Federal Reserve Bank of Minneapolis
Research Department Staff Report 509

March 2015

An Assignment Model of Knowledge Diffusion and Income Inequality*

Erzo G.J. Luttmer

University of Minnesota
and Federal Reserve Bank of Minneapolis

ABSTRACT

Randomness in individual discovery disperses productivities, whereas learning from others keeps productivities together. Long-run growth and persistent earnings inequality emerge when these two mechanisms for knowledge accumulation are combined. This paper considers an economy in which those with more useful knowledge can teach others, with competitive markets assigning students to teachers. In equilibrium, students with an ability to learn quickly are assigned to teachers with the most productive knowledge. This sorting on ability implies large differences in earnings distributions conditional on ability, as shown using explicit formulas for the tail behavior of these distributions.

Keywords: Knowledge diffusion; Growth; Income inequality
JEL Classification: J2, O1, O3, O4

---

*A key step in this paper was reported earlier in Luttmer [2012]. I thank seminar participants at Toulouse, Stanford, the Federal Reserve Bank of Philadelphia, NYU, Cowles, HKUST, the North American Meeting of the Econometric Society in Minneapolis, UCL, Chicago, and UBC for helpful comments. The Toulouse School of Economics provided generous financial support and hospitality while some of the work on this paper was done. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

Useful knowledge can be obtained by processing available information and acquiring new information. The amount of such information is typically enormous, although very little of it may be relevant for the job at hand. A potentially much quicker way to acquire useful knowledge is to learn from others. Some such learning can be by simple observation. But often it requires the participation of someone who is already in possession of the more useful knowledge. In many environments, there are teachers and students engaged in the transfer of useful knowledge.

This paper describes a model of long-run growth in which useful knowledge accumulates as a result of both individual discovery and learning from others. Consumption is produced by teams of managers and workers. Those who have somehow acquired useful knowledge can be managers, and everyone who is not a manager can supply labor. The productivity of a manager in charge of a team of workers evolves over time as the result of a process of random discovery that tends to cause managerial productivities to diverge. In addition, while producing consumption, managers and workers can be assigned to more knowledgeable managers as students. Teaching managers can teach one student at a time, and teachers cannot be students at the same time. The assignment of students to teachers is determined in competitive markets, where students pay teachers for the time it takes to learn what their teachers know.

Everyone is born with a certain ability to learn, but without any useful knowledge. Newborn individuals begin life as workers. Sufficiently fast learners will choose to pay teachers and try to learn something that allows them to start as a manager. This takes an uncertain amount of time, and then their own managerial discovery process begins. Depending on how productive they are at any point in time, they may decide to teach or learn more from others. Managers whose productivities lag behind earn very little and may find it optimal to quit and become workers again. They lose whatever useful knowledge they do have and need to learn from others if they want to become managers again. Because of their comparative advantage in learning, the equilibrium assigns fast learners to the most productive managers in the economy.

At any point in time, the state of the economy is a count of how many managers there are with different abilities to learn and different levels of productivity. Learning from others prevents the distribution of productivities from spreading out indefinitely, and the combination of individual discovery and learning from others causes the distribution of productivities to continuously shift to the right. Stable growth emerges in the long run. Although the allocation of workers to managers and students to teachers is determined

by a more or less standard competitive equilibrium, the state variable in a large economy is an infinite-dimensional capital stock, and the economy turns out to have a continuum of balanced growth paths. These paths are indexed by the growth rate of the economy, and higher growth rates correspond to productivity distributions with thicker right tails, which make learning from others more productive. The long-run growth rate depends crucially on the initial distribution of productivity in the economy. A definite prediction for the long-run growth rate emerges when the initial distribution of productivities has a bounded support, as would be the case in any finite economy: it is the lowest rate among those that are consistent with balanced growth. This growth rate is increasing in the variance of individual discovery rates, in the rate at which the fastest learners can learn, and in their life expectancy. Income inequality also increases with the variance of individual discovery rates. But a faster learning rate makes for a thinner right tail because less knowledgeable individuals can catch up more quickly.

In this economy, ex post earnings inequality is the result of several factors. Because teaching is one-on-one, not all students can be matched with the most knowledgeable teachers. Even among students with the same ability to learn, some will study with more knowledgeable teachers than others. They will have to pay those teachers more, and their subsequent incomes will be higher. Equilibrium only requires that students of the same type are ex ante indifferent across such assignments. The time it takes for students with the same ability and assignment to learn is also uncertain, creating more unequal outcomes. For managers, the individual discovery process generates even more dispersion.

Overall, the factor that dominates the determination of inequality in the right tail of the income distribution is variation in the ability to learn itself. Equilibrium earnings distributions have Pareto-like right tails. An explicit formula obtained for an economy with two types of learners shows that the right tail of the earnings distribution of fast learners is discretely thicker than the right tail of the earnings distribution of slow learners, even if the difference in learning speeds is negligible. Perfect sorting across teachers means that fast learners are assigned to and learn to become like teachers with the most productive knowledge. Instead, slow learners who acquire useful knowledge enter the population of managers at the bottom of the productivity distribution. They can make up for this with successful individual discoveries, but the odds that this pushes them into the right tail of the managerial productivity distribution are extremely small. This shows how the competitive assignment of students to teachers can greatly magnify the inequality in ex post outcomes among individuals with small differences in ability.

If they are not too abundant, fast learners are better off than slow learners also on an ex ante basis. But, unlike the ex post outcomes, the ex ante rents associated with learning ability are continuous: they vanish when differences in ability become negligible.

As in the data, the future income prospects of individuals in this economy are to some extent predictable. The conditional distribution of income growth depends on how fast a worker or a manager can learn. This contrasts with pure random growth models of cities (Gabaix [1999]) or firms (Luttmer [2007]), and is similar to the firm growth model proposed in Luttmer [2011]. Although data on how fast people can learn are typically not available, there are observable characteristics that correlate with future income growth. From the perspective of the model in this paper, these observable characteristics can be seen as proxies for the ability to learn from others. But because there is some indeterminacy in who learns from whom, observable characteristics that are unrelated to current productivity, and that do not reflect any kind of ability, may also predict future income growth.

Related Literature As in the span of control model of Lucas [1978], individuals who are more productive as managers sort into managing those who are less productive. In Lucas [1978], this sorting is instantaneous and mediated by the relative wages of managers and workers. Here it takes time to learn something productive, and quitting as a manager results in a destruction of managerial knowledge. The factor price for managerial services and the wage of a worker still mediate the sorting process, but becoming a manager is not instantaneous, and managers solve a stopping problem to decide when to quit.

This paper is closely related to and motivated by a more recent literature that gives explicit and joint accounts of productive heterogeneity and long-run growth. In Luttmer [2007, 2012], Alvarez, Buera, and Lucas [2008], Lucas [2009], König, Lorenz, and Zilibotti [2012], Lucas and Moll [2014], and Perla and Tonetti [2014], agents randomly find others who are more knowledgeable and imitate instantaneously when they do. With the exception of Luttmer [2007, 2012] and König, Lorenz, and Zilibotti [2012], these papers do not determine the long-run growth rate of the economy but only relate it to an assumption about the thickness of the right tail of the stationary distribution of knowledge. Kortum [1997] pioneered the use of sources of ideas with Pareto-like distributions to generate unbounded growth.

An early version of the random meeting and imitation structure was developed by Jovanovic and Rob [1989]. Randomness in who meets whom limits the speed at which

knowledge can diffuse. Here, the delay in knowledge transmission is not finding others who know more—they are easy to find, in an instant. But it takes an uncertain amount of time to learn from those individuals. Moreover, learning is not just imitation but a process that requires the input of both the individual trying to learn and the individual who has the more useful knowledge. This second feature eliminates an externality. As argued forcefully in Boldrin and Levine [2008], imitation externalities are not an essential ingredient of the process of long-run growth. The results in this paper imply that internalized knowledge transfers and imitation externalities are hard to distinguish based only on their implications for income distributions.

Knowledge transfer is internalized in Chari and Hopenhayn [1991], where unskilled workers can become skilled in using the technology of a particular vintage not only on their own, but also by working in a team with workers who are already skilled in operating that technology. This yields a model of endogenous technology adoption, but not one that produces analytically tractable results for the distribution of knowledge in the economy.[1] In the literature on the international diffusion of knowledge, learning by imitating trading partners involves externalities whereas knowledge flows inside multinational firms may not.[2] For the most part, this literature does not give concrete answers to the question addressed in this paper: how precisely do such knowledge flows determine long-run growth and inequality?

In a recent paper, Jovanovic [2014] does study this question in an overlapping generations economy in which the young are assigned to the old to both produce consumption and learn. What young individuals learn depends not only on what their teachers know but also on an average of what all teachers know. This average effect is an externality that is essential for growth to happen in the model, and an explicit interpretation is not immediately clear. Here there is no such average effect, but teachers with relatively unproductive knowledge can become students instead. Together with randomness in the individual processes of discovery, this sorting of individuals with different levels of knowledge into teaching or learning is what makes long-run growth possible. The fact that managers can exit to become workers also makes the supplies of students and teachers respond to interest rates and beliefs, whereas the assignment problem in Jovanovic [2014] is essentially static.

[1] See Beaudry and Francois [2010] for a recent model, along similar lines, of (a one-time) technology adoption in which unskilled workers learn by working together with skilled managers.

[2] For recent contributions to an extensive literature, see Antras, Garicano, and Rossi-Hansberg [2006], Nocke and Yeaple [2008], Burstein and Monge-Naranjo [2009], Alvarez, Buera, and Lucas [2013], Ramondo and Rodriguez-Clare [2013], and Keller and Yeaple [2013].

Small differences in talent lead to large income differences in the superstar economy of Rosen [1981] and the assignment models of Rosen [1982], Gabaix and Landier [2008] and Terviö [2008]. These are static models in which individual talent has a multiplicative effect on the quality of some consumable output. Here, talent means an ability to learn from others, and the model is explicitly dynamic. The mechanism by which small differences in the ability generate inequality is more akin to the Matthew effect of Merton [1968]: those with the ability to learn quickly shall be given more valuable learning opportunities.

Outline of the Paper Section 2 describes the economy and characterizes the equilibrium tuition schedules that can arise when students and teachers are matched in competitive markets. Section 3 specializes to an economy with individuals who have either high or low abilities to learn from others, and describes the conditions for a balanced growth path. The key insights are Figure 1 and equations (30)-(31). Section 4 illustrates the workings of this economy quantitatively, and Section 5 concludes.

2. An Economy with Competitive Knowledge Diffusion

Consider an economy with a unit measure of dynastic households whose preferences over consumption flows $\{C_t\}_{t\geq 0}$ are determined by the utility function

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

The dynastic subjective discount rate ρ is positive. Every dynastic household is characterized by an immutable ability to learn from others $\lambda \in \Lambda \subset (0,\infty)$. An individual household member dies randomly at a positive rate δ and is then immediately replaced by a successor, with the same ability to learn. The set of household types Λ is finite, and the measure of households of type $\lambda \in \Lambda$ is denoted by $M(\lambda)$. Newborn household members begin life as workers who can supply one unit of labor per unit of time. Over time, workers can learn to become managers, and managers can learn to become more productive managers, in a manner described in detail below. At any given point in time t, there will be a measure $M_t(\lambda, z) \leq M(\lambda)$ of household members who are managers with productivity state variables in $(-\infty, z]$. The remaining $M(\lambda) - M_t(\lambda,\infty)$ type-λ household members are workers. The random death rate δ can be interpreted more generally as a random exit rate that turns managers into workers again—for example, the

project a manager works on may fail and make the knowledge acquired by the manager obsolete.

There is no aggregate uncertainty and markets are complete. The resulting risk-free interest rate satisfies $r_t = \rho + \mathrm{D}C_t/C_t$.[3]

2.1 The Consumption Sector

A manager in productivity state z can hire l units of labor to produce $(e^z/(1-\alpha))^{1-\alpha}(l/\alpha)^\alpha$ units of consumption per unit of time. Wages at time t are w_t, measured in units of consumption, and so a manager in state z earns

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

from producing consumption. Here, v_t can be interpreted as the factor price of one unit of managerial services. The unit cost function for this Cobb-Douglas technology is $v_t^{1-\alpha} w_t^{\alpha}$, and so (1) implies $v_t = 1/w_t^{\alpha/(1-\alpha)}$.

Managers cannot supply labor when they are employing workers, and they must incur a fixed cost of $\phi \geq 0$ units of overhead labor per unit of time to remain active as managers. The wage a manager could earn by becoming a worker again is an opportunity cost that can be viewed as an additional fixed cost associated with being a manager. The amount of variable labor that attains $v(t,z)$ is $(v_t e^z/w_t)\alpha/(1-\alpha)$. Aggregating the output of consumption across managers at the wage that clears the labor market gives

$$C_t = \left(\frac{H_t}{1-\alpha}\right)^{1-\alpha} \left(\frac{1-(1+\phi)N_t}{\alpha}\right)^{\alpha}, \tag{2}$$

where H_t and N_t are defined by

$$H_t = \sum_{\lambda\in\Lambda} \int e^z M_t(\lambda, \mathrm{d}z), \quad N_t = \sum_{\lambda\in\Lambda} M_t(\lambda, \infty). \tag{3}$$

The total supply of labor is $1 - N_t$ and $1 - (1+\phi)N_t$ of this supply can be used as variable labor to produce consumption. The market-clearing wage can be inferred from the fact that, as usual, the compensation of variable labor is a fraction α of output,

$$w_t \left(\sum_{\lambda\in\Lambda} M(\lambda) - (1+\phi)N_t \right) = \alpha C_t. \tag{4}$$

[3]The perfect consumption insurance implications of this model are extreme but help simplify the analysis. Plausible relations between productivity and wealth are absent in this economy. See Nirei and Souma [2007] and Benhabib, Bisin, and Zhu [2011] for recent models that generate thick-tailed income and wealth distributions.

Together, (1)-(4) determine C_t, H_t, N_t, v_t, and w_t in terms of the measure $\{M_t(\lambda, z) : \lambda \in \Lambda\}$ of managers. This measure is the state variable for this economy. Managers can become workers in an instant, and so $M_t(\lambda, z)$ can jump down. But it will take time for workers to become managers again, and so $M_t(\lambda, z)$ cannot jump up. For the purpose of determining aggregate output in this Cobb-Douglas economy, it happens to suffice to know the aggregate stock of managerial human capital H_t. In contrast, the technology for accumulating managerial human capital to be described next depends on the entire distribution $M_t(\lambda, z)$.

Anticipating the balanced growth paths that will be constructed below, suppose the distribution of $z - \kappa t$ happens to be stationary for some κ. Then (3) implies that H_t grows at the rate κ. It follows from (2) and (4) that consumption and wages grow at the rate $(1-\alpha)\kappa$, and $v_t^\alpha w_t^{1-\alpha} = 1$ implies that v_t declines at the rate $\alpha\kappa$. As a result, $v_t e^z / w_t$ is stationary, and so variable employment per manager is stationary as well.

2.2 Knowledge Creation and Transmission

Workers can become managers by learning from incumbent managers. The productivities of incumbent managers evolve stochastically, as a result of idiosyncratic productivity shocks, and because managers can also learn from other managers. Managers can teach workers and other managers, one-on-one, to become as productive as they are themselves. More precisely, a type-λ worker matched with a teaching manager learns the useful knowledge of this teaching manager after a random time, distributed exponentially with mean $1/\lambda$. And the time-t state z_t of a type-λ manager matched with a teaching manager in state $\widetilde{z}_t > z_t$ evolves according to

$$\mathrm{d}z_t = \mu \mathrm{d}t + \sigma \mathrm{d}B_t + (\widetilde{z}_{t-} - z_{t-})\mathrm{d}J_{\lambda,t}, \tag{5}$$

conditional on managerial survival. B_t is a standard Brownian motion, and $J_{\lambda,t}$ is a Poisson jump process with arrival rate λ. The Brownian motions evolve independently across managers, and the Poisson jumps are 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. Workers can supply labor while they learn to become managers, and managers can oversee workers producing consumption while engaged in teaching or learning. But managers who teach cannot be students at the

same time.[4]

The Brownian increments combined with the fact that new and low-z managers can learn from high-z managers will cause the cross-sectional distribution of z to shift to the right, indefinitely. Unlike in some traditional models of human capital transmission (for example, Bils and Klenow [2000]), here there is no mean reversion in human capital. In Gabaix and Landier [2008], managerial skill contributes multiplicatively to the output of a firm and the managerial skill distribution is bounded. Here, the ability of individuals to learn λ is bounded in the population, but the usefulness of what managers can know is not. In this economy, it is not their skill but their useful knowledge, their ideas, that makes managers particularly productive in organizing the production of consumption.[5]

Further, it is important to note that the rate at which managers and workers learn from others is assumed to depend only on their ability type $\lambda \in \Lambda$ and not on what they are trying to learn or what they may already know. This emphasizes the fact that students can acquire useful knowledge without having to know the entire history of thought that gave rise to that useful knowledge. It is easy to imagine a long list of examples of obsolete or simply useless knowledge that is just as difficult to acquire as knowledge that is useful. In this economy, a high z just means that a manager in state z can produce more with a team of workers, not that learning to be like this manager is particularly difficult. But because teaching is one-on-one, and because the supply of managers who can teach high z is limited, not everyone can learn those high z at the same time.

2.3 The Market for Students and Teachers

The assignment of who learns from whom is determined in competitive markets. At any point in time t, students who want to learn from teachers in state z must pay flow tuition $T_t(z) \geq 0$, and potential teachers in state z decide whether to make themselves available at this price or not. Managers can be on different sides of the market at different points in time, depending on how productive they are. Students only pay for the time of their teachers, and students who succeed in adopting the state of their teachers experience capital gains. Markets are complete, and so these capital gains can be hedged in advance. Since there is no aggregate uncertainty, managers and workers

[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 [2014].

[5] Le [2014] explores an important alternative to idiosyncratic Brownian shocks, in which trends in z_t are governed by idiosyncratic Markov chains. This yields unbounded growth even though at any point in time, managerial productivities are bounded after a finite history.

simply maximize the expected present value of their earnings from supplying labor, managing workers, and teaching or learning.

Let $W_t(\lambda)$ and $V_t(z|\lambda)$ be the dynastic present values of the earnings, respectively, of a worker at time t and of a manager in state z at time t, both with learning ability λ. Workers supply labor and can choose to pay tuition and learn to become managers. Thus, $W_t(\lambda)$ is bounded below by the present value of current and future worker wages. Managers can choose at any point in time to become workers again, and hence $V_t(z|\lambda) \geq W_t(\lambda)$. Since wages are positive at all times, it follows that $W_t(\lambda)$ and $V_t(z|\lambda)$ are positive as well. The fact that managerial profits from producing consumption are $v(t,z) = v_t e^z$ will imply that $V_t(z|\lambda)$ is an increasing function of the managerial productivity state z. A lower bound for $V_t(z|\lambda)$ is the expected discounted present value of $\{v_s e^{z_s} - \phi w_s\}_{s \geq t}$ given $z_t = z$, and this present value behaves like e^z for large z. It follows that $V_t(z|\lambda)$ increases without bound as z becomes large. For low enough z, managerial profits are going to be well below the wages of workers, and such unproductive managers will not be able to earn significant or any tuition income from teaching others. Since their ability to learn does not depend on being a manager or a worker, sufficiently unproductive managers will choose to become workers, and hence $V_t(z|\lambda) = W_t(\lambda)$ for all low enough values of z. So $W_t(\lambda) = \min_z \{V_t(z|\lambda)\}$ can be used to recover $W_t(\lambda)$ from $V_t(z|\lambda)$.

With these considerations in mind, fix some time t and conjecture that $W_t(\lambda)$ and $V_t(z|\lambda)$ satisfy

$$\begin{aligned} & V_t(z|\lambda) \text{ is continuous in } z \\ 0 \;&<\; W_t(\lambda) = \min_z \{V_t(z|\lambda)\}, \quad \lim_{z\to\infty} V_t(z|\lambda) = \infty \\ \mathrm{D}V_t(z|\lambda) \;&\geq\; 0 \text{ with equality only if } V_t(z|\lambda) = W_t(\lambda) \\ V_t(z|\lambda') \;&\geq\; V_t(z|\lambda) \text{ when } \lambda' > \lambda. \end{aligned}$$

Implicit in these conditions is the fact that $V_t(z|\lambda) = W_t(\lambda)$ if and only if z is at or below some λ-specific threshold. Given these conjectures about $W_t(\lambda)$ and $V_t(z|\lambda)$, take some tuition schedule $T_t(z) \geq 0$ and define

$$S_t(\lambda) = \sup_z \{\lambda V_t(z|\lambda) - T_t(z)\}. \tag{6}$$

With some abuse of terminology, call this the surplus value of a type-λ student. The actual net expected gain from studying for a type-λ student is $S_t(\lambda) - \lambda W_t(\lambda)$ if the student is a worker, and $S_t(\lambda) - \lambda V_t(z|\lambda)$ if the student is a manager in state z. Since $V_t(z|\lambda) \geq W_t(\lambda)$, these net gains are always larger for workers than for managers with the same ability to learn. Type-λ workers or managers strictly prefer not to be students

if $S_t(\lambda) - \lambda W_t(\lambda) < 0$. But then $T_t(z) \geq \lambda W_t(\lambda) - S_t(\lambda) > 0$ for all z, since (6) implies $T_t(z) \geq \lambda V_t(z|\lambda) - S_t(\lambda)$. So the gain from studying for a type-λ worker can be strictly negative only if tuition is strictly positive at all z, even for arbitrarily low z, and bounded away from zero. This possibility will be ruled out in Lemma 3.

Since $V_t(z|\lambda)$ increases without bound, sufficiently productive managers cannot gain from studying. If $S_t(\lambda) - \lambda V_t(z|\lambda) < 0$, then (6) implies $T_t(z) \geq \lambda V_t(z|\lambda) - S_t(z) > 0$, and so managers who would expect negative net gains from studying can earn strictly positive tuition. Since $V_t(z|\lambda)$ is conjectured to be strictly increasing when $V_t(z|\lambda) > W_t(\lambda)$, essentially all managers strictly prefer to be either a student or a teacher. And if a type-λ manager in state z strictly prefers to teach, then so do all type-λ managers in states $\widetilde{z} \geq z$.

The inequality $T_t(z) \geq \lambda V_t(z|\lambda) - S_t(\lambda)$ holds for all z and λ. If this inequality is strict for some z and all λ, then no students of any type choose to study with teachers in state z. Therefore, if there are managers in state z who choose to teach, then market clearing implies that this inequality has to be an equality for some λ. Only if there are no teachers at z can $T_t(z)$ exceed $\max_{\lambda\in\Lambda}\{\lambda V_t(z|\lambda) - S_t(\lambda)\}$. In that case, lowering tuition by any amount is not going to induce any managers at z to become teachers. And lowering tuition down to $\max_{\lambda\in\Lambda}\{[\lambda V_t(z|\lambda) - S_t(\lambda)]^+\} \geq \max_{\lambda\in\Lambda}\{\lambda V_t(z|\lambda) - S_t(\lambda)\}$ keeps tuition non-negative and does not make any students strictly prefer to select a teacher in state z. That is, such a reduction in the tuition at z will not affect the value of $S_t(\lambda)$ as defined in (6). This implies the following lemma.

Lemma 1 *The tuition schedule can be taken to be of the form*

$$T_t(z) = \max_{\lambda\in\Lambda} \left\{ [\lambda V_t(z|\lambda) - S_t(\lambda)]^+ \right\}, \tag{7}$$

without loss of generality.

An immediate implication of the fact that the value functions $V_t(z|\lambda)$ are strictly increasing when $V_t(z|\lambda) > W_t(\lambda)$ is that the tuition schedule (7) is strictly increasing when positive.

Starting with candidate surplus values $\{S_t(\lambda) : \lambda \in \Lambda\}$ that do not necessarily satisfy (6), one can simply use (7) to define $T_t(z)$. Such a construction immediately implies $S_t(\lambda) \geq \sup_z\{\lambda V_t(z|\lambda) - T_t(z)\}$, but the inequality can be strict. Very large values of some $S_t(\lambda)$ imply that the construction of $T_t(z)$ is not affected by lowering those $S_t(\lambda)$.

Lemma 2 *Given any* $\{S_t^*(\lambda) : \lambda \in \Lambda\}$*, define*

$$T_t(z) = \max_{\lambda\in\Lambda} \left\{ [\lambda V_t(z|\lambda) - S_t^*(\lambda)]^+ \right\}, \quad S_t(\lambda) = \sup_z \left\{ \lambda V_t(z|\lambda) - T_t(z) \right\}.$$

Then $T_t(z) = \max_{\lambda\in\Lambda}\{[\lambda V_t(z|\lambda) - S_t(\lambda)]^+\}$.

The construction of $S_t(\lambda)$ implies $S_t(\lambda) \leq S_t^*(\lambda)$, and that then immediately implies $T_t(z) \leq \max_{\lambda\in\Lambda}\{[\lambda V_t(z|\lambda) - S_t(\lambda)]^+\}$. The reverse inequality follows because $T_t(z) \geq 0$, $S_t(\lambda) \geq \lambda V_t(z|\lambda) - T_t(z)$ for all (λ, z), and hence $T_t(z) \geq \max_{\lambda\in\Lambda}\{[\lambda V_t(z|\lambda) - S_t(\lambda)]^+\}$. Lemma 2 implies that it is without loss of generality to take the surplus values $\{S_t(\lambda) : \lambda \in \Lambda\}$ to be small enough so that the surplus values and the tuition schedule satisfy (6) and (7), respectively.

Since managers die and may choose to become workers, some type of workers will have to be willing to be students if the population of managers is not to die out. That is, there must be at least one type λ for which $S_t(\lambda) - \lambda W_t(\lambda) \geq 0$. If that is indeed the case, then every type of worker is willing to be a student in this economy.

Lemma 3 *Suppose there are* $\lambda \in \Lambda$ *such that* $S_t(\lambda) - \lambda W_t(\lambda) \geq 0$*, and suppose that for such* λ *the supply of type-*λ *managers is strictly positive at all* z *that satisfy* $V_t(z|\lambda) > W_t(\lambda)$*. Then* $\min_z\{T_t(z)\} = 0$*, and hence* $S_t(\lambda) - \lambda W_t(\lambda) \geq 0$ *for all* $\lambda \in \Lambda$*.*

The proof is given in Appendix A. Because managers are subject to Brownian shocks to their productivity states, the supply of type-λ managers will indeed be positive in any state z in which type-λ managers strictly prefer to continue as managers. The basic intuition for Lemma 3 is the fact that there will always be cheap teachers in an economy in which the only alternative use of teaching time is studying. The least productive teachers can only be teaching less productive students of their own type, and the marginal teacher does not have anything to offer to the most productive students of the same type. This means that tuition cannot be positive everywhere.

2.4 Value Functions

At any point in time, the market for students and teachers establishes a tuition schedule $T_t(z)$ of the form (7) and associated surplus values $S_t(\lambda)$ defined by (6). This leads to earnings flows and expected capital gains for the various types of workers and managers. The value of a worker with ability λ must satisfy the Bellman equation

$$r_t W_t(\lambda) = w_t + \max\{0, S_t(\lambda) - \lambda W_t(\lambda)\} + \mathrm{D}_t W_t(\lambda). \tag{8}$$

Workers earn wages w_t and choose to study only if the net gains $S_t(\lambda) - \lambda W_t(\lambda)$ are positive. The value function $V_t(z|\lambda)$ of a type-λ manager has to satisfy the Bellman

equation

$$\begin{aligned} r_t V_t(z|\lambda) &= v_t e^z - \phi w_t + \max\{S_t(\lambda) - \lambda V_t(z|\lambda), T_t(z)\} + \delta[W_t(\lambda) - V_t(z|\lambda)] \\ &\quad + \mathrm{D}_t V_t(z|\lambda) + \mu \mathrm{D}_z V_t(z|\lambda) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} V_t(z|\lambda). \end{aligned} \tag{9}$$

The first two terms on the right-hand side constitute the net revenue from hiring workers to produce output, and the third term represents the gains from being a teacher or a student. At the rate δ, one generation passes and is immediately replaced by a new generation. When this happens, the new generation begins life as a worker, and the dynasty experiences a negative capital gain $W_t(\lambda) - V_t(z|\lambda)$. These value functions also have to satisfy a transversality condition that requires $V_t(z_t|\lambda)$ discounted back to the initial date to converge in mean to zero as t becomes large.

2.5 Equilibrium and a Robust Indeterminacy

At any point in time t, there are measures $M_t(\lambda, z)$ that describe how the managers of the various types λ are distributed across the productivity states z. The number of type-λ workers is $M(\lambda) - M_t(\lambda, \infty)$. These measures can change in an instant only when a positive mass of managers quits. This may happen at some initial date, but not subsequently because there is no aggregate uncertainty. The supplies of managers and workers at time t determine the output of aggregate consumption C_t and the factor prices v_t and w_t. Beliefs about future factor prices $\{v_s, w_s\}_{s>t}$, and future surplus values $\{S_s(\lambda) : \lambda \in \Lambda\}_{s>t}$, together with tuition schedules given by (7), determine the time-t value $W_t(\lambda)$ of a type-λ worker and the time-t values $V_t(z|\lambda)$ of a type-λ manager in state z, for all $\lambda \in \Lambda$ and all z. The market for students and teachers at time t then determines the surplus values $\{S_t(\lambda) : \lambda \in \Lambda\}$, the associated tuition schedule, and the assignment of students and teachers at time t. Together with the productivity dynamics (5), this assignment pins down how the measures $M_t(\lambda, z)$ are evolving at time t, and therefore the prices that will be realized in future factor markets and future markets for students and teachers. In a perfect foresight equilibrium, these prices have to match the beliefs formed at time t.

One can expect the perfect foresight path of equilibrium consumption to be unique. The technologies for producing consumption and for transferring useful knowledge exhibit constant returns to scale, and preferences over consumption sequences have indifference curves that are strictly convex. But the assignment of students to teachers will certainly not be unique: at any point in time, there is a continuum of teachers indexed by their productivity states z, and only a finite number of student types $\lambda \in \Lambda$. This

means students are going to have to be indifferent across teachers with different z, and this implies an assignment that is to some extent random. This will matter for the life histories of individuals in this economy, even though ex ante everyone of a given type faces the same possibilities. This type of indeterminacy is inherent in any environment in which students are more similar than their teachers, and it is a source of ex post income inequality. Random assignment of like students is only one possible mechanism for resolving this indeterminacy. Equilibrium implies that individuals who only differ in terms of payoff irrelevant characteristics are equally wealthy ex ante. But their investment in learning from others and their ex post income histories may very well be correlated with these payoff irrelevant characteristics.

Solving for the perfect foresight equilibrium amounts to finding a fixed point in a space of value functions and productivity measures indexed by time. Finding these value functions and productivity measures requires solving systems of interrelated partial differential equations indexed by equilibrium factor prices and surplus values: the Bellman equations (8)-(9) and the Kolmogorov forward equations that govern the productivity distributions. This is a task that involves confronting approximation and convergence questions that remain unanswered in the existing literature. The rest of the paper will focus instead on balanced growth paths. These balanced growth paths are much more tractable: partial differential equations become ordinary differential equations. Moreover, the assumed Cobb-Douglas technology makes it possible to solve these differential equations analytically (analytical solutions are also available if the technology for producing consumption is Leontief).

3. Balanced Growth in a Two-Type Economy

The assignment of students to teachers is trivial in an economy with only one type of student. And such an economy would fail to account for the fact that there are observable characteristics that can predict future earnings. A two-type economy with $\Lambda = \{\beta, \gamma\}$ and $\gamma > \beta > 0$ is sufficiently rich to illustrate the assignment problem and the impact of small differences in the ability to learn from others, and it can account for the predictability of future earnings.

Conjecture that such an economy has a balanced growth path with a growth rate κ that makes the cross-sectional distribution of $z_t - \kappa t$ independent of time. This gives rise to a constant supply of managers and of variable labor, and the aggregate stock of managerial human capital (3) will grow at the rate κ. The Cobb-Douglas technology (2) then implies that aggregate consumption and wages grow at the rate $(1-\alpha)\kappa$ and that

v_t grows at the rate $-\alpha\kappa$. The implied interest rate is $r_t = \rho + (1-\alpha)\kappa$.

It remains to solve Bellman equations for value functions, Kolmogorov forward equations for stationary densities, and to find an assignment of students to teachers that clears markets.

3.1 The Bellman Equations

Given factor prices of the form $w_t = we^{(1-\alpha)\kappa t}$ and $v_t = ve^{-\alpha\kappa t}$, conjecture that the value functions, tuition schedules, and surplus values are of the form

$$[W_t(\lambda), V_t(z+\kappa t|\lambda), S_t(\lambda), T_t(z+\kappa t)]\, e^{-(1-\alpha)\kappa t} = [W(\lambda), V(z|\lambda), S(\lambda), T(z)]\,,$$

where z now represents the state of a manager detrended by $-\kappa t$. An example of what these functions look like is given in Figure 1. With this conjecture, the Bellman equation (8) for type-λ workers simplifies to

$$W(\lambda) = \max\left\{\frac{w}{\rho}, \frac{w+S(\lambda)}{\rho+\lambda}\right\}. \tag{10}$$

This means that the value of a type-λ worker may exceed the present value of wages, but only if the surplus value $S(\lambda)$ is large enough—more precisely, if and only if $S(\lambda)/\lambda$ exceeds w/ρ. The balanced growth version of the Bellman equation (9) for a type-λ manager becomes

$$\begin{aligned}(\rho+\delta)V(z|\lambda) &= ve^z - \phi w + \delta W(\lambda) + \max\left\{S(\lambda) - \lambda V(z|\lambda), T(z)\right\} \\ &\quad + (\mu-\kappa)\mathrm{D}V(z|\lambda) + \frac{1}{2}\sigma^2\mathrm{D}^2V(z|\lambda).\end{aligned} \tag{11}$$

This differential equation has to hold for all z at which managers strictly prefer to continue as managers. In particular, this means that $V(z|\lambda)$ has to be smooth at thresholds where type-λ managers switch between studying and teaching. As in Lemma 2, the tuition schedule can be taken to be

$$T(z) = \max_{\lambda\in\Lambda}\left\{[\lambda V(z|\lambda) - S(\lambda)]^+\right\}. \tag{12}$$

The option to become a worker again means that $V(z|\lambda) \geq W(\lambda)$ for any z. As already argued, managers with sufficiently low productivities will find it optimal to become workers again. This implies thresholds $b(\lambda)$ so that type-λ managers strictly prefer to continue as managers if and only if $z > b(\lambda)$. By construction, $W(\lambda) = V(z|\lambda)$ for all $z \leq b(\lambda)$. Optimality of the exit decision of a type-λ manager requires that $V(z|\lambda)$ is

also differentiable at $b(\lambda)$. This gives the familiar value-matching and smooth-pasting conditions

$$W(\lambda) = V(b(\lambda)|\lambda), \quad 0 = \mathrm{D}V(b(\lambda)|\lambda) \tag{13}$$

for both $\lambda \in \{\beta, \gamma\}$. Inserting the tuition schedule (12) into the Bellman equation (11) results in a system of piecewise linear second-order differential equations for $\{V(z|\lambda) : \lambda \in \Lambda\}$, with boundary conditions implied by (10) and (13), and parameterized by the factor prices v and w and by the surplus values $\{S(\lambda) : \lambda \in \Lambda\}$.

A recipe for constructing value functions $\{V(z|\lambda) : \lambda \in \Lambda\}$ is to partition $(b(\lambda), \infty)$ into segments on which the types of managers who match as students and teachers do not change. The second-order differential equation (11)-(12) then remains linear on these segments and can be solved explicitly. The resulting solutions will depend on undetermined boundary values and must be smoothly pasted together across segments. The optimal exit thresholds $b(\lambda)$ are then determined by (13). The rest of this section implements this recipe.

3.1.1 Teaching Thresholds

More can be said about who teaches whom. Consumption cannot be produced without managers, and managers die. Along a balanced growth path, there therefore has to be at least one type of worker who is willing to study. Lemma 3 then implies that $S(\lambda) - \lambda W(\lambda) \geq 0$ for both $\lambda \in \{\beta, \gamma\}$. Because the value functions $V(z|\lambda)$ are increasing and unbounded above, it must be that the gains from studying for a type-λ manager $S(\lambda) - \lambda V(z|\lambda) \leq S(\lambda) - \lambda W(\lambda)$ decrease monotonically in z and become negative for large enough z. On the other hand, the tuition such a manager can earn is increasing and unbounded above. There must therefore be thresholds $x(\lambda) \geq b(\lambda)$, defined by

$$S(\lambda) - \lambda V(x(\lambda)|\lambda) = T(x(\lambda)),$$

so that type-λ managers strictly prefer to teach if and only if $z > x(\lambda)$. For all large enough z, these managers will teach type-γ students. To see this, note that an immediate consequence of $\gamma > \beta > 0$ and $V(z|\gamma) \geq V(z|\beta) > 0$ is $\gamma V(z|\gamma) - \beta V(z|\beta) \geq (\gamma - \beta)V(z|\beta) > 0$. This gap will become large for large z, simply because $V(z|\beta)$ increases without bound. Thus, type-γ students determine the tuition schedule for all large enough z. Specifically, there must be a threshold $y \geq \max\{x(\beta), x(\gamma)\}$, defined by

$$y = \sup\{z : T(z) > \gamma V(z|\gamma) - S(\gamma)\},$$

so that teachers in states $z > y$ only teach type-γ students. If $\gamma V(z|\gamma) - S(\gamma)$ crosses $\beta V(z|\beta) - S(\beta)$ only once, as will be the case in the equilibria constructed below, then teachers below y teach only type-β students.

3.1.2 Surplus Values and Managerial Profits in Equilibrium

Given a tuition schedule of the form (12), the value functions and their associated thresholds are functions only of the factor prices $[v, w]$ and the surplus values $[S(\beta), S(\gamma)]$. All prices and value functions can be expressed in units of labor by dividing by w. The resulting value functions $[W(\lambda), V(\cdot|\lambda)]/w$ are then functions only of v/w and $[S(\beta), S(\gamma)]/w$. The variable managerial profits that appear in the Bellman equations are $ve^z/w = e^{\widehat{z}}$. With a change of variables from z to $\widehat{z}$, one obtains Bellman equations that are parameterized only by $[S(\beta), S(\gamma)]/w$. The exit and teaching thresholds for the state variable $\widehat{z}$ are $b(\lambda) + \ln(v/w)$ and $x(\lambda) + \ln(v/w)$ for both $\lambda \in \{\beta, \gamma\}$, and $y + \ln(v/w)$. The Bellman equations therefore imply the equilibrium condition

$$[S(\beta), S(\gamma)]/w \mapsto (v/w)\left[e^{b(\beta)}, e^{b(\gamma)}, e^{x(\beta)}, e^{x(\gamma)}, e^{y}\right]. \tag{14}$$

That is, the Bellman equations relate the managerial profits at the exit and teaching thresholds to the surplus values $[S(\beta), S(\gamma)]/w$. A useful corollary is that the gaps $\{[y - b(\lambda), y - x(\lambda)] : \lambda \in \{\beta, \gamma\}\}$ only depend on $[S(\beta), S(\gamma)]/w$, and not on v/w. Put differently, the shape of the value functions depends only on $[S(\beta), S(\gamma)]/w$, and their location is determined by v/w.

3.1.3 Three Ability Rent Scenarios

Recall from Lemma 3 that $S(\lambda) - \lambda W(\lambda) \geq 0$ for both $\lambda \in \{\beta, \gamma\}$. Workers of both types are willing to study. The Bellman equation for workers (10) implies that this is equivalent to $S(\lambda)/\lambda \geq w/\rho$, and both inequalities are strict at the same time: workers strictly prefer to be students if and only if $S(\lambda)/\lambda$ exceeds w/ρ, and then the value $W(\lambda)$ of being a worker exceeds w/ρ: type-λ workers earn rents from their ability to learn to become managers. An implication of the original definition (6) of the surplus values $S(\lambda)$ is that $S(\lambda)/\lambda$ must be increasing in λ. So fast learners strictly prefer to study as workers when slow learners do, and slow learners do not if fast learners do not. This implies there are three possible scenarios: both types of workers are indifferent between studying or not, slow learners are indifferent but fast learners strictly prefer to study, and a scenario in which both types of workers strictly prefer to study.

In the first scenario, $S(\gamma)/\gamma = S(\beta)/\beta = w/\rho$. This implies that $S(\lambda) - \lambda V(z|\lambda) < 0$ whenever $V(z|\lambda) > W(\lambda) = w/\rho$, and thus managers choose to teach rather than study. Since managers do not study, and since $W(\lambda) = w/\rho$ for both λ, it must be that $V(z|\beta) = V(z|\gamma)$ for all z. The tuition schedule must then be given by $T(z) = \gamma V(z|\gamma) - S(\gamma) = \gamma[V(z|\gamma) - w/\rho]$, since this is strictly greater than $\beta V(z|\beta) - S(\beta) = \beta[V(z|\gamma) - w/\rho]$ whenever $V(z|\gamma) > W(\gamma)$. For slow learners, $S(\beta)$ is attained at the exit threshold, where tuition is zero, whereas fast learners are indifferent across all possible teachers. This can only be an equilibrium if no type-β workers study and all managers are fast learners. Appendix B characterizes this equilibrium and shows that it arises when the ratio $M(\gamma)/M(\beta)$ is high enough.

On the other hand, if type-γ households are sufficiently scarce, then there can be no equilibrium in which both types of workers are indifferent students. One or both types of workers must strictly prefer to study. If $S(\gamma)/\gamma > S(\beta)/\beta > w/\rho$, then both do, and market clearing immediately implies that half of the population is a teacher and half a student. Because only managers can teach, this implies that at least half of the population must be a manager. Attempting to relate such a scenario to data would require the use of a very broad notion of who is a manager. Instead, the focus from here on will be on the remaining scenario in which type-β workers are indifferent students, while type-γ workers strictly prefer to study.

3.1.4 The Scenario $S(\gamma)/\gamma > S(\beta)/\beta = w/\rho$

The value functions and the assignment of students to teachers shown in Figure 1 are constructed for this scenario. Since type-β workers have to be indifferent, the Bellman equation for workers (10) implies

$$S(\beta) = \beta W(\beta), \quad W(\beta) = \frac{w}{\rho}. \tag{15}$$

The ex ante value of type-β workers is simply the present value of their labor income. Some type-β workers may study and become managers with heterogeneous initial productivity states determined by who their teachers were. But the tuition they pay absorbs all the expected gains, and many may be worse off ex post. Type-β managers never choose to be students, because $V(z|\beta) > W(\beta)$ implies $S(\beta) - \beta V(z|\beta) < 0$. Thus, the threshold at which type-β managers become teachers is simply $x(\beta) = b(\beta)$. In contrast, $S(\gamma) > \gamma W(\gamma)$ combined with (10) implies

$$W(\gamma) = \frac{w + S(\gamma)}{\rho + \gamma}, \tag{16}$$

and this will exceed w/ρ precisely when $S(\gamma) > \gamma W(\gamma)$. Relative to type-$\beta$ workers, type-γ workers earn rents from their ability to learn fast. The payoffs of being a student are uncertain for both types of workers, but type-γ workers have a cushion and more of them will gain ex post. Since $S(\gamma) > \gamma V(z|\gamma)$ for $z > b(\gamma)$ close enough to $b(\gamma)$, some type-γ managers will also be students, and the threshold at which they become teachers will satisfy $x(\gamma) > b(\gamma)$. In this scenario, the thresholds are $b(\beta) < y$ and $b(\gamma) < x(\gamma) < y$, where, recall, y is the defined to be the highest managerial state z to which type-β students are assigned. Type-γ students determine the tuition schedule (7) for $z \geq y$. Thus, y is the point where both types of students are willing to pay the same tuition,

$$\gamma V(y|\gamma) - S(\gamma) = \beta V(y|\beta) - S(\beta). \tag{17}$$

To continue the construction, conjecture now that $\beta V(z|\beta) - S(\beta)$ and $\gamma V(z|\gamma) - S(\gamma)$ cross only once, at $z = y$. Then type-β students set the tuition for all $z < y$. In particular, this means that type-γ managers teach type-β students when they switch from studying to teaching at $x(\gamma)$. The threshold $x(\gamma)$ must therefore satisfy

$$S(\gamma) - \gamma V(x(\gamma)|\gamma) = \beta V(x(\gamma)|\beta) - S(\beta). \tag{18}$$

The Bellman equation (11) has to hold at $x(\gamma)$ and y, but smooth solutions that can be constructed on the open intervals $(b(\beta), y)$, $(b(\gamma), x(\gamma))$, $(x(\gamma), y)$, and (y, ∞) will not automatically satisfy the differential equation (11) at $x(\gamma)$ and y. Forcing the solution to be smooth at $x(\gamma)$ and y gives rise to the boundary conditions

$$\lim_{z \uparrow x(\gamma)} \begin{bmatrix} V(z|\gamma) \\ \mathrm{D}V(z|\gamma) \end{bmatrix} = \lim_{z \downarrow x(\gamma)} \begin{bmatrix} V(z|\gamma) \\ \mathrm{D}V(z|\gamma) \end{bmatrix} \tag{19}$$

$$\lim_{z \uparrow y} \begin{bmatrix} V(z|\lambda) \\ \mathrm{D}V(z|\lambda) \end{bmatrix} = \lim_{z \downarrow y} \begin{bmatrix} V(z|\lambda) \\ \mathrm{D}V(z|\lambda) \end{bmatrix}, \quad \lambda \in \{\beta, \gamma\} \tag{20}$$

With continuity and differentiability imposed, twice differentiability is implied because the solutions on the open intervals already satisfy (11). As in more familiar stopping problems (Dixit and Pindyck [1994]), the optimal exit thresholds $b(\beta)$ and $b(\gamma)$ must satisfy the value-matching and smooth-pasting conditions

$$W(\lambda) = V(b(\lambda)|\lambda), \quad 0 = \mathrm{D}V(b(\lambda)|\lambda), \quad \lambda \in \{\beta, \gamma\}. \tag{21}$$

To summarize, the differential equation for $V(z|\beta)$ and $V(z|\gamma)$ is (11)-(12). The values of $S(\beta)$ and $W(\beta)$ are simply given by (15). The surplus value $S(\gamma) > \gamma w/\rho$ immediately

pins down $W(\gamma)$ via (16). This then provides the initial values needed for the value-matching conditions in (21). The remaining boundary conditions are (17)-(20) and the smooth-pasting part of (21).

3.1.5 Constructing the Value Functions

The assumed single-crossing property means that the solution for $V(z|\beta)$ is governed by (11) with $T(z) = \beta V(z|\beta) - S(\beta)$ on $(b(\beta), y)$ and $T(z) = \gamma V(z|\gamma) - S(\gamma)$ on (y, ∞). On both segments, the general solution is a particular solution of (11) plus a linear combination of the two solutions to the homogeneous part of (11). One can verify that one of the homogeneous solutions on (y, ∞) explodes relative to e^z. This cannot be part of the solution because the $V(z|\beta)$ has to converge to a present value that scales with e^z when z becomes large. This leaves three undetermined coefficients: two for the solution on $(b(\beta), y)$ and one for the solution on (y, ∞).

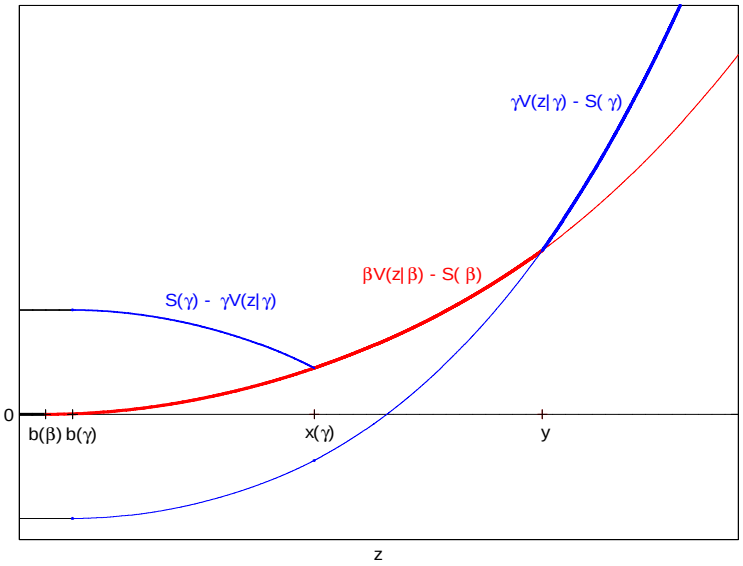


FIGURE 1 Student-Teacher Assignment When $S(\beta) = \beta W(\beta)$

The solution for $V(z|\gamma)$ is governed by (11) with $S(\gamma) - \gamma V(z|\gamma) > T(z)$ on $(b(\gamma), x(\gamma))$, $T(z) = \beta V(z|\beta) - S(\beta)$ on $(x(\gamma), y)$, and $T(z) = \gamma V(z|\gamma) - S(\gamma)$ on (y, ∞). Again, the homogeneous equation has two solutions on each of these segments, and on (y, ∞) one of them can be ruled out because it would cause the value function to diverge from a present value that scales with e^z, for large z. This results in five undetermined coefficients. It is easy to guess particular solutions on the open intervals $(b(\beta), y)$, $(b(\gamma), x(\gamma))$, $(x(\gamma), y)$, and (y, ∞): they can be taken to be present values of flow profits calculated as if the

differential equation holds throughout $(-\infty, \infty)$. So now we have eight undetermined coefficients and the four unknown thresholds $b(\beta)$, $b(\gamma)$, $x(\gamma)$, and y. Determining these coefficients and thresholds requires twelve boundary conditions. These are provided by (17)-(21).

3.2 The Kolmogorov Forward Equations

Continue with the scenario in which ability rents are positive for type-γ workers only. So type-β managers in $(b(\beta), y)$ and type-γ managers in $(x(\gamma), y)$ teach type-β students, and both types of managers in (y, ∞) teach type-γ students. Type-γ managers in $(b(\gamma), x(\gamma))$ are students.

Setting the time derivative of the time-t density of type-β managers to zero in the Kolmogorov forward equation can be used to show that the stationary density of type-β managers must satisfy

$$\begin{aligned} \delta m(\beta, z) &= -(\mu - \kappa)\mathrm{D}_z m(\beta, z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(\beta, z) \\ &\quad + \begin{cases} \beta m(\beta, z), & z \in (b(\beta), x(\gamma)) \\ \beta[m(\beta, z) + m(\gamma, z)], & z \in (x(\gamma), y) \\ 0, & z \in (y, \infty). \end{cases} \end{aligned} \tag{22}$$

By teaching type-β students, type-β managers in $(b(\beta), y)$ are in a sense replicating themselves at the rate β. Type-γ managers in $(x(\gamma), y)$ also teach type-β students, adding a flow $\beta m(\gamma, z)$. In (y, ∞), teaching type-γ students produces more type-γ managers but no additional type-β managers. Type-β managers exit and become workers again when their productivity state crosses $b(\beta)$ from above. This gives rise to the boundary condition

$$m(\beta, b(\beta)) = 0 \tag{23}$$

(see Cox and Miller [1965]). Similarly, the stationary density for type-γ managers has to satisfy

$$\begin{aligned} \delta m(\gamma, z) &= -(\mu - \kappa)\mathrm{D}_z m(\gamma, z) + \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(\gamma, z) \\ &\quad + \begin{cases} -\gamma m(\gamma, z), & z \in (b(\gamma), x(\gamma)) \\ 0, & z \in (x(\gamma), y) \\ \gamma[m(\beta, z) + m(\gamma, z)], & z \in (y, \infty). \end{cases} \end{aligned} \tag{24}$$

Type-γ managers in $(b(\gamma), x(\gamma))$ are students, and they transition into productivity states in (y, ∞) at the rate γ. Type-γ managers in $(x(\gamma), y)$ are teachers of type-β students.

In (y, ∞), everyone teaches type-γ students. Since there is one student per teacher, the flow of new type-γ managers with productivity states in (y, ∞) is γ times the number of teachers in this range, who can be of either type. Exit at $b(\gamma)$ produces the boundary condition

$$m(\gamma, b(\gamma)) = 0. \tag{25}$$

A further set of boundary conditions is implied by the requirement that the flow to the left is continuous everywhere on $(b(\beta), \infty)$ and $(b(\gamma), \infty)$. A discontinuity would imply entry or exit at interior points of these intervals. This continuous-flow requirement says that

$$-(\mu - \kappa)m(\lambda, z) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(\lambda, z) \text{ is continuous at } x(\gamma) \text{ and } y, \text{ for } \lambda \in \{\beta, \gamma\}. \tag{26}$$

The densities $\{m(\lambda, z) : \lambda \in \{\beta, \gamma\}\}$ are supposed to integrable—their integrals represent the numbers of type-β and type-γ managers. A necessary condition for integrability is

$$\lim_{z \to \infty} m(\lambda, z) = 0, \quad \text{for } \lambda \in \{\beta, \gamma\}. \tag{27}$$

Because the densities $\{m(\lambda, z) : \lambda \in \{\beta, \gamma\}\}$ will turn out to have exponential right tails, this condition is also sufficient for integrability.

The equations (22)-(27) define a two-dimensional homogeneous system of piecewise linear second-order differential equations. Both types of managers teach students of the other type, and so the differential equations are interrelated. The system is autonomous, except for the location of the thresholds $b(\beta)$, $b(\gamma)$, $x(\gamma)$, and y. Because of this, the implied type-β and type-γ stationary densities for $z - y$ can only depend on the gaps $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$. The scale of any solution is clearly indeterminate.

3.2.1 Implied Managerial Entry Rates

Integrating (22) and (24) and imposing the boundary conditions (23) and (25)-(27) gives

$$\begin{aligned}
\delta \int_{b(\beta)}^{\infty} m(\beta, z)\mathrm{d}z + \frac{1}{2}\sigma^2 \mathrm{D}_z m(\beta, b(\beta)) &= \beta \left(\int_{b(\beta)}^{y} m(\beta, z)\mathrm{d}z + \int_{x(\gamma)}^{y} m(\gamma, z)\mathrm{d}z \right), \\
\delta \int_{b(\gamma)}^{\infty} m(\gamma, z)\mathrm{d}z + \frac{1}{2}\sigma^2 \mathrm{D}_z m(\gamma, b(\gamma)) &= \gamma \left(\int_{y}^{\infty} [m(\beta, z) + m(\gamma, z)]\mathrm{d}z - \int_{b(\gamma)}^{x(\gamma)} m(\gamma, z)\mathrm{d}z \right).
\end{aligned}$$

The left-hand sides of these two equations represent flows of type-β and type-γ managers exiting, randomly or because their productivities reach the respective exit thresholds $b(\beta)$ and $b(\gamma)$. The right-hand sides represent managerial teaching minus learning by

students who are already managers, or, equivalently, the supply of teaching to students who are not yet managers. If the market for students and teachers clears, then these right-hand sides also represent flows of entering managers. The flows of exiting and entering managers then balance for each type.

3.2.2 Solving the Differential Equation

The differential equation (22)-(27) is parameterized by the unknown balanced growth rate κ. The equation can be solved for stationary densities $m(\beta, z)$ and $m(\gamma, z)$ provided this growth rate is high enough (the inequality (29) below) and $\gamma > \delta$. It is easy to see that $\gamma \leq \delta$ would not be consistent with stationarity. Besides random exit at the rate δ, there is also exit across the thresholds $b(\beta)$ and $b(\gamma)$. So even if all managers taught, at the maximal rate γ, they would not be able to sustain their numbers.

First, note that the differential equation (22) for $m(\beta, z)$ on $(x(\gamma), y)$ and the differential equation (24) for $m(\gamma, z)$ on (y, ∞) are inhomogeneous equations on these intervals, with inhomogeneous terms $m(\gamma, z)$ and $m(\beta, z)$, respectively. A particular solution to (22) on $(x(\gamma), y)$ is simply $-m(\gamma, z)$, because the term $\beta[m(\beta, z) + m(\gamma, z)]$ equals zero for that solution, and $-m(\gamma, z)$ has to solve the second equation in (24). Similarly, $-m(\beta, z)$ is a particular solution to (24) on (y, ∞). On every interval, the homogeneous parts of (22) and (24) are linear second-order differential equations with constant coefficients, and this implies a pair of exponential solutions for both densities on each of the three intervals. If these exponential solutions are distinct, then one can combine these homogeneous solutions with the two particular solutions to construct a twelve-dimensional linear space of solutions. The boundary conditions (23) and (25) at $b(\beta)$ and $b(\gamma)$ provide two linear restrictions. The continuity requirements (26) at $x(\gamma)$ and y provide four more linear restrictions. This leaves a six-dimensional linear space of solutions.

It remains to impose the integrability conditions (27). Consider $m(\beta, z)$ and note that (22) is homogeneous on (y, ∞), with solutions of the form $e^{-\xi_\pm z}$, where $\xi_\pm$ solves the characteristic equation $\delta = (\mu - \kappa)\xi_\pm + \frac{1}{2}\sigma^2\xi_\pm^2$. Since δ is positive, this implies $\xi_+ > 0 > \xi_-$. So the coefficient on $e^{-\xi_- z}$ must be zero, or else $|m(\beta, z)| \to \infty$ as $z \to \infty$. The integrability condition for $m(\beta, z)$ therefore adds another linear restriction. Next, consider the integrability condition for $m(\gamma, z)$. On (y, ∞), the homogeneous part of (24) has solutions of the form $e^{-\zeta_\pm z}$, where $\zeta_\pm$ solves the characteristic equation

$\delta - \gamma = (\mu - \kappa)\zeta_\pm + \frac{1}{2}\sigma^2\zeta_\pm^2$. The roots of this quadratic are

$$\zeta_\pm = \frac{\kappa - \mu}{\sigma^2} \pm \sqrt{\left(\frac{\kappa - \mu}{\sigma^2}\right)^2 - \frac{\gamma - \delta}{\sigma^2/2}}. \tag{28}$$

Because $\gamma > \delta$, these roots will be complex if the balanced growth rate κ is too close to μ. Complex roots would result in solutions for $m(\gamma, z)$ that oscillate around zero and must therefore be ruled out. If the roots $\zeta_\pm$ are real and $\kappa < \mu$, then both $\zeta_\pm$ would be negative, resulting in explosive solutions. A positive and integrable solution for $m(\gamma, z)$ on (y, ∞) can therefore only be constructed if the growth rate κ satisfies $((\kappa - \mu)/\sigma^2)^2 \geq (\gamma - \delta)/(\sigma^2/2)$ and $\kappa > \mu$. This is equivalent to

$$\kappa \geq \mu + \sigma^2\sqrt{\frac{\gamma - \delta}{\sigma^2/2}}. \tag{29}$$

This condition implies $\zeta_+ \geq \zeta_- > 0$ and ensures the integrability of any linear combination of $e^{-\zeta_+ z}$ and $e^{-\zeta_- z}$. The roots ζ_+ and ζ_- are distinct if and only if (29) holds strictly. Thus, the homogeneous part of (24) on (y, ∞) has a two-dimensional space of solutions if (29) holds strictly and a one-dimensional space of solutions if (29) holds with equality.[6]

Suppose that (29) holds strictly. Then the integrability condition (27) generates one linear restriction on the space of solutions. This reduces the dimension of the space of solutions from six to five. Picking a scale reduces it to four. Conjecture now that there is a solution for which not only the flows $-(\mu - \kappa)m(\lambda, z) + \frac{1}{2}\sigma^2 \mathrm{D}_z m(\lambda, z)$ but also the densities $m(\lambda, z)$ themselves are continuous (and thus continuously differentiable). Imposing this continuity condition adds four linear restrictions: one for each $m(\lambda, z)$ at $x(\gamma)$ and y. These four restrictions are just enough to determine the solution for $m(\beta, z)$ and $m(\gamma, z)$ up to a common scale factor. When (29) holds with equality, the two roots ζ_+ and ζ_- merge into one, given by

$$\zeta_\gamma = \sqrt{\frac{\gamma - \delta}{\sigma^2/2}}. \tag{30}$$

[6]Note that the homogeneous part of (22) on $(b(\beta), x(\gamma)) \cup (x(\gamma), y)$ is of the same form as the homogeneous part of (24) on (y, ∞). But if (29) holds weakly, then it must hold strictly when γ is replaced by $\beta \in [\delta, \gamma)$. And the characteristic equation automatically has distinct real roots if $\beta < \delta < \gamma$. The space of homogeneous solutions for $m(\beta, z)$ on $(b(\beta), x(\gamma)) \cup (x(\gamma), y)$ is therefore guaranteed to be four-dimensional.

The homogeneous part of (24) on (y, ∞) is then one-dimensional rather than two-dimensional, suggesting that it might not be possible to construct a smooth solution. But, as in Luttmer [2007] and Appendix B, the continuous solution that can be constructed when (29) holds strictly converges to a continuous solution as κ approaches the lower bound (29) from above. The following proposition summarizes these results.

Proposition 1 *Fix some thresholds $b(\gamma) < x(\gamma) < y$ and $b(\beta) < x(\gamma)$. Suppose $\{\beta, \delta\} \subset (0, \gamma)$ and that the balanced growth rate κ satisfies (29). Then the Kolmogorov forward equation (22)-(27) can be solved for a continuous stationary density $[m(\beta, z), m(\gamma, z)]$, and the solution is unique up to scale. The implied density of $(\lambda, z-y)$ only depends on the threshold gaps $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$.*

The growth rate κ for which (29) holds with equality will be important below. The right tail of $m(\gamma, z)$ will then behave like $e^{-\zeta_\gamma z}$, with a tail index ζ_γ defined by (30). The integrability condition for $m(\beta, z)$ implies that the tail of $m(\beta, z)$ behaves like $e^{-\xi_+ z}$, with $\xi_+ = ((\kappa - \mu)/\sigma^2) + \sqrt{((\kappa - \mu)/\sigma^2)^2 + \delta/(\sigma^2/2)}$. When (29) holds with equality, this says that $m(\beta, z)$ has a tail index $\xi_+ = \zeta_\beta$ given by

$$\zeta_\beta = \zeta_\gamma + \sqrt{\frac{\gamma}{\sigma^2/2}}. \tag{31}$$

This is a revealing formula, for two reasons: it does not depend on $\beta < \gamma$, and it immediately implies $\zeta_\beta > \zeta_\gamma$. For both types of managers, the distribution of e^z restricted to (y, ∞) will be like a Pareto. But (31) implies a significantly thicker right tail for fast learners than for slow learners, even though γ may not exceed β by much. The competitive equilibrium assigns only fast learners to the most productive teachers. The differential equation (22) for $m(\beta, z)$ has no flow into (y, ∞), and (31) shows how this sorting on learning ability magnifies the effect of even small differences in ability on the distribution of managerial productivities.

3.3 Balanced Growth Paths

The value functions and stationary densities consistent with a balanced growth rate κ that satisfies (29) can be combined to construct a balanced growth path. Throughout this section, fix such a κ. Recall from (14) that the Bellman equation maps the surplus values $[S(\beta), S(\gamma)]/w$ into $(v/w)[e^{b(\beta)}, e^{b(\gamma)}, e^{x(\beta)}, e^{x(\gamma)}, e^y]$. In the scenario in which type-β workers are indifferent students, condition (15) pins down the surplus value of type-β

students at $S(\beta)/w = \beta/\rho$, and $x(\beta) = b(\beta)$. The Bellman equation then defines a map

$$\frac{S(\gamma)}{w} \mapsto \frac{v}{w} \times \left[e^{b(\beta)}, e^{b(\gamma)}, e^{x(\gamma)}, e^{y}\right], \tag{32}$$

defined for any $S(\gamma)/w > \gamma/\rho$, and shown in Figure 2. This results in one map $S(\gamma)/w \mapsto ve^{y}/w$. A second such map can be constructed by combining $S(\gamma)/w \mapsto [y - b(\beta), y - b(\gamma), y - x(\gamma)]$ with the stationary densities implied by Proposition 1 and market-clearing conditions.

3.3.1 Student-Teacher Market Clearing

Proposition 1 determines the stationary densities $m(\beta, z)$ and $m(\gamma, z)$ up to scale, taking for granted that the number of workers who want to be students matches the supply of teaching services of incumbent managers. To ensure that this is the case, it remains to impose the market-clearing conditions for students and teachers,

$$M(\beta) - \int_{b(\beta)}^{\infty} m(\beta, z)\mathrm{d}z \;\geq\; \int_{b(\beta)}^{y} m(\beta, z)\mathrm{d}z + \int_{x(\gamma)}^{y} m(\gamma, z)\mathrm{d}z, \tag{33}$$

$$M(\gamma) - \int_{b(\gamma)}^{\infty} m(\gamma, z)\mathrm{d}z \;=\; \int_{y}^{\infty} [m(\beta, z) + m(\gamma, z)]\mathrm{d}z - \int_{b(\gamma)}^{x(\gamma)} m(\gamma, z)\mathrm{d}z. \tag{34}$$

The left-hand sides of (33)-(34) are the numbers of workers who are potential type-β and actual type-γ students, respectively. The right-hand side of (33) is the aggregate of type-β teachers in $(b(\beta), y)$ and type-γ teachers in $(x(\gamma), y)$, all of whom teach type-β students. This market-clearing condition can be a strict inequality because type-β workers are indifferent between studying or not. The right-hand side of (34) is the number of type-β and type-γ managers in (y, ∞), all of whom teach type-γ students, minus the type-γ students who are already managers. One can think of (34) as determining the scale of $[m(\beta, z), m(\gamma, z)]$, and of (33) as a side condition that amounts to a lower bound on how large the supply of type-β households must be for type-β workers to be indifferent students in equilibrium. Type-β workers can be indifferent students in equilibrium only if there are enough type-β households.

3.3.2 The Implied Factor Supplies

Given a stationarity density for $(\lambda, z - y)$ determined by Proposition 1 and (34), one can compute the supply of variable labor used to produce consumption, as well as the stock of managerial human capital relative to e^{y}. First, note that the number of managers is

given by

$$N = \int_{b(\beta)}^{\infty} m(\beta, z)\mathrm{d}z + \int_{b(\gamma)}^{\infty} m(\gamma, z)\mathrm{d}z. \tag{35}$$

The resulting supply of labor is $M(\beta) + M(\gamma) - N$. The aggregate use of overhead labor is ϕN, and so the supply of variable labor is

$$L = M(\beta) + M(\gamma) - (1 + \phi)N. \tag{36}$$

Next, the supply of managerial human capital (3) must satisfy $H_t = He^{\kappa t}$, where

$$He^{-y} = \int_{b(\beta)}^{\infty} e^{z-y} m(\beta, z)\mathrm{d}z + \int_{b(\gamma)}^{\infty} e^{z-y} m(\gamma, z)\mathrm{d}z. \tag{37}$$

Recall that the density of $(\lambda, z - y)$ determined in Proposition 1 only depends on the threshold gaps $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$, and so (35)-(37) only depend on these gaps and not on y itself. Because of the unknown scale factor e^{-y}, (36)-(37) is not enough information to compute aggregate consumption and the factor prices $[v, w]$. But the Cobb-Douglas technology implies $vH/(wL) = (1 - \alpha)/\alpha$, and therefore

$$\frac{ve^{y}}{w} = \frac{1 - \alpha}{\alpha} \frac{L}{He^{-y}}. \tag{38}$$

Combining Proposition 1, the market-clearing condition (34), and (35)-(38) therefore gives an equilibrium condition of the form

$$[y - b(\beta), y - b(\gamma), y - x(\gamma)] \mapsto \frac{ve^{y}}{w}. \tag{39}$$

The shape of the stationary density of managerial productivities pins down the variable profits, expressed in units of labor, of managers at the threshold y.

3.3.3 The Equilibrium Level of a Balanced Growth Path

Figure 2 shows the equilibrium conditions $S(\gamma)/w \mapsto ve^{y}/w$ given by (32) and the composition of (32) and (39). These curves are labeled ve^{y}/w and $((1 - \alpha)/\alpha)L/He^{-y}$, respectively. Their intersection determines $S(\gamma)/w$ and ve^{y}/w. The threshold gaps $y - b(\beta) - y$, $y - b(\gamma)$, and $y - x(\gamma)$ then follow from (32), and the stationary density of $(\lambda, z - y)$ in turn follows from Proposition 1. The distribution of (λ, z) at the initial date $t = 0$ is given. If this distribution is consistent with balanced growth at the rate κ used in this construction, then this initial distribution determines the value of y: it is the y for which the distribution of $(\lambda, z - y)$ at $t = 0$ matches distribution of $(\lambda, z - y)$

implied by Proposition 1. With this solution for y, one can compute H from He^{-y}, and then aggregate consumption and the factor prices $[v, w]$.

A minimal requirement for this construction to work is that the mean of e^{z-y} is finite. Necessary and sufficient for this to be the case for some κ consistent with (29) is the requirement $\zeta_\gamma > 1$. If κ exceeds the lower bound given by (29), then $\zeta_+ > \zeta_\gamma > \zeta_-$, and so the mean of e^{z-y} will not be finite for any κ that satisfies (29) if it is not finite when (29) holds with equality. Suppose therefore that $\zeta_\gamma > 1$. This is equivalent to $\gamma > \delta + \sigma^2/2$, a tightening of the requirement $\gamma > \delta$ needed in Proposition 1 to construct a stationary density.

The construction can still fail if the curve $S(\gamma)/w \mapsto ve^y/w = ((1-\alpha)/\alpha)L/He^{-y}$ implied by the composition of (32) and (39) is everywhere below the curve $S(\gamma)/w \mapsto ve^y/w$ implied by (32) itself, even at $S(\gamma)/w = \gamma/\rho$. This can happen when the ratio

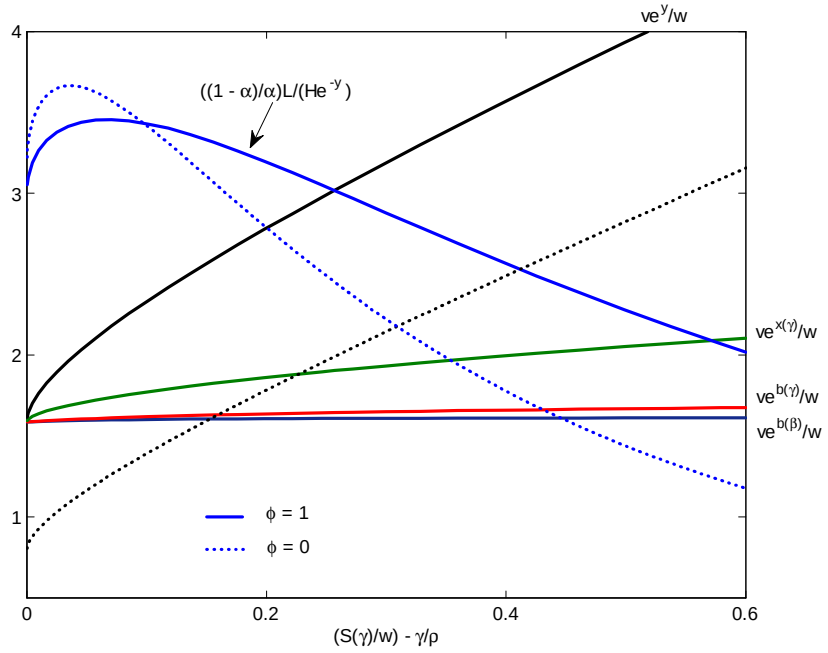


FIGURE 2 Condition (32) and the Composition of (32) and (39).

$M(\gamma)/M(\beta)$ is too high—there are too many fast learners. One can then construct an equilibrium in which ability rents are zero not only for type-β households but also for type-γ households, as outlined in Appendix B. The remaining possibility is that the equilibrium conditions in Figure 2 do intersect for some $S(\gamma)/w > \gamma/\rho$, but the side condition (33) fails. That is, even if they all choose to be students, there are not enough type-β workers to maintain the population of type-β managers required for stationarity. When this happens, $M(\gamma)/M(\beta)$ is still too high. In that case, type-β households must

also earn strictly positive ability rents, and one obtains the scenario in which at least half of the population will be a manager.

3.4 The Emergence of Long-Run Growth

A balanced growth path of the type illustrated in Figure 2 can be constructed if $\gamma > \delta + \sigma^2/2$, type-$\gamma$ households are sufficiently scarce, and the assumed balanced growth rate κ is high enough to satisfy (29). What then determines the long-run growth rate κ?

A rigorous answer to this question can be given in a much simpler economy in which there is only one type of manager, say with learning rate γ, and managers never exit, say because $\phi = 0$ and there is a fixed population of managers who cannot supply labor. With the innocuous further simplification $\mu = 0$, the Kolmogorov forward equation for the time-t density of the state of a manager (not detrended in any way) is given by $\mathrm{D}_t m(t,z) = \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(t,z) - \gamma m(t,z)$ below the median of $m(t,z)$ and $\mathrm{D}_t m(t,z) = \frac{1}{2}\sigma^2 \mathrm{D}_{zz} m(t,z) + \gamma m(t,z)$ above the median. The associated measure $M(t,z)$ then satisfies

$$\mathrm{D}_t M(t,z) = \frac{1}{2}\sigma^2 \mathrm{D}_{zz} M(t,z) - \gamma \min\{M(t,z), M(\gamma) - M(t,z)\}. \tag{40}$$

As in the more general case, there is a continuum of long-run growth rates κ that have to satisfy the bound (29), with associated stationary distributions that are easy to compute. The partial differential equation (40) is a reaction-diffusion equation. A key feature of this particular equation is that the term subtracted on the right-hand side of (40) is a hump-shaped function of $M(t,z)$ with zeros at $M(t,z) = 0$ and $M(t,z) = M(\gamma)$. For equations of this type, Kolmogorov, Petrovskii, and Piskunov [1937] and others have shown that the long-run stationary distribution that emerges from an initial distribution with bounded support is the one associated with κ at the lower bound (29).[7] It is essential for this result that σ is positive. It is not difficult to show that when $\sigma = 0$ (as in related models of Alvarez, Buera, and Lucas [2008], Lucas and Moll [2014], and Perla and Tonetti [2014]), there is a continuum of stationary distributions and growth rates, not only for exponentially de-trended managerial productivities but also for linearly detrended productivities. A bounded initial distribution of productivities in such a

[7] Alternative proofs and generalizations can be found in McKean [1975], Bramson [1983], and a large literature on reaction-diffusion equations. Fisher [1937] formulated the same equation to describe the geographic spread of an advantageous gene. Cavalli-Sforza and Feldman [1981] used the Fisher [1937] interpretation of these equations to describe the geographic spread of a useful idea. Here geography is absent, but there are many ideas, of varying quality. See Staley [2011] and Luttmer [2012] for a more detailed discussion.

world would lead to permanent stagnation because nobody can become more productive than the most productive manager in the initial population.

Applying this idea here results in the prediction that the economy grows, in the long run, at a rate $\kappa = \mu + \sigma^2\zeta_\gamma = \mu + \sigma\sqrt{2(\gamma - \delta)}$. The economy can grow rapidly because managers are able to improve their own productivities at a rapid pace (μ is high), because they can learn quickly from others ($\gamma - \delta$ is high), or because their individual discovery processes are noisy (σ is high). The only learning speed that matters is that of the fastest learners, conditional on their survival. Noisy individual discovery processes produce a lot of dispersion in managerial productivities. Because high and low z can be learned at rates that are unrelated to the level of z, this implies a population with many particularly productive learning opportunities, and therefore rapid growth.

When (29) holds strictly, the right tail of $m(\gamma, z)$ behaves like $e^{-\zeta_- z}$, with $\zeta_- < \zeta_\gamma$ given by (28). The above argument therefore selects the stationary distribution with the thinnest right tail. The expression $\zeta_\gamma = \sqrt{(\gamma - \delta)/(\sigma^2/2)}$ says that rapid growth goes together with a particularly thick-tailed distribution of managerial productivities if it is due to a high σ, but with a thinner tail if due to a high $\gamma - \delta$. Across economies on a balanced growth path, there is no unambiguous relation between growth and inequality: it depends on how economies differ in terms of the underlying parameters.

3.4.1 Relation to Models of Imitation

A similar multiplicity of stationary distributions and balanced growth rates arose in Luttmer [2007]. It was resolved in the same way, by taking the initial distribution of productivities to have bounded support. There, when entrepreneurs can imitate incumbent producers at the rate γ, the resulting growth rate is again $\kappa = \mu + \sigma^2\zeta_\gamma$. But the tail index is $\zeta_\gamma = \sqrt{[(\gamma E/N) - \delta]/(\sigma^2/2)}$, where E is the number of entrepreneurs trying to imitate and N is the number of incumbent firms. The entry rate $\gamma E/N$ is determined in equilibrium and depends on the rate at which consumers discount. The key difference with the assignment model studied here is that, in an economy with imitation, there is no bound on how many can flow into the right tail of the productivity distribution. Here, this flow is limited by the capacity of particularly knowledgeable managers to teach others.

Without any entry or exit, random meetings at a rate γ combined with imitation give rise to the same dynamics as (40), with the last term on the right-hand side replaced by $-\gamma[M(\gamma) - M(t, z)]M(t, z)/M(\gamma)$. That is, a tent is replaced by a parabola in $M(t, z)$. By linearizing this parabola near $M(t, \infty) = M(\gamma)$, one can show that the stationary

distribution will have the same tail index ζ_γ (see Staley [2011] and Luttmer [2012]). The right tail properties of the stationary distribution of managerial productivities cannot be used to distinguish between random meetings with imitation and one-on-one teaching with delayed learning.

4. Quantitative Implications

The type of learning from others modeled in this paper is best thought of not as resulting from formal education but as learning that happens on the job during a career. The ability to successfully engage in such forms of learning probably depends not only on narrowly defined cognitive traits but also on a range of social skills. And unmodeled prior education is likely to affect these social learning speeds as well, by allowing students to acquire a language that they can subsequently use to communicate with and learn from others.

4.1 The Magnification Effect

Modest differences in learning speeds can have rather drastic consequences for who tends to become and succeed as a manager. To illustrate this, consider the following benchmark specification of an economy with a subjective discount rate $\rho = 0.04$, a random exit rate $\delta = 0.04$, and learning speeds $\beta = 0.05$ and $\gamma = 0.06$, with time measured in years. That is, both slow and fast learners are more likely than not (if only just) to learn something useful from someone else before they die or are forced to exit randomly. Consider further an individual discovery process with a standard deviation $\sigma = 0.10$. For US social security data, Guvenen et al. [2015] report a cross-sectional variance of log earnings that increases by .45 between age 25 and age 60. For a pure random walk, such an increase would correspond to a standard deviation of $\sqrt{.45/35} \approx 0.11$. Here, only managers are subject to Brownian productivity shocks, but the dispersion in earnings of both managers and workers is also affected by large positive and negative jumps that occur when students learn from teachers and when managerial projects fail.

The long-run growth rate of the aggregate stock of managerial human capital, in excess of the average rate μ at which managers can improve their own productivities, follows immediately from the assumed rate parameters,

$$\kappa - \mu = \sigma\sqrt{2(\gamma - \delta)} = 0.10\sqrt{2(0.06 - 0.04)} = 0.02.$$

The implied tail indices (30) and (31) of the densities $m(\gamma, z)$ and $m(\beta, z)$ are also

immediate:

$$\zeta_\gamma = \sqrt{\frac{0.06 - 0.04}{(0.1)^2/2}} = 2, \quad \zeta_\beta = 2 + \sqrt{\frac{0.06}{(0.1)^2/2}} \approx 5.5.$$

Workers all earn the same wage, and so the tail index of the overall income distribution will be $\zeta_\gamma = 2$. For the low-inequality 1950s and 1960s, Piketty and Saez [2003] report that the top 10% of US taxpayers by income (not including capital gains) earned about 32.5% of aggregate income. For a Pareto distribution, this would correspond to a tail index of $1/(1 - \ln(.325)/\ln(.1)) \approx 1.95$. The densities $m(\beta, z)$, $m(\gamma, z)$ and $m(\beta, z) + m(\gamma, z)$ are shown in the first panel of Figure 3, and the second panel shows the odds ratio $m(\beta, z)/[m(\beta, z) + m(\gamma, z)]$. The rapid decay of the right tail of $m(\beta, z)$ compared with that of $m(\gamma, z)$ is the immediate reason for the fact that most managers in the right tail are fast learners, as illustrated in the second panel of Figure 3. Although the difference in learning speeds is small, only fast learners are assigned to teachers in the right tail of the productivity distribution, and so only fast learners can jump into the right tail. Slow learners can make it into the right tail, but only as a result of a long string of useful individual discoveries.

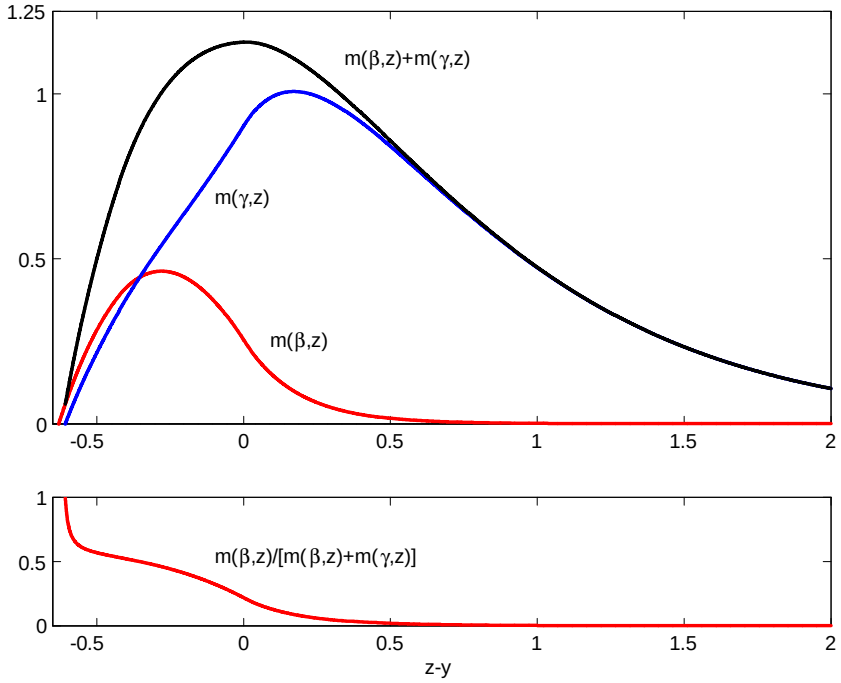


FIGURE 3 The Stationary Densities $m(\lambda, z)$, $\lambda \in \{\beta, \gamma\}$

Figures 2 and 3 were computed for a version of this economy in which 10% of the population is a fast learner, overhead labor equals $\phi = 1$, and the labor share parameter

is $\alpha = 0.60$. As can be seen in these figures, the exit thresholds used by the two types of managers are very similar. In terms of variable profits, these managers earn $ve^{b(\beta)}/w \approx 1.61$ and $ve^{b(\gamma)}/w \approx 1.64$, respectively. This is more than the unit wage earned by workers, but managers also have to cover the fixed cost of $\phi = 1$ units of labor. After fixed costs, their earnings from producing consumption at the exit threshold are below what they could earn as a worker, a reflection of the option value of continuing as a manager. Type-β managers are not earning anything from being teachers at the exit threshold, since they are teaching their own type of students for a tuition equal to $\beta W(\beta) - S(\beta) = 0$. Type-$\gamma$ managers at their exit threshold earn only modest net gains equal to $[S(\gamma) - \gamma W(\gamma)]/w \approx 0.10$ as students, but they can earn the same net gains when studying as workers. The value in units of labor of a type-β worker is simply $1/\rho = 1/.04 = 25$ in this economy. The ability to learn fast results in a value of $W(\gamma)/w = (1 + [S(\gamma) - \gamma W(\gamma)]/w)/\rho \approx 1.1/.04 = 27.5$ for fast learners. This is only about 10% more than $W(\beta)/w$, even though the ex post outcomes for fast and slow learners implied by the tail indices $\zeta_\gamma = 2$ and $\zeta_\beta \approx 5.5$ are very different.

Along the balanced growth path, managers make up about 6.6% of the population. The detrended wages of workers are constant, and so the cross-sectional standard deviation of the continuous part of earnings growth is only about $0.1 \times \sqrt{0.066} \approx 0.026$. Earnings growth in this economy exhibits the combination of stability and large jumps observed in US social security data (Guvenen et al. [2015]). Workers who learn go from earning the wage w as a worker to earning $ve^z - \phi w$ from managing other workers. The distribution of e^z among new managers is just the distribution of e^z among their teachers, who are managers in $(b(\beta), y)$ for slow learners and managers in (y, ∞) for fast learners. For the latter, the distribution of earnings growth of workers who succeed in learning to be managers thus inherits the thick-tailed distribution of e^z. Similarly, every manager is subject to random exit shocks, and such a shock takes a manager from $ve^z - \phi w$ back to w. The left tail of the distribution of earnings growth will therefore also inherit the thick tail of the distribution of e^z. The prevalence of these large positive and negative jumps can be adjusted by changing the gross flows γ and δ by the same amount. This will also affect the tail index ζ_β but leaves the growth rate of the economy and the tail index ζ_γ of the overall earnings distribution unaffected.

Because of the assumed logarithmic preferences, none of this depends on the rate μ at which managers can improve their own productivity. Suppose managers can improve on their own at a rate of $\mu = 0.02$. Then managerial human capital grows at an annual rate of 4%. The assumed labor share parameter of 60% implies that consumption grows

at 1.6% per annum, and the logarithmic utility function then implies an interest rate of 5.6%.

4.2 Ability Rents

Recall from the Bellman equation (10) that the ability rents $(W(\lambda)-w/\rho)/w$ are strictly positive if and only if $S(\lambda)/w > \lambda/\rho$ and strictly increasing in $S(\lambda)/w$ in that case. In the scenario of interest, only fast learners earn ability rents. These rents depend in intuitive ways on technology parameters and the supplies of slow and fast learners.

It is easy to see that the ability rents of fast learners increase with a decline in the Cobb-Douglas labor share parameter α. A reduction in α shifts the curve $ve^y/w = [(1-\alpha)/\alpha]L/He^{-y}$ defined by the composition of (32) and (39) upward, and nothing else changes. Since the equilibrium condition (39) for ve^y/w is upward sloping in $S(\gamma)/w$, the result is immediate. As Figure 2 shows, managers of either type at the thresholds $[b(\beta), b(\gamma), x(\gamma), y]$ will earn more from managing workers as a result of the implied increase in $S(\gamma)/w$. But workers who are slow learners will continue to be indifferent students, and hence it must be that for slow learners these increased managerial earnings are fully absorbed by increases in the equilibrium tuition required to become a manager.

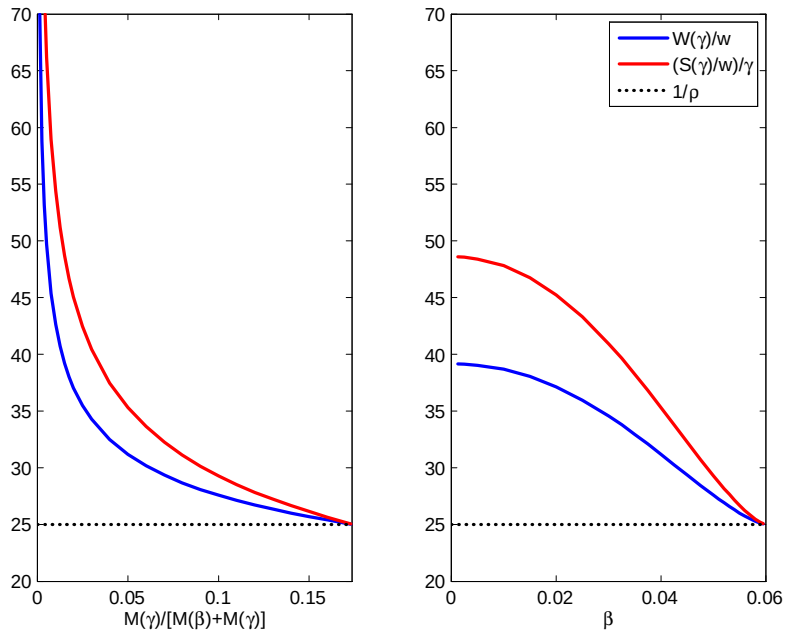


FIGURE 4 Learning Ability Supplies and Rents

It turns out that a reduction in overhead labor costs also increases the ability rents of fast learners. Figure 2 shows how the equilibrium conditions change when the overhead

labor costs are reduced from $\phi = 1$ to $\phi = 0$. There is still an opportunity cost to being a manager when $\phi = 0$, because managers forego the opportunity to supply labor, but there no longer any overhead. The immediate effect is to make it more profitable to be a manager. As with a decline in the labor share parameter, some of the increase in managerial profitability translates into higher ability rents for fast learners.

Figure 4 shows how the ability rents of type-γ agents vary with their supply and with the learning ability β of slow learners. These ability rents are quite sensitive to the supply of fast learners: cutting the supply in half from 10% more than doubles the ability rent $W(\gamma) - w/\rho$ of type-γ workers. Conversely, raising the relative supply to just above 17% eliminates all the ability rents of fast learners. This number matches the threshold (41) above which one can construct an equilibrium with zero ability rents and only type-γ managers. The ability rents of fast learners also disappear, holding fixed the relative supply $M(\gamma)/[M(\beta) + M(\gamma)]$, when slow learners become almost as fast as fast learners. As long as $\beta < \gamma$, the tail indices ζ_γ and $\zeta_\beta = \zeta_\gamma + \sqrt{\gamma/(\sigma^2/2)}$ do not depend on β, and so ex post outcomes for these two types of learners will remain very different when β approaches γ from below. But the second panel of Figure 4 shows that ex ante rents are continuous.

4.3 These Managers Cannot Be Firms

It is tempting to interpret the team of workers employed by a particular manager to be a firm, as in Lucas [1978]. But this interpretation runs into an important empirical difficulty. The parameters used here give rise to an earnings distribution with a tail index $\zeta_\gamma = 2$ that is in line with the US earnings distribution. The Cobb-Douglas technology implies that employment per manager and managerial earnings are both proportional to the managerial productivity state e^z. It therefore follows that the model would predict an employment size distribution of firms with a tail index equal to 2. In US data, the firm size distribution has a tail index of about 1.05 (see, for example, Luttmer [2007]). These are very different distributions. A Pareto-like distribution with a tail index as close to 1 as 1.05 implies a Lorenz curve that is almost rectangular, unlike a Pareto-like distribution with tail index 2 and unlike the Lorenz curve for US earnings.[8] A resolution of this difficulty probably involves taking seriously the fact that large firms are not made

[8] Cagetti and De Nardi [2006] emphasize the prevalence of business owners and entrepreneurs among wealthy households. Jones and Kim [2014] also note the difference between income and firm size distributions. Ai, Kiku, and Li [2014] present a model in which firm size and managerial income distributions have different tail indices.

up of just a single manager with homogeneous workers, but of more intricate associations of heterogeneous managers and workers.

5. Concluding Remarks

The model in this paper considers the opposite extreme of one that is more common in the literature: knowledge diffusion here is all about teaching, whereas much of the literature is about imitation. Teaching actively involves students and teachers, whereas imitation is a more individual activity. Both phenomena play a role in real-world knowledge accumulation, but it is far from obvious how economically important they each are. A fundamental difficulty is that it is often hard to know if two individuals working together are just producing widgets or also transferring knowledge. Very close observation may reveal the answer in specific instances, and one may be able to provide a rough count of how frequent such instances are. But it seems almost incredible that this can be measured with some precision at the aggregate level.

Arbitrarily small differences in learning abilities generate large differences in ex post outcomes, because, with a limited supply of knowledgeable teachers, only fast learners are assigned to the teachers with the most productive knowledge. The paper shows this for an economy with slow and fast learners. More generally, one expects the basic fact that a competitive equilibrium assigns students with faster learning abilities to teachers with more productive knowledge to magnify the dispersion of outcomes that would arise even in an economy with heterogeneous learning abilities and a random assignment of students to teachers. But the ex ante welfare implications for students with different learning abilities will not be as stark as the ex post outcomes, because teachers with productive knowledge to impart are expensive.

The managerial activities of supervising production and teaching others are separable in the economy described in this paper. Managers can teach students while supervising workers who need not be their students. This makes for an extremely tractable model of knowledge transmission and long-run growth, and of the role of both ability and randomness in shaping labor market outcomes. But formal and informal apprentice systems observed in actual economies suggest a complementarity between supervision and teaching. Exploring the effects of such a complementarity on growth and inequality is a worthwhile topic for further research.

A Proof of Lemma 3

Let $\Lambda_+ = \{\lambda \in \Lambda : S_t(\lambda) - \lambda W_t(\lambda) \geq 0\}$ and $\Lambda_- = \Lambda \backslash \Lambda_+$. The maintained assumptions about the value functions imply the existence of finite thresholds $b_t(\lambda) = \min\{z : V_t(z|\lambda) > W_t(\lambda)\}$ and $x_t(\lambda) = \min\{z : \lambda V_t(z|\lambda) \geq S_t(\lambda), z \geq b_t(\lambda)\}$, for all $\lambda \in \Lambda$. Let $x_+ = \min\{x_t(\lambda) : \lambda \in \Lambda_+\}$, and let λ_+ be any type that attains this minimum. These definitions imply that $\lambda_+ V_t(x_+|\lambda_+) - S_t(\lambda_+) = 0$ and therefore $\lambda_+ V_t(z|\lambda_+) - S_t(\lambda_+) > 0$ for all $z > x_+$. Since $x_+ \geq b_t(\lambda_+)$, there is assumed to be a positive supply of type-λ_+ managers at any $z > x_+$. These managers strictly prefer to teach, and so there will have to be students willing to study with managers in all states $z > x_+$. The types $\lambda \in \Lambda_-$ will never be students, not as workers and certainly not as managers.

Suppose now that $T_t(z)$ is strictly positive for all z. For type-λ_+ managers, this implies $T_t(x_+) > 0 = \lambda_+ V_t(x_+|\lambda_+) - S_t(\lambda_+)$, and hence $S_t(\lambda_+) > \lambda_+ V_t(x_+|\lambda_+) - T_t(x_+)$. This inequality will be true as well for all $z > x_+$ close to x_+. Therefore, no type-λ_+ students select teachers in any state $z > x_+$ near x_+. What about a type-λ students with $\lambda \neq \lambda_+$ and $\lambda \in \Lambda_+$? If $x_+ < b_t(\lambda)$, then type-λ workers cannot learn anything from managers in the states $z \in (x_+, b_t(\lambda))$ that would make them viable as a manager. Alternatively, $x_t(\lambda) > x_+ \geq b_t(\lambda)$ and so $\lambda V_t(x_+|\lambda) < S_t(\lambda)$. This implies that type-λ students will not select managers in states $z > x_+$ close to x_+. This rules out these types as students as well. So there are no students of any type who select teachers in the states $z > x_+$ close enough to x_+. This means the markets for students and teachers at $z > x_+$ close to x_+ do not clear. Market clearing therefore requires that $\min_z\{T_t(z)\} = 0$.

B The Scenario $S(\lambda)/\lambda = w/\rho$ for Both $\lambda \in \{\beta, \gamma\}$

In this scenario, as argued in Section 3.1, all managers teach and both teachers and students are fast learners. Define

$$e^{\widehat{z}} = \frac{ve^z}{(1+\phi)w}, \quad \widehat{V}(\widehat{z}) = \frac{V(z|\gamma) - w/\rho}{(1+\phi)w}.$$

Using the fact that $T(z) = \gamma(V(z|\gamma) - w/\rho)$ and $W(\gamma) = w/\rho$, the Bellman equation (11) can be written as

$$(\rho + \delta - \gamma)\widehat{V}(\widehat{z}) = e^{\widehat{z}} - 1 + (\mu - \kappa)\mathrm{D}\widehat{V}(\widehat{z}) + \frac{1}{2}\sigma^2 \mathrm{D}^2 \widehat{V}(\widehat{z}),$$

for all $\widehat{z} > \widehat{b}$, together with the boundary conditions $0 = \widehat{V}(\widehat{b}) = \mathrm{D}\widehat{V}(\widehat{b})$. The solution is reported in Luttmer [2007],

$$\widehat{V}(\widehat{z}) = \frac{1}{\rho + \delta - \gamma} \frac{\omega}{1+\omega} \left(e^{\widehat{z} - \widehat{b}} - 1 - \frac{1 - e^{-\omega(\widehat{z} - \widehat{b})}}{\omega} \right),$$

for $\widehat{z} \geq \widehat{b}$, where

$$e^{\widehat{b}} = \frac{\omega}{1+\omega} \left(1 - \frac{\mu - \kappa + \frac{1}{2}\sigma^2}{\rho + \delta - \gamma} \right), \quad \omega = \frac{\mu - \kappa}{\sigma^2} + \sqrt{\left(\frac{\mu - \kappa}{\sigma^2} \right)^2 + \frac{\rho + \delta - \gamma}{\sigma^2/2}}.$$

The exit threshold for the original state satisfies $ve^{b(\gamma)}/w = (1+\phi)e^{\widehat{b}}$.

There are no type-β managers, and the Kolmogorov forward equation for $m(\gamma, z)$ is

$$0 = -(\mu - \kappa)\mathrm{D}_z m(\gamma, z) + \frac{1}{2}\mathrm{D}^2_{zz} m(\gamma, z) + (\gamma - \delta) m(\gamma, z),$$

for all $z > b(\gamma)$, with the boundary condition $m(\gamma, b(\gamma)) = 0$. The roots of the characteristic equation are $\zeta_+ \geq \zeta_- > 0$, as defined in (28). If $\zeta_+ > \zeta_- > 0$, then the solution is $m(\gamma, z) = Nf(z)$, where

$$f(z) = \zeta_+ \zeta_- \times \frac{e^{-\zeta_-(z-b)} - e^{-\zeta_+(z-b)}}{\zeta_+ - \zeta_-}, \quad z \geq b(\gamma)$$

and N is the yet to be determined measure of managers. Taking the limit $\zeta_\pm \to \zeta_\gamma$ that arises when (29) holds with equality gives

$$f(z) = \zeta_\gamma^2 (z - b(\gamma)) e^{-\zeta_\gamma (z - b(\gamma))}, \quad z \geq b(\gamma).$$

This Gamma density also appears in Luttmer [2007]. A similar limit applies in the general case of Proposition 1.

The mean of $e^{z - b(\gamma)}$ is given by $(\zeta_\gamma/(\zeta_\gamma - 1))^2$, and so the scaled supply of managerial human capital $He^{-b(\gamma)}$ equals N times this mean. The equilibrium values of N and $ve^{b(\gamma)}/w$ follow from

$$(1+\phi)e^{\widehat{b}} = \frac{ve^{b(\gamma)}}{w} = \frac{1-\alpha}{\alpha} \frac{L}{He^{-b(\gamma)}} = \frac{1-\alpha}{\alpha} \frac{M(\beta) + M(\gamma) - (1+\phi)N}{(\zeta_\gamma/(\zeta_\gamma - 1))^2 N}.$$

The Bellman equation pins down the left-hand side, and the right-hand side only depends on N. This leads to a well-defined equilibrium if and only if $M(\gamma) \geq 2N$. All N managers are fast learners and they all teach. So there must be N type-γ workers who are students.

This is only possible if there are at least $2N$ type-γ households. Combining $M(\gamma) \geq 2N$ together with the equilibrium value for N gives

$$\frac{M(\gamma)}{M(\beta)+M(\gamma)} \geq \frac{1}{\frac{1+\phi}{2}\left(1+\frac{\alpha}{1-\alpha}\left(\frac{\zeta_\gamma}{\zeta_\gamma-1}\right)^2 e^{\widehat{b}}\right)}. \tag{41}$$

Nobody earns rents associated with being a student if and only if there are enough type-γ households. If parameters are such that the right-hand side (41) is greater than 1, then some workers will have to strictly benefit from being students, no matter what the skill distribution.

References

[1] Ai, H., D. Kiku, and R. Li, "A Mechanism Design Model of Firm Dynamics: The Case of Limited Commitment, University of Minnesota (2014).

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

[3] Alvarez, F.E., F.J. Buera, and R.E. Lucas, Jr., "Idea Flows, Economic Growth, and Trade," NBER working paper 19667 (2013).

[4] Antras, P., L. Garicano, and E. Rossi-Hansberg, "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, vol. 121, no. 1 (2006), 31-77.

[5] Beaudry, P. and P. Francois, "Managerial Skill Acquisition and the Theory of Economic Development," *Review of Economic Studies*, vol. 77, no. 1 (2010), 90-126.

[6] Benhabib, J., 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.

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

[8] Boldrin, M. and D.K. Levine, "Perfectly Competitive Innovation," *Journal of Monetary Economics*, vol. 55, no. 3 (2008), 435-453.

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

[10] Burstein, A.T. and A. Monge-Naranjo (2009), "Foreign Know-How, Firm Control, and the Income of Developing Countries," *Quarterly Journal of Economics*, vol. 124, no. 1 (2009), 149-195.

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

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

[13] 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.

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

[15] Dixit, A.K. and R.S. Pindyck, *Investment under Uncertainty*, Princeton University Press (1994).

[16] Fisher, R.A., "The Wave of Advance of Advantageous Genes," *Annals of Eugenics*, vol. 7, no. 4 (1937), 355-369.

[17] Gabaix, X., "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, vol. 114, no. 3 (1999), 739-767.

[18] Gabaix, X. and A. Landier, "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, vol. 123, no. 1 (2008), 49-100.

[19] Guvenen, F., F. Karahan, S. Ozkan, J. Song, "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" University of Minnesota (2015).

[20] Jones, C.I. and J. Kim, "A Schumpeterian Model of Top Income Inequality," Stanford University (2014).

[21] Jovanovic, B., "Misallocation and Growth," *American Economic Review*, vol. 104, no. 4 (2014), 1149–1171.

[22] Jovanovic, B. and R. Rob, "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, vol. 56, no. 4 (1989), 569-582.

[23] Keller, W. and S. Yeaple, "The Gravity of Knowledge," *American Economic Review*, vol. 103, no. 4 (2013), 1414-1444.

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

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

[27] Le, T., "Competitive Knowledge Diffusion and Economic Growth," senior thesis, University of Minnesota (2014).

[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, no. 3 (2009), 1-19.

[30] Lucas, R.E., Jr. and B. Moll, "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, vol. 122, no. 1 (2014), 1-51.

[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., "On the Mechanics of Firm Growth," *Review of Economic Studies*, vol. 78, no. 3 (2011), 1042-1068.

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

[34] Merton, R.K., "The Matthew Effect in Science," *Science*, vol. 159, no. 3810 (1968), 56-63.

[35] McKean, H.P., "Application of Brownian Motion to the Equation of Kolmogorov-Petrovskii-Piskunov," *Communications on Pure and Applied Mathematics*, vol. 28, no. 3 (1975), 323-331; "Erratum," vol. 29, no. 5 (1976), 553-554.

[36] 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.

[37] Nocke, V. and S. Yeaple, "An Assignment Theory of Foreign Direct Investment," *Review of Economic Studies*, vol. 75, no. 2 (2008), 529-557.

[38] Perla, J. and C. Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, vol. 122, no. 1 (2014), 52-76.

[39] Piketty, T. and E. Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, vol. 118, no. 1 (2003), 1-39.

[40] Ramondo, N. and A. Rodriguez-Clare, "Trade, Multinational Production, and the Gains from Openness," *Journal of Political Economy*, vol. 121, no. 2 (2013), 273-322.

[41] Rosen, S., "The Economics of Superstars," *American Economic Review*, vol. 71, no. 5 (1981), 845-858.

[42] Rosen, S., "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics*, vol. 13, no. 2 (1982), 311-323.

[43] Staley, M., "Growth and the Diffusion of Ideas," *Journal of Mathematical Economics*, vol. 47, no. 4-5 (2011), 470-478.

[44] Terviö, M., "The Difference That CEOs Make: An Assignment Model Approach," *American Economic Review*, vol. 98, no. 3 (2008), 642-668.