INTERNATIONAL ECONOMIC REVIEW
Vol. 55, No. 3, August 2014

DIVERSE ORGANIZATIONS AND THE COMPETITION FOR TALENT*

BY JAN EECKHOUT AND ROBERTO PINHEIRO[1]

University College London, U.K., and Barcelona,GSE-UPF, Spain; University of Colorado Leeds School, U.S.A.

We propose a theory of firm production that requires diverse inputs. We show that in a competitive labor market, firms differ in their skill composition. Organizations with higher total factor productivity (TFP) are larger and hire from a broader range of skills. Technological progress leads to an increase of all wages and results in downsizing. Quantifying productivity using our model shows that a constant elasticity of substitution (CES) production function generates unbiased estimates of TFP but biased estimates of marginal product and elasticity of substitution across skills. Our model also generates estimates of the TFP distribution based on CEO compensation alone.

1. INTRODUCTION

Organizations are typically composed of a wide variety of skilled agents. Individuals have different levels of training and experience, and there is collaboration between experts, operators, and staff. A basic premise of this article[2] is that diversity of talents within an organization is beneficial for solving problems and optimizing processes, ultimately increasing productivity.[3] The role of diversity in problem solving is important because most workers are involved in activities in which individuals collectively solve an array of problems. These activities range from the very simple, like maintenance, to the extremely complex, such as developing new pharmaceuticals.

We analyze how competition for talent between firms shapes the firms' organization. The building block of our model is a production technology designed to incorporate within-firm diversity while allowing for different skill compositions across firms. To justify our choice of a production technology, we present a microfoundation based on a problem-solving setup similar to that in the knowledge hierarchies literature. In this sense, our model is related to previous work by Garicano (2000) and Garicano and Rossi Hansberg (2006), among others. However, differently from the previous literature, we focus on the macroeconomic implications of the skill composition.

Our objective is to analyze worker and firm productivities and how they are affected by the equilibrium composition of skills within and between firms. This is clearly a very complex problem to solve, and in this article we take a first step by proposing a simple theory of diverse organizations. Firms are viewed as a flexible distribution of heterogeneous workers. We show that in general these distributions are different across firms, depending on the level of total factor productivity (TFP), which represents differences in technology, scarce inputs, and other firm-specific characteristics. These differences in distributions across firms will be instrumental

*Manuscript received November 2010; revised June 2013.

[1] We would like to thank numerous colleagues and seminar participants for insightful discussions and comments. We thank Elizabeth Mellon for excellent research assistance. Eeckhout gratefully acknowledges support from the ERC, Grant 208068 and from Science and Innovation grant ECO2009-12157. Please address correspondence to: Jan Eeckhout, University College London, 30 Gordon Street, WC1H 0AX London, U.K. E-mail: *j.eeckhout@ucl.ac.uk*.

[2] For evidence, see Page (2007).

[3] Companies like Southwest Airlines for example explicitly encourage workers at all levels to get involved in streamlining the production process. By using the viewpoints and experience of the baggage loader as well as the logistics manager, they consistently manage to be among the airlines with the lowest labor cost per miles flown.

for the impact of macroeconomic shocks—for example, technological progress and changes in educational costs—on labor market variables, like the distribution of wages across skills. We are able to address the increase in wage and firm productivity dispersion that occurred in the last 20 years. In particular, our results show how this increased dispersion may relate to the technological progress generated by the introduction of information and communication technologies (ICT) as well as the increase in educational costs. Moreover, we can also show that assuming a restrictive specification of the firm's production function will not only generate skill distributions that are identical across firms, but it will also bias the estimation of workers' marginal product by reducing the variability of marginal product across skills.

A firm's demand for a diverse organization is endogenously driven by the problem-solving technology that embodies both the uncertainty of solving any given problem as well as the differences in problems faced across divisions. We model the problem-solving process in a given sector by a heterogeneous Poisson process in which the solution's arrival rate decreases the more same-skill workers have already tried to solve the problem. Therefore, even though the expected contribution to the value of the problem at hand is increasing with the number of same skilled agents working on it, the marginal contribution is decreasing. This naturally generates a complementarity with other skill levels, since even lower-skilled workers will eventually have a higher marginal productivity than overrepresented high-skill employees. The firm-level technology aggregates the contribution of the different skill categories.

Starting from this simple premise, we build a general framework for studying the demand for skills by firms and their internal organizations while competing in a competitive labor market. The presence of a competitive labor market determines the cost of skills economy-wide and adds a second layer in the trade-off among skills: Although the marginal benefit of hiring an additional worker from a given skill depends on the skill distribution *inside* the firm, the marginal cost is determined *outside* the firm in the labor market. Equilibrium wages depend on the economy-wide skill distribution as well as on the distribution of firm-specific TFP across firms. The interaction between these marginal benefits and costs will ultimately shape the firm's distribution of skills.

Within this framework, we show that if and only if the elasticity of substitution between different skills is constant (CES) will the distribution of skills within different firms be identical. Higher TFP firms are larger, but they have exactly the same skill composition as all other firms. For any non-CES technology, organizations will be diverse. We are unaware of such a result to date. This result is important for two reasons: (i) The CES assumption seems unrealistic in the current context because it requires that the arrival rate of solutions to problems is infinite when any worker embarks on a new problem and (ii) the properties of the CES firm production technology are nonrobust and based on a knife-edge result. This result is important because the CES technology is the workhorse model used to estimate heterogeneity in macroeconomics and empirical labor economics.[4] Since our result establishes that CES implies identical skill distributions across firms with different TFPs, it is not surprising that in assuming a CES specification one finds a small impact of "human capital" variables. In order to show how the technology specification can bias the results on the impact of skill heterogeneity, we provide a simulation exercise. We generate multiple samples of simulated data based on our model and analyze the data through econometric methods similar to the ones used in the labor and IO literature. Our results show that the estimates of marginal productivity and elasticity of substitution across firms show much lower variance across skills and underestimate the marginal product of skills. In this sense, our results represent a cautionary tale to the findings in the literature.

We also show that when there is diversity between firms, the skill distribution of larger firms stochastically dominates the distribution of smaller firms (illustrated in Figure 1). Larger firms hire over a wider range of skills, and as a result they have more levels in their hierarchy.

[4] Examples, using CES and Dixit–Stiglitz style technologies and preferences include Bagger et al. (2010), Fox and Smeets (2011), and Irarrazabal et al. (2009). In a recent paper, Eeckhout, Pinheiro, and Schmidheiny (2014) show that across different cities there is evidence that the production technology is not CES.

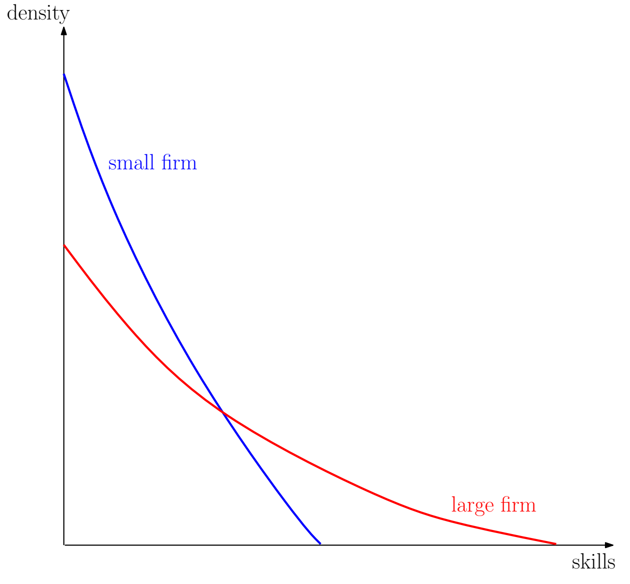

FIGURE 1

STOCHASTIC DOMINANCE OF SKILL DISTRIBUTION IN LARGE FIRMS

The immediate implication is that their highest skilled member—we label her the CEO—is more skilled in larger firms. It also follows that the top of the firm is "leaner," that is, there are proportionately fewer workers at each level. Smaller, low-productivity firms have a larger base of low-skilled workers. Another implication of first-order stochastic dominance in the skill distribution is that there is also stochastic dominance in the wage distribution. Larger firms hire on average more skilled workers and therefore pay on average higher wages. Similar characteristics were found by Caliendo et al. (2012) using data from French manufacturing firms, by Tag (2013) using Swedish data, and by Colombo and Delmastro (1999) using Italian data.

We use our model to analyze how organizations evolve in a changing environment. In recent years, organizations have gone through fundamental changes, not least because the competitive environment in which they operate is changing. We consider the effect of changes in the distribution of TFP on the equilibrium distribution of firms, on the demand for skills, and on wages. We show that due to a stochastic dominance change in the distribution of TFP, more high-TFP firms compete for skills. As a result, the demand for skills increases, thus driving up wages. In equilibrium, the quantity demanded of each skill type is therefore lower. In addition, given there are more high-TFP firms, the skill level of the CEO of a firm that has not changed its TFP will decrease due to increased competition.

We also consider the impact of technological progress on the wage and firm productivity distributions. We show that technological progress that increases skills' productivity increases wages at all skill levels. In particular, if technological progress is skill neutral, more technological progress will imply more wage dispersion, since wages at higher skill levels increase more than the ones at lower skill levels. Finally, if we consider the technological coefficient a component of the TFP, technological progress will imply not only an increase in wage dispersion, but also an increase in dispersion of the TFP distribution. This pattern is empirically corroborated by Faggio et al. (2010), which studied a U.K. firm-level panel data set covering manufacturing and nonmanufacturing firms for the period from 1984 to 2001. That study's evidence not only indicates an increase in TFP, labor productivity, and wage dispersion during the period, but also suggests that the leading cause for the increase in dispersion is the introduction of new technologies. Our model is consistent with those patterns.

We extend the model in several directions. In particular, we consider the case as in Lucas's (1978) span of control model, in which a minimum scale of output is needed. We also analyze the impact of *investment in skills* by ex ante identical agents and show that, in equilibrium, there will be an endogenous distribution of skills. Even with little or no ex ante heterogeneity, there can be considerable ex post inequality as this technology enhances heterogeneity. In equilibrium, if there is scarcity of any one particular input (for example, for specialized welders), the returns to obtaining that skill are high. With increasing investment costs, the ensuing distribution of skills is decreasing in type as the returns in term of wages must be increasing to compensate for higher investments costs. Wages can only be increasing if there is sufficient scarcity in that particular input.

Our work relates to the original literature on matching and assignment problems (see Becker, 1973; Rosen, 1974; Sattinger, 1993; Costrell and Loury, 2004). These papers analyze in different ways the allocation of heterogeneous workers to jobs of different productivities. They all have in common that a firm consists of exactly one job.[5] We relax this one-firm-is-one-job assumption. The distinguishing feature of our analysis is that there is a nondegenerate distribution of jobs *within* the firm. Although in the existing models higher productivity firms hire more productive workers, here we can make a statement about the comparison of distributions of skilled workers across different firms. Therefore, in our model, there is heterogeneity within the firm not only in the sense that workers of different skills collaborate within the firm, but also that firms are free to choose the number of workers at each skill level.

As mentioned above, we build on the literature on knowledge hierarchies. The most general version of this model, presented by Garicano and Rossi-Hansberg (2006), focuses on a problem solving technology in which more skilled workers can solve not only the easier, more routine problems solved by the less skilled workers, but also harder, less common problems. In this sense, all problems are similar, distinguished only by the level of difficulty (easier vs. harder engineering issues, for example). As a result, there is always complementarity among different skill levels and no decreasing returns to scale within a given skill level. The knowledge hierarchies model is well suited to explain the production in industries characterized by one main activity, like manufacturing and consulting. That model is empirically tested by Caliendo et al. (2012).

In contrast to the strict complementarity presented in the basic problem-solving/knowledge hierarchy models, our setup allows for the discussion of changes in the level of complementarity and substitutability across different skill levels. Our model can also generate equilibria in which more productive firms will employ not only highly skilled workers not hired by lower productivity firms, but also lower skilled workers. Therefore, our model is well suited to analyze multidivision firms where tasks in different divisions require different abilities and where there is limited transferability of skills and problems across divisions.[6]

2. THE MODEL

2.1. *Population.* Consider a population consisting of risk-neutral agents endowed with different types and/or levels of skills. There are N different types of skills $i \in \{1, \dots, N\}$. Denote by m_i the measure of agents of skill type i, with $\sum_i m_i = 1$. For ease of exposition, we start out with a discrete number of skills and then take the limit to work with a continuum. There is a measure of entrepreneurs, each of whom is atomless, who own the property rights to a production process $A \in \mathcal{A} \subset \mathbb{R}_+$. This can be interpreted as firm-specific TFP. Let $\mu(A)$ denote the measure of each type A.

[5] Rosen (1982) considers a set of jobs within a one-firm economy and analyzes the competitive equilibrium allocation of heterogeneous workers to each of the jobs in this hierarchical firm. This one-firm economy is formally identical to the standard matching problem where the firm is interpreted as a competitive economy. There is competition within the firm between jobs. The only difference with the standard matching model is that in Rosen's setup, he allows for externalities.

[6] See also Garicano and Hubbard (2005), who focus on independent services that can be sold directly to clients. We focus on the case in which the final product is composed of the tasks undertaken by different divisions (say production versus marketing or distribution) that require the solution to problems that are different in nature.

2.2. *The Technology.* Output y produced by a firm with technology A is equal to $y = AT^{\beta}$, where T is the total value of all tasks and $\beta > 0$ indicates the returns to scale. The total value of the tasks is made up of the contributions by different skill levels, indexed by x_i, There is a weak order[7] on x_i, increasing in i. Workers have the ability to solve one of these tasks corresponding to their type i.[8] The number of tasks solved exhibits decreasing returns in the number of workers of the same ability and is given by $h(n_i)$. The value of these tasks then is $h(n_i)x_i$. The total value of the tasks solved is the sum of the individual tasks, so that we can express the total output as:

$$y = A\left[\sum_{i=1}^{N} h(n_i)x_i\right]^{\beta}. \tag{1}$$

Although this is an arbitrary label, it is intuitive to label the highest skilled agent in a firm as the CEO $i_{\mathrm{CEO}} = \arg\max_i\{x_i|n(x_i) > 0\}$. If the order is weak, different types i have different skills that are equally valuable, for example, a financial and a logistics manager.

Intuitively, the continuous version of the model is very transparent, but the derivation of the equilibrium conditions is much harder. We therefore derive the equilibrium conditions in the context of the discrete model and then take the limit where the grid of types shrinks. The continuous types version then can be represented by the technology

$$y = A\left[\int h(n_i)x_i dF_A(i)\right]^{\beta},$$

where $F_A(i)$ denotes the distribution of skills in firm A. We derive this formally in the Appendix.

2.3. *Microfoundation for $h(n_i)x_i$.* The microfoundation of the value of tasks of type i stems from a problem-solving procedure. We think of these workers as continuously deciding whether to undertake effort to attempt to solve a new problem. Any new problem is randomly circulated among all n_i workers of the same type, and a solution is found probabilistically. If no solution is found by any type i worker, the problem is eventually abandoned. Effort is a choice in $\{0, 1\}$ at a cost q: for $e = 0$, the solution is found with probability 0 at a cost equal to zero; for $e = 1$ she finds the solution with probability $\lambda(n_i)$, where n_i is the measure of same skill i workers who have tried this problem previously. Because workers have common skill sets, an attempt after other agents have tried and failed to find the solution has a lower chance of success. We therefore represent the arrival of a solution probability by a nonhomogeneous Poisson process with arrival rate $\lambda(n_i)$,[9] where $\lambda' < 0$. As mentioned before, the expected number of problems solved is denoted by $h(n_i)$, which is a function of $\lambda(n_i)$ as we derive below when we discuss the workers' incentive problem. Finally, the expected economic value in terms of the numeraire is $h(n_i)x_i$. Observe therefore that the problem-solving should not necessarily be interpreted as a hierarchical process. The organizational form can be equally well understood as a polyarchy of different-but-equal problem solving units (see Sah and Stiglitz, 1986).

We now consider the worker and firm strategies. We first consider the incentive compatibility (IC) and individual rationality (IR) constraints for a given worker of skill level i. Since workers are risk-neutral, there is no reason for risk-sharing. Then, with a disutility of effort q, the IC

[7] When there is equality in skills x_i, we can interpret the skills as being horizontally differentiated.

[8] When we assume a strict order across skills, the model can be extended to the case where type i is able to perform not only his/her own task but also the tasks of any type $j < i$. However, studying a unique data set that contains information about worker and job characteristics, Gautier et al. (2002) showed that more skilled workers—measured by education—were not more productive than less skilled workers executing simpler tasks, which would imply that workers would naturally select themselves at the highest task level they are able to execute.

[9] A formal definition of a nonhomogeneous Poisson process is provided in the Appendix.

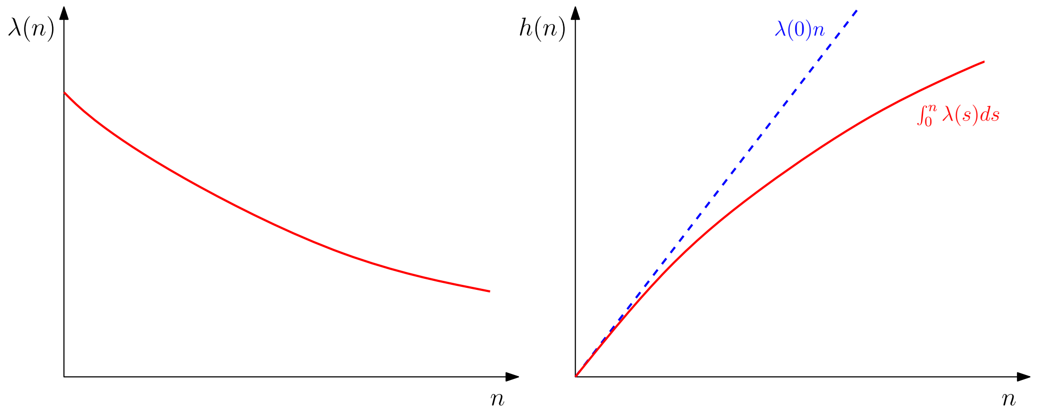

FIGURE 2

(A) THE NONHOMOGENEOUS POISSON ARRIVAL RATE. (B) THE EXPECTED NUMBER OF PROBLEMS SOLVED

constraint for a worker who has skill i and a measure n_i of workers who tried the problem before him/her is given by:

$$\lambda(n_i)\omega(n_i, x_i) + (1 - \lambda(n_i))0 - q \geq 0,$$

where $\omega(n_i, x_i)$ is the worker's payment if the problem is solved. Then for the IR constraint we have:

$$\lambda(n_i)\omega(n_i, x_i) + (1 - \lambda(n_i))0 - q \geq w(x_i),$$

where $w(x_i)$ is the expected wage the worker would obtain from working in a different company, net of effort costs. This outside option will be determined by the labor market equilibrium. We think of workers solving many problems in random order, and the expected return is constant for any individual problem. The hiring decision is therefore based on the expected wage $w(x_i)$ instead of being conditioned on the realization of a problem. In order to maximize profits, the firm pays a wage such that the IR constraint is satisfied with equality, that is,

$$\omega(n_i, x_i) = \frac{w(x_i) + q}{\lambda(n_i)}.$$

Observe that ω is increasing with both x_i and n_i. Therefore, since $\lambda' < 0$, the further in the line of workers to solve a problem, the higher is the compensation per problem solved, even though the expected compensation is the same for all type i workers.

The firm is concerned with the expected number of problems solved, $h(n_i)x_i$, where n_i is the number of workers of skill level i that are hired and provide effort. Since a firm will only hire workers as long as the IC and IR constraints are satisfied, in equilibrium all workers choose effort $e = 1$. As a result, the total number of problems solved can be written as:

$$h(n_i) = \int_0^{n_i} \lambda(s)ds.$$

Adding more agents increases the number of problems solved, that is, $h'(n) > 0$, but it does so at a decreasing rate, that is, $h''(n) < 0$ (since $h' = \lambda$ and $h'' = \lambda'$). Observe that if the Poisson process is homogeneous and λ is a constant, then $h(n) = \lambda n$ is linear (illustrated in Figure 2).

The microfoundation provides compelling reasons to impose restrictions on the admissible functional forms of $h(n)$. In particular:

ASSUMPTION 1.

$$\overline{h}' = \lim_{n \to 0} h'(n) = \lambda(0) < \infty.$$

This assumption captures the idea that under no circumstance do agents have an infinite problem-solving ability and that as a result there must always be a residual amount of unresolved problems. The nonhomogeneous Poisson process cannot have an infinite arrival rate as the number of workers in a skill category becomes small. Notice that this assumption rules out some common production functions, such as the power function. To see this, consider, for example, $h(n) = bn^{\gamma}$. In order to obtain this h function with our nonhomogeneous Poisson process, the arrival rate $\lambda(n)$ must be given by $b\gamma n^{\gamma-1}$. But then as n goes to zero, the arrival rate $\lambda(n) \to \infty$: When there are few agents of a given skill, the solution of a problem arrives infinitely fast. As a result, there is no longer any imperfection in the problem-solving technology. In other words, when no one else is around, the skilled worker is suddenly endowed with an infallible ability to solve any problem immediately. This infinite arrival rate of solutions is still present when there is a positive sunk cost ($a < 0$) such that $h(n) = a + bn^{\gamma}$. Then the arrival rate is exactly the same (the constant does not matter here). However, the presence of $a < 0$ puts a bound on the minimal n that would be chosen, avoiding the n in which the arrival rate is arbitrarily large. The case with $h(n) = a + bn^{\gamma}$ is quite important, since we show in Section 3.2 that any CES production function can be rewritten in this form. We also address the case of CES with a sunk cost in the Online Appendix.

Notice that a violation of Assumption 1 does not necessarily imply identical distributions. As we show in the Online Appendix (Supporting Information), if the production function satisfies decreasing elasticity of substitution, larger firms will hire proportionally more skilled workers than smaller, less productive firms, even though $h'(0) = \infty$. The main difference from the case in which Assumption 1 is satisfied is that now all firms will hire all skills available in the economy. However, this goes against the empirical evidence presented by Caliendo et al. (2012), Ortín-Ángel and Salas-Fumás (2002), and Colombo and Delmastro (1999), among others.[10]

It therefore seems reasonable to focus on processes in which the arrival rate is bounded, thus moving beyond the CES assumption about h. Our assumption is equivalent to assuming that the solution to problems does not arrive at an infinite rate. Given the concavity of h, Assumption 1 is equivalent to imposing that the marginal product of the first worker at any skill level is bounded.

Examples that satisfy this condition include a Poisson process with exponential decay $\lambda(n) = e^{-n}$ (with corresponding $h(n) = 1 - e^{-n}$) or hyperbolic decay $\lambda(n) = \frac{1}{1+n}$ (with corresponding $h(n) = \arctan n$) where in both cases $\overline{h}' = \lambda(0) = 1$, but not the CES technology where $h'(0) \to \infty$.

2.4. *Equilibrium.* Atomless firms act as price takers. Given a vector of wages $w(x)$ and normalizing the output price to 1, firm A's problem is given by:

$$\pi_A = \max_{n_1,\dots,n_N} A\left[\sum_{i=1}^{N} h(n_i)x_i\right]^{\beta} - \sum_{i=1}^{N} n_i w(x_i),$$

where n_i is short for $n_i(A)$, firm A's demand for workers of skill level x_i. A competitive equilibrium of the economy can be defined as follows:

[10] Even though Colombo and Delmastro (1999) and others have pointed to a decrease in skill and level dispersions among firms of different sizes.

DEFINITION 1. In a competitive equilibrium in this economy (1) firms maximize profits π_A, (2) workers choose the job with the highest expected wage offered $w(x_i)$ for a type i, and (3) markets clear.

Before analyzing equilibrium, we derive the elasticity of substitution, which will play a key role in characterizing the equilibrium properties. The elasticity of substitution between inputs n_i and n_j, denoted by σ, is defined as

$$\sigma = \frac{d\ln(n_j/n_i)}{d\ln(TRS(n_i, n_j))},$$

where $TRS = \frac{dy/dn_i}{dy/dn_j}$ is the technical rate of substitution. Then

$$\sigma = -\frac{h'(n_i)}{h''(n_i)}\frac{1}{n_i}.$$

Observe that the elasticity of substitution is independent of β because it measures the change along the isoquant. In fact, the role of β only enters when making comparisons across different isoquants, that is, whether inputs are gross complements or gross substitutes. In the Appendix, we show that when $\beta > 1$ inputs are gross complements and when $\beta < 1$ they are gross substitutes.

3. THE RESULTS

3.1. *Diverse Organizations.* We now proceed by analyzing the problem under Assumption 1. That is, under no circumstance do agents have an infinite problem-solving ability, implying that there will always be a residual amount of unsolved problems.[11] Given the concavity of h, Assumption 1 is equivalent to assuming that the marginal product of the first worker at any skill level is bounded. Also, since the degree of complementarity/substitutability is fully governed by the elasticity of substitution σ, which is independent of β, we will focus on the case where $\beta = 1$. This considerably reduces the notation and allows for closed-form solutions. In the Appendix, we extend our proofs for the case $\beta \neq 1$ and $h = 1 - e^{-n}$. As we will show, the extension for $\beta < 1$ is straightforward, whereas $\beta > 1$ involves more constraints in order to preserve the concavity of the problem; however, in both cases results are preserved.

The seemingly minor restriction on the technology of bounded arrival of new solutions has important implications for the way firms will optimally hire skilled workers. In particular, it will lead to diversity *between* organizations driven by the diversity of skills *within* organizations. To see this effect, observe that the bounded arrival of new solutions means that firms with different levels of TFP will have different marginal returns from hiring any given worker. Inspecting the firm's profit function, we see that the TFP coefficient A multiplies each skill's output. Therefore, for a given identical n_i, the higher the firm's TFP, the more valuable skill i's output:

$$\max_{n_1,\ldots,n_N} A\left[\sum_{i=1}^{N} h(n_i)x_i\right] - \sum_{i=1}^{N} n_i w(x_i). \text{ s.t. } n_i \geq 0, \forall i \in \{1,\ldots,N\}.$$

There are nonnegativity constraints on n_i that will now be binding whenever the marginal product $Ah'(n_i)x_i$ is below the wage rate $w(x_i)$. These constraints are immediately evident from the first-order conditions, which imply:

$$(n_i): h'(n_i) \leq \frac{w(x_i)}{Ax_i}, \quad \forall i \in \{1,\ldots,N\}.$$

[11] This is equivalent to assuming that the solution to problems does not arrive at an infinite rate.

From concavity, we know that the left-hand side of the above inequality decreases in n_i, and the maximum value is achieved when $n_i \to 0 \Rightarrow h'(0) = \overline{h}'$. The right-hand side is constant from price taking. Therefore, if $\overline{h}' < \frac{w(x_i)}{Ax_i}$, firm A does not hire workers of skill i.

Given continuity of the distribution of TFP, for each skill level i there exists a critical firm $\underline{A}(x_i)$ such that only firms with $A \geq \underline{A}(x_i)$ hire workers with skill i. Thus, the critical TFP firm satisfies $\underline{A}(x_i) = \frac{w(x_i)}{\overline{h}' x_i}$. The demand of a firm A for skill i therefore satisfies:

$$n_i(A) = \begin{cases} h'^{-1}\left(\dfrac{w(x_i)}{Ax_i}\right), & \text{if} \quad A \geq \underline{A}(x_i) \\ 0, & \text{otherwise} \end{cases}.$$

Then, from market clearing and substituting demand $n_i(A)$:

$$\sum_{A > \underline{A}(x_i)} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\mu(A) = m_i.$$

For the case of a continuum of skills we take $m_i = F(i) - F(i - \Delta)$ and $\mu(A) = G(A_j) - G(A_j - \Delta)$, dividing both sides by Δ and taking the limit as $\Delta \to 0$. Then the equivalent condition is:

$$\int_{\underline{A}(x_i)}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right) g(A)dA = f(i). \tag{2}$$

It now immediately follows that firms now will not hire on the entire support of skills, but will have a cutoff rule for the highest skill a firm hires given its amount of TFP A. Call the cutoff rule i_{CEO}.

PROPOSITION 1. *Firms with higher A*

(1) Have a larger labor force, and they hire more of all skill types.
(2) Hire from a strictly larger range of skills: supp $f_{\underline{A}} \subset$ supp $f_{\overline{A}}$ for all $\underline{A} < \overline{A}$.

The first part states that firms with higher firm-specific TFP A are larger. The productivity per worker is higher, and therefore at common economy-wide wage rates, it is optimal for them to hire more workers. The second part establishes that because of the bounded marginal product, the support of agents from which low A firms hire is included in the support of high A firms. The question remains how the skill distributions within the different firms compare.

3.2. *Hierarchies.* As we mentioned above, we denote the highest skill type i that a firm with TFP A hires by $i_{\text{CEO}}(A)$, defined by the cutoff rule $i_{\text{CEO}}(A) = \frac{w(x_i)}{\overline{h}' A}$, in which $\overline{h}' = h'(0)$. We can now establish the following proposition:

PROPOSITION 2. *If $f'(i) < 0$, the highest skilled worker $i_{\text{CEO}}(A)$ increases with A and therefore with the size of the firm.*

PROOF. In the Appendix. ■

We further characterize the equilibrium allocation by imposing additional properties on h''' (and therefore on the decay of the arrival process λ''). We can formally establish the following results under the sufficient condition that h''' is not too positive, that is λ is not too convex. That means that there is no sudden drop in the arrival rate followed by a constant arrival. For the

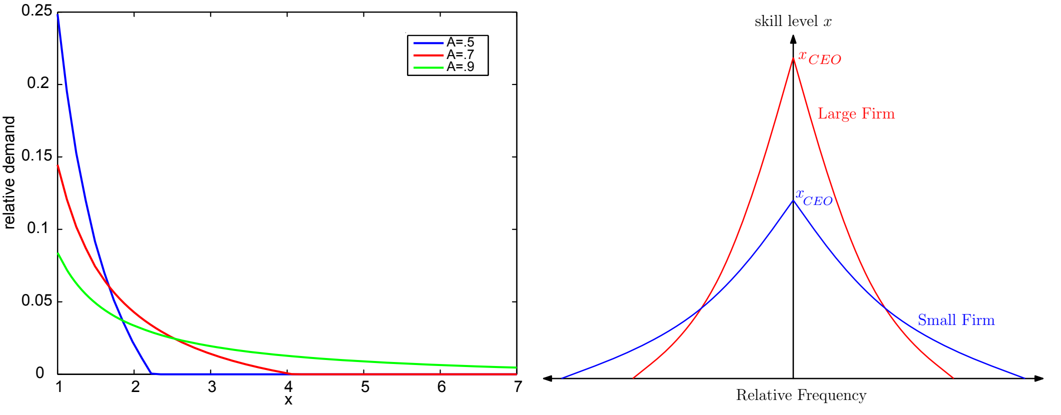

FIGURE 3

(A) THE PDF OF SKILLED WORKERS. (B) ORGANIGRAM IN DIFFERENT FIRMS.

remainder of the results, we maintain this assumption, which is satisfied for a broad class of functions h and all the ones used in examples.

LEMMA 1. *There is single-crossing of the densities:* $\frac{d^2(\frac{n_i(A)}{n(A)})}{dAdx_i} > 0$.

PROOF. In the Appendix. ■

The next result follows from single-crossing of the firm skill densities and the fact that the support of skills hired in smaller firms is included in that of larger firms.

PROPOSITION 3. *(Stochastic Dominance). The skill distribution of larger firms first-order stochastically dominates that of smaller firms.*

Example. In order to see how different TFP firms design their organizations with different skill distributions, consider the following example with the skill distribution Pareto (with support $x > 1$ and with coefficient 1), with the firm TFP distribution uniform on $[0, 1]$, and with exponential decay in the arrival rate of solutions: $\lambda(n) = e^{-n}$. This implies $h(n) = 1 - e^{-n}$. For three different levels of $A = 0.5, 0.7, 0.9$, Figure 3(A) depicts the densities of skills in each firm.

If the skill distribution is taken as the guidance for the organigram of a firm, this has immediate implications for how the chart will look for firms of different sizes. Larger firms will not only hire more skilled workers for their top position of CEO, but they will also have a thinner density of skilled agents at every rank, including at the bottom. The smaller mom-and-pop store will have a broad base of low skilled workers with a CEO who is only moderately more skilled. Figure 3(B) plots the organigram implied by the distribution of skills.

The next proposition then follows immediately from stochastic dominance and the fact that wages are determined competitively, that is, equal skills earn equal wages.

PROPOSITION 4. *Larger firms pay on average higher wages than smaller firms.*

Notice that while in the standard assignment models of the labor market (Sattinger, 1993) more productive firms pay higher wages, those firms are not larger simply by assumption. More importantly, the one-firm-equals-one-job assumption rules out any comparison of firms on the basis of the skill distribution. In our model, more productive firms hire more skilled workers in a stochastic dominance sense. Together with the competitive wage setting and the fact that more productive firms are larger, this implies that average wages are higher.

Skill levels and salaries of the CEOs are higher in larger firms. Since Roberts (1956), it has repeatedly been confirmed that CEO compensation increases with firm size. In particular, the evidence suggests that CEO compensation increases proportionally to a power function of the firm size in a cross section. Most recently, Gabaix and Landier (2008; see also Terviö, 2008) confirm this finding and refer to it as Roberts' law. They find an estimate for $\widehat{\kappa} \simeq 1/3$ where $w \sim S^{\kappa}$. In Section 5 we use CEO compensation to back out the distribution of TFP across different firms.

3.2.1. *Polyarchies.* We now consider a horizontally differentiated economy. All workers are differently skilled in the sense that they are differentially specialized (e.g., logistics versus finance). A logistics manager cannot perform the job of the finance manager in the firm, yet each of the workers is equally successful in generating output. In other words, the value to the firm from employing different types of skills is the same, that is, the order on x is weak, $x_i = x_j$ for all i, j. In the result about hierarchies (Proposition 2), we assume the sufficient condition that the density is decreasing. That is because given the setup of the model, the wage differential is driven by both the productivity differential x and the supply of skilled workers $m(x)$ at the economy-wide level. If all skill classes are equally productive, then the only determinant of wages is the scarcity of skills. The next result formalizes this.

PROPOSITION 5. *(1) The wage decreases as the density of skills $f(m)$ increases and (2) higher A firms hire from a strictly larger range of skills.*

PROOF. The first part of the proof can immediately be verified from market clearing and the first-order condition. The second part is implied by Proposition 1. ■

Those skills i that are more scarce ($f(i) < f(j)$) will necessarily command higher wages, given concavity of h and symmetry in their contribution to output. As a result, wage differentials are entirely driven by the skill distribution. Firms substitute cheaper, more available skills in place of more expensive, scarcer skills, so that workers' marginal contribution to output is equal to the marginal cost. This may explain why scarcity in certain professions (e.g., welders or engineers) can drive up relative wages. Because the higher-TFP firms have a higher marginal product for all n, they will hire those types that a low-A firm hires and other types as well.

3.3. *Relaxing Assumption 1: Identically Distributed Organizations.* We now consider the case in which Assumption 1 is violated. In particular, let us consider $h(n) = bn^{\gamma}$. As we saw before, in order to obtain this h function with our nonhomogeneous Poisson process, the arrival rate $\lambda(n)$ must be given by $b\gamma n^{\gamma-1}$. But then as n goes to zero the arrival rate $\lambda(n) \to \infty$: When there are few agents of a given skill, the solution of a problem arrives infinitely fast. Even though this assumption seems unnatural, the above $h(n)$ function implies a CES production process that is commonly assumed as the aggregate technology in macro models, and we apply it to the firm production. Recall that an often used version of the CES production function is of the form $Y = A[\sum_{i=1}^{N} bn_i^{\gamma} x_i]^{1/\gamma}$. It turns out—as we will establish in the next lemma—that the most general form is

$$Y = A\left[\sum_{i=1}^{N}\left(a + bn_i^{\gamma}\right)x_i\right]^{\beta}.$$

Observe that the CES coefficient γ and the coefficient β for gross substitutes/complements need not be inversely related. The elasticity of substitution is given by $\sigma = -\frac{h'}{h''}\frac{1}{n} = \frac{1}{1-\gamma}$.

LEMMA 2. *The following two statements hold for a, b, and γ constants with $a \in \mathbb{R}$, $b > 0$, $\gamma \in [0, 1]$:*

(1) The elasticity of substitution σ is constant if and only if $h(n_i)$ is of the form $a + bn_i^{\gamma}$.
(2) $L(n)$ is homothetic if and only if $h(\cdot)$ is of the form $a + bn_i^{\gamma}$.

PROOF. In the Appendix. ■

The immediate implication of the firm-level CES technology is that all firms have an identical skill composition.[12]

PROPOSITION 6. *In equilibrium all firms have the same skill distribution $F_A(i)$ equal to the economy's skill distribution $F(i)$ if and only if the production technology is CES with $a \geq 0$.*

PROOF. In the Appendix. ■

The proof documents in detail that this result follows from the homotheticity property of the CES technology, in which the equilibrium allocation depends on the ratio of the inputs, not their value. From the first-order conditions of the firm's problem, it follows that:

$$\frac{n_i}{n_j} = \left(\frac{w(x_j)x_i}{w(x_i)x_j}\right)^{\frac{1}{1-\gamma}}.$$

After solving for the demand and imposing market clearing, the equilibrium allocation of skills j in firm A is given by:

$$n_j(A) = \frac{A^{\frac{1}{1-\gamma\beta}} m(x_j)}{\sum_A A^{\frac{1}{1-\gamma\beta}} \mu(A)}.$$

As a proportion of the total labor force in firm A, $n(A)$, the fraction of j workers is equal to the ratio of those workers in the skill distribution in the market m,

$$\frac{n_j(A)}{n(A)} = \frac{m(x_j)}{m}$$

for every A. As a result, the distribution of skills within the firm is identical to the distribution of skills in the market.

Given identical distributions, we immediately find the following result.

PROPOSITION 7. *Under CES with $a \geq 0$:*

(1) There is full support of the distribution of all firms and
(2) Firms of different sizes pay identical average wages.

The key insight here is that the CES technology, $h(n) = a + bn^{\gamma}$, violates Assumption 1. At zero, the derivative is infinite ($h'(0) = \infty$). The immediate implication of this is that no matter the equilibrium wage, all firms have an infinite marginal return from hiring any skill type. The prediction of the CES production technology is therefore that even the smallest firms will compete for the highest skilled CEO in the economy. With perfect divisibility, those firms will hire only a tiny fraction of that CEO's time. Observe that Lemma 2 establishes necessary and sufficient conditions and that therefore it is also true that identical distributions can arise only under a CES technology.

The prediction of different firms having identical skill distributions may be analytically attractive, but it is not realistic. A small mom-and-pop corner store is unlikely to hire agents as

[12] Below, in Section 4 and Appendix we analyze the case where $a < 0$.

skilled as the CEO of large companies like General Electric, even if only for a tiny fraction of their time. The empirical evidence shows a consistent pattern across countries—among them Spain, France, Italy, Sweden, and the United States—and across industries that firms of different sizes have different numbers of hierarchical layers with different skill levels, measured through both education and worker fixed effects following the methodology developed by Abowd et al. (1999). Similarly, studies analyzing CEO talent—Jung and Subramanian (2012) among others previously mentioned—show that CEO talent and compensation increase with firm size and value.

4. CHARACTERIZATION

In this section, we discuss how the model can be extended and how it can be put to use in interpreting and analyzing several characteristics of the labor market, as well as how these characteristics change as the environment—in particular technology and educational costs—evolves.

4.1. *The Evolution of Diverse Organizations.* The outlook of organizations changes over time. Firms respond to different technological and market conditions, and the equilibrium allocation of skills within the firm and between firms changes. In this section, we analyze how changes in the environment affect the diversity of organizations.

4.1.1. *TFP evolution.* First, we consider an overall change in the productivity of firms. In history, there has been continuous technological change, and while that has affected certain types of firms and certain sectors differently, it is safe to say that all firms have experienced an increase in productivity. Without having to be specific about which type of firms were affected most by the technological change, if TFP has gone up in all firms, it immediately follows that the new distribution of A first-order stochastically dominates the old distribution. That is, we compare a given old economy with distributions of TFP $G(A)$ to an otherwise identical new economy with distribution $G_1(A)$ and where $G_1(\cdot)$ first-order stochastically dominates (FOSD) $G(\cdot)$, that is $G(\cdot) > G_1(\cdot)$ for all A.

PROPOSITION 8. *As the distribution of TFP stochastically dominates wages increase at all skill levels, given A, the equilibrium demand for each skill type i, n_i, decreases, that is, firms become smaller, and the skill type of the CEO i_{CEO} decreases.*

PROOF. In the Appendix. ■

The impact of an increase in the concentration of TFP is an increase in the competition for labor at all skill levels. As a result of the increased competition, wages increase everywhere, which in turn leads to a decrease in equilibrium quantity demanded. For any firm with an unchanged TFP A, that also implies that the skill level of the CEO, x_{CEO}, decreases.

It is key here to observe that the result specifies what happens to the size of a firm that has a constant A. Since the exercise is to consider an increase in A of all firms, as a result of which the size distribution becomes more concentrated, the effect on demand of the average or median firm is ambiguous. There is a trade-off between an increase in the concentration of A and a decrease in the size for a given fixed A. This is consistent with the data on the evolution of the size distribution of firms in the United States (at least for the COMPUSTAT sample of publicly traded firms). This evidence suggests that the employment size distribution has remained relatively stable over time.

4.1.2. *Technological progress.* In addition to changes in the concentration of productive resources, an obvious change is the impact of technological progress in the problem-solving production function, for example, through improved communication. Historically, there have been periods of greater innovation in communication technology, in particular at the turn of 19th century with the development of the telephone and telegraph, and in the last few decades with the advent of the information age. Equally skilled workers now can communicate more easily and more effectively, thus potentially increasing their problem-solving capacity. There may be different channels through which technology affects productivity, and one plausible channel is the arrival rate of the solution of new problems $\lambda(n)$, which we model by means of a monotonic and increasing shift in $\lambda(\cdot)$. Consider a class of functions $\lambda(n;\zeta)$ parameterized by ζ and assume that λ is everywhere increasing in ζ. The immediate implication is that the marginal productivity of every skill group increases since $h'(n;\zeta) = \lambda(n;\zeta)$. Because λ is increasing in ζ everywhere, it therefore follows that $h'(0) = \overline{h}'$ increases.

As we increase the arrival rate of solutions, we can now evaluate the impact on equilibrium. The first implication is that wages will increase unambiguously.

PROPOSITION 9. *As the marginal productivity increases* $\frac{dh'(n;\zeta)}{d\zeta} > 0$, *all wages increase.*

PROOF. In the Appendix. ■

With an increase in the problem-solving ability, all workers become more productive, and in a competitive market this leads to an unambiguous increase in wages at all levels. In contrast, the effect on the distribution of skills within and between firms is ambiguous. Technological change increases the demand by all firms at all skill levels, but the general equilibrium effect from higher prices mitigates and possibly offsets this increased demand. The net effect is ambiguous, and therefore it is also unclear whether the skill level of the CEO will increase or decrease. However, consistent with Gabaix and Landier (2008) and Terviö (2008), wages of the CEOs go up unambiguously.

Finally, let's assume that $h(n;\zeta) = \zeta h(n)$, where ζ is a technological parameter. In this sense, we impose a skill-neutral technological advance. In this case, notice that:

$$h'(n;\zeta) = \zeta h'(n)$$

and

$$\frac{dh'(n;\zeta)}{d\zeta} = h'(n) > 0.$$

If the technological progress assumes this particular form, we can show the following result.

PROPOSITION 10. *If* $h^*(n;\zeta)$ *is linearly increasing in a skill-neutral technological progress, higher skills have a higher salary boost due to technological progress.*

PROOF. In the Appendix. ■

Notice that the production function is now:

$$A\sum_{i=1}^{N} \zeta h(n_i)x_i = A\zeta \sum_{i=1}^{N} h(n_i)x_i.$$

Therefore, we can see ζ as a component of the TFP. In this case, an increase in ζ will be seen as an increase in the standard deviation of the TFP distribution. Consequently, technological progress will generate

(1) An increase in the dispersion of the TFP distribution;
(2) An increase in the wage dispersion that benefits highly skilled workers.

This pattern is empirically corroborated by the evidence in Faggio et al. (2010), which studied a U.K. firm-level panel data set covering manufacturing and nonmanufacturing firms for the period from 1984 to 2001. Their evidence not only indicates an increase in TFP, labor productivity, and wage dispersion during the period, but also suggest that the leading cause for the increase in dispersion is the introduction of new technologies. Our model can clearly speak to these patterns.

4.2. *Investment in Skills: Endogenous Heterogeneity.* Consider an economy with ex ante identical agents and a technology in which each agent can choose to invest in education to obtain a level of skills i. The cost of education is given by

$$C(x_i) = K + c(x_i),$$

consisting of a fixed cost $K \geq 0$ and a strictly convex variable cost $c(x_i)$, where $c\,(0) = 0$. Without loss of generality, we normalize the workers' net utility to zero. Considering that in equilibrium all skill are supplied, we have that:

$$w(x_i) = K + c(x_i), \quad \forall x_i \in (0, \overline{x})\,.$$

Observe that $w'(x_i) = c'(x_i) > 0$ and $w''(x_i) = c''(x_i) > 0$. Therefore, in equilibrium the wage function must be increasing and convex.

Considering the case in which we have $h(n_i)$ strictly increasing and concave but $h'\,(0) = \overline{h}' < \infty$, we obtained from the labor market equilibrium that:

$$\int_{\underline{A}(x_i)}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right) g(A)dA = f\,(i) \qquad (\star),$$

where

$$\underline{A}(x_i) = \frac{w(x_i)}{\overline{h}' x_i}.$$

In the previous sections, equilibrium was determined by an exogenous distribution of skills $f\,(i)$ and an endogenous wage schedule. Now, $w(x_i)$ is exogenously pinned down by the cost function $C(x_i)$, whereas $f\,(i)$ is determined endogenously in ($\star$). Therefore, an increase in educational costs to acquire higher skill levels—for example, higher college tuition costs—naturally generates a more dispersed wage distribution in this setup.

With diversity in the production technology, ex ante identical agents have incentives to take on different levels of investment. Because the marginal productivity at all skill levels is decreasing, in equilibrium agents will choose investment levels that are different, yet obtain the same net utility. More costly investment will necessarily lead to higher skills and thus higher wages, and for agents to be indifferent the equilibrium measure of agents investing must eventually be decreasing in skills. For agents to be willing to bear the higher cost, wages must increase sufficiently, which is the case when there are increasingly fewer workers who obtain higher skills. This then leads to the following result.

PROPOSITION 11. *The equilibrium distribution of skills is always unimodal and has a long right tail. When there is no fixed cost of investment ($K = 0$), the density is everywhere downward sloping.*

PROOF. In the Appendix. ■

It is worth pointing out here that the properties of the distribution are derived in the context of a competitive market, and that no externalities are needed to generate a nondegenerate distribution of firms.[13] In the Appendix, we explicitly solve an example with investment.

5. QUANTIFYING TFP AND LABOR PRODUCTIVITY

In recent years, increasingly there have been attempts to measure total productivity and labor productivity.[14] Observing the productivity of a firm or an economy is central to understanding the firm's individual profitability and the aggregate state of the economy. In this section, we first derive the aggregate distribution of TFP from CEO compensation only, and then we simulate an economy to investigate the impact of the production technology on estimates of TFP and of the marginal productivity of labor across firms.

5.1. *The Aggregate Distribution of TFP.* Since TFP is not directly observable, all estimation procedures build, implicitly or explicitly, on a theoretical model in the background. Even though the empirical literature has evolved in order to take into account the biases created by the endogeneity of input choices and the correlation between the inputs choices and the total factor productivity,[15] TFP is still estimated as a residual after the estimate of the production function parameters.

By design, our model allows firms to be different in their individual TFP A, and to illustrate the theory, our objective is to derive the distribution of A within the economy. Of course, if we had enough detailed firm-level data on the skill and wage distribution, we could back out the distribution of A.[16] In the absence of such detailed data, we can nonetheless obtain further information on the distribution of TFP using data from aggregate distributions, in particular the aggregate distribution of employment (number of workers per firm) or the aggregate distribution of earnings of CEOs.

Suppose that the underlying model of the economy is ours, in which organizations differ in the skill distribution of their workers. We know then that the production technology is not CES, and we can use information on the distribution of compensation of the highest skilled worker in each firm to pin down the distribution of TFP across different firms. From the firm's problem, at each skill level i the first-order condition holds, including at i_{CEO}. By definition, the CEO is the type for whom $n(x_{\text{CEO}}) = 0$, $h'(n)$ is evaluated at zero, which is common to all firms. This allows us to identify A from CEO characteristics only:

$$A = \frac{w(x_{\text{CEO}})}{h'(0)x_{\text{CEO}}}.$$

Instead of using the CEO skill level i_{CEO}, we can also use the investment decision as in Section 4.2. Let the convex cost of investment function be denoted by $C(x) = bx^{\theta}$ where $\theta > 1$ and $b > 0$ is a constant. Then in equilibrium $bx^{\theta} = w(x)$ and we can write $A = \kappa w(x_{\text{CEO}})^{1-1/\theta}$, where $\kappa = \frac{b^{1/\theta}}{h'(0)}$ is a constant.

Using Compustat Executive Compensation Data, we obtain the distribution of $w(x_{\text{CEO}})$, up to a constant κ, provided we know the θ. Recall that θ measures the curvature of the investment cost. In Figure 4, we plot the estimated TFP distribution for values $\theta = 2$ and $\theta = 3$. Irrespective of the horizontal scale, which is pinned down by the constant κ, and even for different θ's, what

[13] For a framework with spillovers from technology adoption and the ensuing endogenous heterogeneity of ex ante identical agents, see, for example, Eeckhout and Jovanovic (2012).

[14] Bartelsman and Doms (2000) review the empirical literature that uses longitudinal microlevel data sets, which follow large numbers of establishments or firms over time. They conclude that the most significant finding is the degree of heterogeneity in productivity across establishments and firms in nearly all industries examined.

[15] As well as issues with entry and exit of firms, see Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), among others

[16] With an impressive matched employer employee data set, this is exactly what Caliendo et al. (2012) do.

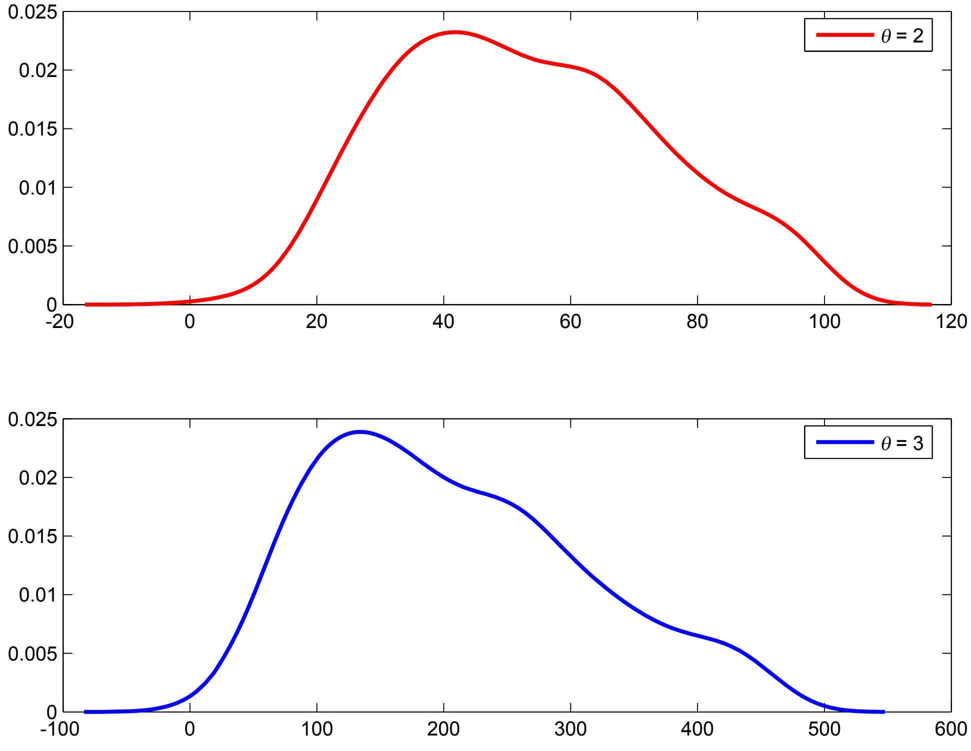

FIGURE 4

IMPLIED TFP DISTRIBUTION (A) ACROSS DIFFERENT FIRMS FOR DIFFERENT INVESTMENT COSTS (θ).

is surprising is the extent to which there is heterogeneity in TFP across firms. This is consistent with the findings reported in Bartelsman and Doms (2000), and together with the evidence that skill distributions differ between firms, this provides support for the hypothesis that firm technology is not CES.

5.2. *The Impact of Technology on Measured TFP and Worker Productivity.* Next, we investigate what the impact is on estimated TFP or worker productivity when the technology is misspecified. Much of the empirical literature assumes CES or Cobb–Douglas production functions. Our results obtained above show that these production functions generate identical skill distributions across firms, a fact that goes against the characteristics of most data sets collected to study firms' skill dispersion. We study the impact of these restrictive functional form assumptions on the estimates of firms' TFP, the marginal product of skills, and the elasticity of substitution of skills across different firms.

Next, we simulate repeated panel data samples.[17] We consider a simplified version of the model described in Section 2 in which there are three skill levels and three different levels of TFP. We consider that all firms are ex ante homogeneous up to the differences in TFP. The production function for a firm with TFP type i is given by:

$$A_i \left[h(n_1^i)x_1 + h(n_2^i)x_2 + h(n_3^i)x_3\right],$$

[17] The use of simulated data allows us to provide a best case scenario for the exercise, since it allows us to generate a data set for a best case scenario: there is no entry or exit of firms or workers, and therefore no attrition or selection bias, all firms are single-product, and the market is competitive.

TABLE 1
ELASTICITY OF SUBSTITUTION (FROM SIMULATED DATA)

	Low TFP	Medium TFP	High TFP
σ_{12}	0.104	0.099	0.097
σ_{13}	0.104	0.099	0.097
σ_{23}	0.232	0.211	0.198

where $h(n) = 1 - e^{-n}$. Therefore, the firm's problem is given by:

$$\max_{n_1^i, n_2^i, n_3^i} A_i \sum_{j=1}^{3} [1 - e^{-n_j^i}] x_j - \sum_{j=1}^{3} n_j^i w(x_j) \text{s.t. } 0 \leq n_j^i \leq m(x_j) \text{ , } \forall j,$$

where $m(x_j)$ is the measure of workers of skill level j in the economy. Solving the firm's problem, we obtain:

$$n_j^{i*} = \begin{cases} \ln\left(\dfrac{A_i x_j}{w(x_j)}\right) & \text{if } \frac{A_i x_j}{w(x_j)} \geq 1 \\ 0 & \text{otherwise} \end{cases}.$$

The system of equations that pins down the equilibrium is then given by:

$$\begin{cases} n_j^{i*} = \max\left\{\ln\left(\frac{A_i x_j}{w(x_j)}\right), 0\right\} & \forall\, i, j \in \{1, 2, 3\} \\ \sum_{i=1}^{3} N_i n_j^{i*} = m(x_j) & \forall\, j \in \{1, 2, 3\}, \end{cases}$$

where N_i is the measure of firms with TFP type i in the economy.

5.2.1. *Parameterization.* In order to simulate the data, we pin down parameters used in the simulations as reported in Table 1.

	Firms				Workers		
	$i = 1$	$i = 2$	$i = 3$		$j = 1$	$j = 2$	$j = 3$
A_i	1,000	1,500	2,000	x_j	60	75	90
N_i	990	3,990	990	m_j	60,000	30,000	15,000

The source of variability included in the data is measurement errors. We included an additive i.i.d. noise $\upsilon \sim N(0, \gamma^2)$ to the optimal demands by firms for each skill level. In the results presented below, we assume $\gamma = 0.1$. We also include measurement errors in the output data. We assume an additive i.i.d. noise $\varepsilon \sim N(0, \sigma)$ with $\sigma = 10$.[18]

We generate 100 independent samples of 5,970 firms. Each sample consists of a balanced panel of 20 periods. The use of repeated samples follows Berk et al. (1999) and allows us to have some idea of the robustness of our results across multiple estimations of the econometric methods.

Once the data is generated, we estimate the production function parameters (and consequently TFP) using the following methods: (1) pooled OLS with clustered error by firm,

[18] We also generated data using AR(1) shocks and population growth. Since results were similar, we decided to omit them.

TABLE 2
AVERAGE MARGINAL PRODUCT

	Simulated Data	Fixed Effects			Dynamic GMM			GMM System		
		Low A	Med A	High A	Low A	Med A	High A	Low A	Med A	High A
MP_1	3.81	1.47	2.15	2.82	56.07	82.16	107.43	1,134.26	1,662.03	2,173.42
MP_2	724.91	1.81	2.54	3.23	64.06	89.94	114.45	1,298.39	1,823.04	2,319.81
MP_3	1,0731.47	2.19	2.82	3.40	59.89	76.77	92.68	1,101.50	1,411.53	1,704.31

(2) linear panel with fixed effects, and (3) GMM methods, in particular GMM dynamic and system.

Results for the pooled OLS with clustered errors by firm are overall poor, in particular since this method suffers from a simultaneity bias that is corrected by the other methods. Therefore, results from this method will be omitted. The other two methods, fixed effects and GMM methods, address potential biases generated by the fact that all skill demands are pinned down jointly by the firm after observing the TFP. The difference between the fixed effects and the GMM methods comes from the way in which these potential biases are taken into account. Although the fixed effects model assumes that firm characteristics (i.e., TFP) are constant over time and therefore avoids the bias by working with the data in first differences, the GMM methods instrument the endogenous regressors (skill demands) using past observations of the same variables. For more details, please see Van Beveren (2012).[19]

Notice that most of the methods developed to estimate TFP distributions and other production function parameters assume a CES or Cobb–Douglas production function. Here we discuss the results for the Cobb–Douglas case.[20] Therefore, we generate the data using the production function:

$$Y = A\left[h\left(n_1\right)x_1 + h\left(n_2\right)x_2 + h\left(n_3\right)x_3\right],$$

where $h(n) = 1 - e^{-n}$. Yet, we apply the estimation methods that assume a Cobb–Douglas production function (in natural logs):

$$\ln Y = \ln A + \beta_1 \ln n_1 + \beta_2 \ln n_2 + \beta_3 \ln n_3. \tag{3}$$

We use the methods and simulated data described above to recover the parameters $\{\beta_i\}_{i=1}^{3}$ and through the residual the distribution of TFP across firms in the data. We then compare the estimated TFP distributions against the underlying TFP distribution in the model that generated our samples, as well as measures of marginal productivity and elasticity of substitution across skills and firms.

First, notice that the elasticity of substitution in the underlying model, given the equilibrium and the parameters used in the simulation, is presented in Table 1. Under the maintained assumption of a Cobb–Douglass production technology, the elasticity is constant and equal to 1.[21]

In Table 2, we evaluate the marginal product of each skill and therefore these skills' contributions to output. The first column reports the marginal product for each skill type from the

[19] Most of the latest applications use the models developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). However, these models depend on capital inputs. In addition they were developed to deal with entry and exit decisions, which are not an issue in our case.

[20] We also have estimates for the trans-log production function, which is a common generalization that accounts for CES estimates. Since results were similar we omitted them, but they are available upon request.

[21] Notice that even a general CES production function will not significantly improve the estimates, given the fact that elasticities are not only varying across skills, but also decreasing with firm TFP.

simulated data. Note that given perfect competition in the labor market, marginal productivity is identical across firms of different TFP and equal to the market wage at each skill level. Next, we compare these numbers to the estimates obtained from the different methods. The following columns present the averages of the estimates obtained across multiple samples. Keep in mind that the estimates are based on a technology (Cobb–Douglas) different from the technology that generated the data. This mismatch of the model and estimated production functions implies that the estimates of marginal productivity are not necessarily equal across different TFP firms.

Observe that the estimates are far from the ones derived from the simulated data. In particular, the estimation methods are completely off when it comes to the variation of productivity across different skill types, and they tend to equalize the contribution of heterogeneous skills. They overestimate the contribution of lower skills and underestimate the contribution of the higher ones, reducing the standard deviation of the marginal products. In addition, the estimates of the regression coefficients show a fairly erratic behavior across different samples, which raises concerns about how much these methods can say about the contribution of any given skill to the output.

Once we obtain the estimates for the skill coefficients β, TFP is estimated as the residual. Our results show that all methods deliver similar results. They tend to overestimate TFP, in particular for larger firms. However, the bias is small. Once we demean the estimates, the estimated TFP distribution is close to the actual distribution.

In summary, if a researcher is only interested in an estimate of the TFP distribution, the biases in the estimates using a standard Cobb–Douglas or CES production functions, though real, are a minor concern. If instead the goal is to estimate the marginal productivity across skills and therefore calculate the contribution of different skills to output, then the estimates obtained are heavily biased. The marginal product of the low skilled is overestimated and that of the high skilled is underestimated. Moreover, the point estimates exhibit too much variation across sample estimates, thus compromising their interpretation. Therefore, we see these results as a cautionary tale emphasizing the importance of the production function specification in empirical work studying the role of different skills and firm organization to output.

6. CONCLUDING REMARKS

There is evidence that diversity within groups and organizations is beneficial for the performance of those groups (see, for example, Page, 2007). In this article, we have embedded a stylized notion of diversity in an equilibrium framework and have shown that within-group diversity induces between-group diversity. Consequently, in equilibrium, firms differ in their skill distributions. In particular, more productive firms will have skill distributions that first-order stochastically dominate the skill distributions of less productive firms. Moreover, highly productive firms are larger, employ more-skilled CEOs, and hire from a broader set of skills. These patterns have been empirically corroborated in exercises using data sets from several countries. We have highlighted that assuming CES production technologies for firms is convenient but highly specific; CES is necessary and sufficient for having identical firms. We have also analyzed the impact of changes in the environment and shown that technological progress in this framework leads to downsizing. When technological progress is skill-neutral, we see an increase in wage dispersion.

We have used the model to investigate some quantitative features of productivity within and across firms. Because the model has a sharp prediction about the productivity of the CEO, we used CEO compensation to derive the induced TFP distribution of firms. We have also simulated the economy and shown that assuming a CES production function in empirical exercises does not substantially affect estimates of TFP across firms. However, in pinning down worker productivities, assuming CES heavily biases the estimates: It underestimates the impact of high skills and overestimates that of low skills.

APPENDIX

A.1. *Production Function Microfoundation and Characteristics*

A.1.1. *Definition of a nonhomogeneous Poisson process*

DEFINITION 2. $[N(t), t \geq 0]$ is said to be a nonhomogeneous Poisson process with intensity function $\lambda(t)$, if:

(i) $N(0) = 0$
(ii) $[N(t), t \geq 0]$ has independent increments;
(iii) $\Pr[2 \text{ or more events in} (t, t+h)] = o(h)$
(iv) $\Pr[\text{ exactly 1 event in } (t, t+h)] = \lambda(t)h + o(h)$.

Then if we let,

$$h(t) = \int_0^t \lambda(s)ds,$$

it can be shown that:

$$\Pr[N(t) = n] = e^{-h(t)} \frac{[h(t)]^n}{n!}, \quad n \geq 0.$$

Or, in other words, $N(t)$ has a Poisson distribution with mean $h(t)$. $h(t)$ is said to be the *mean value function* of the process.

Alternatively, we can define a nonhomogeneous Poisson Process as follows.

DEFINITION 3. $N(t)$ is a nonhomogeneous Poisson process with arrival rate $\lambda(t)$ if it is a counting process such that:

(i) The increments are independent;
(ii) $N(0) = 0$;
(iii) $P(N(v) - N(u) = n) = \frac{(\int_u^v \lambda(t)dt)^n}{n!} e^{-\int_u^v \lambda(t)dt}$.

Derivation of the continuous case. First of all, we rewrite the model with Δs, that is, we assume that $x_{i+1} = x_i + \Delta$. Then, we use a partition/refinement argument, which delivers a Riemann integral,[22] taking $\Delta \to 0$.[23] By doing this transformation, we can obtain a partition $\mathcal{P}$ and an associated set of points $\mathcal{X}$ in which $\mathcal{X}_i \in I_i$, where I_i is an interval in the partition $\mathcal{P}$. Then, $S[(\mathcal{P}, \mathcal{X}), f]$ is defined by:

$$S[(\mathcal{P}, \mathcal{X}), f] = \sum_{i=1}^{N-1} h(n(\mathcal{X}_i)) \mathcal{X}_i |I_i|.$$

A function f is integrable if and only if:

$$\lim_{|\mathcal{P}| \to 0} S[(\mathcal{P}, \mathcal{X}), f] = \int_{\underline{x}}^{\overline{x}} h(n(x)) x dx$$

for any $(\mathcal{P}, \mathcal{X})$.

[22] A function is Riemann integrable if it is continuous almost everywhere, that is, it is discontinuous in at most a zero measure set.

[23] A necessary condition for this proof is that $n(x)$ is piecewise continuous, which is satisfied in equilibrium.

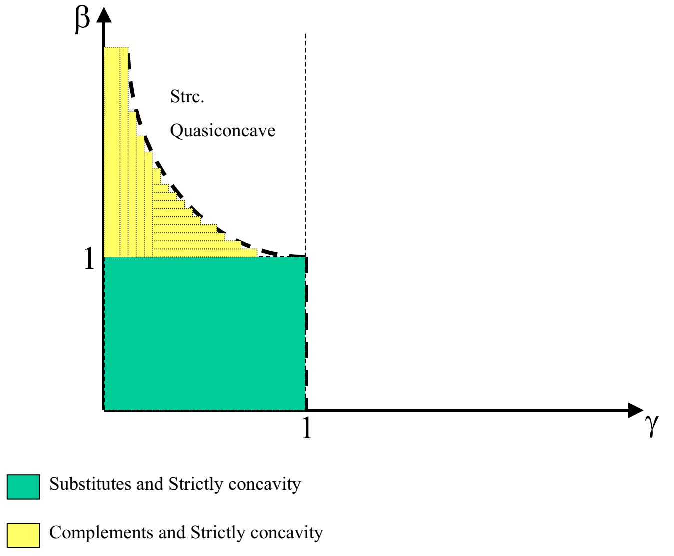

FIGURE A1

COMPLEMENTS AND SUBSTITUTES

Gross complements and gross substitutes. From the firm's objective function, we derive

$$\frac{\partial^2 \pi}{\partial n_i \partial n_j} = A\beta\left(\beta - 1\right)\left[\sum_{i=1}^{N} h(n_i)x_i\right]^{\beta-2} h'(n_i)h'(n_j)x_ix_j.$$

Notice that $\frac{\partial^2 \pi}{\partial n_i \partial n_j} > 0 \Longleftrightarrow \beta > 1$. Therefore, β determines whether x_i and x_j are gross complements or substitutes.

CLAIM 1. *If $\beta > 1$, inputs are gross complements. If $\beta < 1$ they are gross substitutes.*

For example, let $h(n_i) = n_i^{\gamma}$; then we can summarize this in terms of the parameter values for $\beta \in \mathbb{R}^+$ and $\gamma \in [0, 1]$. The firm's problem is well defined for $\beta < 1/\gamma$ (a sufficient condition for concavity is $\gamma\beta < 1$). Then the yellow area in Figure A1 is the range of parameters where inputs in production are complements, and the green area is where they are substitutes.

Concavity of the firm's objective function.

LEMMA 3. If $\gamma\beta < 1$, then the firm's objective function as defined generally above is strictly concave, whenever $a \geq 0$.

PROOF. Notice that:

$$\frac{\partial^2 \pi}{\partial n_i^2} = k\beta\left(\beta - 1\right)\left[\sum_{i=1}^{N}\left(a + bn_i^{\gamma}\right)x_i\right]^{\beta-2} b^2\gamma^2 n_i^{(\gamma-1)2}x_i^2 + k\beta\left[\sum_{i=1}^{N}\left(a + bn_i^{\gamma}\right)x_i\right]^{\beta-1} b\gamma\left(\gamma - 1\right)n_i^{\gamma-2}x_i.$$

Rearranging,

$$\frac{\partial^2 \pi}{\partial n_i^2} = k\beta\left[\sum_{i=1}^{N}\left(a + bn_i^{\gamma}\right)x_i\right]^{\beta-2} b\gamma n_i^{\gamma-2}x_i\left\{\left(\beta - 1\right)bn_i^{\gamma}\gamma x_i + \left(\gamma - 1\right)\left[\sum_{i=1}^{N}\left(a + bn_i^{\gamma}\right)x_i\right]\right\}.$$

Then, $\frac{\partial^2\pi}{\partial n_1^2} < 0$ if we have:

$$k\beta\left[\sum_{i=1}^{N}\left(a+bn_i^{\gamma}\right)x_i\right]^{\beta-2} b\gamma n_1^{\gamma-2}x_1\left\{(\beta-1)\,bn_1^{\gamma}\gamma x_1+(\gamma-1)\left[\sum_{i=1}^{N}\left(a+bn_i^{\gamma}\right)x_i\right]\right\}<0,$$

which implies:

$$(\beta-1)\,bn_1^{\gamma}\gamma x_1+(\gamma-1)\left[\sum_{i=1}^{N}\left(a+bn_i^{\gamma}\right)x_i\right]<0.$$

Rearranging, we have:

$$(\gamma\beta-1)\,bn_1^{\gamma}x_1+(\gamma-1)\left[a\sum_{i=1}^{N}x_i+b\sum_{i=1}^{N}n_i^{\gamma}x_i\right]<0.$$

From our assumption that $h'(\cdot)>0$, we must have $b>0$. However, initially we do not have any assumptions on a. If we consider $a\geq 0$, we notice that a sufficient condition would be $\gamma\beta<1$ (We are already assuming by concavity of $h(\cdot)$ that $\gamma<1$). To get Inada conditions, we necessarily have $a=0$. If $a<0$, then we would not have strict concavity holding for all $\mathbf{n}$.

Let us now consider the second principal minor. Then, our condition is given by:

$$k^2\beta^2\left[\sum_{i=1}^{N}\left(a+bn_i^{\gamma}\right)x_i\right]^{2\beta-3} b^2\gamma^2 n_1^{\gamma-2}n_2^{\gamma-2}x_1x_2(\gamma-1)\left\{\begin{array}{c}(\gamma\beta-1)\,b\left(n_1^{\gamma}x_1+n_2^{\gamma}x_2\right)\\ +(\gamma-1)\left[a\sum_{i=1}^{N}x_i+b\sum_{i=3}^{N}n_i^{\gamma}x_i\right]\end{array}\right\}>0.$$

Again, for the case in which $a\geq 0$, $\gamma\beta<1$ is a sufficient condition, since $\gamma<1$.

Let us now consider the third principal minor. Then, our condition is given by:

$$k^3\beta^3\left[\sum_{i=1}^{N}\left(a+bn_i^{\gamma}\right)x_i\right]^{3\beta-4} b^3\gamma^3 n_1^{\gamma-2}n_2^{\gamma-2}n_3^{\gamma-2}x_1x_2x_3(\gamma-1)^2$$
$$*\left\{\begin{array}{c}(\gamma\beta-1)\,b\left(n_1^{\gamma}x_1+n_2^{\gamma}x_2+n_3^{\gamma}x_3\right)\\ +(\gamma-1)\left[a\sum_{i=1}^{N}x_i+b\sum_{i=4}^{N}n_i^{\gamma}x_i\right]\end{array}\right\}<0.$$

Then, again, for the case in which $a\geq 0$, $\gamma\beta<1$ is a sufficient condition. We also can see the pattern for these conditions, meaning that $\gamma\beta<1$ is a sufficient condition for any N and $a\geq 0$. Therefore, $\gamma\beta<1$ is a sufficient condition for strict concavity of the objective function whenever $a\geq 0$. ∎

Relation to Lucas' span of control model. One obvious way to relax the property that $h'(0)=\infty$ while maintaining the CES formulation is to allow for a technology that has a fixed cost of initial investment. Suppose a is negative, and one allows for the possibility that the firm decides not to produce with a particular skill level when output of that skill is negative. Then a can be considered as a fixed cost that only is incurred in the case of positive output. This obviously truncates the production function and renders the production set nonconvex. The production technology then is:

$$y=A\left[\sum_{i=1}^{N}\max\{a+bn_i^{\gamma},0\}x\right]^{\beta}.$$

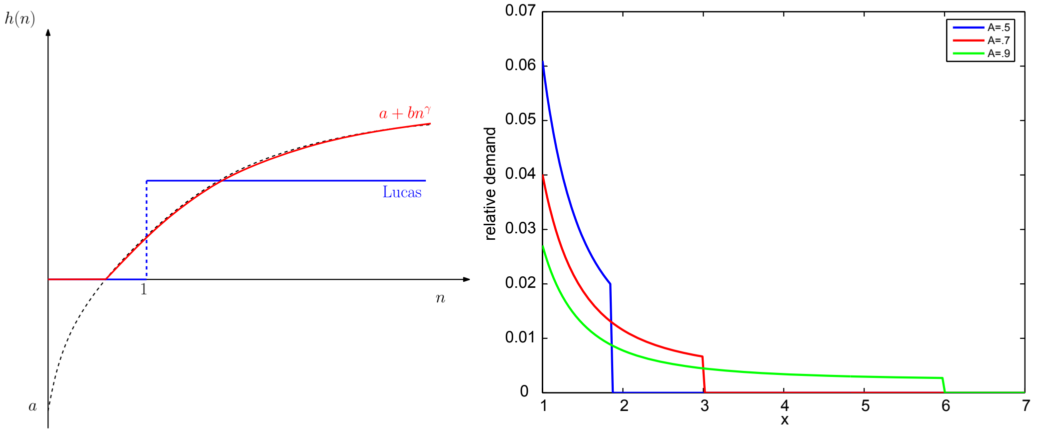

FIGURE A2

(A) SPAN OF CONTROL: $h(n)$ VERSUS LUCAS (1978). (B) SKILL DENSITIES, CES, AND SUNK COSTS.

This formulation is reminiscent of Lucas' (1978) span of control technology of the manager (though not of the workers): a firm needs exactly one manager, no more no less. In our version, the firm needs to incur a fixed cost a before hiring any skilled worker, which will require a minimum scale of $(-\frac{a}{b})^{1/\gamma}$. There is, however, no maximum scale (illustrated in Figure A2).

In equilibrium, firms will now differ, but due to the nonconvexity in the production technology, the size distribution of skills of all firms is truncated at the top. This is illustrated below for an example where $h(n) = -0.5 + n^{1/2}$, $\beta = 1$, skills are distributed according to the Pareto with coefficient 1, and the firm TFP distribution is uniform on $[0, 1]$.

We derive under plausible conditions that the highest skilled worker has a higher type in larger firms than in smaller firms. This implies that the distribution of higher k firms has fat tails at the top as long as the skill distribution has decreasing density.

Consider the production function we used above $h(n) = a + bn^\gamma$, where $a < 0$. A firm will hire a type x if for that type, the equilibrium n^* yields positive output: $h(n^*) = a + b(n^*)^\gamma$, where we derived n earlier as:

$$n^*(k) = \frac{A^{\frac{1}{1-\gamma\beta}} m(x)}{\sum_k A^{\frac{1}{1-\gamma\beta}} \mu(k)}.$$

The firm's decision problem is therefore to choose n^* as long as $h^* = a + b\left(\frac{A^{\frac{1}{1-\gamma\beta}} m(x)}{\sum_A A^{\frac{1}{1-\gamma\beta}} \mu(k)}\right)^\gamma > 0$. A firm with capital $\underline{A}(x_i)$ will therefore be indifferent between hiring and not hiring provided

$$\underline{A}(x_i) = \left(-\frac{a}{b}\right)^{\frac{1-\gamma\beta}{\gamma}} \left[\frac{1}{m(x)} \sum_{A \in \overline{A}(x)} A^{\frac{1}{1-\gamma\beta}} \mu(A)\right]^{1-\gamma\beta}.$$

The only caveat is of course that the summation over A is for all A actively hiring workers of type x. $\overline{A}(x)$ denotes the set of firms actively hiring type x workers.

PROPOSITION 12. *Let the elasticity of substitution σ be constant, and there is a fixed cost of employing one skill type ($a < 0$); then (1) higher A firms hire more workers; (2) the support of skills hired in lower A firms is included in the support of skills of higher A firms; (3) when the skill density is decreasing, higher A firms hire more skilled workers.*

The important characteristic of this technology is the nonconvexity that leads to a minimum size of each skill level. That implies that even at the top, there is a collection of individuals all with the same skill level. Instead of one CEO, there is a board of top skilled directors who run the firm. In the example below, we derive the equilibrium skill distribution in different A firms. The plot of the density for skills is in Figure A2(B). At the highest level, there is a sharp drop in the density due to the minimum size employed. This is illustrated in the following example.

Let skills be distributed according to the Pareto with location 1 and coefficient 1. Then the cdf is $P(x) = x^{-1}$ and the density is $p(x) = x^{-2}(= m(x))$. Let the distribution of firms be uniform, $\mu = 1$ for $A \in [0, 1]$. Let $h(n) = a + n^{1/2}$, and $\beta = 1$. We have:

$$h(n) = \begin{cases} a + n^{\frac{1}{2}} & \text{if} \quad n > 0 \\ 0 & \text{if} \quad n = 0, \end{cases}$$

where $a < 0$. From previous calculations, we obtain:

$$n_x(A) = \frac{A^2 x^{-2}}{\int_{\overline{\mathcal{A}}(x)} A^2 dA}.$$

Define $A(x_i) = \{A \in \overline{\mathcal{A}} | h(n_x(A)) = 0\}$. Therefore, there exists a threshold such that if $A < A(x_i)$, $\max\{0, a + n_x(A)^{\frac{1}{2}}\} = 0$. This implies that $\overline{\mathcal{A}} = [A(x_i), 1]$. Solving for $A(x_i)$,

$$a + \left[\frac{A(x_i)^2}{x^2 \int_{A(x_i)}^1 A^2 dk} \right]^{\frac{1}{2}} = 0,$$

and rearranging, we have:

$$3A(x_i)^2 = (-ax)^2 [1 - A(x_i)^3], \tag{A.1}$$

which defines $\underline{A}(x_i)$. From the implicity function theorem, we have:

$$\frac{dA(x_i)}{dx} = \frac{2a^2 x [1 - A(x_i)^3]}{3A(x_i)[2 + a^2 x^2 A(x_i)]} > 0.$$

CLAIM 2. *$x_i \to \infty$ as $\underline{A}(x_i) \to 1$.*

PROOF. Assume that there is a $x^* \in \mathbb{R}$ such that $\underline{A}(x^*) = 1$. But then, from (1) we must have:

$$\underbrace{3A(x^*)}_{=3} - (-ax^*)^2 \left[\underbrace{1 - A(x^*)}_{=0} \right] = 0$$

$$3 = 0$$

which is a contradiction. Then, we cannot have $A(x^*) = 1$ for x^* finite. Since $\frac{d\underline{A}(x_i)}{dx_i} > 0, \forall \underline{A}(x_i) \in (0, 1)$, we must have $\underline{A}(x_i) \to 1$ as $x_i \to \infty$. ■

CLAIM 3. $\underline{A}(1) > 0$, *that is, some firms shut down in equilibrium.*

PROOF. From (A.1), we have:

$$3\underline{A}(1)^2 = (-a)^2 [1 - \underline{A}(1)^3].$$

Now, observe that the LHS of this equality is strictly increasing in $\underline{A}(1)$, whereas the RHS is strictly decreasing. But if $A(1) = 0$, we have LHS < RHS, so we must have that $\underline{A}(1) > 0$. ■

The fact that A is increasing in x_i of course also implies that the larger firms A have higher cutoff types for their highest skilled employee. The maximum quality of x that a given A firm hires is:

$$x_{\text{CEO}}(A) = \frac{\sqrt{3}A}{-a\,(1-A^3)^{\frac{1}{2}}}$$

and is increasing in A. The lowest firm that has positive profits in this market

$$x = \frac{\sqrt{3}A}{0.5\,(1-A^3)^{\frac{1}{2}}}$$
$$A = 0.25.$$

Finally, we also verify that the demand in the right tail is in fact decreasing as x increases:

$$\frac{dn_x(A)}{dx} = \frac{d\left\{\frac{3A^2}{x^2[1-A(x)^3]}\right\}}{dx} = \frac{-3A^2\left\{2x\left[1-A(x)^3\right]-3A(x)^2\frac{dA(x)}{dx}x^2\right\}}{x^4\left[1-A(x)^3\right]^2}.$$

Substituting $\underline{A}(x)$ and rearranging, we have

$$\frac{dn_x\,(A)}{dx} = \frac{-12xA^2}{x^4\,[2+a^2x^2\underline{A}(x)]\left[1-\underline{A}(x)^3\right]} < 0.$$

So, the demand is strictly decreasing in x, for a given k and a cutoff rule is optimal.

For this example, we now explicitly have the measure of skills within a firm

$$n(x \mid A) = \frac{3A^2}{x^2\left[1-\underline{A}(x)^3\right]},$$

where $\underline{A}(x)$ solves (A.1). Normalizing this measure to sum up to one, we obtain the firm's distribution of skills. Larger firms hire more workers of all skill types, but from simple comparison of the normalized densities, we see that the low A firms hire proportionally more low skilled workers. The high A firm's skill distribution is therefore heavy in the tail and skewed to the right.

A.2. *Proofs*

A.2.1. *Proof of Proposition 2.* Taking the total derivative of (2) with respect to x_i, we have:

$$-h'^{-1}\left(\frac{w(x_i)}{\underline{A}(x_i)x_i}\right)g\,(A(x_i))\,\frac{dA(x_i)}{dx_i} + \int_{A(x_i)}^{\overline{A}} \frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i}\frac{g(A)}{A}dA = f'(x_i).$$

The first term on the LHS vanishes, since $\underline{A}(x_i) = \frac{w(x_i)}{\overline{h}' x_i} \Rightarrow h'^{-1}(\frac{w(x_i)}{\underline{A}(x_i)x_i}) = h'^{-1}(\overline{h}') = 0$. Then, we have:

$$\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} \int_{\underline{A}(x_i)}^{\overline{A}} \frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)} \frac{g(A)}{A} dA = f'(x_i)$$

$$\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} = \frac{f'(x_i)}{\int_{\underline{A}(x_i)}^{\overline{A}} \frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)} \frac{g(A)}{A} dA}.$$

If $f'(x_i) < 0$, we have $\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} > 0$ because $h'' < 0$. So, we have that $\frac{d\underline{A}(x_i)}{dx_i} > 0$, that is, the higher the skill, the higher the amount of TFP that the firm must have to consider it optimal to hire this worker. It immediately follows that the highest level of skills that a firm A hires $x_{\text{CEO}}(A)$ is increasing in A. ∎

A.2.2. *Proof of Lemma 1.* Observe that:

$$\frac{dn_i(A)}{dx_i} = \frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right) A} \frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} < 0.$$

Therefore, as x_i increases, $n_i(A)$ decreases. Also note that, as we should expect, $n_i(A)$ increases with A:

$$\frac{dn_i(A)}{dA} = -\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right) A^2} \frac{w(x_i)}{x_i} > 0.$$

So, firms with more capital hire more workers of all skills.

Now consider the distribution of skills. Define

$$n(A) = \int_{\underline{x}}^{x_{\text{CEO}}(A)} n_i(A) di,$$

where $x_{\text{CEO}}(A)$ is the x such that $A = \frac{w(x)}{\overline{h}' x}$. Substituting $n_i(A)$, we have:

$$n(A) = \int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right) dx_i.$$

Then:

$$\frac{n_i(A)}{n(A)} = \frac{h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)}{\int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right) dx_j}.$$

Taking the derivative with respect to x_i,

$$\frac{d\left(\frac{n_i(A)}{n(A)}\right)}{dx_i} = \frac{\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)A}\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i}}{\int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)dx_j} < 0.$$

And the cross-derivative is:

$$\frac{d^2\left(\frac{n_i(A)}{n(A)}\right)}{dAdx_i} = \frac{\left\{\begin{array}{c}\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} * \left[\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)A}\right]^2 A\frac{h'''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}\frac{w(x_i)}{x_iA^2}\int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)dx_j \\ +\int_{\underline{x}}^{x_{\text{CEO}}(A)}\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)\right)}\frac{1}{A^2}\frac{w(x_j)}{x_j}dx_j\left[\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)A}\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i}\right]\end{array}\right\}}{\left[\int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)dx_j\right]^2}$$

$$= \frac{\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i}\left[\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)A^3}\right]\left\{\begin{array}{c}\frac{h'''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}{\left[h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)\right]^2}\frac{w(x_i)}{x_i}\int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)dx_j \\ +\int_{\underline{x}}^{x_{\text{CEO}}(A)}\frac{1}{h''\left(h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)\right)}\frac{w(x_j)}{x_j}dx_j\end{array}\right\}}{\left[\int_{\underline{x}}^{x_{\text{CEO}}(A)} h'^{-1}\left(\frac{w(x_j)}{Ax_j}\right)dx_j\right]^2}.$$

So, if $h'''(\cdot) < 0$, we have that $\frac{d^2\left(\frac{n_i(A)}{n(A)}\right)}{dAdx_i} > 0$ and we obtain our single-crossing property. ■

A.2.3. *Proof of Lemma 2.* Part 1. Assume that σ is a constant. Then, we have that:

$$h''(n_i) + \frac{1}{\sigma n_i}h'(n_i) = 0$$

is a homogeneous second order linear differential equation. Considering $h'(n_i) = g(n_i)$ we reduce it to a first order ODE. Solving it, we obtain:

$$h'(n_i) = h'(n_0)\, e^{-\int_{n_0}^{n_i}\frac{1}{\sigma y}dy},$$

where $h'(n_0)$ is the initial condition. Taking the integral on both sides, we obtain:

$$h(n_i) - h(n_0) = h'(n_0)\int_{n_0}^{n_i} e^{-\int_{n_0}^{z}\frac{1}{\sigma y}dy}dz.$$

Then, notice that:

$$-\int_{n_0}^{z}\frac{1}{\sigma y}dy = \frac{1}{\sigma}\int_{z}^{n_0}\frac{1}{y}dy = \frac{1}{\sigma}\ln y|_z^{n_0} = \frac{1}{\sigma}\ln\frac{n_0}{z}.$$

Substituting back, we have:

$$e^{-\int_{n_0}^{z}\frac{1}{\sigma y}dy} = \left[e^{\ln\left(\frac{n_0}{z}\right)}\right]^{\frac{1}{\sigma}} = \left(\frac{n_0}{z}\right)^{\frac{1}{\sigma}}.$$

Substituting back again, we have:

$$h(n_i) - h(n_0) = h'(n_0) \int_{n_0}^{n_i} \left(\frac{n_0}{z}\right)^{\frac{1}{\sigma}} dz$$
$$h(n_i) - h(n_0) = h'(n_0)\, n_0^{\frac{1}{\sigma}} \int_{n_0}^{n_i} z^{-\frac{1}{\sigma}} dz.$$

Solving the integral, we obtain:

$$h(n_i) - h(n_0) = h'(n_0)\, n_0^{\frac{1}{\sigma}} \left[\frac{\sigma}{\sigma - 1} z^{\frac{\sigma-1}{\sigma}} \Big|_{n_0}^{n_i} \right].$$

Then, rearranging, we have:

$$h(n_i) = h(n_0) - \frac{\sigma}{\sigma - 1} h'(n_0)\, n_0 + \frac{\sigma}{\sigma - 1} h'(n_0)\, n_0^{\frac{1}{\sigma}} n_i^{\frac{\sigma-1}{\sigma}}.$$

Therefore:

$$h(n_i) = a + b n_i^{\gamma},$$

where:

$$a := h(n_0) - \frac{\sigma}{\sigma - 1} h'(n_0)\, n_0$$
$$b := \frac{\sigma}{\sigma - 1} h'(n_0)\, n_0^{\frac{1}{\sigma}}$$
$$\gamma := \frac{\sigma - 1}{\sigma}.$$

The reverse argument—show that if $h(n_i) = a + bn_i^{\gamma}$, σ is constant—is trivial and it will be omitted.

Part 2. We know that, by definition, $L(\mathbf{n}; \mathbf{x})$ is homotetic if for any $i, j \in \{1, \ldots, N\}$ and for any $t > 0$, we have that:

$$\frac{\frac{\partial L(\mathbf{n};\mathbf{x})}{\partial n_i}}{\frac{\partial L(\mathbf{n};\mathbf{x})}{\partial n_j}} = \frac{\frac{\partial L(t\mathbf{n};\mathbf{x})}{\partial n_i}}{\frac{\partial L(t\mathbf{n};\mathbf{x})}{\partial n_j}}.$$

But then, we should have:

$$\frac{h'(n_i)}{h'(n_j)} = \frac{h'(tn_i)}{h'(tn_j)}.$$

Rearranging,

$$\frac{h'(tn_j)}{h'(n_j)} = \frac{h'(tn_i)}{h'(n_i)}.$$

Since this must always be satisfied, we must have:

$$\frac{h'(tn_i)}{h'(n_i)} = c,$$

where c is a constant. But then, we must have:

$$h'(tn_i) = ch'(n_i),$$

since the function $f(\beta) = t^\beta$, with $t > 0$, is continuous and has image on $(0, \infty)$, by mean value theorem we have that there is a $(\gamma - 1) \in (0, \infty)$ such that $t^{(\gamma-1)} = c$. Therefore, we have:

$$h'(tn_i) = t^{\gamma-1}h'(n_i).$$

Therefore, $h'(\cdot)$ is a homogeneous function of degree $\gamma - 1$.

Since $h(\cdot)$ is a univarite function, it is easy to see that it must be of the form $dn_i^{\gamma-1}$, where d is a constant (note that $h(n_i) = h(n_i * 1) = n_i^{\gamma-1}h(1) = dn_i^{\gamma-1}$, where $d = h(1)$). But then, we have:

$$h(n_i) = \int h'(n_i)dn_i = \int dn_i^{\gamma-1}dn_i = \frac{d}{\gamma}n_i^\gamma + a.$$

Define $b = \frac{d}{\gamma}$, so we have:

$$h(n_i) = a + bn_i^\gamma.$$

Again, the reverse argument is trivial and it will be omitted. ■

A.2.4. *Proof of Proposition 6.* (⇒) From the first-order conditions of the firm's problem:

$$\frac{n_i}{n_j} = \left(\frac{w(x_j)x_i}{w(x_i)x_j}\right)^{\frac{1}{1-\gamma}}.$$

Substituting back, we obtain the demand for labor quality x_j as a function of wages:

$$n_j(A) = \left(A\beta\gamma b^2\right)^{\frac{1}{1-\gamma\beta}} \left(\frac{x_j}{w(x_j)}\right)^{\frac{1}{(1-\gamma)}} \left[\sum_{i=1}^{N} \left(\frac{x_i}{w(x_i)^\gamma}\right)^{\frac{1}{1-\gamma}}\right]^{\frac{\beta-1}{1-\gamma\beta}}.$$

Market clearing satisfies:

$$\sum_A n_j(A)\mu(A) = m(x_j),$$

where $m(x_j) = F(x_j) - F(x_{j-1})$ is the measure of worker type x_j. Substituting for the equilibrium quantity of $n_j(A)$ and solving for $w(x_j)$, we obtain the equilibrium wages:

$$w(x_j) = \frac{x_j}{m(x_j)^{1-\gamma}} \left[\sum_{i=1}^{N} \left(\frac{x_i}{w(x_i)^\gamma}\right)^{\frac{1}{1-\gamma}}\right]^{\frac{(\beta-1)(1-\gamma)}{1-\gamma\beta}} \left[\sum_A \left(A\beta\gamma b^2\right)^{\frac{1}{1-\gamma\beta}} \mu(A)\right]^{1-\gamma}.$$

Now, substituting in the demand for wages, we obtain the equilibrium allocations:

$$n_j(A) = \frac{A^{\frac{1}{1-\gamma\beta}} m(x_j)}{\sum_A A^{\frac{1}{1-\gamma\beta}} \mu(k)}.$$

Then, looking at the total labor force of a firm with capital A, we have:

$$n(A) = \sum_{j=1}^{N} n_j(A) = \frac{A^{\frac{1}{1-\gamma\beta}} m}{\sum_A A^{\frac{1}{1-\gamma\beta}} \mu(A)}.$$

where $m \equiv \sum_{j=1}^{N} m(x_j)$.

To see this, from the first-order conditions we get that the fraction of quality j workers in terms of the total number of workers is given by:

$$\frac{n_j(A)}{n(A)} = \frac{A^{\frac{1}{1-\gamma\beta}} m(x_j)/\sum_A A^{\frac{1}{1-\gamma\beta}} \mu(A)}{A^{\frac{1}{1-\gamma\beta}} m/\sum_A A^{\frac{1}{1-\gamma\beta}} \mu(A)} = \frac{m(x_j)}{m}$$

for every A. Therefore, the distribution of workers inside a firm is exactly the same as the one in any other firm and mimics the distribution in the market.

($\Leftarrow$) If firms have identical distributions, we have:

$$\frac{n_i(A)}{\sum_s n_s(A)} = \frac{n_i(A')}{\sum_s n_s(A')}, \quad \forall i,\ \forall A, A'.$$

Therefore, for two different TFPs, A and A', we have:

$$\frac{n_i(A)}{n_i(A')} = \underbrace{\frac{\sum_s n_s(A)}{\sum_s n_s(A')}}_{=c},$$

where c is a constant. From the above expression, we obtain:

$$n_i(A) = cn_i(A').$$

Similarly,

$$n_j(A) = cn_j(A').$$

From F.O.C.s, we obtain:

$$\frac{h'(n_i(A))\,x_i}{h'(n_j(A))\,x_j} = \frac{w(x_i)}{w(x_j)}$$

and

$$\frac{h'(n_i(A'))\,x_i}{h'(n_j(A'))\,x_j} = \frac{w(x_i)}{w(x_j)}.$$

Combining these expressions,

$$\frac{h'\left(n_i(A)\right)}{h'\left(n_j(A)\right)}=\frac{h'\left(n_i(A')\right)}{h'\left(n_j(A')\right)}.$$

Substituting $n_i(A)$ and $n_j(A)$, we have:

$$\frac{h'\left(n_i(A')\right)}{h'\left(n_j(A')\right)}=\frac{h'\left(cn_i(A')\right)}{h'\left(cn_j(A')\right)}$$

As we saw, this implies that $h(\cdot)$ is homothetic. From Lemma 1, we must have that the production function is CES. ■

A.2.5. *Proof of Proposition 8.* Recall that the market clearing condition for skill type i is:

$$\int_{\underline{A}(x_i)}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)dG(A)=f(x_i).$$

Since $h'(\cdot)$ is strictly decreasing, $h'^{-1}(\cdot)$ is also strictly decreasing. Then $h'^{-1}(\frac{w(x_i)}{Ax_i})$ is strictly increasing in A. Then, by the definition of first order stochastic dominance, $G_1(\cdot)$ FOSD $G(\cdot)$ means that we have:

$$\int_{\underline{A}(x_i)}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)dG_1(A)\geq\int_{\underline{A}(x_i)}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)dG(A).$$

Shifting from G to G_1 increases the LHS of the market clearing condition. Since the RHS is a constant, wages must adjust in equilibrium. Since

$$\frac{d}{dw(x_i)}\left\{\int_{\frac{w(x_i)}{\overline{h}'x_i}}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)dG(A)\right\}=\int_{A(x_i)}^{\overline{A}}\frac{1}{Ax_ih''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}dG(A)<0,$$

it follows that changes from $G(\cdot)$ to $G_1(\cdot)$ generate (1) higher wages for every skill level x_i, (2) higher cutoffs $A(x_i)$ and therefore lower x_{CEO}, and (3) lower demand n_i in all firms. To see the effect on demand, observe that from F.O.C.s we have:

$$n_i(A)=\begin{cases}h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right), & \text{if } A\geq A(x_i)\\ 0, & \text{otherwise}\end{cases}.$$

As $w(x_i)$ increases, we have that $n_i(A)$ decreases since $h'^{-1}(\cdot)$ is strictly decreasing. Therefore, each firm of each type demands less from every skill. ■

LEMMA 4. *If $h(n;\zeta)$ is strictly concave and twice differentiable and $h'(n;\zeta)$ is monotonic in the parameter ζ, then $h'^{-1}(n,\zeta)$ is also monotone in ζ in the same direction.*

PROOF. From the definition of $h'^{-1}(n;\zeta)$, we have:

$$h'\left(h'^{-1}(n;\zeta);\zeta\right)=n$$

Applying total derivative.

$$h''\left(h'^{-1}(n;\zeta);\zeta\right)\frac{dh'^{-1}(n;\zeta)}{d\zeta}+\frac{dh'\left(h'^{-1}(n;\zeta);a\right)}{d\zeta}=0.$$

Rearranging.

$$\frac{dh'^{-1}(n;\zeta)}{d\zeta}=-\frac{1}{h''\left(h'^{-1}(n;\zeta);\zeta\right)}*\frac{dh'\left(h'^{-1}(n;\zeta);\zeta\right)}{d\zeta}.$$

Looking at the right-hand side of the above expression, the first term is positive, since $h''(\cdot)<0$. Since $h'(\cdot)$ is monotonic everywhere, it is also when evaluated at $h'^{-1}(n;\zeta)$. Therefore, inversion preserves the monotonic increase or decrease in ζ. ■

A.2.6. *Proof of Proposition 9.* Consider the market clearing condition

$$\int_{\frac{w(x_i)}{h'(0;\zeta)x_i}}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i};\zeta\right)dG(A)=f(x_i).$$

Let us consider the case in which $\frac{dh'(n;\zeta)}{d\zeta}>0$; then, we have:

$$\frac{d}{d\zeta}\left\{\int_{\frac{w(x_i)}{h'(0;\zeta)x_i}}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i};\zeta\right)dG(A)\right\}=\int_{\frac{w(x_i)}{h'(0;\zeta)x_i}}^{\overline{A}}\frac{dh'^{-1}\left(\frac{w(x_i)}{Ax_i};\zeta\right)}{d\zeta}dG(A)+$$

$$\underbrace{h'^{-1}\left(h'(0)\right)}_{=0}g\left(\frac{w(x_i)}{h'(0;\zeta)x_i}\right)*\frac{w(x_i)}{\left[h'(0;\zeta)\right]^2x_i}\frac{dh'(0;\zeta)}{d\zeta}$$

$$=\int_{\frac{w(x_i)}{h'(0;\zeta)x_i}}^{\overline{A}}\frac{dh'^{-1}\left(\frac{w(x_i)}{Ax_i};\zeta\right)}{d\zeta}dG(A)>0.$$

In Lemma 4, we establish that h'^{-1} is increasing in ζ. As a result, an increase in ζ raises the left-hand side, while keeping the right-hand side constant. So we need to change wages to preserve the equality. Notice that:

$$\frac{d}{dw(x_i)}\left\{\int_{\frac{w(x_i)}{\overline{h}'x_i}}^{\overline{A}} h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)dG(A)\right\}=\int_{A(x_i)}^{\overline{A}}\frac{1}{Ax_ih''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}dG(A)<0.$$

Therefore, from implicit function theorem, we have:

$$\frac{dw(x_i)}{d\zeta}=-\frac{\int_{\frac{w(x_i)}{h'(0;\zeta)x_i}}^{\overline{A}}\frac{dh'^{-1}\left(\frac{w(x_i)}{Ax_i};\zeta\right)}{d\zeta}dG(A)}{\int_{\frac{w(x_i)}{h'(0;\zeta)x_i}}^{\overline{A}}\frac{1}{Ax_ih''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)}dG(A)}>0.$$ ■

A.2.7. *Proof of Proposition 10.* From the firm's problem, we have:

$$Ah^{*\prime}(n_i;\zeta)x_i=w(x_i)$$

$$n_i = h^{*\prime -1}\left(\frac{w(x_i)}{Ax_i};\zeta\right).$$

Then, let's calculate $h^{*\prime -1}(n;\zeta)$:

$$\begin{aligned} h^{*\prime}\left(h^{\prime -1}(n;\zeta);\zeta\right) &= n \\ \zeta h^{\prime}\left(h^{\prime -1}(n;\zeta)\right) &= n \\ h^{\prime -1}(n;\zeta) &= h^{\prime -1}\left(\frac{n}{\zeta}\right). \end{aligned}$$

Therefore:

$$n_i = h^{\prime -1}\left(\frac{w(x_i)}{A\zeta x_i}\right).$$

Then, from Equation (2), we have:

$$\int_{\frac{w(x_i)}{\zeta x_i h'(0)}}^{\overline{A}} h^{\prime -1}\left(\frac{w(x_i)}{A\zeta x_i}\right) g(A)dA = f(x_i) \qquad (\blacktriangle).$$

Then, from the implicit function theorem, we have:

$$\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} = \frac{f'(x_i)}{\int_{\frac{w(x_i)}{\zeta x_i h'(0)}}^{\overline{A}} \frac{1}{h''\left(h^{\prime -1}\left(\frac{w(x_i)}{A\zeta x_i}\right)\right)} \frac{g(A)}{A} dA} > 0,$$

given $f'(x_i) < 0$.

Now, back to ($\blacktriangle$), taking the total derivative with respect to ζ, we have:

$$\int_{\frac{w(x_i)}{\zeta x_i h'(0)}}^{\overline{A}} \frac{1}{\zeta}\left[-\frac{w(x_i)}{\zeta x_i} + \frac{d\left(\frac{w(x_i)}{x_i}\right)}{d\zeta}\right] \frac{1}{h''\left(h^{\prime -1}\left(\frac{w(x_i)}{A\zeta x_i}\right)\right)} \frac{g(A)}{A} dA = 0.$$

Since from (2), we know that $w(x_i)$ does not depend on any A in particular, we have:

$$\frac{1}{\zeta}\left[-\frac{w(x_i)}{\zeta x_i} + \frac{d\left(\frac{w(x_i)}{x_i}\right)}{d\zeta}\right] \int_{\frac{w(x_i)}{\zeta x_i h'(0)}}^{\overline{A}} \frac{1}{h''\left(h^{\prime -1}\left(\frac{w(x_i)}{A\zeta x_i}\right)\right)} \frac{g(A)}{A} dA = 0.$$

Once $h''(\cdot) < 0$, the equality is satisfied only if:

$$\frac{d\left(\frac{w(x_i)}{x_i}\right)}{d\zeta} = \frac{w(x_i)}{\zeta x_i} > 0.$$

Finally,

$$\frac{d^2\left(\frac{w(x_i)}{x_i}\right)}{dx_i d\zeta} = \frac{1}{\zeta}\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} > 0.$$

Notice that:

$$\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} = \frac{\frac{dw(x_i)}{dx_i}x_i - w(x_i)}{(x_i)^2}.$$

Therefore, if $\frac{d\left(\frac{w(x_i)}{x_i}\right)}{dx_i} > 0 \Rightarrow \frac{dw(x_i)}{dx_i} > 0$. Similarly, $\frac{d\left(\frac{w(x_i)}{x_i}\right)}{d\zeta} = \frac{1}{x_i}\frac{dw(x_i)}{d\zeta}$. Finally,

$$\frac{d\left(\frac{w(x_i)}{x_i}\right)}{d\zeta dx_i} = \frac{1}{(x_i)^2}\left\{\frac{dw(x_i)}{d\zeta dx_i}x_i - \frac{dw(x_i)}{d\zeta}\right\}.$$

Therefore, if $\frac{d\left(\frac{w(x_i)}{x_i}\right)}{d\zeta dx_i} > 0 \Rightarrow \frac{dw(x_i)}{d\zeta dx_i} > 0$. ■

A.2.8. *Proof of Proposition 11.* Given the equilibrium condition ($\star$), the derivate of $\underline{A}(x_i)$ with respect to x_i is:

$$\underline{A}'(x_i) = \frac{1}{\overline{h}'(x_i)^2}\left[w'(x_i)x_i - w(x_i)\right].$$

Therefore, we have that $\underline{A}'(x_i) > 0$ if and only if $w'(x_i)x_i - w(x_i) > 0$. Since $w\colon (0, \overline{x}) \to \mathbb{R}$ is a strictly convex function, we know that:

$$w(z) > w(y) + w'(y)(z - y), \qquad \forall z, y \in (0, \overline{x})$$

Taking in our case $z = 0$ and $y = x_i$,[24] we have that:

$$w(0) > w(x_i) + w'(x_i)(0 - x_i),$$

since $w(0) = a$; rearranging, we obtain:

$$w'(x_i)x_i - w(x_i) > a.$$

If $a = 0$, we have that $w'(x_i)x_i - w(x_i) > 0 \Rightarrow \underline{A}'(x_i) > 0$. Therefore, if there is no sunk cost, the threshold is always increasing on x_i.

Consider now the case in which $a > 0$. Notice that:

$$\frac{d}{dx_i}\left(w'(x_i)x_i - w(x_i)\right) = w''(x_i)x_i > 0.$$

Therefore, whenever we have $\underline{A}'(x_i) < 0$, there is a threshold x^*, such that for every $x > x^*$, $\underline{A}'(x) > 0$.

Once we have the properties of $\underline{A}(x_i)$, we can derive the characteristics of the equilibrium distribution $f(x_i)$. From ($\star$) we have:

$$f'(x_i) = \int_{\underline{A}(x_i)}^{\overline{A}} \frac{1}{h''\left(h'^{-1}\left(\frac{w(x_i)}{Ax_i}\right)\right)} * \frac{1}{Ax_i^2}\left[w'(x_i)x_i - w(x_i)\right] g(A)dA.$$

[24] In principle the most rigourous argument takes $z = \varepsilon > 0$ but arbitrarily small and then uses a continuity argument to obtain the result.

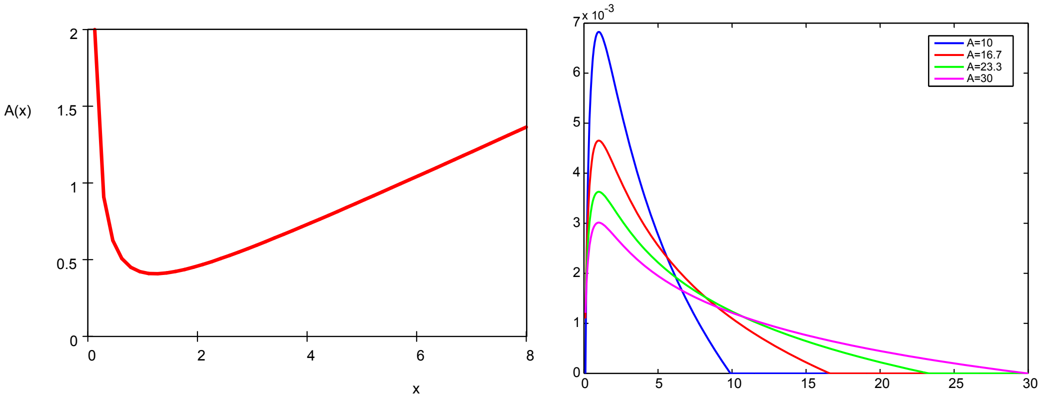

FIGURE A3

(A) THE EQUILIBRIUM FUNCTION $\underline{A}(x_i)$. (B) THE ENDOGENOUS DEMAND FOR SKILLS: $n(x)$

Therefore, whether $f'(x_i) < 0$ depends on $w'(x_i)x_i - w(x_i)$ and, consequently, on the presence or not of a fixed cost a. If $a = 0$, we have that $w'(x_i)x_i - w(x_i) > 0 \Rightarrow f'(x_i) < 0, \forall x_i$. Similarly as before, since $w'(x_i)x_i - w(x_i)$ is increasing in x_i, even if $f'(x_i)$ is positive for some x_i there is a threshold x^* such that $\forall x > x^*$, $f'(x) < 0$. Notice that the threshold is the same for both conditions. This establishes the proposition. ∎

A.3. *Additional Example*

A.3.1. *Example: Investment in skills.* Consider an economy with exponential decay of problem solving ability $h(n_i) = 1 - e^{-\gamma n_i}$, and $c(x_i) = cx_i^2$ and A is exponentially distributed with parameter λ. Then, we have $h'(n_i) = \gamma e^{-\gamma n_i}$ and $\overline{h}' = h'(0) = \gamma$. In this case, we can easily calculate a critical value x^*, which is defined as $x^* = \sqrt{\frac{3}{2}}$. Notice that all firms with TFP above a given threshold will hire a given skill, but now each TFP level has a minimum and a maximum skill thresholds:

$$\underline{A}(x_i) = \frac{K + cx_i^2}{\gamma x_i} \quad \text{and} \quad \underline{A}'(x_i) = \frac{cx_i^2 - K}{\gamma x_i^2}.$$

We now consider the parameter values $K = \frac{1}{2}$, $c = \frac{1}{3}$, $\gamma = 2$. The graph of $\underline{A}(x_i)$ is given in Figure A3.

In order to calculate the density of skills in this economy, we will derive the demand for each skill per TFP. From our previous calculations, we obtain that:

$$n_A(x_i) = \frac{1}{\gamma}\left[\ln A - \ln\left(\frac{K + cx_i^2}{\gamma x_i}\right)\right].$$

Using the expression we obtained for $\underline{A}(x_i)$ above, $A = \frac{K+cx_i^2}{\gamma x_i}$ and $-cx_i^2 + \gamma A x_i - K = 0$, we can get the minimum and the maximum skill thresholds for a given A: $x_{\text{CEO}}(A)$ and $x_{Janitor}(A)$. From the case in which $\Delta = 0$, we obtain the minimum company in activity, the one in which

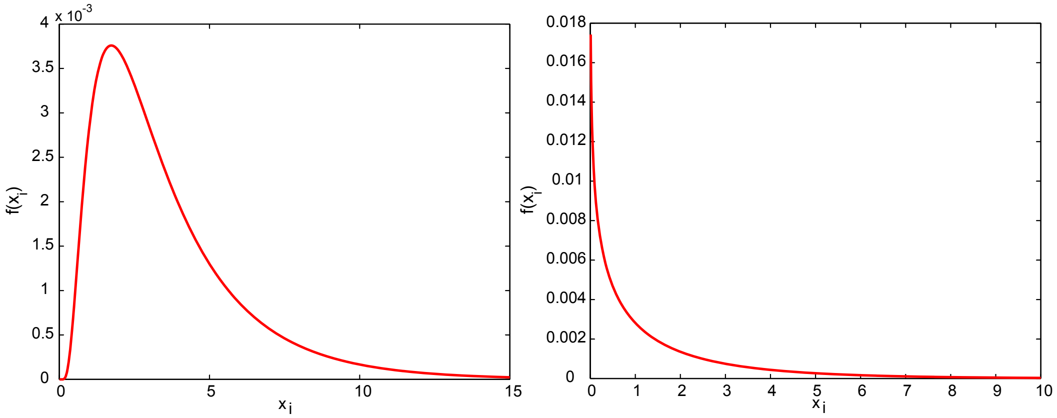

FIGURE A4

INVESTMENT LEADS TO ENDOGENOUS DISTRIBUTION OF SKILLS. THE DISTRIBUTION IS UNIMODAL WHEN THERE IS A FIXED COST OF INVESTMENT (LEFT) AND OTHERWISE EVERYWHERE DOWNWARD SLOPING (RIGHT)

$\gamma^2 A^2 - 4Kc = 0$. For our parameters $\underline{A} = \sqrt{\frac{1}{6}} \approx 0.40825$. Then, solving the equation above, we have:

$$x_{\text{CEO}}(A) = \frac{\gamma A + \sqrt{\Delta}}{2c} \quad \text{and} \quad x_{Janitor}(A) = \frac{\gamma A - \sqrt{\Delta}}{2c}.$$

In our example $x(\underline{A}) = \sqrt{\frac{3}{2}} = x^*$, as we should expect. Graphically, the demand for different type TFP firms in this example is given in Figure A3(B).

We can now derive the distribution of skills. Recall that:

$$f(x_i) = \int_{\underline{A}(x_i)}^{\infty} n_A(x_i) g(A) dA = \int_{\underline{A}(x_i)}^{\infty} \left[-\frac{1}{\gamma} \ln\left(\frac{K + c x_i^2}{A \gamma x_i} \right) \right] \lambda e^{-\lambda A} dA.$$

Rearranging, and using the definition of $\underline{A}(x_i) = \frac{K + c x_i^2}{\gamma x_i}$, we obtain:

$$f(x_i) = -\frac{1}{\gamma} \ln\left(\underline{A}(x_i)\right) e^{-\lambda \underline{A}(x_i)} + \frac{1}{\gamma} \int_{\underline{A}(x_i)}^{\infty} \lambda e^{-\lambda A} \ln A dA.$$

For the second term, using integration by parts, we have:

$$\begin{aligned} \int_{\underline{A}(x_i)}^{\infty} \lambda e^{-\lambda A} \ln A dA &\overset{I.P.}{=} \left\{ -e^{-\lambda A} \ln A \Big|_{\underline{A}(x_i)}^{\infty} \right\} + \int_{\underline{A}(x_i)}^{\infty} \frac{e^{-\lambda A}}{A} dA \\ &\overset{L.H.}{=} e^{-\lambda A} \ln \underline{A}(x_i) + \int_{\underline{A}(x_i)}^{\infty} \frac{e^{-\lambda A}}{A} dA. \end{aligned}$$

since from de L'Hôpital rule, the first term evaluated at ∞ is zero:

$$\lim_{A \to \infty} \frac{\ln A}{e^{\lambda A}} =_{L.H.} \lim_{A \to \infty} \frac{\frac{1}{A}}{\lambda e^{\lambda A}} = \lim_{A \to \infty} \frac{1}{\lambda A e^{\lambda A}} = 0.$$

Then the density of skills is given by:

$$f(x_i) = \frac{1}{\gamma} \int_{\underline{A}(x_i)}^{\infty} \frac{e^{-\lambda A}}{A} dA.$$

This is one form of the exponential integral that does not have a closed form solution, but it is a well used numerical integral. The graph of the density is given in the left panel of Figure A4. The right panel of the figure plots the density for the same example when there is no sunk cost of investment ($K = 0$). In that case, the density is everywhere downwards sloping.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Appendix: Additional Material

REFERENCES

Abowd, J. M., F. Kramar, and D. N. Margolis, "High Wage Workers and High Wage Firms,"*Econometrica* 67(2) (1999), 251–333.

Antràs, P., L. Garicano, and E. Rossi-Hansberg, "Offshoring in a Knowledge Economy," *Quarterly Journal of Economics*, 121(1) (2006), 31–77.

Arellano, M., and S. Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies* 58(2) (1991), 277–97.

Bagger, J., B. J. Christensen, and D. T. Mortensen, "Wage and Productivity Dispersion: Labor Quality or Rent Sharing?" mimeo, 2010.

Bartelsman, E., and M. E. Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," mimeo, Federal Reserve Board, 2000.

Berk, J., R. C. Green, and V. Naik, "Optimal Investment, Growth Options, and Security Returns," *Journal of Finance* LIV(5) (1999), 1553–607.

Cahuc, P., F. Marque, and E. Wasmer, "A Theory of Wages and Labor Demand with Intra-firm Bargaining and Matching Frictions," *International Economic Review* 49(3) (2008), 943–72.

Caliendo, L., F. Monte, and E. Rossi-Hansberg, "The Anatomy of French Production Hierarchies," mimeo, Princeton, 2012.

Colombo, M., and M. Delmastro, "Some Stylized Facts on Organization and Its Evolution," *Journal of Economic Behavior & Organization* 40 (1999), 255–74.

Costrell, R. M., and G. C. Loury, "Distribution of Ability and Earnings in a Hierarchical Job Assignment Model," *Journal of Political Economy* 112(6) (2004), 1322–63.

Crawford, V., and A. Kelso, "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica* 50 (1982), 1483–504.

Eeckhout, J., and B. Jovanovic, "Occupational Choice and Development," *Journal of Economic Theory* 147 (2012), 657–83.

———, R. Pinheiro, and K. Schmidheiny, "Spatial Sorting," *Journal of Political Economy* (2014), forthcoming.

Faggio, G., K. Salvanes, and J. Van Reenen, "The Evolution of Inequality in Productivity and Wages: Panel Data Evidence," *Industrial and Corporate Change* 19(6) (2010) 1919–51.

Faingold, E., "Building a Reputation under Frequent Decisions," mimeo, Yale, 2005.

Ferreira, D., and R. K. Sah, "Who Gets to the Top? Generalists versus Specialists in Managerial Organizations," *RAND Journal of Economics* 43(4) (2012), 577–601.

Fox, J. T., and V. Smeets, "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review* 52(4) (2011), 961–89.

Gabaix, X., and A. Landier, "Why Has CEO Pay Increased so Much?" *Quarterly Journal of Economics* 123(1) (2008), 49–100.

Garicano, L., "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy* 108(5) (2000), 874–904.

———, and T. Hubbard, "Managerial Leverage Is Limited by the Extent of the Market: Hierarchies, Specialization and the Utilization of Lawyers' Human Capital," CEPR Discussion Paper 4924, 2005.

——— and E. Rossi-Hansberg, "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 121(4) (2006), 1383–435.

GAUTIER, P., G. J. VAN DEN BERG, J. C. VAN OURS, AND GEERT RIDDER, "Worker Turnover at the Firm Level and Crowding out of Lower Educated Workers," *European Economic Review* 46(3) (2002), 523–38.

HAWKINS, W., "Competitive Search, Efficiency, and Multi-Worker Firms," mimeo, University of Rochester, 2006.

HELPMAN, E., O. ITSKHOKI, AND S. REDDING, "Wages, Unempoyment and Inequality with Heterogeneous Firms and Workers," mimeo, 2008.

HONG, L., AND S. PAGE, "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory* 97 (2001) 123–63.

———, AND ———, "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers," *Proceedings of the National Academy of Sciences* 101(46) (2004), 16385–9.

IRANZO, S., F. SCHIVARDI, AND E. TOSETTI, "Skill Dispersion and Firm Productivity: An Analysis with Employer-Employee Matched Data," *Journal of Labor Economics* 26(2) (2008), 247–85.

IRARRAZABAL, A., A. MOXNES, AND K. H. ULLTVEIT-MOE, "Heterogeneous Firms or Heterogeneous Workers? Implications for the Exporter Premium and the Impact of Labor Reallocation on Productivity," CEPR Discussion Paper DP7577, 2009.

JUNG, H. W. H., AND A. SUBRAMANIAN, "CEO Talent, CEO Compensation and Product Market Competition," mimeo, 2012.

LENTZ, R., AND D. MORTENSEN, "An Empirical Model of Growth Through Product Innovation," *Econometrica* 76(6) (2008), 1317–73.

LEVINSOHN, J., AND A. PETRIN, "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies* 70(2) (2003), 317–41.

LOPES DE MELO, R., "Sorting in the Labor Market: Theory and Measurement," mimeo, Yale, 2009.

LUCAS, R., "On the Size Distribution of Business Firms," *Bell Journal of Economics* 9(2), (1978) 508–23.

LUTTMER, E., "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics* 122(3) (2007), 1103–44.

MÖLLER, M., "Sorting across Labor Tournaments," mimeo, 2009.

OLLEY, G. S., AND A. PAKES, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica* 64 (1996), 1263–97.

ORTÍN-ÁNGEL, P., AND V. SALAS-FUMÁS, "Compensation and Span of Control in Hierarchical Organizations," *Journal of Labor Economics* 20(4) (2002), 848–76.

PAGE, S. E., *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton, NJ: Princeton University Press, 2007).

PARROTTA, P., D. POZZOLI, AND M. PYTLIKOVAL, "Does Labor Diversity Affect Firm Productivity," mimeo, 2010.

RADNER, R., "Hierarchy: The Economics of Managing," *Journal of Economic Literature* 30 (1992), 1382–415.

ROBERTS, D. R., "A General Theory of Executive Compensation Based on Statistically Tested Propositions," *Quarterly Journal of Economics* 70 (1956), 270–96.

ROSEN, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82(1), (1974), 34–55.

———, "Authority, Control, and the Distribution of Earnings," *Bell Journal of Economics* 13 (1982), 311–23.

ROSSI-HANSBERG, E., AND M. WRIGHT, "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review* 97(5) (2007), 1639–66.

SAH, R. K., AND J. E. STIGLITZ, "The Architecture of Economic Systems: Hierarchies and Polyarchies," *American Economic Review* 76(4) (1986), 716–27.

SATTINGER, M., "Comparative Advantage and the Distribution of Earnings and Abilities," *Econometrica* 43(3) (1975), 455–68.

———, "Differential Rents and the Distribution of Earnings," *Oxford Economic Papers* 31(1) (1979), 60–71.

———, "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature* 31(2) (1993), 831–80.

SHIMER, R., AND L. SMITH, "Assortative Matching and Search," *Econometrica* 68 (2000), 343–69.

SMITH, E., "Search, Concave Production and Optimal Firm Size," *Review of Economic Dynamics* 2 (1999) 456–71.

STOLE, L. A., AND Z. JEFFREY, "Organizational Design and Technology Choice under Intrafirm Bargaining," *The American Economic Review* 86(1) (1996), 195–222.

TAG, J., "Production Hierarchies in Sweden," IFN Working Paper 963, 2013.

TERVIÖ, M., "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review* 98(3) (2008), 642–68.

TEULINGS, C. N., "The Wage Distribution in a Model of the Assignment of Skills to Jobs," *Journal of Political Economy* 103(2) (1995), 280–315.

———, “Comparative Advantage, Relative Wages, and the Accumulation of Human Capital,” *Journal of Political Economy* 113(2) (2005), 425–61.

VAN BEVEREN, I., “Total Factor Productivity Estimation: A Practical Review,” *Journal of Economic Surveys* 26(1) (2012), 98–128.

WOOLDRIDGE, J. M., “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables,” *Economics Letters* 104(3) (2009), 112–4.