An Assignment Model of Knowledge
Diffusion and Income Inequality[*]

Erzo G.J. Luttmer

ABSTRACT

Randomness in individual discovery tends to spread out productivities in a population, while learning from others keeps productivities together. In combination, these two mechanisms for knowledge accumulation give rise to long-term growth and persistent income inequality. This paper considers a world in which those with more useful knowledge can teach those with less useful knowledge, with competitive markets assigning students to teachers. In equilibrium, students who are able to learn quickly are assigned to teachers with the most productive knowledge. The long-run growth rate of this economy is governed by the rate at which the fastest learners can learn. The income distribution reflects learning ability and serendipity, both in individual discovery and in the assignment of students to teachers. Because of naturally arising indeterminacies in this assignment, payoff irrelevant characteristics can be predictors of individual income growth. Ability rents can be large when fast learners are scarce, when the process of individual discovery is not too noisy, and when overhead labor costs are low.

Keywords: Knowledge diffusion; Growth; Inequality
JEL classification: L2, O3, O4

1. Introduction

When some know more than others, it may be efficient to assign the latter to follow the instructions of more their more knowledgeable peers, as in Lucas [1978]. But if the wages of knowledgeable agents are high enough, there are also incentives for less knowledgeable agents to learn. They might be able to do this on their own, by going through all the available information, acquiring more information, and updating. The amount of available 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 transfers of useful knowledge. When knowledgeable teachers are scarce one expects students to have to pay.

This paper describes an economy in which useful knowledge accumulates as a result of both individual discovery and diffusion across individuals. Consumption is produced by teams of managers and workers. Everyone who is not a manager can supply labor, and those who have somehow acquired useful knowledge can be managers. 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. While guiding workers to produce consumption, managers can also teach other managers or workers what they know. Managers and workers can be students while they produce consumption, but managers cannot be students and teachers at the same time. Teaching is one-on-one, and so there is a constraint on how many workers and managers can learn from more knowledgeable managers. Students have to pay teachers for the time it takes to learn what teachers know. The assignment of students to teachers is determined in competitive markets.

Individuals are born with a certain ability to learn, but without any useful knowledge. They begin life as workers. Slow learners may not find it optimal to pay teachers for the time it takes to learn something that would allow them to do better as a manager. Sufficiently fast learners will pay the tuition and try to learn something that allows them to start as a manager. This takes an uncertain amount of time, and then their own discovery process begins. Depending 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

1

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 productivity levels. 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 shift to the right, at a pace that stabilizes 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 has 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. Thick tails make learning more productive. Thus the long-run growth rate depends crucially on the initial distribution of productivity in the economy, at least in principle. But a definite prediction for the long-run growth rates 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, income 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 further heterogeneity. 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. With time, the effects of differences in learning rates accumulate. Moreover, learning from others is expensive, and only sufficiently fast learners will consistently pay the price. Slow learners may not particularly want to become managers, or once managers may not want to pay to learn more. As a result, the right tail of the income distribution is almost exclusively made

2

up of fast learners.

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]. While data on how fast people can learn is 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 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.

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 show that internalized knowledge transfers and imitation externalities are hard to distinguish based only on their implications for income and wealth 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 while knowledge flows inside multinational firms may not.[2] This literature mostly 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, while the assignment problem in Jovanovic [2014] is essentially static.

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

---

[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].

competitive markets. Section 3 specializes to an economy with individuals who have either high or low abilities to learn, and describes the conditions for a balanced growth path. Section 4 illustrates the workings of this economy quantitatively, and Section 5 concludes.

2. Learning Ability and Uncertain Discovery

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 $\lambda \in \Lambda \subset (0, \infty)$. 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 $M_t(z|\lambda) \leq M(\lambda)$ of household members who are managers with productivity state variables in $(-\infty, z]$. The remaining $M(\lambda) - M_t(\infty|\lambda)$ type-λ household members are workers.

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

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

---

[3]The perfect consumption insurance implications of this model are extreme. Plausible relations between wealth and productivity are absent in this economy. As usual, market incompleteness would present a significant analytical complication.

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[4]

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

where H_t and M_t are defined by

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

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

$$w_t(1 - (1+\phi)M_t) = \alpha C_t. \tag{4}$$

Together, (1)-(4) determine C_t, H_t, M_t, v_t and w_t in terms of the measure $\{M_t(z|\lambda) : \lambda \in \Lambda\}$ of managers. This measure is the state variable for this economy. Managers can become workers in an instant and so $M_t(z|\lambda)$ can jump down. But it will take time for workers to become managers again, and so $M_t(z|\lambda)$ cannot jump up. For the purpose of determining aggregate output in this Cobb-Douglas economy, it suffices 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 $M_t(z|\lambda)$.

Anticipating the balanced growth paths that will be constructed, suppose the distribution of $z - \kappa t$ happens to be stationary for some κ. Then (3) implies that H_t grows

---

[4]This familiar aggregation result is a special feature of the Cobb-Douglas technology. More generally, one can take the output of a manager in state z employing l units of labor to be $e^z G(1, l)$, where G exhibits constant returns to scale. This is consistent with balanced growth. When G is Leontief, as in Houthakker [1955-1956], this remains analytically tractable, although there will no longer be a single variable H_t that represents the "aggregate stock of managerial human capital." This would invalidate commonly used growth accounting procedures.

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 the 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. A type-λ worker matched with a teaching manager will learn to adopt the productivity of the teaching manager after a random time, distributed exponentially with mean $1/\lambda$. 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

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

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.

In a large economy, the distribution of managerial human capital at any point in time will have a support that is unbounded above because of the Brownian increments. Combined with the fact that low-z managers can learn from high-z managers, these Brownian increments will cause the distribution to shift to the right, indefinitely. It is important that there is no mean-reversion in human capital, as there is, for example, in Bils and Klenow [2000]. 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 also bounded in the population, but the usefulness of what managers may know is not. In this economy, it is the quality of their knowledge, their ideas, rather than a managerial skill, that makes managers particularly productive in directing the production of consumption.[5]

---

[5]One can generalize (5) by replacing $\widetilde{z}_{t-} - z_{t-}$ with $\max\{\widetilde{z}_{t-} - z_{t-} - \Delta, 0\}$ for some positive Δ, as in Luttmer [2007], and this does not induce mean reversion. 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

Workers can supply labor while they learn to become managers. Managers can also oversee workers producing consumption while engaged in teaching or learning. But managers who teach cannot be students at the same time.[6] It is important to note that the rate at which workers or managers learn is assumed to depend only on their ability type $\lambda \in \Lambda$, and not on what they 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 more difficult. 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 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 $v(t, z) = v_t e^z$

---

after a finite history.

[6]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].

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

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

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. Figure 1 below displays the value functions $V_t(z|\lambda)$ for an economy with two types and Figure 2 illustrates the equilibrium conditions that will be examined in the remainder of this section.

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.

Type-λ managers in state z strictly prefer to teach if and only if $T_t(z) > S_t(\lambda) - \lambda V_t(z|\lambda)$, and then the equilibrium tuition schedule must be of the form (7) at z. Because $T_t(z) \geq \max_{\lambda \in \Lambda}\{\lambda V_t(z|\lambda) - S_t(\lambda)\}$ in any case, this says that type-λ managers in state z strictly prefer to teach if and only if $\lambda V_t(z|\lambda) - S_t(\lambda) + \max_{\mu \in \Lambda}\{\mu V_t(z|\mu) - S_t(\mu)\} > 0$. The left-hand side of this inequality is no less than two times $\lambda V_t(z|\lambda) - S_t(\lambda)$, and so $\lambda V_t(z|\lambda) > S_t(\lambda)$ suffices to ensure that any type-λ managers in state z strictly prefer to teach. All sufficiently productive managers choose to teach. There will be some type of manager who strictly prefers to teach at z if and only if

$$\max_{\mu \in \Lambda}\{\mu V_t(z|\mu) - S_t(\mu)\} > \min_{\lambda \in \Lambda}\{S_t(\lambda) - \lambda V_t(z|\lambda)\}$$

This inequality holds automatically at any z if the surplus values $S_t(\lambda)$ are zero, and becomes less and less likely to hold as the surplus values $S_t(\lambda)$ increase.

10

Lemma 1 together with the definition (6) of $S_t(z)$ characterizes equilibrium tuition schedules. 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 $T_t(z) = \max_{\lambda \in \Lambda}\left\{[\lambda V_t(z|\lambda) - S_t(\lambda)]^+\right\}$.

The construction of $S_t(\lambda)$ implies $S_t(\lambda) \leq S_t^*(\lambda)$, and that then immediately implies $T_t(z) \leq \max_{\lambda \in \Lambda}\left\{[\lambda V_t(z|\lambda) - S_t(\lambda)]^+\right\}$. 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.

2.3.1 All Worker Types are Willing to be Students

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 $S_t(\lambda) - \lambda W_t(\lambda) \geq 0$. The following lemma proves that 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 any z that satisfies* $V_t(z|\lambda) > W_t(\lambda)$. *Then* $\inf\{T_t(z)\} = 0$ *and* $S_t(\lambda) - \lambda W_t(\lambda) \geq 0$ *for all* $\lambda \in \Lambda$.

Proof Let $\Lambda_+ = \{\lambda \in \Lambda : S_t(\lambda) - \lambda W_t(\lambda) \geq 0\}$ and $\Lambda_- = \Lambda\backslash\Lambda_+$. So Λ_+ is the set of types who anticipate non-negative gains from studying as workers, and the claim is that $\Lambda_- = \varnothing$. 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 $z > x_+$.

11

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 $z > x_+$. If Λ_- is non-empty, then there are λ with $S_t(\lambda) - \lambda W_t(\lambda) < 0$, and this can only be because tuition is positive everywhere. The types $\lambda \in \Lambda_-$ will never be students, not as workers, and certainly not as managers. For type-λ_+ managers positive tuition 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, type-λ_+ students do not select teachers in any state $z > x_+$ near x_+. Consider a type $\lambda \neq \lambda_+$ in Λ_+, so that $x_t(\lambda) > x_+$. 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, $x_t(\lambda) > x_+$ may arise because $\lambda V_t(x_+|\lambda) < S_t(\lambda)$, which 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. This is all because a non-empty Λ_- implies positive tuition everywhere and types who never want to be students. So Λ_- must be empty. Tuition must be zero at x_+ as well, or else the above argument can be repeated for $\Lambda_+ = \Lambda$.∎

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.

There are many tuition schedules of the form (7), indexed by surplus values $\{S_t(\lambda) : \lambda \in \Lambda\}$ that satisfy (6). In equilibrium, these surplus values must adjust to ensure the numbers of students and teachers on each side of every market match. The equilibrium surplus values will, of course, depend on the precise distribution of agents at a given point in time.

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

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

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

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

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.

2.4.1 Varying the Overhead Labor Parameter ϕ

The effective fixed cost for a manager is $1 + \phi$ units of labor, in a sense that can be made precise by showing how the value function depends on ϕ. In this section only, write $W_t(\lambda; \phi)$ and $V_t(z|\lambda; \phi)$ for the value functions for type-λ workers and managers when overhead labor is $\phi \geq 0$. Let U_t be the time-t present value of current and future wages. Then U_t must satisfy

$$r_t U_t = w_t + \mathrm{D}_t U_t$$

and a transversality condition. Subtracting this from the Bellman equations (8)-(9) shows that $W_t(\lambda; \phi) - U_t$ and $V_t(z|\lambda; \phi) - U_t$ must satisfy the same Bellman equations, with three modifications: the wage w_t disappears from (8), the fixed cost $-\phi w_t$ in (9) is replaced by $-(1 + \phi)w_t$, and the surplus values $S_t(\lambda)$ that appear in the tuition schedule (7) and in the Bellman equation (9) are replaced by $S_t(\lambda) - \lambda U_t$. Note that $v_t e^z - (1+\phi)w_t = (1+\phi)[v_t e^{z - \ln(1+\phi)} - w_t]$, and divide the analog of (8)-(9) for $W_t(\lambda; \phi) - U_t$ and $V_t(z|\lambda; \phi) - U_t$ by $1 + \phi$. The resulting Bellman equations then imply

$$\begin{aligned}
W_t(\lambda; \phi) - U_t &= (1 + \phi)\left[W_t(\lambda; 0) - U_t\right], \\
V_t(z|\lambda; \phi) - U_t &= (1 + \phi)\left[V_t(z - \ln(1 + \phi)|\lambda; 0) - U_t\right],
\end{aligned} \tag{10}$$

13

for all $\lambda \in \Lambda$. This says that one can infer the value functions for arbitrary $\phi \geq 0$ from those for $\phi = 0$, simply by shifting these value functions to right by $\ln(1+\phi)$, multiplying the result by $1 + \phi$, and subtracting ϕU_t. In particular, exit thresholds computed for $\phi = 0$ shift to the right by $\ln(1 + \phi)$ for alternative values ϕ of these overhead labor costs.

2.5   Equilibrium and a Generic Indeterminacy

At any point in time t, there are measures $M_t(z|\lambda)$ that describe how the managers of the various types λ are distributed across the productivity states z. The number of type-λ workers is $M(\lambda) - M_t(\infty|\lambda)$. 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 $M_t(z|\lambda)$ 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 expects a perfect foresight equilibrium to be essentially 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. No formal proof will be attempted here. 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. An example of such an assignment will be sketched below.

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 is amenable to calibration. Conjecture that such an economy has a balanced growth path with a growth rate κ, to be determined, that makes the cross-sectional distribution of $z_t - \kappa t$ independent of time. Aggregate consumption and wages then grow at the rate $(1 - \alpha)\kappa$ and the implied interest rate is $r_t = \rho + (1 - \alpha)\kappa$. Write $\theta = \mu - \kappa$ for the drift of $z_t - \kappa t$ and let $\{m(\cdot|\lambda) : \lambda \in \Lambda\}$ be the stationary density of this de-trended state variable, normalized so that $m(z|\lambda)$ integrates to the supply of type-λ managers.

3.1 Value Functions

Given the factor prices $w_t = we^{(1-\alpha)\kappa t}$ and $v_t = ve^{-\alpha t}$, conjecture further that the value functions $W_t(\lambda)$, $V_t(\cdot|\lambda)$, and the tuition schedules $T_t(\cdot)$ are of the form

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

15

Note well that z now represents the state of a manager de-trended by $-\kappa t$. With this conjecture, the Bellman equation (8) for type-λ workers simplifies to

$$W(\lambda) = \max\left\{\frac{w}{\rho}, \frac{w + S(\lambda)}{\rho + \lambda}\right\}. \tag{11}$$

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

$$\begin{aligned}(\rho + \delta)V(z|\lambda) &= ve^z - \phi w + \delta W(\lambda) + \max\left\{S(\lambda) - \lambda V(z|\lambda), T(z)\right\} \\ &\quad + \theta \mathrm{D}V(z|\lambda) + \frac{1}{2}\sigma^2\mathrm{D}^2V(z|\lambda).\end{aligned} \tag{12}$$

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 any thresholds where type-λ managers switch between studying and teaching. As in Lemma 2, the tuition schedule can be taken to be

$$T(z) = \max_{\lambda \in \Lambda}\left\{[\lambda V(z|\lambda) - S(\lambda)]^+\right\}. \tag{13}$$

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 $z \leq b(\lambda)$ and optimality of the exit decision of a type-λ manager requires that $V(z|\lambda)$ is differentiable at $b(\lambda)$. This implies the familiar value-matching and smooth-pasting conditions

$$W(\lambda) = V(b(\lambda)|\lambda), \quad 0 = \mathrm{D}V(b(\lambda)|\lambda) \tag{14}$$

for both $\lambda \in \{\beta, \gamma\}$.

Inserting the tuition schedule (13) into the Bellman equation (12) results in a system of piecewise linear second-order differential equations for $\{V(z|\lambda) : \lambda \in \Lambda\}$ with boundary conditions implied by (11) and (14), 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 conjecture segments of $(b(\lambda), \infty)$ on which the types of managers who match as students and teachers do not change. The second-order differential equation (12)-(13) then remains linear and can be solved explicitly on these segments. The resulting solutions will depend on boundary values and can be smoothly pasted together, and then (14) can be used to find the optimal exit thresholds $b(\lambda)$.

3.1.1 Teaching Thresholds

More can be said about who teaches whom. Consumption cannot be produced without managers, and managers die. So there has to be at least one type of worker who is willing to study in equilibrium. 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 are therefore 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 $z > x(\lambda)$. An immediate consequence of $\gamma > \beta$ and $V(z|\gamma) \geq V(z|\beta)$ is $\gamma V(z|\gamma) - \beta V(z|\beta) > (\gamma - \beta) V(z|\beta)$. 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. More 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, then teachers below y teach only type-β students.

3.1.2 Scaling Properties and an Equilibrium Condition

The value functions and their associated thresholds are functions of the factor prices $[v, w]$ and the surplus values $[S(\beta), S(\gamma)]$. Two scaling properties help characterize this dependence. First, dividing the Bellman equations (11)-(14) by w simply expresses the value functions in units of labor instead of in units of consumption. So $[v, S(\beta), S(\gamma)]/w$ pins down the value functions $[V(z|\beta), V(z|\gamma)]/w$. Second, multiplying v by the constant e^{Δ} only affects the Bellman equation (11)-(14) by changing the managerial flow revenues from producing consumption from ve^z to $ve^{z+\Delta}$. Such a multiplication therefore shifts both $V(z|\beta)$ and $V(z|\gamma)$ to the left by Δ. In particular, multiplying v by e^{Δ} reduces the thresholds $b(\lambda)$, $x(\lambda)$ and y by Δ, but leaves the variable profits $ve^{b(\lambda)}/w$, $ve^{x(\lambda)}/w$ and ve^y/w at these thresholds unchanged.

These scaling properties imply that the Bellman equation maps the surplus values $[S(\beta), S(\gamma)]/w$ into the threshold gaps $y - b(\beta)$, $y - b(\gamma)$, $y - x(\beta)$, and $y - x(\gamma)$. Changing

17

the level of v/w simply shifts the value functions and their thresholds, and so the gaps between the thresholds can only depend on $[S(\beta), S(\gamma)]/w$. When v/w is replaced by ve^y/w using the threshold y implied by v/w, the new threshold above which teachers only teach type-γ students is 0 by construction. In other words, forcing this threshold to be 0 given $[S(\beta), S(\gamma)]/w$ allows one to back out ve^y/w. The Bellman equation therefore implies the equilibrium condition

$$[S(\beta), S(\gamma)]/w \rightarrow [y - b(\beta), y - b(\gamma), y - x(\beta), y - x(\gamma), ve^y/w]. \tag{15}$$

The surplus values $[S(\beta), S(\gamma)]/w$ determine the gaps between the thresholds, as well as the variable profits ve^y/w of the marginal teacher of type-γ students (and hence also the variable profits at all other thresholds.) Of course, given v/w this just pins down the threshold y. But the map (15) does not depend on v/w, and it will turn out to be convenient to express the overall equilibrium conditions in terms of ve^y/w.

The map (15) does depend on the overhead labor parameter ϕ. As already argued in (10), the overhead labor parameter ϕ shifts thresholds computed for zero overhead labor to the right by $\ln(1+\phi)$. So the threshold gaps do not depend on ϕ, and $ve^{b(\lambda)}/w$, $ve^{x(\lambda)}/w$ and ve^y/w are proportional to $1 + \phi$, holding fixed the surplus values $[S(\beta), S(\gamma)]/w$.

3.1.3 Only Two Possible Scenarios

Recall from Lemma 3 that $S(\lambda) - \lambda W(\lambda) \geq 0$. Workers of both types are willing to study. The Bellman equation for workers (11) 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. An implication of the original definition (6) of these surplus values 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 suggests there are three possible scenarios: both types of workers strictly prefer to study, slow learners are indifferent while fast learners strictly prefer to study, and both types are indifferent. But this last scenario is not possible. If both types of workers are indifferent students, then $S(\lambda) = \lambda W(\lambda) = \lambda w/\rho$ for both λ, and this is inconsistent with $V(z|\gamma) > V(z|\beta) > w/\rho$ and the definition (6) of $S(\lambda)$.

On the other hand, if both $S(\beta) - \beta W(\beta)$ and $S(\gamma) - \gamma W(\gamma)$ are strictly positive, then all workers choose to be students. Managers who do not teach also choose to be students. Market clearing then immediately implies that half of the population is a teacher and half a student. Because only managers can teach, this implies that at least

18

half the population must be a manager. Attempting to relate such a scenario to data then requires the use of a very broad notion of who is a manager. The more interesting scenario arises when $S(\gamma) - \gamma W(\gamma) > 0 = S(\beta) - \beta W(\beta)$, so that workers who are fast learners choose to be students, while slow learners may or may not. In this scenario there is no restriction on how small the population of managers can be equilibrium.

3.2 Abundant Type-β Households

Consider, therefore, the scenario in which all type-γ workers and only some type-β workers try to become managers. An example of the type of value functions and of the assignment of students to teachers that arises in this scenario is shown in Figures 1 and 2. Since type-β workers have to be indifferent, the Bellman equation for workers (11) implies

$$S(\beta) = \beta W(\beta), \quad W(\beta) = \frac{w}{\rho}. \tag{16}$$

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 (11) implies

$$W(\gamma) = \frac{w + S(\gamma)}{\rho + \gamma}, \tag{17}$$

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 value of z selected by type-β students. For $z \geq y$, type-γ students determine the tuition schedule (7). Thus y is the point where both types of students are willing to pay the same tuition,

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

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 $z < y$. 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

$$S(\gamma) - \gamma V(x(\gamma)|\gamma) = \beta V(x(\gamma)|\beta) - S(\beta). \tag{19}$$

The Bellman equation (12) 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 (12) at $x(\gamma)$ and y. Forcing the solution to be continuous and differentiable at $x(\gamma)$ and y gives rise to the boundary conditions

$$\lim_{z \uparrow x(\gamma)} \begin{bmatrix} V(z|\gamma) \\ DV(z|\gamma) \end{bmatrix} = \lim_{z \downarrow x(\gamma)} \begin{bmatrix} V(z|\gamma) \\ DV(z|\gamma) \end{bmatrix} \tag{20}$$

$$\lim_{z \uparrow y} \begin{bmatrix} V(z|\lambda) \\ DV(z|\lambda) \end{bmatrix} = \lim_{z \downarrow y} \begin{bmatrix} V(z|\lambda) \\ DV(z|\lambda) \end{bmatrix}, \quad \lambda \in \{\beta, \gamma\} \tag{21}$$

With continuity and differentiability imposed, twice differentiability is implied because the solutions on the open intervals already satisfy (12). 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

$$W(\lambda) = V(b(\lambda)|\lambda), \quad 0 = DV(b(\lambda)|\lambda), \quad \lambda \in \{\beta, \gamma\}. \tag{22}$$

To summarize, the differential equation for $V(z|\beta)$ and $V(z|\gamma)$ is (12)-(13). The values of $S(\beta)$ and $W(\beta)$ are simply given by (16). The surplus value $S(\gamma) > \gamma w/\rho$ immediately pins down $W(\gamma)$ via (17). This then provides the initial values needed for the value-matching conditions in (22). The remaining boundary conditions are (18)-(21), and the second half of (22).

3.2.1   Constructing The Value Functions

The assumed single-crossing property means that the solution for $V(z|\beta)$ is governed by (12) 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 plus a linear combination of the two solutions to the homogeneous equation. One can verify that one of the homogeneous solutions on (y, ∞) explodes relative to e^z and 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. The solution for $V(z|\gamma)$ is governed by (12) 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, ∞).
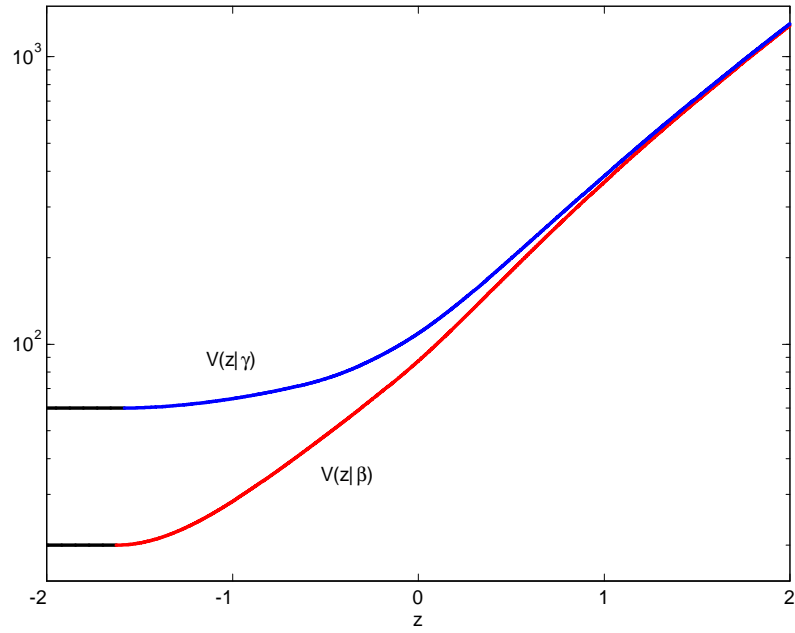
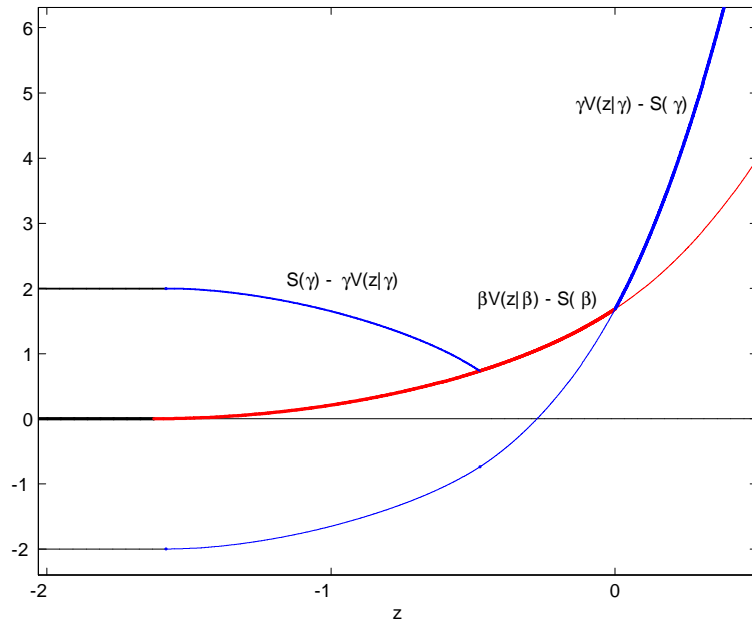FIGURE 1 The value functions $V(z|\lambda)$, $\lambda \in \{\beta, \gamma\}$

FIGURE 2 Assignment when $S(\gamma) - \gamma W(\gamma) > 0 = S(\beta) - \beta W(\beta)$

21

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. To determine these coefficients and thresholds requires twelve boundary conditions. These are provided by (18)-(22).

3.2.2 An Equilibrium Restriction on Thresholds and Relative Prices

The scaling properties of the value function say that $[S(\beta), S(\gamma)]/w$ determines ve^y/w and the thresholds relative to y, as in (15). Here, the surplus value of type-β students is fixed at $S(\beta)/w = \beta/\rho$. So the factor price ve^y/w of the marginal teacher of type-γ students and the threshold gaps $[y - b(\beta), y - b(\gamma), y - x(\gamma)]$ really only vary with $S(\gamma)/w$. Put differently, the Bellman equation implies that there is a one-dimensional curve in the space of threshold gaps $[y - b(\beta), y - b(\gamma), y - x(\gamma)]$ and relative prices $[ve^y, S(\gamma)]/w$ that is consistent with an equilibrium value function and tuition schedule. It is convenient to parameterize this curve by picking values for $y - x(\gamma)$ and backing out $[y - b(\beta), y - b(\gamma)]$ and $[ve^y, S(\gamma)]/w$,

$$ y - x(\gamma) \mapsto [y - b(\beta), y - b(\gamma), ve^y/w, S(\gamma)/w]. \tag{23} $$

The boundary conditions for $V(z|\beta)$ and $V(z|\gamma)$ give rise to twelve equations that are linear in eight undetermined coefficients and two relative prices $[ve^y, S(\gamma)]/w$. So this linear system is overdetermined by two equations. But this system also depends non-linearly on the three threshold gaps $[y - b(\beta), y - b(\gamma), y - x(\gamma)]$. Given $y - x(\gamma)$, the two gaps $y - b(\beta)$ and $y - b(\gamma)$ can be varied to make the linear system consistent. In this way, the problem of finding the curve (23) can be reduced to solving two non-linear equations in $y - b(\beta)$ and $y - b(\gamma)$, given $y - x(\gamma)$. The factor price ve^y/w will turn out to be an increasing function of the gap $y - x(\gamma)$.

3.3 The Kolmogorov Forward Equations

Continue with the scenario in which $S(\beta) = \beta W(\beta)$, so that $x(\beta) = b(\beta)$. Fix thresholds $b(\beta) < y$ and $b(\gamma) < x(\gamma) < y$, with $b(\beta) < x(\gamma)$ because of the conjectured single-crossing property of the tuition flows $\lambda V(z|\lambda) - S(\lambda)$. Then type-$\beta$ 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 density of type-β managers to zero, the Kolmogorov forward equation for the stationary density $m(z|\beta)$ becomes

$$\delta m(z|\beta) = -\theta \mathrm{D}m(z|\beta) + \frac{1}{2}\sigma^2 \mathrm{D}^2 m(z|\beta) + \begin{cases} \beta m(z|\beta), & z \in (b(\beta), x(\gamma)) \\ \beta[m(z|\beta) + m(z|\gamma)], & z \in (x(\gamma), y) \\ 0, & z \in (y, \infty) \end{cases} .$$

(24)

By teaching type-β students, type-β managers in $(b(\beta), y)$ teach 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 $\beta m(z|\gamma)$. Teaching type-γ students produces more type-γ managers, but no additional type-β managers in (y, ∞). Type-β managers exit and become workers again when their productivity state crosses $b(\beta)$ from above. This gives rise to the boundary condition

$$0 = m(b(\beta)|\beta) \tag{25}$$

(see Cox and Miller [1966]). Similarly, the stationary density $m(z|\gamma)$ for type-γ managers has to satisfy

$$\delta m(z|\gamma) = -\theta \mathrm{D}m(z|\gamma) + \frac{1}{2}\sigma^2 \mathrm{D}^2 m(z|\gamma) + \begin{cases} -\gamma m(z|\gamma), & z \in (b(\gamma), x(\gamma)) \\ 0, & z \in (x(\gamma), y) \\ \gamma[m(z|\beta) + m(z|\gamma)], & z \in (y, \infty) \end{cases} .$$

(26)

Type-γ managers in $(b(\gamma), x(\gamma))$ are students, and they transition into productivity states to the right of this interval at the rate γ. Type-γ managers in $(x(\gamma), y)$ are teachers of type-β students. In (y, ∞), all managers are teachers of 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. These teachers can be of either type. Exit at $b(\gamma)$ produces the boundary condition

$$0 = m(b(\gamma)|\gamma). \tag{27}$$

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

$$-\theta m(z|\lambda) + \frac{1}{2}\sigma^2 \mathrm{D}m(z|\lambda) \text{ is continuous at } x(\gamma) \text{ and } y, \text{ for } \lambda \in \{\beta, \gamma\}. \tag{28}$$

23

The equations (25)-(28) 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 $z - y$ only depend on the gaps $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$. The scale of the solution is clearly indeterminate. It is determined elsewhere, by a market clearing condition.

3.4   Constructing Stationary Densities

The differential equation (25)-(28) can be solved for all high enough balanced growth rates κ. The solution is explicit. First, note that the differential equation (24) for $m(z|\beta)$ on $(x(\gamma), y)$ and the differential equation (26) for $m(z|\gamma)$ on (y, ∞) are inhomogeneous equations on these intervals, with inhomogeneous terms $m(z|\gamma)$ and $m(z|\beta)$, respectively. A particular solution to (24) on $(x(\gamma), y)$ is simply $-m(z|\gamma)$, because the term $\beta[m(z|\beta) + m(z|\gamma)]$ equals zero for that solution, and $-m(z|\gamma)$ has to solve the second equation in (26). Similarly, $-m(z|\beta)$ is a particular solution to (26) on (y, ∞). On every interval, the homogeneous parts of (24) and (26) 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 (25) and (27) at $b(\beta)$ and $b(\gamma)$ provide two linear restrictions. The continuity requirements (28) at $x(\gamma)$ and y provide four more linear restrictions. Thus we are left with a six-dimensional linear space of solutions.

Conjecture now that there is a solution for which $m(z|\beta)$ is in fact differentiable on $(b(\beta), \infty)$, and $m(z|\gamma)$ on $(b(\gamma), \infty)$. This adds two differentiability restrictions at both $x(\gamma)$ and y. These restrictions are also linear, and they reduce the space of solutions down to a two-dimensional linear space. One dimension is simply the scale of the solution. To see where the remaining restriction comes from, note that (24) is homogeneous on (y, ∞), with solutions of the form $e^{-\omega_\pm z}$, where $\omega_\pm$ solves the characteristic equation $0 = \theta\omega_\pm + \frac{1}{2}\sigma^2\omega_\pm^2 - \delta$. The roots are $\omega_\pm = -(\theta/\sigma^2) \pm \sqrt{(\theta/\sigma^2)^2 + \delta/(\sigma^2/2)}$. These roots are real and distinct, and a positive δ implies $\omega_+ > 0 > \omega_-$. The negative root would imply that $|m(z|\beta)| \to \infty$, and so the coefficient on $e^{-\omega_- z}$ must be zero. This integrability condition is the remaining linear restriction, suggesting a stationary density that is, given the assumed differentiability, unique up to scale.
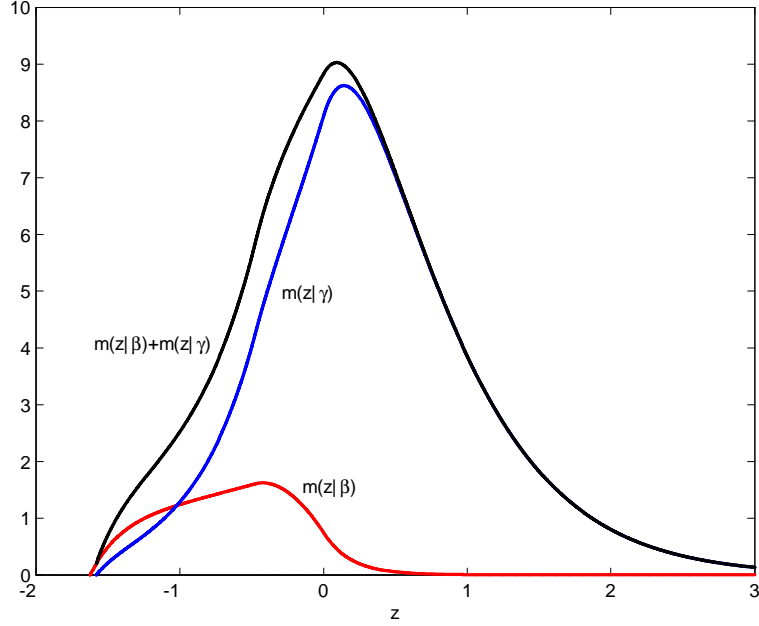
24

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

Ensuring that this solution is well defined requires $m(z|\gamma)$ to be integrable on (y, ∞) as well. The homogeneous equation for (26) on this interval again has solutions of the form $e^{-\zeta_{\pm}z}$, where $\zeta_{\pm}$ now solves the characteristic equation $0 = \theta\zeta_{\pm} + \frac{1}{2}\sigma^2\zeta_{\pm}^2 + \gamma - \delta$. The roots of this quadratic equation are $\zeta_{\pm} = -(\theta/\sigma^2) \pm \sqrt{(\theta/\sigma^2)^2 - (\gamma - \delta)/(\sigma^2/2)}$. Suppose $\gamma > \delta$ so that type-γ students can learn faster than they die. Suppose further that $\kappa > \mu$ so that $\theta < 0$. Under these conditions, if the two roots $\zeta_{\pm}$ are real, then they are both positive. Complex roots result in densities that are not of one sign and must therefore be ruled out. The condition for real roots is $(\theta/\sigma^2)^2 \geq (\gamma - \delta)/(\sigma^2/2)$. Given $\theta = \mu - \kappa < 0$, this is equivalent to

$$\kappa \geq \mu + \sigma\sqrt{2(\gamma - \delta)}. \tag{29}$$

The two roots $\zeta_{\pm}$ are distinct if and only if this inequality is strict. The boundary conditions, integrability conditions, and the assumed differentiability of the stationary density pin down a unique solution when ζ_+ and ζ_- are indeed distinct. But when these two roots merge, the dimension of the linear space of general solutions is reduced by one, suggesting that it might not be possible to find a differentiable solution. In fact, it can be shown that letting κ approach its lower bound (29) from above, and thus let ζ_+ and ζ_- converge to $\zeta = \sqrt{(\gamma - \delta)/(\sigma^2/2)}$, results in a solution that remains well defined and differentiable.

25

Proposition 1 *Suppose $\gamma > \delta > 0$ and fix some balanced growth rate κ that satisfies (29). Also fix some thresholds $b(\beta) < y$ and $b(\gamma) < x(\gamma) < y$. Then the Kolmogorov Forward equations can be solved for a differentiable stationary density $[m(z|\beta), m(z|\gamma)]$ that is unique up to scale. The density for $z - y$ depends only on $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$.*

A stationary density constructed as in Proposition 1 for a given κ and given thresholds implies certain aggregate measures of type-β and type-γ individuals. Market clearing conditions have to be imposed to ensure that these match $M(\beta)$ and $M(\gamma)$. Type-β students are indifferent across teachers in $(b(\beta), y)$ and type-γ students across teachers in (y, ∞). Market clearing therefore only requires that the overall numbers of students and teachers match. This gives rise to the conditions

$$M(\beta) - \int_{b(\beta)}^{\infty} m(z|\beta)\mathrm{d}z \ \geq \ \int_{b(\beta)}^{y} m(z|\beta)\mathrm{d}z + \int_{x(\gamma)}^{y} m(z|\gamma)\mathrm{d}z, \tag{30}$$

$$M(\gamma) - \int_{x(\gamma)}^{\infty} m(z|\gamma)\mathrm{d}z \ = \ \int_{y}^{\infty} [m(z|\beta) + m(z|\gamma)]\mathrm{d}z. \tag{31}$$

In the light of Proposition 1, these market clearing conditions only depend on the gaps $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$. The left-hand sides of (30)-(31) are the numbers of potential type-β and actual type-γ students, respectively. They are all workers and managers who are not teachers. The right-hand side of (30) 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 (31) is the number of type-β and type-γ managers in (y, ∞), all of whom teach type-γ students. These two market clearing conditions add up to the requirement that the number of type-β teachers in $(b(\beta), \infty)$ and the number of type-γ teachers in $(x(\gamma), \infty)$ is no more than half the total population $M(\beta) + M(\gamma)$. The total number of teachers can be less than half of the overall population, and much less if $M(\beta)$ is large. One can think of (31) as determining the scale of $[m(z|\beta), m(z|\gamma)]$, and of (30) 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.[7]

---

[7]In the alternative scenario in which type-β workers strictly prefer to be students, (30) has to hold with equality. Proposition 1 determines $[m(z|\beta), m(z|\gamma)]$ up to scale, and so market clearing in that scenario implies an equilibrium restriction on the possible threshold gaps. But the equality of $S(\beta)$ and $\beta W(\beta)$ is no longer an equilibrium condition.

3.4.1 Exit Rates

The exit rate of type-λ managers is given by

$$\texttt{type-}\lambda \texttt{ exit rate } = \delta + \frac{\frac{1}{2}\sigma^2 \mathrm{D}m(b(\lambda)|\lambda)}{\int_{b(\lambda)}^{\infty} m(z|\lambda)\mathrm{d}z}.$$

The first term is the random exit rate, and the second is the flow of type-λ managers who cross the exit threshold $b(\lambda)$ from above, relative to the total number of type-λ managers. Stationarity requires that the flows of exiting and entering managers balance for each type. By integrating (24) and (26), imposing the boundary conditions (25), (27) and (28), and using the market clearing conditions (30)-(31), one can verify that this is not an extra condition for stationarity but implied by those already given.

3.4.2 Sorting on Payoff-Irrelevant Characteristics

Observe that the forward equations (24) and (26) are silent about who exactly is assigned to learn from whom. For example, type-γ students are indifferent across all teachers in (y, ∞). To determine stationary distributions, one does not need to know which type-γ workers or which type-γ managers in $(b(\gamma), y)$ are assigned to teachers in some particular state $z > y$, just that all teachers in states $z > y$ are matched with one of those students.[8] Given a particular mechanism for assigning students to teachers, one can say more. For example, suppose there is a payoff-irrelevant individual characteristic $\psi \in \Psi$, with Ψ some finite ordered set. Managers and workers are then of types $(\lambda, \psi) \in \Lambda \times \Psi$, and one possible mechanism for assigning, say, type-(γ, ψ) managers in states $z \in (b(\gamma), y)$ to teaching managers in states $z \in (y, \infty)$ is to partition (y, ∞) into a number of segments equal to the number of types in Ψ, and then assign students to these segments ranked by their payoff-irrelevant characteristic ψ. One can then restate the forward equation (24)-(28) and the market clearing conditions (30)-(31) for densities $m(z|\lambda, \psi)$ and so construct stationary densities that depend on not only the ability to learn λ, but also, non-trivially, on an individual characteristic ψ that is unrelated to ability.[9]

---

[8]More generally, without a particular mechanism, it is not possible to describe the transition probabilities for managers indexed by (λ, z).

[9]Plausible theories of statistical discrimination typically imply ex ante welfare consequences for individuals treated differently in equilibrium. See Fang and Moro [2010] for a recent survey. Here, sorting on payoff irrelevant characteristics has no ex ante welfare consequences, and the assumption of perfect insurance markets implies that ex post income differences do not result in ex post differences in consumption.

3.5 Balanced Growth Paths

It remains to ensure that the thresholds used to construct stationary densities are consistent with the Bellman equation. Fix some growth rate κ that satisfies (29), as well as positive threshold gaps $y - b(\beta)$, $y - b(\gamma)$, and $y - x(\gamma)$. Proposition 1 and the market clearing condition (31) pin down the stationary densities of $z - y$ for the two types $\lambda \in \{\beta, \gamma\}$, provided $M(\beta)$ is large enough. The number of managers is then given by

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

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

$$L = M(\beta) + M(\gamma) - (1 + \phi)N.$$

The aggregate supply of managerial human capital H is determined by

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

Observe that N, L and He^{-y} only depend on the stationary densities for $z - y$, and not on the actual location of the threshold y. Of course, this means that H scales with e^y. Given these factor supplies, marginal products determine factor prices. Specifically, $[vH, wL] = [1 - \alpha, \alpha]H^{1-\alpha}L^{\alpha}$, which scales with $e^{(1-\alpha)y}$. In particular

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

Thus, given stationarity and a conjecture about the growth rate κ, there is a mapping

$$[y - b(\beta), y - b(\gamma), y - x(\gamma)] \mapsto \frac{ve^y}{w} \tag{32}$$

that is determined by market clearing conditions in the market for students and teachers, and equilibrium conditions in the consumption sector. This mapping is constructed for thresholds that have to satisfy $\min\{y - b(\beta), y - b(\gamma)\} > y - x(\gamma) > 0$, but the construction of (32) does not require that these thresholds are consistent with the Bellman equation.
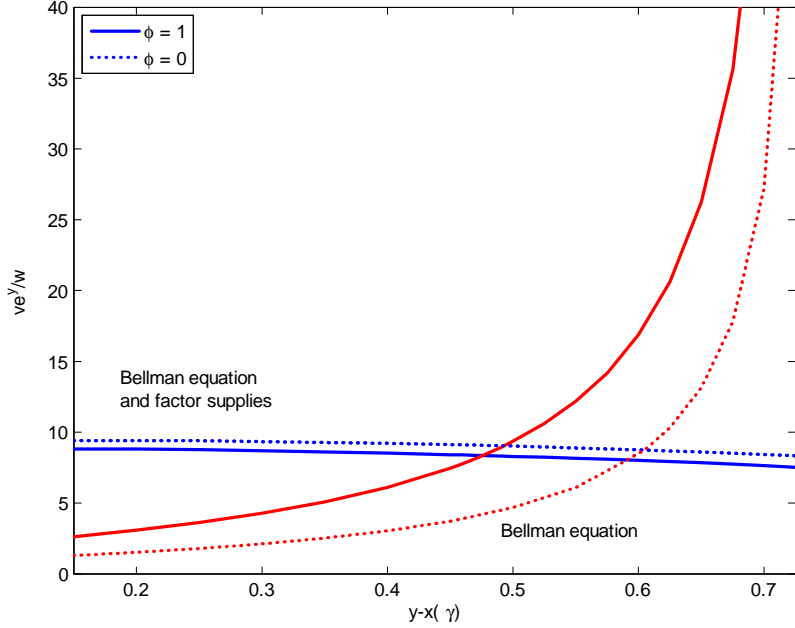
FIGURE 4 The Equilibrium Conditions (23) and $\{(23),(32)\}$

Now recall from (23) that the Bellman equation determines a mapping from $y - x(\gamma)$ to not only the other two threshold gaps $y - b(\beta)$ and $y - b(\gamma)$, but also to the relative price ve^y/w. This mapping only relies on equilibrium restrictions on the shape of the tuition schedule, and on optimality conditions for the exit decision faced by managers. So there are two distinct ways to compute ve^y/w as a function of $y - x(\gamma)$: directly via the Bellman equation, or via the threshold gaps implied by the Bellman equation, the resulting stationary distributions of $z - y$, and factor market clearing in the consumption sector. Figure 4 shows the two equilibrium conditions. Along a balanced growth path, $y - x(\gamma)$ has to be such that these two calculations match. Given the equilibrium value of $y - x(\gamma)$, the threshold y follows from lining up the stationary distribution with the historical distribution at the initial date.

3.6 Determining the Balanced Growth Rate

This construction works for any balanced growth rate κ that satisfies the lower bound (29), provided $M(\beta)$ is large enough, and provided that the mean of e^{z-y} is finite. If $M(\beta)$ is too small, then the balanced growth path will instead be one in which (30) holds with equality. Type-β students are then scarce, and the gains from studying $S(\beta) - \beta W(\beta)$ for such students will be positive, as they are for type-γ students. As already noted, this can only happen if at least half the population is a manager. The

29

requirement that the mean of e^{z-y} is finite implies an additional lower bound on κ, one that is not necessarily implied by (29).

So what then determines the long-run growth rate κ? A rigorous answer to this question can be given in a simpler economy in which there is only one type of manager, say of type γ, and managers never exit, say because $\phi = 0$ and managers cannot supply labor. The resulting Kolmogorov forward equation is $D_t m(t,z) = -\theta D_z m(t,z) + \frac{1}{2}\sigma^2 D_{zz} m(t,z) - (\delta+\gamma)m(t,z)$ below the median of $m(t,z)$ and $D_t m(t,z) = -\theta D_z m(t,z) + \frac{1}{2}\sigma^2 D_{zz} m(t,z) - (\delta-\gamma)m(t,z)$ above the median. For this equation, Kolmogorov, Petrovskii and Piscounov [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). A similar multiplicity of stationary distributions and balanced growth rates arose in Luttmer [2007]. There it was resolved in the same way, by assuming that the initial distribution of productivities has a bounded support, as it does in any finite economy. It is essential for this result that σ is positive. It is not difficult to show that at $\sigma = 0$ there is a continuum of stationary distributions and growth rates, not only for exponentially de-trended managerial productivities, but also for linearly de-trended productivities. But a bounded initial distribution of productivities in such a world would lead to permanent stagnation (relative to θ) because nobody can be more productive than the most productive individual in the initial population.[10]

Applying this idea here results in the prediction that this economy grows, in the long run, at a rate $\kappa = \mu + \sigma\sqrt{2(\gamma - \delta)}$. 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.) What matters is the learning speed 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. Recall that at the lower bound (29), the right tail of $m(z|\gamma)$ behaves like $e^{-\zeta z}$, with $\zeta = \sqrt{(\gamma - \delta)/(\sigma^2/2)}$. The right tail declines slowly for low values of ζ. So rapid growth goes together with a

---

[10]Fisher [1937] formulated the same equation to describe the geographic spread of an advantageous gene. Alternative proofs and generalizations can be found in McKean [1975], Bramson [1982], and a large literature on reaction-diffusion equations. 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.

particularly thick-tailed distribution of managerial productivities if it is due to a high σ, and with a thinner tail if due to a high $\gamma - \delta$.

3.6.1   Time-Varying Noise and Learning Parameters

Beyond the precise formulation of this model, a comparison across balanced growth paths and the basic workings of the model predict that periods of high σ are periods of rising inequality, and periods of high γ are periods of declining inequality. If σ and γ are not fixed parameters but determined in part by effort allocation choices, then such a negative co-movement between σ and γ is to be expected. Because the aggregate growth rate is increasing in both parameters, the effects of such negative co-movement in σ and γ on overall growth may be limited. But the rising inequality during the late 19th and early and late 20th centuries, and the declining inequality during the mid 20th century in parts of the Western world could be the result of such variation in σ and γ.

4.   Quantitative Implications

As a benchmark specification, consider an economy with a subjective discount rate $\rho = 0.05$, a random death rate $\delta = 0.04$, slow and fast learning rates $[\beta, \gamma] = [0.025, 0.075]$, and an individual discovery process with a standard deviation $\sigma = 0.125$. Slow learners are more likely to die before they learn something useful, while the odds are that fast learners succeed before they die.[11] As managers, the productivities of both types, conditional on no learning, are almost as volatile as the aggregate stock market in the US. Assume that managers have to hire one worker as overhead labor, so that $\phi = 1$. Take the labor share parameter of the Cobb-Douglas technology to be $\alpha = 0.60$, which is somewhat below conventional values of this statistic for the US economy. The actual labor share in this economy will in fact exceed α, because labor is also required to cover fixed costs.

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

$$\kappa - \mu = \sigma\sqrt{2(\gamma - \delta)} = 0.125 \times \sqrt{2(0.075 - 0.04)} \approx 0.033.$$

If managers cannot improve, on average, their own log productivities, then $\mu = 0$,

---

[11] The death rate δ can also be interpreted as the rate at which the project a manager works on disappears randomly, making the knowledge acquired by the manager obsolete.

resulting in a long-run growth rate of managerial human capital of a little over 3% per annum. The implied consumption growth rate is about 1.3% and log utility then implies an interest rate of 6.3%. Both rates will be slightly lower in an alternative benchmark specification in which the drift of productivity, $\mu + \sigma^2/2$, equals zero. The implied tail index of the right tail of the distribution of managerial productivities is also immediate

$$\zeta = \frac{\kappa - \mu}{\sigma^2} = \sqrt{\frac{\gamma - \delta}{\sigma^2/2}} = \sqrt{\frac{0.075 - 0.040}{(0.125)^2/2}} \approx 2.1.$$

If managers are CEOs, then the employment size distribution of firms will inherit this tail index, because of the Cobb-Douglas technology. In US data, the firm size distribution has a tail index of about 1.05 (see, for example, Luttmer, [2007]). So this benchmark economy produces a firm size distribution with a much thinner tail than observed in the data.

A lower fast learning rate γ and a higher death rate δ, combined with a higher variance σ^2 of the idiosyncratic productivity shocks, would result in a thicker tail, without necessarily changing the growth rate κ. But CEO wealth dominates the right tail of the US wealth distribution, and that distribution also has a much thinner tail than the firm size distribution.[12] In an incomplete markets version of the model, firm size and entrepreneurial or managerial wealth are likely to be related. The thicker tail of the US firm size distribution could be driven by a combination of heterogeneity in managerial productivities and a within-firm replication mechanism, as in Luttmer [2011]. Alternatively, a more elaborate model of the CEO span of control could be constructed by combining features of the Beckmann [1958] model of city hierarchies and the Garicano [2000] and Garicano and Rossi-Hansberg [2006] models of hierarchies in the organization of production. Replication and hierarchies are plausible interpretations of the very thick tail of the US firm size distribution, and the above parameter choices for the process of managerial knowledge accumulation leave room for such interpretations.

Figures 1 through 4 were computed for a version of this economy in which 10% of the population is a type-γ fast learner. As can be seen in Figures 3 and 4, the exit thresholds used by the two types of managers are very similar. In terms of variable profits, these managers earn $ve^{b(\beta)}/w \approx 1.65$ and $ve^{b(\gamma)}/w \approx 1.72$, respectively. This is more than

---

[12]See Castañeda, Díaz-Giménez, and Ríos-Rull [2003] and Piketty and Saez [2003] for evidence on US income and wealth distributions, and Nirei and Souma [2007] and Benhabib, Bisin and Zhou [2011] for recent models that generate thick-tailed income and wealth distributions. Cagetti and DeNardi [2006, 2008] and Jones and Kim [2014] emphasize the prevalence of business owners and entrepreneurs among wealthy households.

the unit wage earned by workers, but managers also have to cover the fixed cost $\phi = 1$. After fixed costs, their earnings from producing consumption at the exit threshold are negative, as expected, because 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 significant net gains equal to $S(\gamma) - \gamma W(\gamma) \approx 2.00$ 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 $1/\rho = 1/.05 = 20$ in this economy. The ability to learn fast results in a value of $W(\gamma)/w = (1 + [S(\gamma) - \gamma W(\gamma)]/w)/\rho \approx 3.00/.05 = 60.0$ for fast learners. This reflects expected net gains from trying to become a manager. As already emphasized, learning from others is an investment with uncertain returns.

About 7.4% of the population are managers along the balanced growth path. It is clear from the stationary densities displayed Figure 3 that most managers are type-γ managers, and virtually all managers above the mode of the distribution are fast learners. Since the population of type-β individuals is so much larger than the population of type-γ individuals, it has to be the case that only a small fraction of type-β workers choose to study in this equilibrium. Those who succeed in becoming managers do not continue to study but teach what they have learnt. This severely reduces upward mobility of type-β managers.

4.1 Varying Overhead Labor and the Ability Distribution

Figure 4 also shows how the equilibrium conditions change when the overhead labor costs are reduced to zero. There is still an opportunity cost to being a manager, in the form of foregone wages from supplying labor, but no additional overhead. As argued earlier, the Bellman equation implies that the variable profits ve^y/w are proportional to $1 + \phi$, given $y - x(\gamma)$. So the equilibrium condition (23) shifts down with a reduction in ϕ. None of the threshold gaps implied by the Bellman equation given $y - x(\gamma)$ depend on ϕ, and so neither do the stationary densities of $z - y$. A reduction in ϕ therefore leaves N and He^{-y} unaffected, given $y - x(\gamma)$. But it increases the supply of variable labor L because the number of managers does not change and less of the labor supplied by those who do not manage is soaked up by overhead. This increase in variable labor implies an increase in the factor price ve^y/w of managers at the threshold y, and hence an upward shift in (32). This explains how the equilibrium conditions in Figure 4 shift, and it immediately follows that $y - x(\gamma)$ increases with a reduction in ϕ. The net effect of a reduction in overhead labor is to increase ve^y/w and the value of a fast learner,

33

from $W(\gamma) \approx 60$ to $W(\gamma) \approx 70$. Doubling overhead labor instead, to $\phi = 2$, would result in $W(\gamma) \approx 50$.
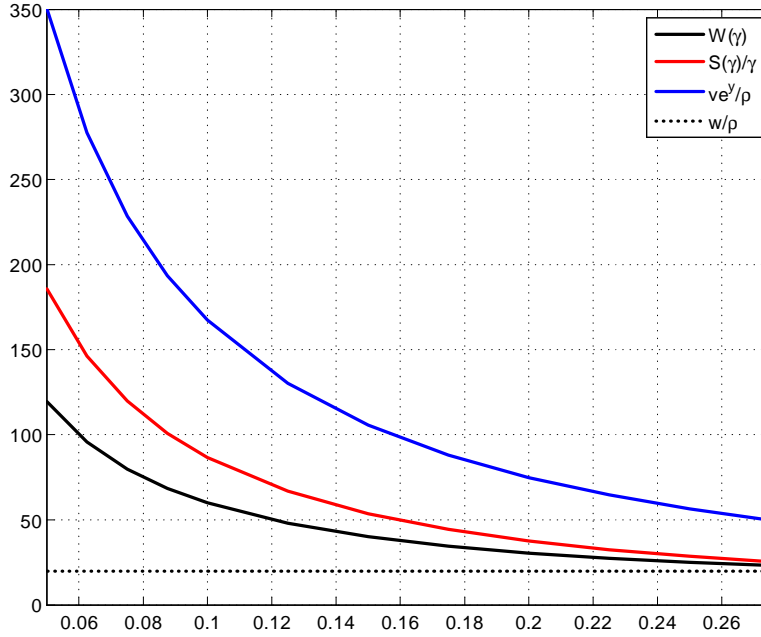

FIGURE 5 Ability Rents as a Function of $M(\gamma)/[M(\beta) + M(\gamma)]$.

Figure 5 shows how the ability rents of type-γ agents in this economy vary with their supply, holding fixed all other parameters. These ability rents are quite sensitive to the supply of fast learners: cutting the supply in half from 10% roughly doubles the value $W(\gamma)$ of type-γ workers, and raising it to 25% drives $W(\gamma)$ down to almost $W(\beta)$. One can imagine that some of the heterogeneity in learning speeds can be attributed not to innate ability but to certain types of general education. The historical expansion of formal education observed almost everywhere would then imply a reduction in $W(\gamma)$ relative to $W(\beta)$. But the growth rate of the economy and the thickness of the right tail of the income distribution are not affected by such changes in the relative supplies of type-γ and type-β individuals.

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, while much of the literature is about imitation. Teaching actively involves students and teachers, while imitation is a more individual activity. Both phenomena play a role in real-world knowledge

34

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.

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.

REFERENCES

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

[2] Alvarez, F.E., F.J. Buera, and R.E. Lucas, Jr., "Idea Flows, Economic Growth, and Trade," NBER working paper no. 19667 (2013).

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

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

[5] Beckmann, M.J., "City Hierarchies and the Distribution of City Size," *Economic Development and Cultural Change*, vol. 6, no. 3 (1958), 243-248.

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

[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 Innovations," *Journal of Monetary Economics*, vol. 55 (2008), 435-453.

[9] Burstein, A. 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.

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

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

[12] Cagetti, M. and M. De Nardi, "Wealth Inequality: Data and Models," *Macroeconomic Dynamics*, vol. 12, (2008), 285-313.

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

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

[15] 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.

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

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

[18] Fang, H. and A. Moro, "Theories of Statistical Discrimination and Affirmative Action: A Survey," *Handbook of Social Economics, Volume 1A,* J. Benhabib, M.O. Jackson, and A. Bisin, editors, Elsevier.

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

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

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

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

[23] Garicano, L. and E. Rossi-Hansberg, "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, vol. 121, no. 4 (2006), 1383-1435.

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

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

[26] Jovanovic, B. and R. Rob, "The Growth and Diffusion of Knowledge," *Review of Economic Studies*, vol. 56, no. 4 (1983), 569-582.

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

[28] Keller, W. and S. Yeaple , "Gravity of Knowledge," *American Economic Review* (2013)

[29] 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).

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

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

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

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

[34] 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.

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

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

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

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

[39] 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.

[40] Nocke, V. and S. Yeaple, "An Assignment Theory of Foreign Direct Investment," *Review of Economic Studies*, vol. 75 (2008), 529-557.

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

[42] Piketty, T. and E. Saez, "Income Inequality in the United States, 1993-1998," *Quarterly Journal of Economics*, vol. 118, 1-39.

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

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