge transfer is one-on-one between two managers, takes an uncertain amount of time, and a manager who learns becomes like the manager from whom he or she learns. Active managers whose knowledge lags behind too much quit to earn wages again as inactive managers, and possibly learn something new.

A competitive market determines the assignment of students to teachers—inactive managers and active managers who want to learn and more productive active managers who can transfer their useful knowledge at a price. More knowledgeable managers can charge more tuition than less knowledgeable managers, and active managers with little useful knowledge choose to pay tuition to learn from others. Because their incumbency does allow these active managers to learn more quickly, they pay the higher tuition to learn from the most productive active managers. An intermediate range of active managers helps inactive managers become active managers. If active managers did not also learn on their own at stochastically varying rates, all managers would eventually learn what the most knowledgeable manager in the initial population already knew from the start. But the fact that active managers also accumulate knowledge independently results in a distribution of useful knowledge across managers that keeps shifting to the right.

This economy has a continuum of balanced growth paths. Along any balanced growth path, the right tail of the distribution of managerial human capital is Pareto, and the thicker the right tail of this Pareto distribution, the faster the economy grows. More heterogeneity means larger knowledge gaps between students and teachers, and this in turn implies more rapid growth. But there is a minimal growth rate that is an increasing function of the variance of the shocks that affect the individual knowledge accumulation of active managers, and increasing in the rate at which active managers (because of ability and occupational status, the most rapid learners in this economy) can learn from other active managers. In a stripped-down version of the economy, it can be shown that this minimal growth rate will be the long-run growth rate of the economy if the initial distribution of managerial productivities has bounded support.

An interesting question that arises is: what happens to workers when a technological improvement lowers the overhead labor needed to support active managers? The constant return to scale technology for producing consumption is taken to be Cobb-Douglas in managerial human capital and variable labor inputs. Because of the presence of overhead labor, the overall factor share of labor exceeds the usual Cobb-Douglas factor share parameter, and it does so by more the higher is the ratio of aggregate fixed labor over aggregate variable labor. The direct effect of a reduction in the overhead labor per active manager is to lower this ratio. But there is also an indirect effect driven by an extensive margin that induces more managers to become active, which tends to increase aggregate fixed labor. The direct effect turns out to dominate the indirect effect, lowering the ratio of aggregate fixed labor over aggregate variable labor, and thus lowering the factor share of labor. Technical progress that lowers the need for overhead labor supplied by workers hurts workers relative to managers. Quantitative explorations of the model suggest this can account for the labor share declines documented by Karabarbounis and Neiman [2013].

The economy sketched in this paper can be contrasted with an economy in which knowledge is not transferred from teachers to students, but by random imitation, as in Jovanovic and Rob [1989], Luttmer [2007], Alvarez, Buera and Lucas [2008], Lucas [2009], Lucas and Moll [2013], and Perla and Tonetti [2012]. In these economies, producers randomly meet other producers and copy the other when the other is more knowledgeable. In Luttmer [2007] it is only potential entrants who can copy, while Lucas and Moll [2013] and Perla and Tonetti [2012] focus economies without entry in which everyone can copy everyone else. This also causes the distribution of useful knowledge to move to the right, and if the initial distribution has a sufficiently thick right tail, then this process of copying from others can continue forever, resulting in long-run growth. This builds on Kortum's [1997] use of thick-tailed Pareto and Fréchet distributions to generate constant long-run growth rates with a growing population of researchers. As already emphasized in Luttmer [2012], adding small individual shocks to productivity will ensure that long-run growth happens even if the initial distribution has bounded support. Moreover, only one of the many possible long-run growth rates and only one of the many associated stationary distributions is accessible from such an initial distribution, providing a definite prediction for what determines long-run growth that is independent of the details of the initial distribution of useful knowledge. The same happens here. The partial differential equations that govern the distribution of useful knowledge over time are reaction-diffusion equations in both economies, with very similar "reaction" terms: in the shape of a tent when learning is one-on-one and in the shape of a parabola (that also governs the familiar logistic differential equation of epidemiology) in the case of random imitation.[1]

There is, however, one crucial difference. In the random-imitation version of the economy described in this paper, as in the economy with entrants copying incumbents in Luttmer [2007], the long-run growth rate depends on how many in the economy are trying to copy and participate in the random meeting process that allows useful knowledge to spread. The more the better. There are no limitations on how many can copy one (particularly productive) manager, and it is only the fact that meetings are random that limits the pace at which useful ideas can spread. Here instead, there are bandwidth limitations: a very productive manager cannot broadcast the knowledge that makes the manager productive (with everyone listening and comprehending immediately) but must teach others one-on-one. The speed with which knowledge can diffuse is therefore inherently limited, and having more trying to learn only lowers the quality of the marginal

[1] In independent work, Staley [2011] has already added Brownian noise to the productivities of individual producers in the Lucas [2009] random imitation economy and shown how the resulting reaction-diffusion equations with a parabola reaction term can be used to determine the long-run growth rate and the productivity distribution.

teacher but does not affect how many catch up with managers in the right tail.

In closely related work, König, Lorenz, and Zilibotti [2012] consider an economy in which producers face a trade-off not between producing and copying, as in Lucas and Moll [2012], but between innovating and copying. Firms that innovate succeed following random delays that are independent across firms. Because of this some firms will always be ahead of others. Growth never comes to a halt, and will be faster than it would be if there imitation were impossible.

The economy in this paper is one in which growth is driven by human capital accumulation, in the presence of a fixed factor. As in Rebelo [1991], this is made possible by assuming the technology is Cobb-Douglas. But balanced growth does not crucially rely on this. The Cobb-Douglas technology implies that managerial human capital can be measured in efficiency units. If the technology is not Cobb-Douglas, this is no longer the case and there is no single variable that can be interpreted as the aggregate stock of human capital. But the distribution of managerial skills can be de-trended to generate a stationary distribution, and this enough for balanced growth.

1.1 Outline

The rest of this paper is organized as follows. Section 2 highlights the problematic assumptions needed to build a theory of long-run growth based on knowledge transfer alone, without some amount of individual experimentation. Section 3 describes the economy with students and teachers and derives differential equations for the value function and the stationary distribution of human capital. Section 4 explicitly solves these equations and constructs balanced growth paths, and Section 5 provides a detailed characterization of the equilibrium when active and inactive managers can learn at the same rate. Section 6 shows why initial distributions with a bounded support imply a unique long-run growth rate. Some quantitative implications are considered in Section 7 and Section 8 concludes.

2. Two Motivating Examples and a Solution

Consider a large population of managers with heterogeneous attributes $z \in (-\infty, \infty)$. At time t, the cross-sectional distribution of these attributes is $F(t, z)$. High-z managers are more productive, creating incentives for low-z managers to learn from high-z managers.

As in Alvarez, Buera and Lucas [2008] and Lucas [2009], suppose managers randomly observe other managers at a rate β. When a manager observes another manager with a higher z, the manager with the lower z can copy or adopt the higher z at no cost. Then $F(t,z)$ evolves according to

$$\mathrm{D}_t F(t,z) = -\beta F(t,z)[1 - F(t,z)]. \tag{1}$$

Per unit of time, $\beta F(t,z)$ managers with attributes in $(-\infty, z]$ observe other managers. A fraction $1 - F(t,z)$ observe managers in (z,∞) and adopt their attributes. For every z, (1) is a logistic differential equation in t. Solving these differential equations gives

$$F(t,z) = \frac{1}{1 + \left(\frac{1}{F(0,z)} - 1\right) e^{\beta t}} \tag{2}$$

as long as $F(0,z) \in (0,1]$, and $F(t,z) = 0$ for all t otherwise. Observe that $F(t,z)$ is strictly decreasing in t whenever $F(0,z)$ is positive. If the support of the initial distribution does not have a finite upper bound then $F(t,z)$ keeps shifting to the right. For certain special initial distributions, this happens at a constant rate. Consider distributions of the form $F(0,z) = 1/(1+T(z))$. Then

$$\begin{aligned}
T(z) &= e^{-\zeta z}, \quad z \in (-\infty,\infty) \Rightarrow F(t,z) = F(0, z - (\beta/\zeta)t), \\
T(z) &= z^{-\zeta}, \quad z \in (0,\infty) \quad\; \Rightarrow F(t,z) = F(0, ze^{-(\beta/\zeta)t}), \\
T(z) &= (-z)^{\zeta}, \; z \in (-\infty, 0] \quad \Rightarrow F(t,z) = F(0, ze^{(\beta/\zeta)t}),
\end{aligned}$$

for any positive ζ. Clearly, the assumed initial distribution affects the long-run rate at which $F(t,z)$ moves to the right. It is not difficult to show that exponential growth at the rate β/ζ happens in the long run precisely when $[1 - F(0,z)]z^{\zeta}$ converges to a positive constant as z becomes large.

Notice that these invariant distributions are related to the extreme value distributions $\exp(-T(z))$. This connection can be traced back to the fact that $F(t,z)$ is also the distribution of the maximum of N_t independent draws from $F(0,z)$ when N_t follows a pure birth process with birth rate βN_t. The distribution of N_t implied is the geometric distribution $\{e^{-\beta t}(1 - e^{-\beta t})^{n-1}\}_{n=1}^{\infty}$ and a direct computation of the mean of $\{[F(0,z)]^n\}_{n=1}^{\infty}$ gives (2). Copying better draws from a continuously updated distribution $F(t,z)$ at a constant average rate acts like sampling repeatedly from the same initial $F(0,z)$ and taking the maximum, at an average rate that is exponential in time.

Instead of having managers observe other managers at random, suppose that managers below the median are assigned to managers above the median at any point in time. In every assignment, the manager below the median learns to adopt the attribute of the manager above the median at an average rate β. Let x_t be the median of $F(t,z)$. Then $\mathrm{D}_t F(t,z) = -\beta F(t,z)$ for $z < x_t$ because managers in $(-\infty, z)$ learn at the rate β, and $\mathrm{D}_t[1-F(t,z)] = \beta[1-F(t,z)]$ for $z > x_t$ because managers in (z,∞) replicate themselves at the rate β. Another way to write this is

$$\mathrm{D}_t F(t,z) = -\beta \min\{F(t,z), 1-F(t,z)\}. \tag{3}$$

With this alternative mechanism for transferring managerial attributes, the quadratic in $F(t,z)$ on the right-hand side of (1) is replaced by a tent. It is easy to see that the time-t median is determined by $\frac{1}{2} = e^{\beta t}[1-F(0,x_t)]$. This immediately establishes a connection between the right tail of the initial distribution and the rate at which the distribution $F(t,z)$ moves to the right. Solving (3) gives

$$F(t,z) = \begin{cases} e^{-\beta t} F(0,z), & z \in (-\infty, x_0), \\ \frac{1}{2}\frac{1/2}{e^{\beta t}[1-F(0,z)]}, & z \in (x_0, x_t), \\ 1 - e^{\beta t}[1-F(0,z)], & z \in (x_t, \infty). \end{cases}$$

As in the case of random copying, one can use this to show that there are three types of invariant distributions, with associated linear, exponential, and negative exponential trends.

In both these examples, there is an initial distribution of the managerial attribute z, and how this distribution evolves over time depends critically on the shape of this initial distribution. An initial distribution with an exponential right tail implies linear growth and an initial distribution that follows a power law in the right tail implies exponential growth. This dependence on initial conditions is a natural outcome of the assumption, in both economies, that improvements of managerial attributes in the population can only arise from learning from other managers. In what follows, some amount of experimentation by individual managers is added to (1) and (3). The particular way in which experimentation is added —the increments in z that arise from experimentation are *i.i.d.* over time and across managers— results in a definite prediction for the rate at which the distribution of managerial attributes improves over time.

2.1 Two Agents Imitating at Random Times

To illustrate, consider an economy with only two managers with attributes $\{z_{n,t}\}_{n=1}^{2}$. Most of the time, these attributes evolve as independent Brownian motions with variance $\sigma^2 t$. But at random times $\{\tau_j\}_{j=1}^{\infty}$, the manager with the lower productivity has the opportunity to learn from the other manager and become as productive. Suppose the inter-arrival times $\tau_{j+1} - \tau_j$ are independent and identically distributed with an exponential distribution that has a mean $1/\beta$. At time τ_j, after the less productive manager has learned from the more productive manager, the two managers will have the common attribute $y_j = \max_n\{z_{n,\tau_j -}\}$. At the next opportunity to learn, this yields $y_{j+1} = y_j + \sigma \max_n\{B_{n,\tau_{j+1}} - B_{n,\tau_j}\}$, where $B_{n,t}$ is the standard Brownian motion that drives the productivity of manager $n \in \{1, 2\}$. Using integration-by-parts, it is not difficult to verify that the maximum of two independent standard Brownian motions has a mean equal to $\sqrt{t/\pi}$ at time t. The density of $\tau_{j+1} - \tau_j$ is exponential with mean $1/\beta$. Therefore

$$\mathrm{E}\left[y_{j+1} - y_j | y_j\right] = \sigma \int_0^\infty \sqrt{t/\pi}\beta e^{-\beta t}\mathrm{d}t = \frac{\sigma}{\sqrt{\beta\pi}} \int_0^\infty \sqrt{a} e^{-a}\mathrm{d}a = \frac{1}{2}\frac{\sigma}{\sqrt{\beta}}.$$

The last equality uses $\Gamma(3/2) = \frac{1}{2}\sqrt{\pi}$. The average time that elapses between τ_{j+1} and τ_j is $1/\beta$, and so per unit of time this increment equals $\frac{1}{2}\sigma\sqrt{\beta}$. One can also show that the growth rates $\{(Y_{j+1} - Y_j)/(\tau_{j+1} - \tau_j)\}_{j=1}^\infty$ have a mean equal to $\sigma\sqrt{\beta}$, which exceeds $\frac{1}{2}\sigma\sqrt{\beta}$ because of Jensen's inequality. Below, it will be argued that the trend in a very large population is in fact $\sigma\sqrt{2\beta} > \frac{1}{2}\sigma\sqrt{\beta}$.[2] But the main point of this simple example is that, because of the random shocks that move around the individual productivities of managers in between times they can copy from each other, the process of mutual learning never stops, as it would in a world without individual experimentation. The rest of this paper shows how one can turn this into a theory of long run balanced growth that is not based on an assumption that there is no bound on what is already known at some initial date.

3. A Knowledge Economy

There is a unit measure of infinitely-lived households whose preferences over consumption flows C_t are given by

$$\int_0^\infty e^{-\rho t} \ln(C_t)\mathrm{d}t,$$

[2] Staley [2011] makes the important observation that this can be used to account for scale effects (Jones [2005] discusses the literature on scale effects in idea-driven models of growth.) Growth is slow when there are a few agents experimenting and imitating, and accelerates to a finite limit when the population becomes large.

where ρ is a positive subjective discount rate. Households can trade in a complete set of markets and face interest rates r_t. The usual intertemporal Euler condition requires

$$r_t = \rho + \frac{\mathrm{D}C_t}{C_t}.$$

A fraction $1 - A$ of households are workers who supply labor inelastically, normalized at one unit of labor per worker. The other A households have an ability to accumulate the human capital required to manage workers. Some of these potential managers are active managers who actually employ workers. Other managers are inactive managers who do not employ workers but may do so (again) in the future. Active managers cannot supply labor, but inactive managers have one unit of labor that they supply inelastically, just as workers do.

Active managers vary in how skilled they are in their managerial task, and the skills of individual active managers evolve stochastically over time. Managerial skills are tied to being active managers: a manager who stops managing workers loses his or her particular managerial skills. Active managers can only remain active by maintaining a support staff of $\phi \geq 0$ units of labor. The wages inactive managers can earn and the overhead labor cost of being active imply that active managers whose skills have not kept up with the skills of other active managers may find it optimal to become inactive.

3.1 Product Markets

Active managers can hire workers or inactive managers at time-t wages w_t and produce a flow of consumption subject to decreasing returns. The technology is Cobb-Douglas with a labor share parameter α. The revenues earned by type-z manager at time t are

$$v(t, z) = \max_l \left\{ \left(\frac{e^z}{1 - \alpha} \right)^{1-\alpha} \left(\frac{l}{\alpha} \right)^{\alpha} - w_t l \right\} = v_t e^z.$$

Here v_t is the factor price of one unit of managerial services. From the unit cost function $v_t^{1-\alpha} w_t^{\alpha} = 1$, this factor price is given by $v_t = 1/w_t^{\alpha/(1-\alpha)}$. The fact that inactive managers can earn wages w_t represents an opportunity cost that can be viewed as an additional fixed cost associated with being an active manager. After accounting for both types of fixed costs, the profits of a type-z active manager are $v(t, z) - (1 + \phi)w_t$. It will be convenient to write $(1 + \phi)l(t, z)$ for the amount of variable labor that solves the profit maximization problem of a type-z manager,

$$(1 + \phi)l(t, z) = \frac{\alpha}{1 - \alpha} \frac{v_t e^z}{w_t}.$$

So $l(t,z)$ is the amount of variable labor used by a type-z manager relative to the total fixed labor input. The measure of type-z active managers at time t is denoted by $M(t,z)$. Every manager who is not an active manager supplies labor. Aggregating the output of consumption across managers at the wage that clears the labor market gives

$$C_t = \left(\frac{H_t}{1-\alpha}\right)^{1-\alpha}\left(\frac{1-(1+\phi)M_t}{\alpha}\right)^{\alpha}, \tag{4}$$

where H_t and M_t are defined in terms of $M(t,z)$ by

$$H_t = \int e^z M(t,\mathrm{d}z), \quad M_t = M(t,\infty). \tag{5}$$

The total supply of labor is $1-M_t$ and $1-(1+\phi)M_t$ of this supply can be used as variable labor to produce consumption. The number of active managers is $M_t \leq A$ and the remainder $A-M_t$ is the number of inactive managers. The market-clearing wage can be inferred from the fact that the compensation of variable labor is a fraction α of output,

$$w_t(1-(1+\phi)M_t) = \alpha C_t. \tag{6}$$

Together, (4)-(6) and the cost function $v_t^{1-\alpha}w_t^{\alpha} = 1$ determine C_t, H_t, M_t, v_t and w_t in terms of the measure $M(t,z)$ of active managers.

The measure $M(t,z)$ is the state variable for this economy. Managers can stop being active managers in an instant and so $M(t,z)$ can jump down. But it will take time for inactive managers to become active again, and so $M(t,z)$ cannot jump up. For the purposes of determining aggregate output it suffices to know the aggregate stock of managerial human capital H_t, but the technology for accumulating managerial human capital to be described next depends on the entire distribution $M(t,z)$.[3]

3.2 Managerial Learning and Teaching

Active and inactive managers can learn from active managers. The type of an active manager can change because of manager-specific random shocks, or because a manager

[3] Production functions of the type $e^z G(1,l)$, where G exhibits constant returns to scale, are consistent with balanced growth and much of the paper applies. In the Leontief case $v(t,z) = \max\{0, e^z - w_t\}$ and this implies analytically tractable solutions. The fact that one can define an aggregate stock of human capital H_t as in (5) is special because $e^{(1-\alpha)z}G(1,l) = G(e^z, l)$ in the Cobb-Douglas case. But this plays only a minor role and one needs to keep track of the distribution of z in any case, to describe how knowledge is transmitted.

learns from another manager. As long as a manager remains active, the type z_t of a manager evolves according to

$$\mathrm{d}z_t = \mu \mathrm{d}t + \sigma \mathrm{d}B_t + \mathrm{d}J_t,$$

where B_t is a standard Brownian motion, and J_t is a process that jumps when a manager learns something useful from another manager. The Brownian motions evolve independently across managers and the arrival of jumps is independent as well. The drift μ may be interpreted as learning-by-doing. The Brownian increments may be the result of a changing environment that affects the usefulness of what a manager knows how to do. Alternatively, a manager may be in charge of a project and have to make irreversible decisions about how the project is operated.

A manager running a project with a team of workers can also teach another active or inactive manager how to run a similar project. This transfer of knowledge takes time: while a type-z' manager and a type-$z < z'$ manager are matched, the type-z manager learns how to be a type-z' manager randomly at some positive rate γ. Similarly, when a type-z' manager and an inactive manager are matched, the inactive manager learns how to become an active type-z' manager randomly at the positive rate β. Although managers can continue to oversee workers while they are teaching or learning from others, they cannot, at the same time, teach or learn from more than one manager.[4]

The rate at which managers learn does not depend on what they learn. A high z just means that a type-z manager can produce more, not that learning to be like this manager is more difficult. The fact that the supply of managers who can teach high z is limited means that not everyone can learn those high z, and those who do will have to pay. The maintained assumption will be that $\gamma > \beta > 0$. Both types of managers learn at a positive rate, but being active and employing workers provides active managers with a context in which they can learn more quickly.

Managers maximize the present value of their earnings from managing and teaching or learning. Let U_t and $V(t, z) \geq U_t$ be the present values, respectively, for an inactive manager and for an active manager of type z at time t. Inactive managers supply labor and may learn to become active managers, and active managers can choose to become inactive to supply labor themselves. So U_t and $V(t, z) \geq U_t$ will be bounded below by the time-t present value W_t of the current and future wages of workers. The value $V(t, z)$ will turn out to be increasing in the managerial type z. The fixed costs active managers

[4] A natural assumption would be that knowledge transfer interferes with overseeing workers as well, as it does in the random imitation environment of Lucas and Moll [2012].

must pay to remain active will induce sufficiently unproductive managers to give up the advantage of incumbency in learning implied by $\gamma > \beta$ and become inactive again. To replace these managers, active managers will have to teach inactive managers.

3.2.1 The Tuition Schedule

There are competitive markets for the teaching time of the various types of managers. Students become like their teachers, and so the tuition charged by an active manager will depend on the value V of this manager. Write $T_t(V) \geq 0$ for the flow tuition of an active manager whose value at time t is $V \geq 0$. Managers earn or pay tuition, depending on whether they choose to spend their time teaching or learning.

Given a tuition schedule $T_t(V)$, define

$$S_t(\lambda) = \sup_{V \geq 0} \left\{ \lambda V - T_t(V) \right\}, \tag{7}$$

for any non-negative learning rate λ. This is the convex conjugate of the tuition schedule $T_t(\cdot)$. It represents the flow value of learning for any manager who is paying the tuition of some optimally selected active manager, and who is learning to become like that manager at the rate $\lambda \in \{\beta, \gamma\}$. Observe that $S_t(\lambda)$ is a convex function of λ, and $T_t(V) \geq 0$ ensures that $S_t(\lambda)/\lambda$ is (weakly) increasing in λ. In particular, the assumption $\gamma > \beta$ implies $S_t(\gamma)/\gamma \geq S_t(\beta)/\beta$. Fast students can gain more than slow students, and the gain is more than proportional in the learning rate λ. An example of the type of tuition schedule that can arise in equilibrium is shown in Figure 1.

To determine who teaches whom and begin to characterize equilibrium tuition schedules $T_t(V)$, observe that $S_t(\gamma) \geq \gamma V - T_t(V)$, and that this inequality has to hold with equality for all active managers who teach other active managers. It follows that the tuition schedule must satisfy $T_t(V) = \gamma V - S_t(\gamma)$ for all values V associated with active managers who teach other active managers. Similarly, $T_t(V) = \beta V - S_t(\beta)$ for all V associated with active managers who teach inactive managers. Inactive managers who contemplate studying with active managers who also teach active managers would gain

$$\beta V - T_t(V) = \beta V - [\gamma V - S_t(\gamma)] = S_t(\gamma) - (\gamma - \beta)V.$$

Since $\gamma > \beta$, this is decreasing in V, and so inactive managers can only be studying with active managers who have the lowest value V among active managers teaching other active managers. There will therefore be a threshold managerial value V so that active managers above this threshold teach active managers and active managers below this value teach inactive managers, if they teach at all. These considerations suggest a piecewise linear tuition schedule of the form

$$T_t^*(V) = \max\left\{0, \max_{\lambda \in \{\beta,\gamma\}} \{\lambda V - S_t(\lambda)\}\right\}. \tag{8}$$

For a tuition schedule $T_t(V)$ of this form, (7) implies $S_t(0) = 0$ and the right-hand side of (8) can also be written as a maximum of $\lambda V - S_t(\lambda)$ over $\lambda \in \{0, \beta, \gamma\}$. That is, the value of learning $S_t(\gamma)$ and the tuition schedule $T_t(V)$ are convex conjugates of each other, and one can recover the tuition schedule $T_t(V)$ from the learning values $S_t(\lambda)$ originally defined by it in (7). This will be one possible equilibrium tuition schedule in this economy.

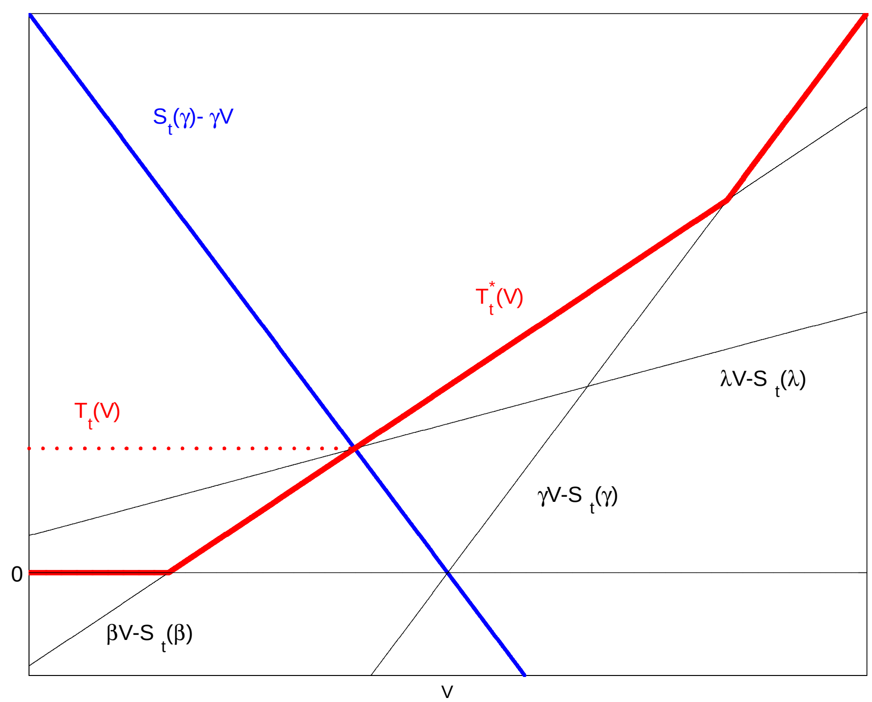

FIGURE 1 The Tuition Schedules $T_t(V)$ and $T_t^*(V)$

Although tuition is positive for high enough V, not all active managers who could earn positive tuition may want to teach. An active manager with value V can also choose to be a student and earn flow payoffs equal to $S_t(\gamma) - \gamma V$. If $S_t(\gamma) - \gamma V > T_t(V)$ then an active manager with value V will not teach but choose to learn from higher-value active managers instead. A range of possible equilibrium tuition schedules arises from the fact that one can raise the tuition schedule (8) all the way to

$$\max\{S_t(\gamma) - \gamma V, T_t^*(V)\} = \max\{\beta V - S_t(\beta), |\gamma V - S_t(\gamma)|\}$$

without changing the incentives of active managers to teach. Recall from (7) that $S_t(\gamma)/\gamma \geq S_t(\beta)/\beta$. If this is a strict inequality, then $\beta V - S_t(\beta)$ is strictly positive when V solves $\gamma V - S_t(\gamma) = 0$. Managers at the margin between teaching active managers and learning from other active managers will then strictly prefer to teach inactive managers. This creates an intermediate range of V where active managers strictly prefer to teach inactive managers for a tuition $T_t^*(V) = T_t(V) = \beta V - S_t(\beta)$. Inactive managers can be assigned to these managers, leaving active managers with values V low enough that $S_t(\gamma) - \gamma V > \beta V - S_t(\beta)$ to study with other active managers. So nobody pays $T_t(V)$ when $T_t(V) > T_t^*(V)$.

It is not possible for an equilibrium tuition schedule to be such that $S_t(\gamma)/\gamma \leq U_t$. Since $V \geq U_t$ for all active managers, this would mean that none of them want to be students, and $S_t(\beta)/\beta \leq S_t(\gamma)/\gamma$ implies the same for inactive managers. Markets cannot clear if nobody wants to be a student. If $S_t(\beta)/\beta \leq U_t < S_t(\gamma)/\gamma$ then some active managers with $V \geq U_t$ close to U_t will want to study, but no inactive managers do. This possibility can only arise under special circumstances: when there are active managers with $V \geq U_t$ near U_t but no active managers in the range of V where $\beta V - S_t(\beta)$ dominates $|\gamma V - S_t(\gamma)|$. If there are some active managers in this range, then it must be that $S_t(\beta)/\beta \geq U_t$. Figure 2 shows the knife-edge scenario $S_t(\beta)/\beta = U_t$ in which inactive managers are indifferent between studying and not studying. This will turn out to be the equilibrium in economies with an abundance of agents who can learn fast of enough to be managers.

There are now two thresholds, a lower threshold defined by $S_t(\gamma) - \gamma V = \beta V - S_t(\beta)$ and an upper threshold defined by $\beta V - S_t(\beta) = \gamma V - S_t(\gamma)$. Low-value active managers study with high-value active managers, and active managers in between the two thresholds teach inactive managers. Given an increasing and continuous value function $V(t, z)$, this implies productivity thresholds $x_t < y_t$ defined by $V(t, x_t) = [S_t(\gamma) + S_t(\beta)]/(\gamma + \beta)$ and $V(t, y_t) = [S_t(\gamma) - S_t(\beta)]/(\gamma - \beta)$, respectively. Active managers with $z \in [b_t, x_t]$ study, active managers in $[x_t, y_t]$ teach inactive managers, and active managers in $[y_t, \infty)$ teach other active managers. Student managers matched with managers of any type $z > y_t$ will have to pay higher tuition in an amount that equals all they can expect to gain from studying with these more productive managers.[5] An inactive manager who successfully learns how to produce becomes a active manager of some type $z \in [x_t, y_t]$

[5] In the case of Ivy league colleges, the payment is literally in the form of tuition. In the case of many organizations with highly skilled employees (consultancies, investment banks, law firms, medical schools) inexperienced professional employees seem to pay their dues in part with long work hours.

and immediately starts teaching other inactive manager. Active managers who fall behind and enter $[b_t, x_t]$ become students and study with active managers in $[y_t, \infty)$. They may improve on their own, learn and jump into $[y_t, \infty)$, or choose to exit because fixed costs are too high.

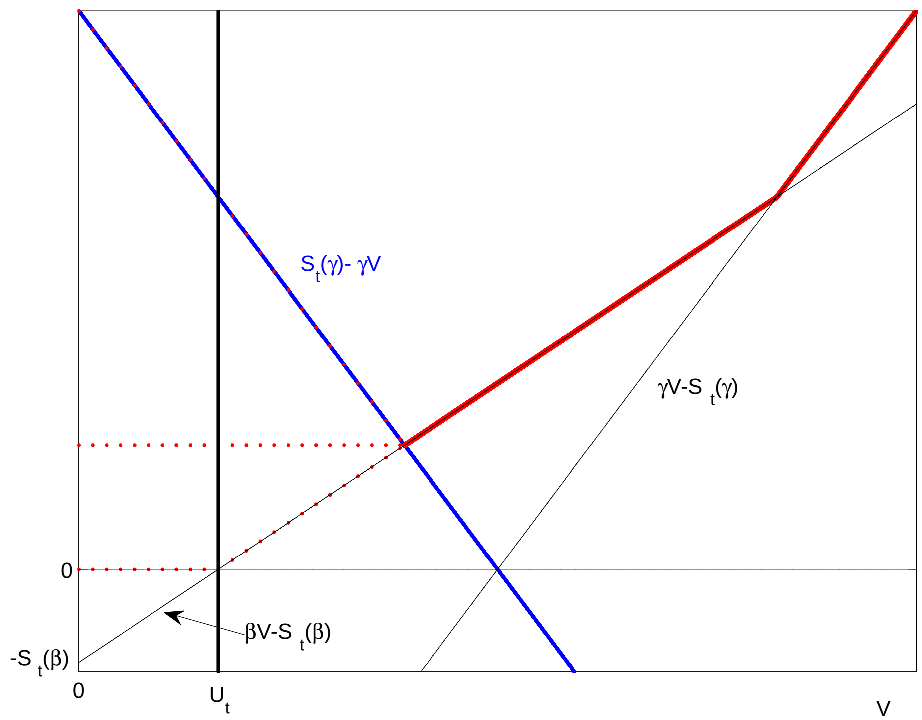

FIGURE 2 The Case $S_t(\beta) - \beta U_t = 0$

The assumption $\gamma > \beta$ gives active managers a learning advantage over inactive managers. If $\beta > \gamma$ instead, then inactive managers would be matched with the most productive active managers and enter the right tail of the productivity distribution, rather than work themselves up, one way or another, as active managers. The assumption that workers cannot learn at all in this economy can be relaxed by assuming that they can learn at sufficiently low rates $\lambda \in (0, \beta)$. The equilibrium tuition schedule $\max\{S_t(\gamma) - \gamma V, T_t^*(V)\}$ is strictly positive, and this implies that $S_t(\lambda) < 0$ when $\lambda > 0$ is small enough, as illustrated in Figure 1. For such low learning rates, studying with managers is too expensive and workers find it optimal instead to supply labor at all times. Obviously, one can allow for heterogeneity in learning abilities and potential managers as well.

In a stationary environment there will have to be some inactive managers who are students, since there will always be some active managers whose productivities lag behind and choose to quit. Inactive managers must therefore weakly prefer to pay tuition and be students. One implication of $\gamma > \beta$ is that the least productive active manager will find it optimal to pay tuition and learn when inactive managers are indifferent between learning and not learning and, a fortiori, if inactive managers all choose to learn.

3.2.2 Bellman Equations

Workers are always employed and earn wages w_t. The present value W_t of their labor earnings must therefore satisfy

$$r_t W_t = w_t + \mathrm{D}W_t. \tag{9}$$

Inactive managers also earn these wages, and they earn income by investing in their human capital. The present value U_t of their earnings has to satisfy

$$r_t U_t = w_t + \max\{0, S_t(\beta) - \beta U_t\} + \mathrm{D}U_t. \tag{10}$$

At a rate β, the value of an inactive manager jumps from U_t to some $V(t,z)$ with $z \in [x_t, y_t]$, for an expected capital gain of $\beta[V(t,z) - U_t]$. In exchange, the inactive manager pays the flow tuition $T_t(V(t,z))$, for a net expected gain of $\beta[V(t,z) - U_t] - T_t(V(t,z)) = S_t(\beta) - \beta U_t$. Inactive managers will not choose to learn from active managers if this is negative.

Given a tuition schedule $T_t(\cdot)$, the Bellman equation for an incumbent manager is

$$r_t V(t,z) = v(t,z) - \phi w_t + \max\{S_t(\gamma) - \gamma V(t,z), T_t(V(t,z))\} + \mathcal{A}V(t,z) \tag{11}$$

as long as $V(t,z) > U_t$, where

$$\mathcal{A}V(t,z) = \mathrm{D}_t V(t,z) + \mu \mathrm{D}_z V(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} V(t,z)$$

is the expected capital gain from changes in the state (t,z) not generated by learning from other managers. The manager can abandon the project and inactive manager again. This will be optimal below some optimally chosen exit threshold b_t that satisfies $V(t,b_t) = U_t$.

The one-on-one teaching technology implies that students have to be matched with active managers who teach. The tuition schedule $T_t(\cdot)$ has to clear this market. Simply knowing that there are some inactive managers and some active managers who are students tells us that the teaching schedule can be taken to be of the form (8). This tuition schedule is parameterized by $S_t(\beta)/\beta < S_t(\gamma)/\gamma$. With some abuse of terminology, $\{S_t(\beta), S_t(\gamma)\}$ will be referred to as the tuition schedule as well. Given $S_t(\beta)$ and $S_t(\gamma)$, the two Bellman equations (10) and (11) (with appropriate boundary conditions) determine the value U_t of an inactive manager and the value function $V(t, z)$ of active managers. Implicit in U_t and $V(t, z)$ are the thresholds b_t, x_t and y_t that determine what active managers do. At any point in time, the values of $S_t(\beta)$ and $S_t(\gamma)$ have to clear the markets for managerial time.

3.2.3 Surplus Value Functions

The fact that inactive managers cannot only supply labor but also learn from active managers implies that $U_t \geq W_t$. The fact that active managers can choose to become inactive at any time implies that $V(t, z) \geq U_t$ for all active managers who choose to remain active. Subtracting (9) from (10) gives a Bellman equation for the present-value surplus $U_t - W_t$ of inactive managers. It is of the same form as (10), except that there is no wage and the gain from learning is $\max\{0, S_t(\beta) - \beta W_t - \beta(U_t - W_t)\}$. Similarly, subtracting (9) from (11) gives a Bellman equation for the present-value surplus $V(t, z) - W_t$ of an active manager of type z. This Bellman equation is of the same form as (11), with a fixed cost equal to $(1 + \phi)w_t$ and a gain from learning equal to $S_t(\gamma) - \gamma W_t - \gamma(V(t, z) - W_t)$.

3.3 Market Clearing and Population Dynamics

Take as given some smooth process for the exit barrier b_t and the net gain $S_t(\beta) - \beta U_t$ from learning for inactive managers. Suppose the measure $M(t, z)$ of active managers at time t has a strictly positive density $m(t, z)$ on (b_t, ∞) and recall that this means that the net gains from learning $S_t(\beta) - \beta U_t$ have to be non-negative in any equilibrium. Write E_t for the measure of inactive managers who study with active managers. Some inactive managers may choose not to study, but only if the net gain from studying is zero. This gives rise to the complementary slackness condition

$$E_t + M_t \leq A, \quad \text{w.e. if } S_t(\beta) > \beta U_t, \tag{12}$$

where

$$M_t = \int_{b_t}^{\infty} m(t, z)\mathrm{d}z. \tag{13}$$

If the net gains from learning are strictly positive, then E_t is pinned down by $E_t = A - M_t$, but not otherwise. At the initial date, $m(0, z)$ may jump down as some active managers instantaneously choose to become inactive. Subsequently, the two thresholds $x_t < y_t$ have to adjust so that the markets for students and teachers clear,

$$\begin{aligned} E_t &= \int_{x_t}^{y_t} m(t,z)\mathrm{d}z, && (14) \\ \int_{b_t}^{x_t} m(t,z)\mathrm{d}z &= \int_{y_t}^{\infty} m(t,z)\mathrm{d}z. && (15) \end{aligned}$$

Because active managers quit when their types cross b_t from above, it must be that the density at b_t satisfies the boundary condition

$$0 = m(t, b_t), \tag{16}$$

(see Cox and Miller [1965]). The flow to the left at any $z \in [b_t, \infty)$ is equal to $-\mu m(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(t,z)$, and hence (16) implies that the flow of active managers who quit is equal to $\frac{1}{2}\sigma^2 \mathrm{D}_z m(t, b_t)$. Markets clear and so E_t inactive managers are matched with active managers, learning at the rate β. The number of active managers therefore follows

$$\mathrm{D}M_t = -\frac{1}{2}\sigma^2 \mathrm{D}_z m(t, b_t) + \beta E_t. \tag{17}$$

Again because markets clear, the density of active managers evolves according to

$$\begin{aligned} \mathrm{D}_t m(t,z) &= -\mu \mathrm{D}_z m(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(t,z) - \gamma m(t,z), \quad z \in (b_t, x_t), \\ \mathrm{D}_t m(t,z) &= -\mu \mathrm{D}_z m(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(t,z) + \beta m(t,z), \quad z \in (x_t, y_t), \\ \mathrm{D}_t m(t,z) &= -\mu \mathrm{D}_z m(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(t,z) + \gamma m(t,z), \quad z \in (y_t, \infty). \end{aligned} \tag{18}$$

The first two terms on the right-hand side of (18) describe the dynamics of $m(t,z)$ that results from the drift and diffusion of z, as in the usual Kolmogorov forward equation. To see the third term in each of these equations, note that active managers in (b_t, x_t) learn and therefore jump out of (b_t, x_t) at the rate γ, active managers in (x_t, y_t) teach inactive managers to become like themselves at the rate β, and active managers in (y_t, ∞) teach other active managers to become like themselves at the rate γ.

The differential equation (18) only determines what happens on the open intervals (b_t, x_t), (x_t, y_t) and (y_t, ∞). As described, it could be that there is a flow of managers who enter or exit at the thresholds x_t and y_t. To rule this out requires two further boundary conditions at x_t and y_t,

$$\begin{aligned} -\mu m(t, x_t-) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(t, x_t-) &= -\mu m(t, x_t+) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(t, x_t+), && (19) \\ -\mu m(t, y_t-) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(t, y_t-) &= -\mu m(t, y_t+) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(t, y_t+). && (20) \end{aligned}$$

This says that the flow to the left $-\mu m(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(t,z)$ of managers across some point z is continuous at the thresholds x_t and y_t. This rules out entry and exit at these thresholds.

The equations (12)-(20) define a differential inclusion that governs the joint evolution of E_t, $m(t,z)$ and the thresholds $y_t > x_t$, given a process for the exit threshold b_t the surplus $S_t(\beta) - \beta U_t$. From one instant to the next, the changes in $m(t,z)$ implied by (18) do not automatically guarantee that the market clearing conditions (14) and (15) continue to be satisfied. In general, this will require an adjustment of both x_t and y_t, and this in turn will then impact subsequent changes in $m(t,z)$ via (18). In equilibrium, the resulting process for the thresholds x_t and y_t will have to match what managers want to do given the tuition schedule $[S_t(\beta), S_t(\gamma)]$.

3.4 Random Imitation Versus Competitive Diffusion

The two motivating examples (1) and (3) show the relation between random imitation and teaching in the absence of individual experimentation or entry and exit. To relate (3) to (18) and facilitate a comparison with random copying by active and inactive managers, consider the right cumulative distribution

$$R(t,z) = \int_z^\infty m(t,u)\mathrm{d}u.$$

Differentiating both sides of this equation with respect to t and using (18) and (19)-(20) to evaluate the integral gives

$$\mathrm{D}_t R(t,z) - \left(-\mu \mathrm{D}_z R(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} R(t,z)\right)$$

$$= \begin{cases} \gamma R(t,y_t) + \beta[R(t,x_t) - R(t,y_t)] - \gamma[R(t,z) - R(t,x_t)], & z \in (b_t, x_t), \\ \gamma R(t,y_t) + \beta[R(t,z) - R(t,y_t)], & z \in (x_t, y_t), \\ \gamma R(t,z), & z \in (y_t, \infty). \end{cases} \tag{21}$$

$$= \min\{\min\{\gamma R(t,z), \gamma R(t,y_t) + \beta[R(t,z) - R(t,y_t)]\}, \gamma R(t,y_t) + \beta[R(t,x_t) - R(t,y_t)] - \gamma[R(t,z) - R(t,x_t)]\}.$$

This generalizes the tent that appears in (3). An example of the right-hand side of (21) is shown in Figure 2 below. Note that the market clearing conditions (14) and (15) are $E_t = R(t,x_t) - R(t,y_t)$ and $R(t,b_t) - R(t,x_t) = R(t,y_t)$, respectively. The right-hand side of (21) starts near 0 when z is large and then increases with $R(t,z)$ as long as $R(t,z)$ is below the measure $R(t,x_t)$ of all managers who teach. When $R(t,z)$ begins to include managers who study with other managers, the right-hand side of (21) decreases with $R(t,z)$, down to the flow βE_t of entering entrepreneurs that is attained when $R(t,z)$ reaches $R(t,b_t)$.

In the alternative scenario of random imitation, suppose the E_t inactive managers observe incumbent managers at the rate β and that active managers randomly observe other active managers at the rate γ. Inactive managers adopt whatever they observe, but active managers in $[b_t, z]$ adopt only when they observe active managers in $[z,\infty)$. The fraction of all managers in $[z,\infty)$ is $R(t,z)/R(t,b_t)$, and so random sampling implies

$$\mathrm{D}_t R(t,z) - \left(-\mu \mathrm{D}_z R(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} R(t,z)\right) = \left(\beta E_t + \gamma[R(t,b_t) - R(t,z)]\right) R(t,z)/R(t,b_t). \tag{22}$$

The right-hand side of (22) is also shown in Figure 2. Just as in the initial motivating examples (1) and (3), the right-hand sides of (22) and (21) are quadratic and tent-like in $R(t,z)$, respectively. Note that both right-hand sides are tied down at 0 when $R(t,z) = 0$ and at βE_t when $R(t,z) = R(t,b_t)$.

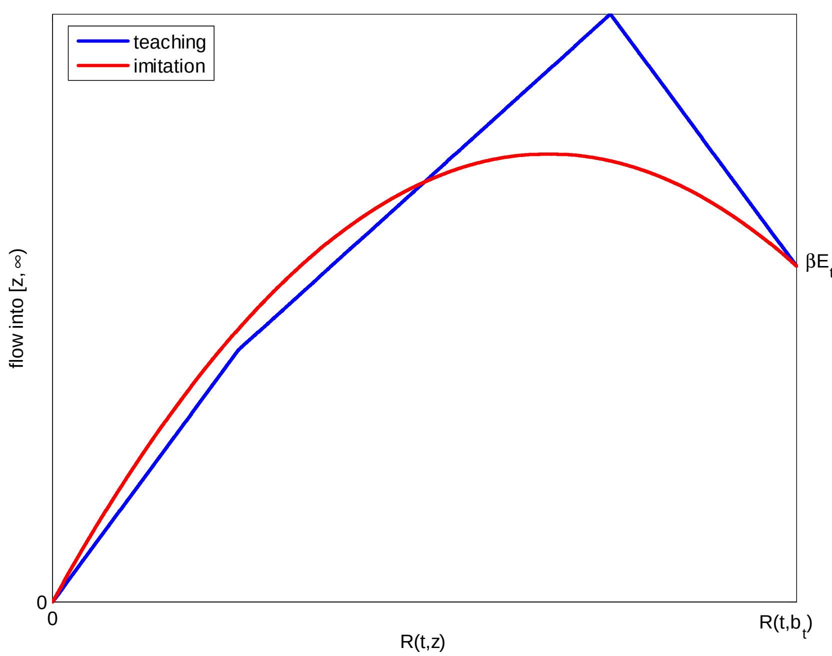

FIGURE 2 The Right-hand Sides of (21) and (22).

Despite the strong similarity between (21) and (22), there is an important difference. For large z, when $R(t,z)$ is small, the right-hand side of (21) behaves like $\gamma R(t,z)$, while in the right-hand side of (22) behaves like $\gamma R(t,z)+\beta E_t/R(t,b_t)$. Transitions into the right tail of the productivity distribution that are the result of students learning from teachers are completely restricted by the available number of teachers, while an increase in the rate $\beta E_t/R(t, b_t)$ of random copying by entrants will increase the number of transitions into the right tail.

4. Balanced Growth Paths

Learning by active and inactive managers keeps the distribution of managerial productivities from spreading out too much: managers with low human capital learn from more productive managers, either as active managers, or, if their potential profits are too low, as inactive managers. Conjecture that there is a growth rate κ so that $z - \kappa t$ is stationary. Then the aggregate stock of managerial human capital H_t grows at the rate κ, and thus the factor price $v_t = v(t, 0)$ of a manager who supplies one unit of managerial services declines at the rate $\alpha\kappa$. Aggregate consumption C_t and wages w_t grow at the rate $(1 - \alpha)\kappa$. Since $w_t(1 + \phi)l(t, z)/\alpha = v_t e^z/(1 - \alpha)$ it follows that the distributions of employment and variable profits $v(t, z)/w_t$ in units of labor across active managers are stationary as well. Constant consumption growth implies constant interest rates, and income flows measured in units of labor are discounted at the subjective discount rate ρ.

These considerations suggests a balanced growth path with prices and value functions that satisfy

$$\begin{aligned} [w_t, W_t, U_t, S_t(\beta), S_t(\gamma)] &= [w, W, U, S(\beta), S(\gamma)]\, e^{(1-\alpha)\kappa t}, \\ [v(t, z), V(t, z)] &= [v(z - \kappa t), V(z - \kappa t)]\, e^{(1-\alpha)\kappa t}, \end{aligned}$$

and with a stationary productivity distribution of the form

$$[b_t, x_t, y_t, m(t, z)] = [b + \kappa t, x + \kappa t, y + \kappa t, Mf(z - \kappa t)],$$

where $f(\cdot)$ is a probability density function on $[b, \infty)$ and M is the measure of active managers. Variable employment relative to fixed labor along such a balanced growth path is given by

$$l(t, z) = \frac{1}{1 + \phi}\frac{\alpha}{1 - \alpha}\frac{v(t, z)}{w_t} = \frac{1}{1 + \phi}\frac{\alpha}{1 - \alpha}\frac{e^{z-\kappa t}}{w^{1/(1-\alpha)}} = l(z - \kappa t). \tag{23}$$

Note well that the arguments of the stationary density $f(\cdot)$, of variable labor relative to fixed labor $l(\cdot)$, and of the profit and value functions $v(\cdot)$ and $V(\cdot)$ are now de-trended (or time-0) managerial productivities. The thresholds b, x and y are thresholds for these de-trended productivities. Of course, the drift of de-trended productivities is $\mu - \kappa$.

4.1 Managerial Choices Given κ and a Tuition Schedule

Since W_t and U_t grow at the same rate as consumption and wages, and since $r_t = \rho + \mathrm{D}C_t/C_t$, it follows from (9)-(10) that the values of workers and inactive managers are given by

$$W = \frac{w}{\rho}, \quad U = W + \max\left\{0, \frac{S(\beta) - \beta W}{\rho + \beta}\right\}. \tag{24}$$

The surplus of an inactive manager $U - W$ is the present value of the flow gain from learning discounted at a rate that takes into account the fact that learning takes time. Together with the tuition schedule (8), the Bellman equation for active managers (11) simplifies to

$$\begin{aligned}\rho V(z) \quad = \quad & v(z) - \phi w + \max\left\{\beta V(z) - S(\beta), |\gamma V(z) - S(\gamma)|\right\} \\ & + (\mu - \kappa)\mathrm{D}V(z) + \frac{1}{2}\sigma^2 \mathrm{D}^2 V(z)\end{aligned} \tag{25}$$

for all z above the exit threshold b. At this exit threshold the value function must satisfy the value matching and smooth pasting conditions

$$U = V(b), \quad 0 = \mathrm{D}V(b). \tag{26}$$

Exit at b forces $U = V(b)$ and far away from this exit threshold $V(z)$ must behave like $v(z) \propto e^z$. Fixing b and omitting the smooth pasting condition $0 = \mathrm{D}V(b)$, this produces a second-order differential equation for $V(z)$ with boundary conditions at b and at infinity. These two boundary conditions together with smooth pasting conditions at the thresholds x and y are enough to determine $V(\cdot)$ on $[b, \infty)$. The smooth pasting condition $0 = \mathrm{D}V(b)$ then serves to determine the optimal b. Explicit calculations are given in the appendix. Given a tuition schedule that satisfies $S(\beta)/\beta < S(\gamma)/\gamma$ and the resulting value function $V(\cdot)$, the thresholds x and y can be inferred from

$$V(x) = \frac{S(\beta) + S(\gamma)}{\beta + \gamma}, \quad V(y) = \frac{S(\gamma) - S(\beta)}{\gamma - \beta}. \tag{27}$$

The condition $S(\beta)/\beta < S(\gamma)/\gamma$ ensures that $V(x) < V(y)$ and then the fact that $V(\cdot)$ is increasing implies $x < y$.

Inactive managers will not choose to learn if $S(\beta) < \beta W$. With a finite threshold $b > -\infty$ some active managers will choose to quit all the time, and thus the population of active managers will forever decline unless $S(\beta) \geq \beta W$. The tuition schedule will have to satisfy this conditions along a balanced growth path. If the condition holds as a strict inequality, then all inactive managers pay tuition and try to become active managers. If

$S(\beta) = \beta W$ then $U = W$ and some inactive managers may learn and others may not. The fact that $S(\gamma)/\gamma > S(\beta)/\beta$ implies that $S(\gamma) - \gamma U > S(\beta) - \beta U$ and so the least productive active managers will want to be students. Given $S(\beta) \geq \beta W$, note that (24) implies that $S(\beta) - \beta U = (S(\beta) - \beta W)\rho/(\rho + \beta)$. So the balanced growth version of the complementary slackness condition (12) can be stated in term of $S(\beta) - \beta W$ instead of $S(\beta) - \beta U$.

4.1.1 Scaling Properties of the Value Function

The surplus value function $V(\cdot) - W$ also satisfies (27), with a fixed cost $(1 + \phi)w$ instead of ϕw, and with a tuition schedule determined by $S(\beta) - \beta W$ and $S(\gamma) - \gamma W$. Multiplying $[w, v(\cdot), S(\beta) - \beta W, S(\gamma) - \gamma W]$ by a positive constant only scales the solution for $U - W$ and $V(\cdot) - W$ by that same constant, with no effect on the thresholds b, x and y. Therefore, given κ and $[v(\cdot), S(\beta) - \beta W, S(\gamma) - \gamma W]/[(1 + \phi)w]$, the Bellman equation determines $(U - W)/[(1 + \phi)w]$, $(V(\cdot) - W)/[(1 + \phi)w]$ and the thresholds $b < x < y$. The behavior of managers and entrepreneurs along a balanced growth path is completely determined by κ and $[v(\cdot), S(\beta) - \beta W, S(\gamma) - \gamma W]/[(1 + \phi)w]$. In turn, note that variable profits relative to fixed costs can be written as

$$\frac{v(z)}{(1 + \phi)w} = \frac{v(b)}{(1 + \phi)w} \times e^{z-b}.$$

Consider the change of variables $u = z - b$. The exit threshold for the scaled and shifted surplus value function $(V(b + u) - W)/[(1 + \phi)w]$ is then at $u = 0$ by construction, and the other thresholds are at $x - b$ and $y - b$. The Bellman equation for this alternative value function only depends on $[S(\beta) - \beta W, S(\gamma) - \gamma W]/[(1 + \phi)w]$, and choosing an exit threshold b for the original value function amounts to choosing $v(b)/[(1 + \phi)w]$ for the scaled and shifted surplus value function. Since $v(b)/[(1 + \phi)w] = l(b)(1 - \alpha)/\alpha$, this means that $l(b)$ and the distances $x - b$ or $y - b$ are functions only of $[S(\beta) - W, S(\gamma) - W]/[(1 + \phi)w]$. That is, the Bellman equation produces a map

$$[S(\beta) - \beta W, S(\gamma) - \gamma W]/[(1 + \phi)w] \mapsto [l(b), x - b, y - b]. \tag{28}$$

Importantly, given a "surplus" tuition schedule measured in units of fixed labor, the amount of variable labor relative to fixed labor used by the least productive manager does not depend on $(1 + \phi)w$. Neither do the distances that separate managers who are about to exit from managers at the margin between studying and teaching, and from managers at the margin between teaching entrepreneurs and teaching active managers.

4.2 Stationary Productivity Distributions

Along a balanced growth path, the measure of inactive managers choosing to study is $E_t = E$, and the density of active managers is $m(t,z) = Mf(z-\kappa t)$, with an exit barrier $b_t = b + \kappa t$. Market clearing requires that all students are matched with teachers. This means that

$$\frac{E}{M} = \int_x^y f(z)\mathrm{d}z, \tag{29}$$

$$\int_b^x f(z)\mathrm{d}z = \int_y^\infty f(z)\mathrm{d}z. \tag{30}$$

Imposing $\mathrm{D}M_t = 0$ in (17) and using the market clearing condition (14) for inactive managers gives

$$\beta \int_x^y f(z)\mathrm{d}z = \frac{1}{2}\sigma^2 \mathrm{D}f(b). \tag{31}$$

The conjectured stationarity of de-trended managerial human capital means that $Mf(z) = m(t, z+\kappa t)$ and thus $\mathrm{d}m(t, z+\kappa t)/\mathrm{d}t = 0$. Using this in (18) gives

$$0 = -(\mu-\kappa)\mathrm{D}f(z) + \frac{1}{2}\sigma^2\mathrm{D}^2 f(z) - \gamma f(z), \quad z \in (b,x), \tag{32}$$

$$0 = -(\mu-\kappa)\mathrm{D}f(z) + \frac{1}{2}\sigma^2\mathrm{D}^2 f(z) + \beta f(z), \quad z \in (x,y), \tag{33}$$

$$0 = -(\mu-\kappa)\mathrm{D}f(z) + \frac{1}{2}\sigma^2\mathrm{D}^2 f(z) + \gamma f(z), \quad z \in (y,\infty). \tag{34}$$

The boundary conditions (16)-(20) remain essentially the same,

$$f(b) = 0, \tag{35}$$

and

$$-(\mu-\kappa)f(x-) + \frac{1}{2}\sigma^2\mathrm{D}f(x-) = -(\mu-\kappa)f(x+) + \frac{1}{2}\sigma^2\mathrm{D}f(x+), \tag{36}$$

$$-(\mu-\kappa)f(y-) + \frac{1}{2}\sigma^2\mathrm{D}f(y-) = -(\mu-\kappa)f(y+) + \frac{1}{2}\sigma^2\mathrm{D}f(y+). \tag{37}$$

The requirement that $f(z)$ is probability density provides the boundary condition

$$\int_b^\infty f(z)\mathrm{d}z = 1. \tag{38}$$

Note that E/M is simply defined in terms of $f(\cdot)$ by the market clearing condition (29). Integrating (32)-(34) using (31) and (36)-(37) gives (30). This redundancy arises because the dynamics of $m(t,\cdot)$ is such that $\mathrm{D}M_t = \int_{b_t}^\infty \mathrm{D}_t m(t,z)\mathrm{d}z$ holds by construction, and therefore imposing $\mathrm{D}M_t = 0$ implies $\int_{b_t}^{\infty} \mathrm{D}_t m(t,z)\mathrm{d}z = 0$ and vice versa. Thus one can omit (29)-(30) and focus on solving (31)-(37) for $f(\cdot)$ and the thresholds x and y, with (38) left to ensure that $f(\cdot)$ integrates to 1.

Observe that $z > b$ only appears as the argument of $f(\cdot)$. Thus the differential equation for $f(\cdot)$ is autonomous and the stationary densities of $u = z - b$ do not depend on $b > -\infty$. The exit barrier b is simply a variable that determines the location of a density that does not otherwise depend on b. The threshold b does not pin down the location of stationary densities of z in the special case of $b = -\infty$. In that case, if $f(z)$ is a stationary density, then so is $f(z - c)$ for any constant c. Note that the boundary conditions (30)-(31) force $x = y$ to be the median when $b = -\infty$. In other words, if $b = -\infty$ then the median $x = y$ of the stationary density is indeterminate. As will be shown next, even if $b > -\infty$ is fixed, the threshold x remains indeterminate. This then implies a continuum of stationary densities given κ and b.

4.2.1 Stationary Densities Indexed by Turnover

Given a trend κ and an exit threshold b we now have a system of equations (31)-(38) that must be solved for the density of active managers $f(z)$, and the thresholds x and y. The differential equations (32)-(34) are linear second-order equations with constant coefficients. The characteristic polynomial for (32) automatically has two real solutions because γ is positive. Given that $\beta \in (0, \gamma)$, the characteristic polynomials for (33) and (34) each have two real solutions if and only if

$$\left(\frac{\mu - \kappa}{\sigma^2}\right)^2 \geq \frac{\gamma}{\sigma^2/2} \tag{39}$$

holds as a strict inequality. To see this, note that $e^{-\zeta z}$ solves (34) on (y, ∞) when ζ solves the quadratic equation $0 = \gamma + (\mu - \kappa)\zeta + \frac{1}{2}\sigma^2\zeta^2$. This quadratic has distinct real solutions if and only if the discriminant condition (39) holds. Both solutions for ζ are positive when $\kappa > \mu$, thereby ensuring that $e^{-\zeta z}$ is integrable on (y, ∞). With $\kappa > \mu$, (39) puts a lower bound on κ,

$$\kappa \geq \mu + \sigma\sqrt{2\gamma}. \tag{40}$$

Note that e^z has a finite mean if and only if the roots ζ satisfy $\zeta > 1$. This is implied by (39) if the underlying parameters satisfy $\gamma > \sigma^2/2$.

Under these conditions, the three differential equations (32)-(34) each have two exponential solutions, and linear combinations of these exponential solutions are also solutions. There are therefore six coefficients (two for each of the three differential equations) that need to be determined, as well as the two thresholds x and y. These eight variables are restricted by only five boundary conditions: (31) and (35)-(38). This leaves a three-dimensional family of solutions.

In all that follows, impose the further restriction that $f(z)$ is continuous at x and y. Because of (36)-(37), this implies that $f(z)$ is differentiable as well, and so the solutions on (b, x), (x, y) and (y, ∞) are smoothly pasted together. The two additional restrictions at x and y increase the number of boundary conditions from five to seven, leaving only one degree of freedom. As it turns out, this degree of freedom can be parameterized by the length $x - b$ of the interval (b, x) of managers who learn from other managers. To see this fix some positive $x - b$ and observe from (30)-(32), (35) and (38) that any stationary density has to satisfy

$$\begin{aligned} 0 &= -(\mu - \kappa)\mathrm{D}f(z) + \frac{1}{2}\sigma^2\mathrm{D}^2 f(z) - \gamma f(z), \quad z \in (b, x), \\ 0 &= f(b), \quad \frac{1}{2}\sigma^2\mathrm{D}f(b) = \beta\left(1 - 2\int_b^x f(z)\mathrm{d}z\right). \end{aligned}$$

This differential equation has two distinct exponential solutions on (b, x), and precisely one linear combination of these solutions will satisfies the two boundary conditions at b. This determines $m(z)$ on (b, x) given b and $x - b$. Next, choose some $y > x$. The density on (x, y) and (y, ∞) is then determined by the four linear restrictions implied by the smooth pasting conditions at x and y. This follows as long as (39) holds as a strict inequality, so that (34) does indeed have two distinct exponential solutions on (y, ∞). Although these two solutions merge when (39) becomes an equality, one can show that the limiting density on (x, y) and (y, ∞) continues to satisfy the smooth-pasting conditions at x and y. The density on (b, x) and (x, y) can now be used to evaluate both sides of the boundary condition (31). Imposing this boundary condition results in a non-linear equation in $y - x$ that has a unique solution. A proof of the following proposition appears in the appendix.

Proposition 1 *Fix some $\kappa > \mu$ that satisfies (39), as well as some positive $x - b$. Then there is one and only one smooth stationary distribution for $z - b$. The resulting distribution for e^z has right tail probabilities that behave like $e^{-\zeta z}$, with a tail index given by*

$$\zeta = \frac{\kappa - \mu}{\sigma^2} - \sqrt{\left(\frac{\kappa - \mu}{\sigma^2}\right)^2 - \frac{\gamma}{\sigma^2/2}}.$$

When (39) holds with equality this tail index reduces to $\zeta = \sqrt{\gamma/(\sigma^2/2)}$.

The right tail of the stationary distribution for e^z is a Pareto distribution, with a tail that gets thinner as κ approaches the lower bound (39) from above. Growth rates above this lower bound are associated with distributions of managerial productivities that have thicker right tails, allowing low-productivity managers to learn more from high-productivity managers.

The slope of the stationary density at b allows one to infer E/M from $\beta E/M = \frac{1}{2}\sigma^2 \mathrm{D}f(b)$. Together with (29)-(31) and (38) this can also be expressed as

$$\frac{E}{E+M} = \frac{1}{2}\left(1 + \frac{\beta}{\sigma^2/2}\int_b^x \frac{f(z)\mathrm{d}z}{\mathrm{D}f(b)}\right)^{-1}, \tag{41}$$

which only depends on the differential equation that governs $f(z)$ on (b, x) and no longer on the normalizing constant that makes $f(z)$ into a probability density. An easy calculation given in the appendix shows that $E/(E+M)$ is a decreasing function of the distance $x - b$, and hence so is the entry rate E/M. Lowering $x - b$ decreases the number of relatively less productive managers who are learning from more productive managers. This increases the flow of less productive managers across the exit threshold. At the same time, a lower $x - b$ also increases the number of incumbent managers who can teach inactive managers. This makes a steady state with a larger number of inactive managers self-sustaining. In this way, productivity distributions with low managerial turnover and productivity distributions with high managerial turnover are both consistent with market clearing and stationarity.

To summarize, given $b > -\infty$ and a $\kappa > \mu$ that satisfies (39) there is a one-dimensional continuum of stationary densities $f(z)$ indexed by $x - b$. Since the implied densities for $z - b$ do not depend on b, one can write these solutions as

$$f(z) = f(z - b|\kappa, x - b) \tag{42}$$

Note that $y - x$ can be inferred from $f(\cdot|\kappa, x - b)$ and the differential equations (32)-(34).

4.3 Labor Market Clearing

Recall that the balanced growth value (24) of U and the fact that $S(\beta) \geq \beta W$ implies that $S(\beta) - \beta U$ is a positive fraction of $S(\beta) - \beta W$. The balanced growth version of the complementary slackness condition (12)-(13) can therefore be stated as

$$E + M \leq A, \quad \text{w.e. if } S(\beta) - \beta W > 0. \tag{43}$$

Given a stationary density $f(z) = f(z-b|\kappa, x-b)$, the remaining condition for a balanced growth path is the labor market clearing condition

$$1 = (1+\phi)M \int_0^\infty (1 + l(b)e^u) f(u|\kappa, x-b) \mathrm{d}u. \tag{44}$$

Every active manager takes up $1+\phi$ units of labor by not supplying labor and to cover overhead. Variable labor per active manager of type z is $(1+\phi)l(b)e^{z-b}$, and hence (44) follows. Since $f(\cdot|\kappa, x-b)$ does not depend on b, one can use the labor market clearing condition (44) to determine $l(b)$ as a function of $x-b$ and M. The solution is guaranteed to be positive if M satisfies $(1+\phi)M < 1$. If $S(\beta) > \beta W$ then the measure of active managers M itself is determined by (43) together with the ratio $E/(E+M)$ given in (41). Since $E/(E+M)$ only depends on $x-b$, this makes $l(b)$ a function only of $x-b$ as long as $S(\beta) > \beta W$. But the value of M implied by $x-b$ and any $S(\beta) > \beta W$ may violate $(1+\phi)M < 1$. In that case a balanced growth path must have $S(\beta) - \beta W = 0$, so that M can be taken to be in the range $(0, (1 - E/(E+M))A) \cap (0, 1/(1+\phi))$.

4.4 Constructing Balanced Growth Paths

Fix a growth rate $\kappa > \mu$ that satisfies the discriminant condition (39). Proposition 1 establishes that there are many stationary distributions for $z - b$, indexed by positive thresholds $x-b$. Associated with every particular $x-b$ is a unique threshold $y-b > x-b$ that separates managers who teach inactive managers and managers who teach other active managers. Pick some $x-b$ and its associated $y-b$. The smooth pasting conditions of the Bellman equation at b, x and y then imply a set of linear equations that determine the vector of implicit "prices" $[v(b), S(\beta) - \beta W, S(\gamma) - \gamma W]/[(1+\phi)w]$, and hence $l(b)$. By construction, $v(b)$ is positive and $S(\gamma) - \gamma W > S(\beta) - \beta W$, but $S(\beta) - \beta W$ may not be positive.

If $S(\beta) - \beta W$ is negative, then the initial pick of $x-b$ is not consistent with balanced growth because no inactive managers would choose to learn. Suppose instead that $S(\beta) - \beta W$ happens to be zero. Then the complementary slackness conditions (43) admit a range of possible M. All components of the labor market clearing condition (44) are already pinned down, except for M. So one can solve this labor market clearing condition for M and check if it is in the range implied by the complementary slackness conditions (43) and the ratio $E/(E+M)$ already determined by the stationary density. If it is, then the initial guess $x - b$ is associated with a balanced growth path, one in which not necessarily all inactive managers choose to learn from active managers.[6] But

[6] A corollary is that the same balanced growth path is an equilibrium for any A larger than the one

it may also be that the M that clears the labor market given the stationary density and the $l(b)$ implied by the Bellman equation causes $E + M$ to exceed A. In that case one can consider thresholds $x - b$ for which the Bellman equation produces a strictly positive $S(\beta) - \beta W$ at the implied $y - b$. The complementary slackness conditions (43) then determine M given the ratio $E/(E + M)$ implied by the stationary density associated with $x - b$. The labor market condition (44) can now be solved for $l(b)$. The Bellman equation already generates a $l(b)$ given $[x - b, y - b]$, and these two solutions have to match. Imposing this requirement generates an equilibrium condition for $x - b$.

As stated, this algorithm does not obviously rule out multiple balanced growth paths given κ. More importantly, it may be that for any positive $x - b$ the associated vector of thresholds $[x - b, y - b]$ implied by stationarity is such that the Bellman equation produces a negative $S(\beta) - \beta W$. This happens when $\beta - \delta \in (0, \gamma - \delta)$ is too small. In that case, any attempt to construct a balanced growth path will fail because inactive managers do not want to study. Below, it will be shown that a balanced growth path can be constructed for every $\kappa > \mu$ that satisfies (39) in the special case of $\gamma = \beta$.[7]

Suppose the construction proposed so far produces a stationary density and a $l(b)$. The actual density $f(0, z)$ of managerial human capital at the initial date is a given. Imposing $f(0, z) = f(z - b|\kappa, x - b)$ then delivers the exit threshold b. Combining the $l(b)$ and b so obtained determines wages at the initial date via $l(b) = (\alpha/(1-\alpha))e^b/((1+\phi)w^{1/(1-\alpha)})$. Wages then imply the present value of worker labor earnings $W = w/\rho$, and from this $[S(\gamma), S(\beta)]$ follows.

This construction implies a simple comparative static for an across the board improvement in managerial productivity. Starting from a balanced growth path, an unforeseen proportional improvement in productivity across all active managers shifts the initial distribution $f(0, z)$ to the right. Along the balanced growth path $f(0, z) = f(z - b|\kappa, x - b)$ and this implies an equal shift to the right in b. Because of $l(b) = (\alpha/(1 - \alpha))e^b/((1 + \phi)w^{1/(1-\alpha)})$, this implies an increase in worker wages with an elasticity $1 - \alpha$. The change from one balanced growth path to the other is instantaneous.

used to check the complementary slackness conditions (43). Inactive managers are abundant and earn no surplus over workers.

[7] A small flow of inactive managers who succeed in learning a sufficiently useful z (relative to the current distribution) without having to pay active managers can restore existence of a balanced growth path. A natural mechanism for this is imitation. Another possibility, more distant from the model under consideration, is that new industries can arise that do not build on the knowledge accumulated in existing industries.

5. When All Managers Learn at the Same Rate

To describe more explicitly how balanced growth paths are determined and show how they depend on the fixed cost parameter ϕ, consider the special case in which $\gamma = \beta$. In such an economy, active and inactive managers learn at the same rate, and it does not matter how these managers are assigned to the various managers who teach. As before, there will be a threshold x so that active managers in the interval (b, x) are students. But the intervals (x, y) and (y, ∞) merge into a single interval (x, ∞) of managers who teach. There is only one value of learning $S(\beta)$, and the scaling properties of the value function imply a relation $(S(\beta) - \beta W)/[(1+\phi)w] \mapsto [l(b), x - b]$. Since ϕ only affects the Bellman equation via $(S(\beta) - \beta W)/[(1+\phi)w]$, this relation cannot depend on ϕ.

If the growth rate κ is at its lower bound $\kappa = \mu + \sigma\sqrt{2\beta}$ then the value function is well defined and the tail index of the managerial productivity distribution is greater than 1 under the very weak conditions that ρ is positive and $\beta > \sigma^2/2$, which is just the requirement that $\zeta > 1$. There is no need to assume an upper bound on the learning rate. Although the drift of e^{z_t} may very well exceed the effective discount rate $\rho - \beta$ for a manager who teaches, the fact that $\kappa - \mu$ is positive means that z_t has a negative drift. This implies the eventual exit of even the most productive managers, and it can be shown that this suffices to ensure that present values are finite. The following proposition is given for the more restrictive assumption that the present value of e^{z_t} is finite when $\mathrm{d}z_t = (\mu - \kappa)\mathrm{d}t + \sigma \mathrm{d}B_t$ forever.[8]

Proposition 2 *Fix some $\kappa > \mu$ such that $\rho > \beta$ and $\rho > \beta + \mu - \kappa + \sigma^2/2$. The value function is well defined, strictly increasing, and convex for any positive $(S(\beta) - \beta W)/[(1+\phi)w]$. The distance $x - b$ and the ratio $l(b)$ of variable over fixed labor of the least productive manager are both strictly increasing in $(S(\beta) - \beta W)/[(1+\phi)w]$, with $x - b$ varying from 0 to ∞ and $l(b)$ varying in the bounded range*

$$l(b) \in \frac{\alpha}{1-\alpha}\left(1 - \frac{1}{\frac{\kappa-\mu}{\sigma^2} + \sqrt{\left(\frac{\kappa-\mu}{\sigma^2}\right)^2 + \frac{\rho-\beta}{\sigma^2/2}}}, 1 - \frac{1}{\frac{\kappa-\mu}{\sigma^2} + \sqrt{\left(\frac{\kappa-\mu}{\sigma^2}\right)^2 + \frac{\rho+\beta}{\sigma^2/2}}}\right)$$

The variable-over-fixed labor ratio $l(b)$ is proportional to the variable profits $v(b)$ of the least productive manager. At the exit barrier b these profits cannot be enough to

[8] As in the case $\gamma > \beta$ treated in the appendix, both the value function and the stationary density can be determined analytically by solving a system of equations that are linear given $x - b$. The proofs of the next two propositions involve lengthy calculations of derivatives with respect to $x - b$ that will be made available in an on-line appendix.

cover overhead labor and the time of the manager, because of the option value of being an active manager. The option to remain an active manager is valuable since inactive managers can only become active again after a stochastic delay. This has to be true for any $(S(\beta)-\beta W)/[(1+\phi)w]$, and hence the bounded range for $v(b)$ and $l(b)$. The length of the interval (b,x) over which it is optimal to be a student expands without bound as the value of learning becomes large. A very high value of $(S(\beta)-\beta W)/[(1+\phi)w]$ means that managers who teach other managers have to deliver a very valuable improvement in productivity, and hence they must be very productive themselves.

The value $(S(\beta)-\beta W)/[(1+\phi)w]$ of learning in units of fixed labor only appears in the construction of the value function. By varying this value one can therefore trace out an equilibrium relation between $x-b$ and $l(b)$. Proposition 2 implies that $l(b)$ is an increasing and bounded function of $x-b$, as illustrated in Figure 3 below. When $x-b$ is large, it must be that $(S(\beta)-\beta W)/[(1+\phi)w]$ is large, and this in turn means that managers learning from each other is a source of managerial incomes that tends to dominate the variable profits $v(b)\propto l(b)$ that the least productive managers earn by managing workers. Such high returns from learning and teaching reduce the importance of the option to remain an active manager, and hence bring the variable profits $v(b)<(1+\phi)w$ closer to the fixed cost $(1+\phi)w$ of being an active manager.

As in the discussion leading up to Proposition 1, the stationary densities are indexed by growth rates $\kappa>\mu$ that have to satisfy (39), and by $x-b$. The resulting stationary densities $f(z)=f(z-b|\kappa,x-b)$ can be characterized as follows.

Proposition 3 *Fix some $\kappa>\mu$ such that (39) holds. The associated stationary density $f(z)=f(z-b|\kappa,x-b)$ and the fraction of active managers fraction $M/(E+M)=\mathcal{M}(\kappa,x-b)$ implied by (41) satisfy*

$$\lim_{\Delta\downarrow 0}\int_0^\infty e^u f(u|\kappa,\Delta)\mathrm{d}u=\frac{\beta}{\beta+\mu-\kappa+\frac{1}{2}\sigma^2},\qquad \lim_{\Delta\downarrow 0}\mathcal{M}(\kappa,\Delta)=\frac{1}{2},$$

and

$$\frac{\partial}{\partial\Delta}\int_0^\infty e^u f(u|\kappa,\Delta)\mathrm{d}u>0,\qquad \frac{\partial}{\partial\Delta}\mathcal{M}(\kappa,\Delta)>0$$

for all positive $\Delta=x-b$. The mean of e^u grows without bound as $x-b$ becomes large.

The first of these derivatives simply says that the measure M of active managers, as a fraction of $E+M$, is increasing in $x-b$. A larger distance $x-b$ implies less turnover and hence a smaller stock of inactive managers trying to re-enter. The other derivative implies that the mean of variable labor relative to variable labor employed by the least productive active manager is increasing in $x-b$.

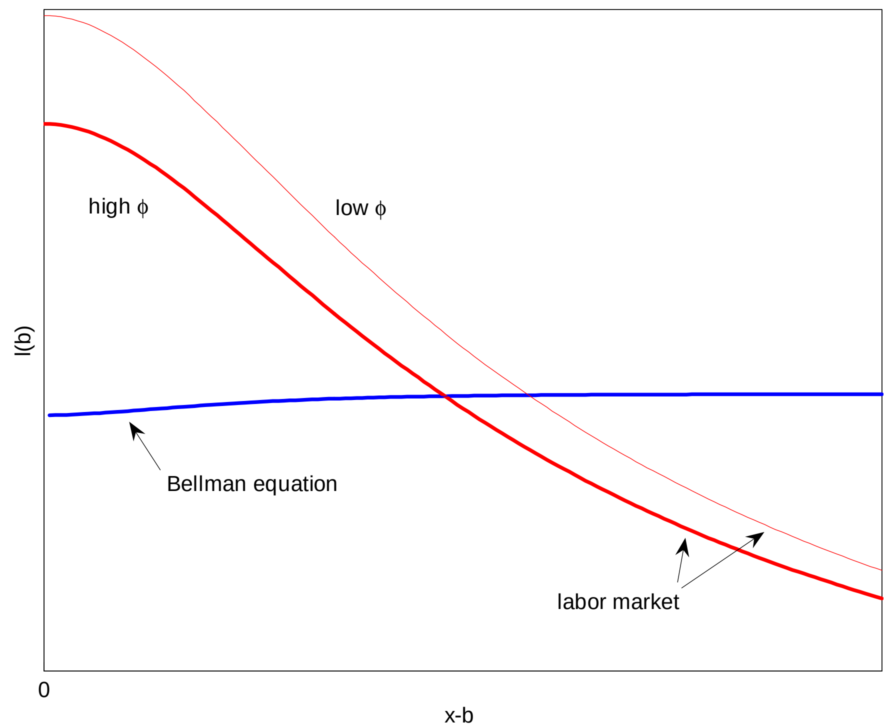

FIGURE 3 The Effect of a Reduction in Fixed Costs

Proposition 3 can be used together with the complementary slackness relation (43) and the labor market clearing condition (44) to derive another equilibrium relation between $x-b$ and $l(b)$. If $S(\beta)-\beta W>0$ then the continuity of the value function implies that some active managers want to be students, and hence it must be that $x-b$ is positive. Alternatively, if $S(\beta)-\beta W=0$ then active managers strictly prefer to be teachers, and hence $x-b=0$. Using the fact that $S(\beta)-\beta W$ and $x-b$ as positive at the same time together with $M/(E+M)=\mathcal{M}(\kappa,x-b)$ and the complementary slackness and labor market clearing conditions (43)-(44) gives

$$1\leq(1+\phi)A\mathcal{M}(\kappa,x-b)\int_0^\infty(1+l(b)e^u)f(u|\kappa,x-b)\mathrm{d}u,\quad \text{w.e. if } x-b>0.$$

Proposition 3 implies that the right-hand side of this inequality is increasing in $x-b$. If $S(\beta)-\beta W$ is positive, this gives $l(b)$ as a decreasing function of $x-b$. More variable labor relative to the variable labor employed by the least productive active manager is only consistent with market clearing if the least productive manager employs fewer workers. The $x-b\downarrow 0$ limits say that no fewer than half of all managers can be active and that there is a lower bound on how much variable labor the average active manager uses relative to the least productive active manager. In turn, this implies a finite upper bound on $l(b)$ that applies as long as $S(\beta) - \beta W$ is positive. When all managers are either active to learning to become active, there is only so much variable labor per unit of fixed labor that the least productive manager can be using in any equilibrium. On the other hand, if $S(\beta) - \beta W = 0$ then inactive managers are indifferent between trying to become active managers and just supplying labor. In that case $x - b = 0$ since active managers will strictly prefer to be teachers, and then $l(b)$ can be anywhere above the finite upper bound obtained for positive $S(\beta) - \beta W$.

Figure 3 shows the functions $x - b \mapsto l(b)$ implied by the Bellman equation and Proposition 2 and by the labor market clearing condition together with Proposition 3. The balanced growth path for a given $\kappa > \mu$ that satisfies (39) is determined by the intersection of these two functions. Since one function is decreasing and the other increasing, it follows that there can be only one pair $[x - b, l(b)]$ that is consistent with balanced growth for a given κ. The fact that the mean of e^{z-b} given $x - b$ grows without bound implies that $l(b)$ converges to zero as $x - b$ becomes large. Thus there will be an intersection, and it will be unique. In turn this means only one candidate stationary density, and forcing $m(0, z) = Mf(z - b|\kappa, x - b)$ then delivers the threshold b. The path of wages along this balanced growth path is implied by $l(b) = (\alpha/(1-\alpha))e^b/((1 + \phi)w^{1/(1-\alpha)})$, as before. The intersection between the two curves in Figure 3 may occur at $x - b$. In that case $S(\beta) - \beta W$ is zero as well. Inactive managers are then indifferent between trying to become active managers or not, and the population of managers engaged in learning may be smaller that the population A of all managers. The lower bound reported in Proposition 2 then determines the equilibrium use of variable labor by the least productive active managers.

5.1 Fixed Costs and Factor Shares

Propositions 2 and 3 can be used to determine the effects of a change in the overhead labor cost parameter ϕ. By Proposition 2, the upward-sloping function $x - b \mapsto l(b)$ implied by the Bellman equation does not depend on ϕ. But Proposition 3 and the labor market clearing condition (44) imply that a decline in ϕ raises the downward-sloping function $x - b \mapsto l(b)$. Given $x - b$ and the implied population of active managers, less overhead labor per active manager can only imply more variable labor per active manager. As illustrated in Figure 3, a decline in ϕ then leads to an increase in both $x - b$ and $l(b)$. By Proposition 3, this means more active managers $M = \int_0^\infty m(u|\kappa, x - b)\mathrm{d}u$, and a higher mean of variable labor relative to variable labor of the least productive active manager. Together with the labor market clearing condition (44) this in turn implies a decline in $(1+\phi)M$. That is, the ability of managers to be active with less overhead labor leads to an increase in the number of active managers, but the overall amount of fixed labor (overhead and managerial time) will still decline.

By Proposition 2, an increase in $x-b$ and $l(b)$ along the equilibrium relation $x-b \mapsto l(b)$ determined by the Bellman equation implies an increase in the surplus value of learning $(S(\beta)-\beta W)/[(1+\phi)w]$, measured in units of fixed labor. The factor shares of workers and of labor overall are

$$\frac{w}{C}\begin{bmatrix} 1-A \\ 1-M \end{bmatrix} = \frac{\alpha}{1-(1+\phi)M}\begin{bmatrix} 1-A \\ 1-M \end{bmatrix},$$

respectively. The denominator $1-(1+\phi)M$ is simply the supply of labor used as variable labor to produce consumption. A fraction $1-A$ of the population are workers. Inactive managers also supply labor and so the overall supply of labor is $1-M$. Note that the factor share of labor exceeds the Cobb-Douglas share parameter α, but this need not be the case for the factor share of workers. A decline in ϕ lowers $(1+\phi)M$, which tends to lower both factor shares. The factor share of labor will decline by more than the factor share of workers because fewer managers are inactive when overhead labor per active manager declines.

To summarize, technical progress that lowers the amount of overhead labor required by active managers hurts workers relative to active managers. In contrast, a one-time improvement in the productivity of all active managers leaves the number of active managers and the cross-sectional distribution of their productivities unaffected, and has no effect on factor shares. The distributional implications of these two types of technical change are very different.

6. Determining the Long-Run Growth Rate

As anticipated by the two initial examples (1) and (3), the economy has a continuum of balanced growth paths with different stationary distributions of managerial productivities and different growth rates. This raises the question: which of these many different growth rates is likely to occur? If the initial distribution of managerial productivity is not a stationary distribution associated with a certain balanced growth rate, what will happen?

A detailed answer to this question can be given for a simplified economy in which there are no fixed costs of any kind: no overhead labor, and inactive managers who cannot supply labor. The aggregate supply of labor is then simply $1-A$. Aggregate consumption and wages can be computed from the measure of managerial types, as in (4)-(5). The key simplification that arises from the absence of fixed costs is the fact that managers never quit and all managers remain active, learning at the common rate γ. The assignment of who learns from whom becomes trivial: managers below the median x_t are assigned to learn from managers above the median and the particular assignment of managers below the median to managers above the median does not matter. The tuition schedule simplifies to $T(t,z) = \max\{0, \gamma V(t,z) - S_t(\gamma)\}$ and $S_t(\gamma)$ will adjust so that the value function satisfies $\gamma V(t, x_t) = S_t(\gamma)$.

The dynamics of the distribution of types is still given by (21), but now $E_t = 0$ and $b_t = -\infty$. Because b_t no longer plays a role, the distribution of productivities evolves exogenously, unaffected by prices or beliefs. The differential equation (21) reduces to

$$\mathrm{D}_t R(t,z) = -\mu \mathrm{D}_z R(t,z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} R(t,z) + Q(R(t,z)), \tag{45}$$

where $Q(\cdot)$ is given by

$$Q(R) = \gamma \max\{A - R, R\}. \tag{46}$$

When managers imitate at random instead of learn from the managers to whom they are assigned, the function $Q(R)$ becomes the quadratic $Q(R) = \gamma(A-R)R/A$. Thus (45)-(46) generalizes the two initial examples (1)-(3).

The differential equation (45) is a reaction-diffusion equation. The properties of its solutions depend very much on the shape of $Q(\cdot)$. Here, $Q(R) \geq 0$, $Q(H) = Q(0) = 0$, and $\mathrm{D}Q(0) > 0$ while $\mathrm{D}Q(A) < 0$. Under these conditions, (45) has a stationary solution of the form $R(t,z) = R(z - \kappa t)$ as long as $\kappa \geq \mu + \sqrt{2\mathrm{D}Q(0)}$. The reasoning is essentially that used to derive the restriction (39) on κ. Note that $R(t,z) = R(z-\kappa t)$ and (45) implies $0 = -(\mu - \kappa)\mathrm{D}R(z) + \frac{1}{2}\sigma^2 \mathrm{D}^2 R(z) + Q(R(z))$, or

$$\mathrm{D}R(z) = -f(z), \quad \mathrm{D}f(z) = \frac{(\mu - \kappa) f(z) + Q(R(z))}{\sigma^2/2}.$$

The phase diagram is shown in Figure 4, for (46) and the quadratic case that arises from imitation. Since $Q(0) = 0$, a linear approximation of $Q(R)$ near $R = 0$ yields

$$0 \approx \mathrm{D}Q(0)R(z) - (\mu - \kappa)\mathrm{D}R(z) + \frac{1}{2}\sigma^2 \mathrm{D}^2 R(z),$$

for $R(z)$ small. This equation is exact for (46) but approximate for $Q(R) = \gamma(A - R)R/A$. This is now a linear second-order differential equation with solutions of the form $e^{-\zeta z}$, where ζ solves the characteristic equation $0 = \mathrm{D}Q(0) + (\kappa - \mu)\zeta + \frac{1}{2}\sigma^2\zeta^2$. Complex roots to this equation lead to solutions that fail to be non-negative. The quadratic has real solutions if and only if $((\kappa - \mu)/\sigma^2)^2 \geq \mathrm{D}Q(0)/(\sigma^2/2)$, with $\kappa > \mu$ to ensure that ζ is positive. So the condition for real solutions is $\kappa \geq \mu + \sigma\sqrt{2\mathrm{D}Q(0)}$, which generalizes (39).

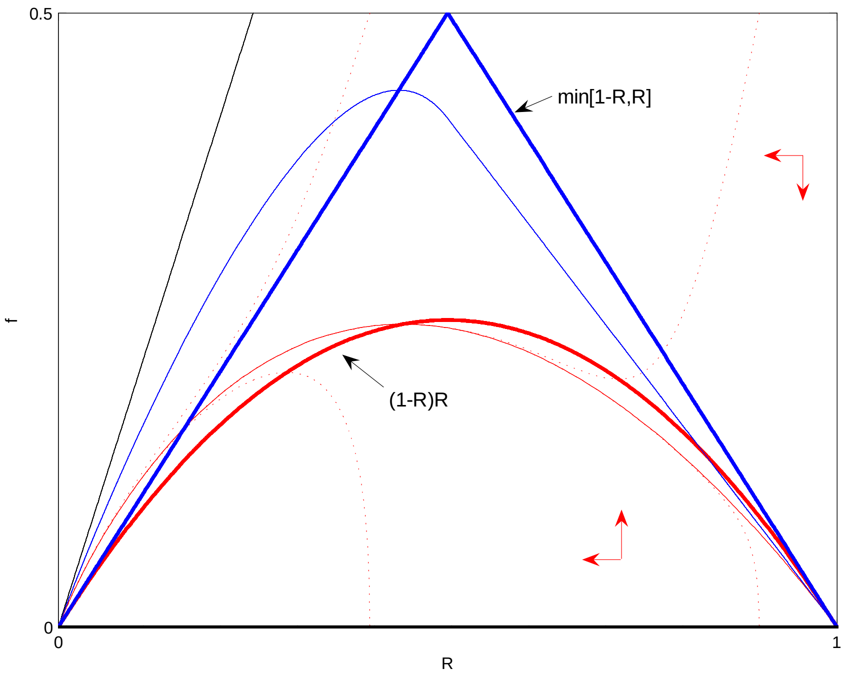

FIGURE 4 Imitation (red) versus Teaching (blue)

More importantly, an initial distribution $R(0, z)$ with bounded support converges to a distribution $R(z - \kappa t)$ with $\kappa = \mu + \sqrt{2\mathrm{D}Q(0)}$. That is, although there are many stationary distributions and many associated long-run growth rates, an economy that starts out with a limited amount of heterogeneity converges to one with the lowest possible long-run growth rate (40). Both random matching at the rate γ and students below the median learning from teachers above the median lead to $\mathrm{D}Q(0) = \gamma$. Both therefore lead to the same long-run growth rate.

As already indicated in the discussion surrounding (22), when there are fixed costs the equilibrium entry rate in an economy with random imitation does influence the long-run growth rate. As suggested by Figure 2 and detailed in Luttmer [2012], the basic reason is that the slope of $Q(\cdot)$ near zero depends on the equilibrium entry rate when there is random imitation. But this effect does not depend on the ability of active managers to imitate other active managers. If only entrants can imitate, then the quadratic $Q(\cdot)$ becomes linear with a slope determined by the entry rate (set γ equal to zero in (22)).

This is then essentially the economy introduced in Luttmer [2007], where equilibrium entry and growth rates are determined jointly.

6.1 Traditional Interpretations of (45)-(46)

The differential equation (45) was used by Fisher [1937] and Kolmogorov, Petrovskii and Piskunov [1937] to study the geographic spread of an advantageous gene. The equation is often known as the Fisher-KPP equation or simply the KPP equation. It is also known as a reaction-diffusion equation after its applications in chemistry. In the interpretation of Fisher and Kolmogorov, Petrovskii and Piskunov, $R(t, z)$ is the fraction of the population at time t in location z that carries the advantageous gene. The diffusion term reflects random movement along the line of individual in the population, and the logistic term $\gamma(A - R)R/A$ arises when individuals carrying the advantageous gene replicate at a higher rate than individual not carrying this gene. McKean [1976] showed that $1 - R(t, z)$ can also be interpreted as the distribution of a the maximum of a branching Brownian motion. For $\sigma = 0$ this corresponds to the interpretation of (3) given earlier. Bramson [1983] gave precise conditions on initial distributions that give rise to the alternative stationary distributions that arise when κ is above its lower bound. Volpert, Volpert, and Volpert [1994] provide a textbook treatment of reaction-diffusion equations.

7. Quantitative Implications

Managers have finite careers and lives. By teaching, they can transmit their knowledge to others, including to successors and descendants. To account account for the fact that they may not be around long enough, suppose that both workers and managers die randomly at some average rate δ. Assume new workers and managers are born at the same average rate so that the population remains constant. To avoid potentially important dynamic efficiency issues that arise in an economy with overlapping generations, assume that households form infinitely-lived dynasties. Newborn individuals may very well inherit their ability to learn. Or part of this ability could be the result of prior education. But managers are assumed not to inherit any useful knowledge. They have to begin life as inactive managers.[9]

[9] In Bils and Klenow [2000], new generations use time in school and time on the job to build on the collective human capital of an earlier generation of teachers. Here new generations can learn from anyone alive who has something useful to teach.

The only effect of random death on the Bellman equations (9)-(11) is to raise the effective discount rate from r_t to $r_t + \delta$. The subjective discount rate ρ that appears in the Bellman equations for a balanced growth path increases to $\rho + \delta$.[10] The continuous rejuvenation of the population of managers does affect the shape of the distribution of managerial productivities. Specifically, the dynamics of the stock of inactive managers (17) has an extra term $\delta(A - M_t)$ on the right-hand side because active managers who die are replaced by newborn managers who are inactive. The terms $-\gamma m(t, z)$ and $\gamma m(t, z)$ in the differential equations (18) for $m(t, z)$ change to $-(\gamma + \delta)m(t, z)$ for $z \in (b_t, x_t)$ and to $(\gamma - \delta)m(t, z)$ for $z \in (y_t, \infty)$. The replication rates for managers in (b_t, x_t) and (y_t, ∞) no longer add up to zero. But Proposition 1 can be extended to show that one can still solve for stationary distributions as long as $\gamma \geq \beta > \delta$ and the lower bound (39) on κ is modified to $\kappa \geq \mu + \sigma\sqrt{2(\gamma - \delta)}$. The number of active managers tends to shrink because of exit at the threshold b and random death. To overcome this, active managers in (x, y) must be able to replicate themselves, by transferring their knowledge to inactive managers at the rate β, more quickly than the random rate δ at which they die. If the gap $\beta - \delta$ is small, the stationary distributions (indexed by κ and $x - b$) are such that most of the active managers teach inactive managers. The long-run growth rate exceeds the average rate μ at which active managers can improve their own productivity only to the extent that their ability to replicate exceeds their death rate. At the lower bound for κ, the tail index of the distribution of e^z is $\zeta = \sqrt{(\gamma - \delta)/(\sigma^2/2)}$. The aggregate managerial demand for labor is finite only if this exceeds 1.

7.1 Notes on Identification

Aggregate consumption growth can be used to infer $(1-\alpha)\kappa$ and interest rates then reveal ρ. The share parameter α can be computed, in principle, from aggregate consumption and the wages earned by anyone who supplies variable labor. Average career or life spans pin down δ. Because variable labor depends on the de-trended states of individual managers (23), one can identify the learning rates β and γ, the drift $\mu - \kappa$ of de-trended managerial productivities, and the diffusion coefficient σ from sample paths of the numbers of workers assigned to individual managers. More indirectly, one piece of dynamic information can be used together with the stationary distribution of employment per manager to identify these parameters. For example, the growth rate $\kappa - \mu = \sigma\sqrt{2(\gamma - \delta)}$

[10] An alternative specification that is likely to be important involves the possibility that the knowledge of individual managers becomes obsolete at random times. The random death rate δ cannot be interpreted in this way because obsolescence does not affect workers and inactive managers.

and the tail index $\zeta = \sqrt{(\gamma - \delta)/(\sigma^2/2)}$ imply $[\sigma^2, \gamma - \delta] = (\kappa - \mu)[1/\zeta, \zeta/2]$. Concretely, away from the exit threshold b, variable employment among surviving active managers grows at the rate $g = \mu - \kappa + \frac{1}{2}\sigma^2 = \frac{1}{2}\sigma^2 - \sigma\sqrt{2(\gamma - \delta)}$. Combining this with the formula for the tail index ζ allows one to infer

$$\left[\sigma^2, \kappa - \mu, \gamma - \delta\right] = -\frac{g}{\zeta - \frac{1}{2}} \times \left[1, \zeta, \frac{1}{2}\zeta^2\right].$$

The maintained assumption $\zeta > 1$ implies that g is negative. The fact that the distribution of managerial productivities moves to the right more quickly than the rate at which individual managers can improve their own productivities means employment among incumbent active managers has to decline, even among surviving managers.[11] The rate at which this happens helps identify σ^2, $\kappa - \mu$ and $\gamma - \delta$. More detailed features of the stationary distribution of employment across managers can be used to infer $\beta - \delta$ as well. An alternative is to use information about entry and exit rates into active management.

The drift μ now follows as a residual constructed from κ and $\kappa - \mu$. By assumption, there is no growth in total factor productivity in this economy. This assumption attributes all of μ to managerial learning-by-doing. Without a direct measurement of the aggregate stock of managerial human capital H_t or of the managerial services e^{z_t} provided by at least some managers, it is not possible to distinguish constant growth in total factor productivity from trend growth in managerial human capital.

7.2 Benchmark Specifications

Consider an economy with a subjective discount rate $\rho = 0.03$ and a random death rate $\delta = 0.04$, with time measured in years. This implies average careers of only 25 years, but realized careers can be much longer. Suppose $A = 0.2$, so that 20% of the population is a manager, either active or inactive. The US has approximately 6 million employer firms and a private-sector labor force of about 100 million. So managers here are not just owners of businesses and CEOs. The Bureau of Labor Statistics (BLS) in its Current Employment Survey (CES) reports that about 4/5 of all private sector employees are production and non-supervisory workers. So $A = 0.2$ could be interpreted as the number of supervisory workers. Average weekly earnings of production and non-supervisory workers are about 80% of average weekly earnings of the average private sector employee in the CES. Combined with $A = 0.2$ this implies a factor share for production and non-supervisory workers of 64%. The actual share should be lower

[11] It is very easy to modify this economy to allow the population of households to grow, and then g need no longer be negative.

because business owners and CEOs are not in the CES. In all specifications that follow, the Cobb-Douglas share parameter is $\alpha = 0.5$. As already emphasized, fixed costs imply that the actual factor share of labor in this economy will exceed α. Economies with these baseline parameters are shown in Table 1. Managers who teach other managers are referred to as mentors in Table 1, and managers who teach inactive managers (who themselves are employed supplying labor) are called supervisors.

In a first benchmark specification, fix the rate at which both active and inactive managers learn to be $\gamma = \beta = 0.05125$. Managers trying to learn break through very infrequently. It takes them on average almost 20 years, if they survive. Recall that the tail index of the managerial employment distribution is $\zeta = \sqrt{(\gamma - \delta)/(\sigma^2/2)}$ when κ is at the lower bound (39) associated with long-run growth from bounded initial conditions. At $\sigma = 0.05$, this yields $\zeta = 3$. If all active managers were business owners or CEOs, this would imply a Pareto-like employment size distribution of firms with the same tail index, one that is much too high compared to the $\zeta \approx 1$ found in US data (Luttmer [2007]). But here firms must be aggregations of managers, and this can result in a firm size distribution with a thicker tail than is implied by $\zeta = 3$. The implied contribution to long-run growth in the stock of managerial human capital $\kappa - \mu = \sigma\sqrt{2(\gamma - \delta)}$ equals a modest 0.0075. Given $\alpha = 0.5$, the contribution to consumption growth is only half.

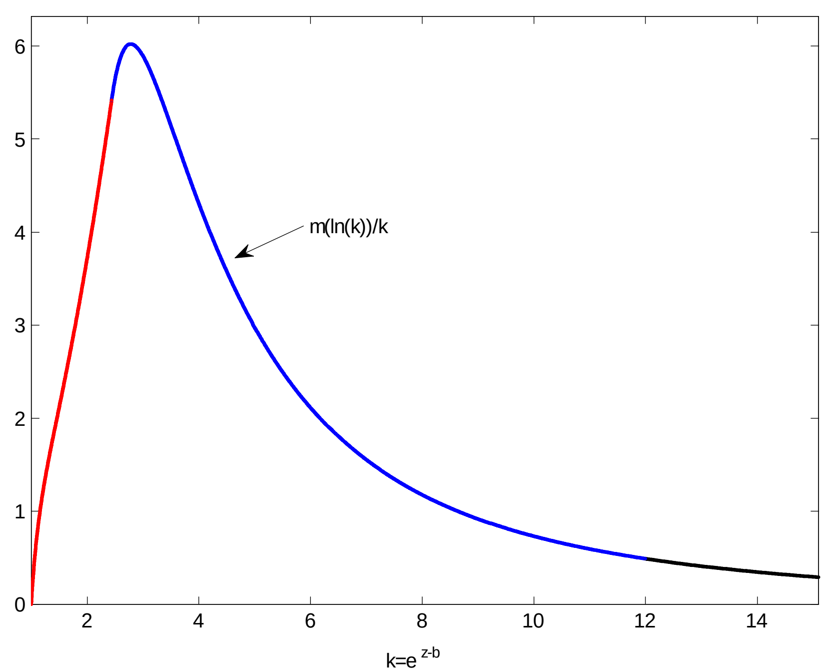

FIGURE 5 The Stationary Density of $k = e^{z-b}$

The stationary density of $k = e^{z-b}$ is shown in Figure 5. This density displays the typical Pareto shape in the right tail, but it has a mode strictly above the minimum, as in other economies with fixed costs (Luttmer [2007]). Figure 6 shows the value function and managerial flow profits from producing consumption and transferring human capital. The flow profits from transferring knowledge rise more quickly with the type of the manager than the flow profits from producing consumption. High-productivity managers earn more from teaching lower-productivity managers than they do from overseeing their own production team.

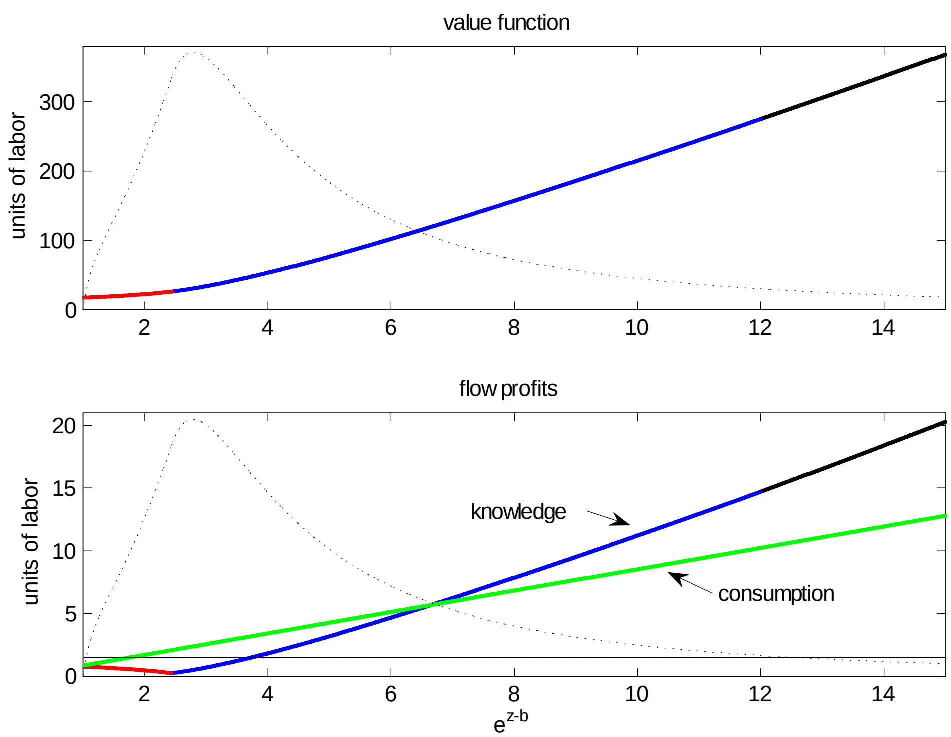

FIGURE 6 The Value Function and Flow Profits

The first column of Table 1 shows the properties of this economy when the overhead input requirement is $\phi = 2$. Including the value of the manager's outside option as an inactive manager, it takes as much as 3 units of labor to remain active. Managers on the verge of exiting earn only 2.726 units of labor from providing their managerial services. They are still willing to remain active because they lose their human capital upon exit and can become active again only after an uncertain delay. The flow gain $\gamma V(t,z) - S_t(\gamma)$ from learning for the marginal active manager is only 0.135 units of labor per annum. Since $\beta = \gamma$, active status does not help, and marginal active managers do not remain active to preserve a learning advantage. At least half of all managers must be active, since every inactive manager is also a student in this economy. In equilibrium, the fraction of managers who are active is only slightly above one half. Since active and inactive managers account for 20% of the population, this means that the average active manager manages roughly 9 employees.

The market value of the investment that results from managers teaching other managers is given by

$$\begin{aligned} I_t &= \gamma\left\{\int_{y_t}^{\infty}[V(t,z)-U_t]m(t,z)\mathrm{d}z - \int_{b_t}^{x_t}[V(t,z)-U_t]m(t,z)\mathrm{d}z\right\} \\ &\quad +\beta\int_{x_t}^{y_t}[V(t,z)-U_t]m(t,z)\mathrm{d}z. \end{aligned}$$

Improving the quality of managerial human capital involves a simultaneous destruction of low-quality capital and creation of high-quality capital. For managers teaching other managers, the implied market value of investment is measured by the first two terms in this equation, and the third term accounts for inactive managers learning to become active managers. In this benchmark economy, this measure of aggregate investment is as high as 43% of aggregate consumption. This amounts to 7% of the aggregate stock H_t of human capital, which together with learning-by-doing at the rate $\mu + \sigma^2/2$ makes up for random death and active managers who choose to become inactive to ensure that the stock of human capital grows at the rate κ.

Among managers who teach other managers, the implied Gini coefficient for earnings from managing workers is only $1/(2\zeta - 1) = 0.20$, but a substantial part of the earnings of these managers is derived from their teaching activities, and these earnings are more skewed. Even considering only earnings from managing workers, the Gini rises to 0.298 among all active managers, because there are many active managers whose human capital is too low to make teaching active managers profitable. Once the labor earnings of inactive managers are included the Gini shoots up to 0.500. This is low compared to the earnings Gini of 0.64 in the 2007 Survey of Consumer Finances reported by Díaz-Giménez, Glover, and Ríos-Rull [2011], but the Gini calculated here only covers the 20% of all households who are managers. Because of the high overhead input requirement $\phi = 2$, about 24% of all labor in the economy is used to cover overhead. The result is a labor share (in the output of consumption) of .65, well above the Cobb-Douglas share parameter $\alpha = 0.5$, but very close to the number implied by CES data on production and non-supervisory workers.

The second column of Table 1 shows how the balanced growth path changes when the labor input requirement ϕ is cut in half. As was already shown analytically, the number active managers rises and the amount of overhead labor they use declines. But the increase in the number of active managers is very small, and the share of labor used to cover overhead therefore drops by almost 50%. The effect on the factor share of labor is to cut the difference between it and the Cobb-Douglas share parameter α in half. This means a decline in the labor share of about 8 percentage points, 3 percentage points more than the decline documented by Karabarbounis and Neiman [2013] for a wide range of countries. The increased amount of labor that can be used to produce consumption raises per capita consumption, measured in units of labor. There is only a small increase in the Gini coefficient for managers, but the fact that the factor share of workers has declined substantially implies that earnings inequality has increased—worker earnings measured in units of labor are constant and equal to 1.

The effects of an increase in both the diffusion coefficient σ and the managerial learning rates $\gamma = \beta$ are shown in the third column of Table 1. Both of these changes improve the extent $\kappa - \mu$ to which learning from others enhances the growth rate of the aggregate stock of human capital. The learning rates $\gamma = \beta$ increase from .05125 to .06 while the diffusion coefficient doubles. The result is a more significant contribution $\kappa - \mu$ to human capital growth of 2% per annum. At the same time, the distribution of managerial human capital and wealth becomes more dispersed. The Gini for managerial profits rises significantly and profits of the 1% active manager increase by 85%. A further decline of overhead labor input requirements is shown in the fourth column of Table 1, with ϕ again cut in half. The main effect of this is again to lower the factor share of labor, which is now only 3.5 percentage points above the Cobb-Douglas share parameter. The effect of a small learning advantage for active managers is shown in the fifth column of Table 1. This makes it more costly for managers to become inactive, and as a result the least productive active manager now employs less than one unit of variable labor. Although the learning differential is very small, the flow gain from learning $S(\beta) - \beta U$ is now significantly lower than the analogous flow gain $S(\gamma) - \gamma V(b)$ for a marginal but still active manager. A somewhat larger learning differential would drive these gains to zero and some inactive managers would not choose to learn, in effect becoming workers. Small further changes in the diffusion coefficient σ also significantly affect the balanced growth path. An increase from $\sigma = 0.100$ to $\sigma = 0.125$ lowers the tail index of the managerial human capital distribution from $\zeta = 2$ to $\zeta = 1.6$, with the attendant rise in both the tail Gini $1/(2\zeta - 1)$ and the other Gini coefficients. In this last example, the present value of earnings of the 1% active manager are 180 times what they are for a worker. The increase in inequality in this economy relative to the first column of Table

1 is illustrated by the Lorenz curves for the earnings from managing workers displayed in Figure 7.

Table 1 Benchmark Specifications

	$\phi = 2$	$\phi = 1$	$\phi = 1$	$\phi = 0.5$	$\phi = 0.5$	$\phi = 0.5$
learning rate inactive managers	0.051	0.051	0.060	0.060	0.058	0.060
active managers	0.051	0.051	0.060	0.060	0.060	0.060
diffusion coefficient	0.050	0.050	0.100	0.100	0.100	0.125
growth rate differential	0.008	0.008	0.020	0.020	0.020	0.025
tail index	3.000	3.000	2.000	2.000	2.000	1.600
Gini all managers	0.500	0.519	0.593	0.593	0.603	0.673
active managers	0.298	0.307	0.485	0.493	0.528	0.635
mentors	0.200	0.200	0.333	0.333	0.333	0.455
variable labor at exit	2.726	1.849	1.667	1.292	0.853	1.127
marginal supervisor	3.199	3.578	2.217	2.330	2.077	1.356
marginal mentor	13.769	11.064	15.516	10.771	10.230	32.285
factor price manager at exit	2.726	1.849	1.667	1.292	0.853	1.127
flow gain from learning at exit	0.135	0.680	0.170	0.402	0.753	0.052
inactive manager	0.135	0.680	0.170	0.402	0.239	0.052
investment in learning	0.582	0.702	1.081	1.177	1.216	1.361
consumption per capita	1.367	1.552	1.568	1.655	1.651	1.691
earnings* inactive managers	1.135	1.680	1.170	1.402	1.239	1.052
median active manager	4.456	5.851	4.343	4.663	4.506	3.970
1% active manager	62.689	69.110	128.68	128.91	135.94	180.06
number of workers	0.800	0.800	0.800	0.800	0.800	0.800
active managers	0.105	0.112	0.108	0.115	0.116	0.103
supervisors	0.095	0.088	0.092	0.085	0.084	0.097
mentors	0.005	0.012	0.008	0.015	0.016	0.003
overhead labor/all labor	0.236	0.126	0.121	0.065	0.066	0.058
factor share of workers	0.585	0.515	0.510	0.483	0.485	0.473
of labor	0.654	0.572	0.569	0.535	0.535	0.531

($\alpha = 0.50$, $\delta = 0.04$, $\rho = 0.03$; all market values in units of labor)

*annuity value of earnings: $(\rho + \delta)U/w$ and $(\rho + \delta)V(z)/w$

In all scenarios shown in Table 1, the number of inactive managers is large. These are individuals with the ability to learn who do not manage workers but supply labor themselves. In the equilibria considered here, the value of learning for these managers is positive and all of them try to become active managers. They have to be matched with active managers to learn, and so most active managers teach inactive managers. Only a small fraction of the population of active managers teaches other active managers. A reduction in the random death rate would result in fewer inactive managers, freeing up active managers to teach other active managers.

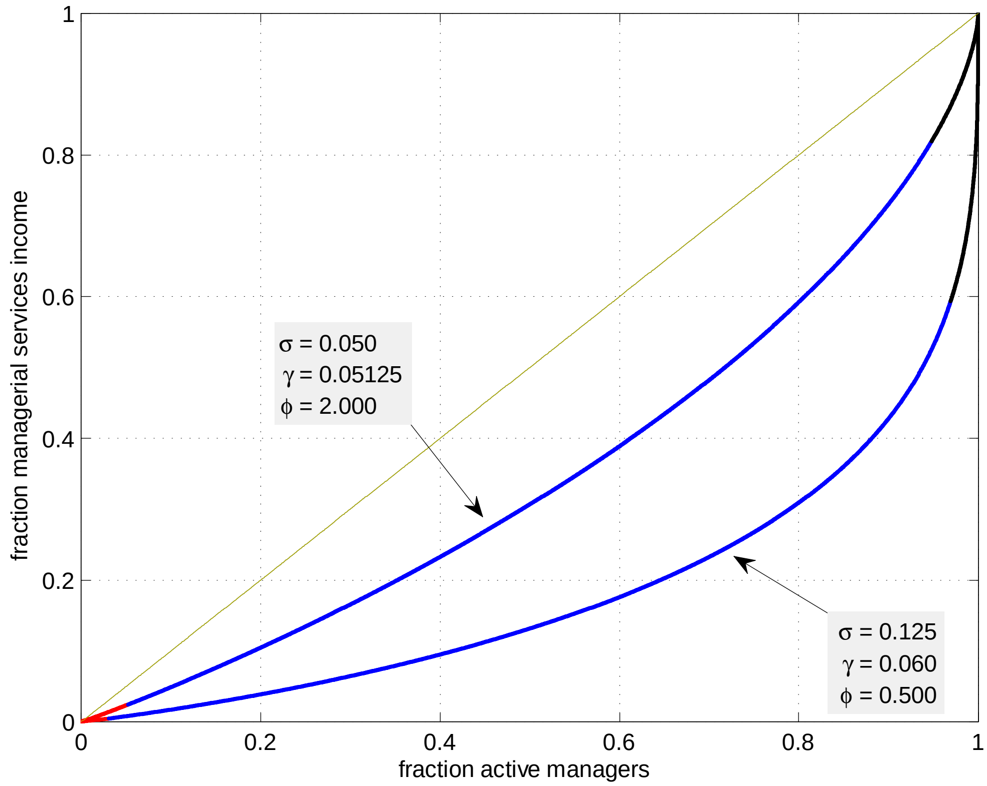

FIGURE 7 Lorenz Curves for Active Managers

8. CONCLUDING REMARKS

One way to leverage the useful knowledge of a manager is to assign more workers to follow the manager's instructions, directly as in Lucas [1978], or indirectly via hierarchies, as in Garicano [2000]. This paper has explored another: assign agents to learn from the manager, so they no longer need instructions.

The growth rates and stationary distributions of productivity in an economy in which knowledge is traded in competitive markets and in an economy with random copying are very similar. In both cases, the initial distribution of knowledge matters. But there is a sense in which it is harder to change growth rates in the competitive economy than it is in an economy with random imitation. With random imitation, the flow of agents who can enter the right tail of the productivity distribution is essentially unlimited: an increase in the number of entrepreneurs, say because entry costs decline, will increase that flow and increase the growth rate of the economy (Luttmer [2012]). This is not the case in the economy with a Leontief teaching technology described here. In this economy, the flows into the right tail are limited by the teaching capacity of those already in that right tail. An increase in the number of entrepreneurs will just crowd out incumbent managers trying to learn from more productive incumbent managers.

A Solving the Bellman Equation

Write $v = 1/w^{\alpha/(1-\alpha)}$ so that flow revenues are ve^z. Write $\theta = \mu - \kappa$ and recall that $U = S(\beta)/(\rho+\beta)$. The Bellman equation then reduces to

$$\begin{aligned}
(\rho+\gamma)V(z) &= ve^z - w\phi + S(\gamma) + \theta \mathrm{D}V(z) + \frac{1}{2}\sigma^2\mathrm{D}^2V(z), \quad z \in (b,x), \\
(\rho-\beta)V(z) &= ve^z - w\phi - S(\beta) + \theta \mathrm{D}V(z) + \frac{1}{2}\sigma^2\mathrm{D}^2V(z), \quad z \in (x,y), \\
(\rho-\gamma)V(z) &= ve^z - w\phi - S(\gamma) + \theta \mathrm{D}V(z) + \frac{1}{2}\sigma^2\mathrm{D}^2V(z), \quad z \in (y,\infty),
\end{aligned} \tag{47}$$

with the boundary conditions

$$V(b) = \frac{S(\beta)}{\rho+\beta}, \quad V(x) = \frac{S(\beta)+S(\gamma)}{\beta+\gamma}, \quad V(y) = \frac{S(\gamma)-S(\beta)}{\gamma-\beta} \tag{48}$$

and a condition that requires $V(z)$ to behave like e^z for large z. Furthermore, the value function has to be differentiable at b, x and y. At b this means $\mathrm{D}V(b) = 0$ and at x and y this requires left- and right-derivatives to be the same. Note that the effective discount rates $\rho+\gamma$, $\rho-\beta$ and $\rho-\gamma$ are positive by assumption.

The solutions of the homogeneous equation $\lambda V(z) = \theta \mathrm{D}V(z) + \frac{1}{2}\sigma^2\mathrm{D}^2V(z)$ are $\{e^{\psi_-(\lambda)z}, e^{\psi_+(\lambda)}\}$, where $\psi_\pm(\lambda)$ solves the quadratic $\lambda = \theta\psi + \frac{1}{2}\sigma^2\psi^2$,

$$\psi_\pm(\lambda) = -\frac{\theta}{\sigma^2} \pm \sqrt{\left(\frac{\theta}{\sigma^2}\right)^2 + \frac{\lambda}{\sigma^2/2}}.$$

Since $\theta < 0$ and $\lambda > 0$, this satisfies $\psi_+(\lambda) > 0 > \psi_-(\lambda)$. Particular solutions of (47) are present values calculated as if flow revenues continue forever. The resulting general solutions are

$$
\begin{aligned}
V(z) &= \frac{ve^z}{\rho+\gamma-\theta-\frac{1}{2}\sigma^2} - \frac{\phi w - S(\gamma)}{\rho+\gamma} + F_+ e^{\psi_+(\rho+\gamma)(z-b)} + F_- e^{\psi_-(\rho+\gamma)(z-b)},\ z \in (b,x), \\
V(z) &= \frac{ve^z}{\rho-\beta-\theta-\frac{1}{2}\sigma^2} - \frac{\phi w + S(\beta)}{\rho-\beta} + G_+ e^{\phi_+(\rho-\beta)(z-x)} + G_- e^{\psi_-(\rho-\beta)(z-x)},\ z \in (x,y), \\
V(z) &= \frac{ve^z}{\rho-\gamma-\theta-\frac{1}{2}\sigma^2} - \frac{\phi w + S(\gamma)}{\rho-\gamma} + H_+ e^{\psi_+(\rho-\gamma)(z-y)} + H_- e^{\psi_-(\rho-\gamma)(z-y)},\ z \in (y,\infty).
\end{aligned}
$$

This describes the solution as a function of the thresholds $[b, x, y]$ and the coefficients $[F_+, F_-, G_+, G_-, H_+, H_-]$. Viewed as a function of $z-b$ and taking as given $x-b$ and $y-b$, the value function is linear in e^b, the coefficients $[F_+, F_-, G_+, G_-, H_+, H_-]$ and the parameters $[S(\beta), S(\gamma)]$. Instead of solving for $x-b$ and $y-b$, one can take these values as given and back out the implicit tuition schedule $[S(\beta), S(\gamma)]$. The requirement that $V(z)$ behaves like e^z for large z forces $H_+ = 0$. There are three additional boundary conditions in (48) and differentiability at b, x and y yields five more. So there are nine equations to be solved for the nine parameters e^b, $[S(\beta), S(\gamma)]$ and $[F_+, F_-, G_+, G_-, H_+, H_-]$. These are all linear equations.

B Stationary Densities

The general solutions to the second-order differential equations (32)-(34), with the boundary condition $m(b) = 0$ imposed, are given by

$$
m(z) = \begin{cases} A\left(e^{\xi_+(z-b)} - e^{\xi_-(z-b)}\right), & z \in (b,x) \\ B_+ e^{-\omega_+(z-x)} + B_- e^{-\omega_-(z-x)}, & z \in (x,y) \\ C_+ e^{-\zeta_+(z-y)} + C_- e^{-\zeta_-(z-y)}, & z \in (y,\infty) \end{cases}
$$

where

$$
\xi_\pm = \frac{\mu-\kappa}{\sigma^2} \pm \sqrt{\left(\frac{\mu-\kappa}{\sigma^2}\right)^2 + \frac{\gamma}{\sigma^2/2}},
$$

$$
\omega_\pm = -\frac{\mu-\kappa}{\sigma^2} \pm \sqrt{\left(\frac{\mu-\kappa}{\sigma^2}\right)^2 - \frac{\beta}{\sigma^2/2}}, \quad \zeta_\pm = -\frac{\mu-\kappa}{\sigma^2} \pm \sqrt{\left(\frac{\mu-\kappa}{\sigma^2}\right)^2 - \frac{\gamma}{\sigma^2/2}}.
$$

The fact that $\mu-\kappa$ is negative and γ positive implies that $\xi_+ > 0 > \xi_-$. Condition (39) implies $\zeta_+ > \zeta_- > 0$ and $\beta \in (0,\gamma)$ ensures that $\omega_+ > \omega_- > 0$ as well.

The boundary condition (31) becomes

$$
\frac{1}{2}\sigma^2(\xi_+ - \xi_-)A = \beta\left(\left(\frac{1-e^{-\omega_+(y-x)}}{\omega_+}\right)B_+ + \left(\frac{1-e^{-\omega_-(y-x)}}{\omega_-}\right)B_-\right). \tag{49}
$$

The required continuity of $m(z)$ and $\mathrm{D}m(z)$ at x and y amounts to, respectively,

$$\begin{bmatrix} 1 & -1 \\ -\xi_+ & \xi_- \end{bmatrix} \begin{bmatrix} Ae^{\xi_+(x-b)} \\ Ae^{\xi_-(x-b)} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \omega_+ & \omega_- \end{bmatrix} \begin{bmatrix} B_+ \\ B_- \end{bmatrix}, \tag{50}$$

$$\begin{bmatrix} 1 & 1 \\ \zeta_+ & \zeta_- \end{bmatrix} \begin{bmatrix} C_+ \\ C_- \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \omega_+ & \omega_- \end{bmatrix} \begin{bmatrix} B_+e^{-\omega_+(y-x)} \\ B_-e^{-\omega_-(y-x)} \end{bmatrix}. \tag{51}$$

The normalization (38) pins down the scale of $[A, B_+, B_-, C_+, C_-]$. The fact that $\omega_+ > \omega_- > 0$ implies that the matrix on the right-hand side of (50) is non-singular. So (50) determines $[B_+, B_-]/A$ in terms of $x-b$. Given $[B_+, B_-]/A$, condition (49) is a nonlinear equation that can be solved for $y-x$. Condition (51) then determines $[C_+, C_-]/A$.

When (39) holds with equality, $\zeta_+ = \zeta_- = \zeta = -(\mu-\kappa)/\sigma^2 = \gamma/(\sigma^2/2)$ and the matrix on the left-hand side of (51) is singular. But we are only interested in the density $C_+e^{-\zeta_+(z-y)} + C_-e^{-\zeta_-(z-y)}$ and its derivative $-(C_+\zeta_+e^{-\zeta_+(z-y)} + C_-\zeta_-e^{-\zeta_-(z-y)})$, and these converge as $\zeta_+ \downarrow \zeta$ and $\zeta_- \uparrow \zeta$. So see this, note that

$$\begin{aligned} &\begin{bmatrix} e^{-\zeta_+(z-y)} & e^{-\zeta_-(z-y)} \\ -\zeta_+e^{-\zeta_+(z-y)} & -\zeta_-e^{-\zeta_-(z-y)} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \zeta_+ & \zeta_- \end{bmatrix}^{-1} \\ &\quad = \frac{1}{\zeta_+ - \zeta_-} \begin{bmatrix} e^{-\zeta_+(z-y)} & e^{-\zeta_-(z-y)} \\ -\zeta_+e^{-\zeta_+(z-y)} & -\zeta_-e^{-\zeta_-(z-y)} \end{bmatrix} \begin{bmatrix} -\zeta_- & 1 \\ \zeta_+ & -1 \end{bmatrix} \\ &\quad \to \begin{bmatrix} 1+\zeta(z-y) & -(z-y) \\ -\zeta^2(z-y) & \zeta(z-y)-1 \end{bmatrix} e^{-\zeta(z-y)} \end{aligned}$$

as $\zeta_+ \downarrow \zeta$ and $\zeta_- \uparrow \zeta$. It then follows from (51) that the density $C_+e^{-\zeta_+(z-y)} + C_-e^{-\zeta_-(z-y)}$ and its derivative converge as $\zeta_+ \downarrow \zeta$ and $\zeta_- \uparrow \zeta$, even though the coefficients $[C_+, C_-]$ do not.

It remains to show that there is a unique $y-x$ that solves (49)-(50). Eliminating $[B_+, B_-]/A$ from these two conditions gives

$$\begin{aligned} \frac{1}{2}\sigma^2\left(\frac{\omega_+ - \omega_-}{\beta}\right) &= \left(\left(\frac{1-e^{-\omega_-(y-x)}}{\omega_-}\right)\omega_+ - \left(\frac{1-e^{-\omega_+(y-x)}}{\omega_+}\right)\omega_-\right)\frac{e^{\xi_+(x-b)} - e^{\xi_-(x-b)}}{\xi_+ - \xi_-} \\ &\quad + \left(\frac{1-e^{-\omega_-(y-x)}}{\omega_-} - \frac{1-e^{-\omega_+(y-x)}}{\omega_+}\right)\frac{\xi_+e^{\xi_+(x-b)} - \xi_-e^{\xi_-(x-b)}}{\xi_+ - \xi_-}. \end{aligned}$$

The right-hand side is zero at $y-x=0$ and strictly increasing in $y-x$. To guarantee a unique $y-x$, it suffices to show that the right-hand side dominates the left-hand side for large $y-x$. Using the fact that $\omega_+\omega_- = \beta/(\sigma^2/2)$ and $\omega_+ + \omega_- = -(\xi_+ + \xi_-)$, this requirement can be written as

$$1 < -(\xi_+ + \xi_-)\frac{e^{\xi_+(x-b)} - e^{\xi_-(x-b)}}{\xi_+ - \xi_-} + \frac{\xi_+ e^{\xi_+(x-b)} - \xi_- e^{\xi_-(x-b)}}{\xi_+ - \xi_-} = \frac{\xi_+ e^{\xi_-(x-b)} - \xi_- e^{\xi_+(x-b)}}{\xi_+ - \xi_-}.$$

The right-hand side is a strictly convex function of $x - b$, equal to 1 with a zero slope at $x - b = 0$. Thus the inequality holds for any positive $x - b$, and hence (49)-(50) has a unique solution for $y - x$ for any positive $x - b$.

Note that for $z \in (b, x)$

$$\frac{m(z)}{\mathrm{D}m(b)} = \frac{e^{\xi_+(z-b)} - e^{\xi_-(z-b)}}{\xi_+ - \xi_-}$$

and hence

$$\int_b^x \frac{m(z)\mathrm{d}z}{\mathrm{D}m(b)} = \frac{\frac{e^{\xi_+(x-b)}-1}{\xi_+} - \frac{e^{\xi_-(x-b)}-1}{\xi_-}}{\xi_+ - \xi_-}$$

is increasing in $x - b$.

References

[1] Alvarez, F.E., F.J. Buera, and R.E. Lucas, Jr., "Models of Idea Flows," NBER working paper 14135 (2008).

[2] Becker, G.S. and N. Tomes, "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy*, vol. 87, no. 6 (1979), 1153-1189.

[3] Beaudry, P. and P. Francois, "Managerial Skill Acquisition and the Theory of Economic Development," *Review of Economic Studies*, vol. 77 (2010), 90-126.

[4] Bils, M. and P.J. Klenow, "Does Schooling Cause Growth?" *American Economic Review*, vol. 90, no. 5 (2000), 1160-1183.

[5] Benhabib, J. and A. Bisin and S. Zhu, "The Distribution of Wealth and Fiscal Policy in Economies with Finitely Lived Agents," *Econometrica*, vol. 79, no. 1 (2011), 123-157.

[6] Bramson, M., *Convergence of Solutions of the Kolmogorov Equation to Travelling Waves*, Memoirs of the American Mathematical Society, vol. 44, no. 285 (1983).

[7] Cagetti, M. and C. DeNardi,"Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, vol. 114, no 5 (2006), 835-870.

[8] Castañeda, A., J. Díaz-Giménez, and J.-V. Ríos-Rull, "Accounting for the U.S. Earnings and Wealth Inequality," vol. 111, no. 4 (2003), 818-857.

[9] Cavalli-Sforza, L.L. and M.W. Feldman, *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton University Press (1981).

[10] Chari, V.V. and H. Hopenhayn, "Vintage Human Capital, Growth, and the Diffusion of New Technology," *Journal of Political Economy*, vol. 99, no.6 (1991), 1142-1165.

[11] Chatterjee, S. and E. Rossi-Hansberg, "Spin-Offs and the Market for Ideas," *International Economic Review*, vol. 53, no. 1 (2012), 53-93.

[12] Cox, D.R. and H.D. Miller, *The Theory of Stochastic Processes*, Chapman and Hall, London, 1965.

[13] Díaz-Giménez, J., A. Glover, and J.-V. Ríos-Rull, "Facts on the Distribution of Earnings, Income, and Wealth in the United States: 2007 Update," *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 34, no. 2 (2011), 2-31.

[14] Eeckhout, J. and B. Jovanovic, "Knowledge Spillovers and Inequality," *American Economic Review*, vol. 92, no. 5 (2002), 1290-1307.

[15] Franco, A. and D. Filson, "Spin-outs: Knowledge Diffusion through Employee Mobility," *RAND Journal of Economics*, vol. 37, no. 4 (2006), 841-860.

[16] Fisher, R.A., "The Wave of Advance of Advantageous Genes," *Annals of Eugenics*, vol. 7 (1937), 355-369.

[17] Garicano, L. "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, vol. 108, no. 5 (2000) , 874-904.

[18] Hassler, J. and J.V. Rodríguez Mora, "Intelligence, Social Mobility, and Growth," *American Economic Review*, vol. 90, no. 4 (2000), 888-908.

[19] Houthakker, H.S., "The Pareto Distribution and the Cobb-Douglas Production Function in Activity Analysis," *Review of Economic Studies*, vol. 23, no. 1 (1955-1956), 27-31.

[20] Huggett, M., G. Ventura, and A. Yaron, "Sources of Lifetime Inequality," *American Economic Review*, vol. 101, no. 7 (2011), 2923-2954.

[21] Jones, C.I.,"Growth and Ideas," *Handbook of Economic Growth Volume 1B*, edited by P. Aghion and S.N. Durlauf (2005), Elsevier.

[22] Jovanovic, B. and G.M. MacDonald, "Competitive Diffusion," *Competitive Diffusion*, vol. 102, no. 1 (1994), 24-52.

[23] Karabarbounis, L. and B. Neiman, "The Global Decline of the Labor Share," University of Chicago (2013).

[24] Kolmogorov, A.N., I.G. Petrovskii and N.S. Piskunov, "A Study of the Diffusion Equation with Increase in the Amount of Substance, and its Application to a Biological Problem," in *Selected Works of A.N. Kolmogorov I*, V.M. Tikhomirov, Editor, Kluwer Academic Publishers 1991 (1937).

[25] Kortum, S., "Research Patenting and Technological Change," *Econometrica*, vol. 65, no. 6 (1997), 1389-1419.

[26] König, M., J. Lorenz, and F. Zilibotti, "Innovation vs Imitation and the Evolution of Productivity Distributions," CEPR Discussion Papers 8843 (2012).

[27] Lagakos, D., B. Moll, T. Porzio, N. Qian, "Experience Matters: Human Capital and Development Accounting," working paper (2013).

[28] Lucas, R.E., Jr., "On the Size Distribution of Business Firms," *Bell Journal of Economics*, vol. 9, no. 2 (1978), 508-523.

[29] Lucas, R.E., Jr.,"Ideas and Growth", *Economica*, vol. 76 (2009), 1-19.

[30] Lucas, R.E., Jr. and B. Moll, "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, forthcoming (2013).

[31] Luttmer, E.G.J., "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics*, vol. 122, no. 3 (2007), 1103-1144.

[32] Luttmer, E.G.J., "Eventually, Noise and Imitation Implies Balanced Growth," *Minneapolis Federal Reserve Working Paper 699* (2012).

[33] Manuelli, R.E. and A. Seshadri, "Human Capital and the Wealth of Nations," working paper (2010).

[34] McKean, H.P., "Applications of Brownian motion to the Equation of Kolmogorov-Petrovskii-Piskunov," *Communications in Pure and Applied Mathematics*, vol. 28 (1975), 323—331; "Erratum," vol. 29 (1976), 553-554.

[35] Nirei, M. and W. Souma, "A Two-Factor Model of Income Distribution Dynamics," *Review of Income and Wealth*, vol. 53, no. 3 (2007), 440-459.

[36] Perla, J. and C. Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, forthcoming (2013).

[37] Rustichini, A. and J.A. Schmitz, Jr., "Research and Imitation in Long-Run Growth," *Journal of Monetary Economics*, vol. 27 (1991), 271-292.

[38] Staley, M., "Growth and the Diffusion of Ideas," *Journal of Mathematical Economics*, vol. 47 (2011), 470-478.

[39] Volpert, A.I., V.A. Volpert, and V.A. Volpert, *Traveling Wave Solutions of Parabolic Systems*, Translations of Mathematical Monographs vol. 140, American Mathematical Society (1994).