Diverse Organizations and the Competition for Talent*

Jan Eeckhout and Roberto Pinheiro†

September, 2008

Abstract

Organizations benefit from a diverse composition of skills. The basic premise of this paper is that *within* the organization, workers of different skills improve the problem-solving ability as the return to adding similar workers is decreasing. We show that when firms compete for talent in the labor market, in equilibrium organizations will differ *between* each other. We find that organizations with higher Total Factor Productivity (TFP) are larger and hire from a broader range of skills. This implies that there are more levels within the organization hierarchy and that their CEO is more skilled. We also find that the skill distribution in high productivity firms first-order stochastically dominates, implying that there are proportionately fewer workers at each level, i.e. they are leaner at the top, whereas the low productivity firms have a wider base of low skilled workers. Our model provides a benchmark for analyzing diverse organizations in a competitive labor market and how they evolve as the economic environment changes. For example, in our model merger waves lead to downsizing of the original firms firms and an adjustment of the skill distribution. When investment in skills is endogenized, we show that the equilibrium skill distribution has a long right tail, even if ex ante all agents are identical.

*We would like to thank numerous colleagues and seminar participants for insightful discussions and comments. In particular, we greatly benefited from comments by Belén Jerez, Boyan Jovanovic, Philipp Kircher, Esteban Rossi-Hansberg and Randy Wright.

†Pinheiro: University of Pennsylvania, rpinheir@econ.upenn.edu. Eeckhout: UPF Barcelona and University of Pennsylvania, eeckhout@ssc.upenn.edu.

1 Introduction

Organizations typically are 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. The basic premise of this paper is that diversity of talents within an organization is beneficial for solving problems and optimizing processes, which ultimately increases productivity. 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. We believe the role of diversity in problem solving is important because even in the manufacturing industry, only a small fraction of workers earn a living actually manufacturing goods. The vast majority of the labor force is involved in design, planning, and services. The firm is considered to be a group of individuals who collectively solve problems such as the design of consumer electronics, the development of software, finding new pharmaceuticals, providing management consulting services,... The building block of our economy is therefore a production technology at the firm level that is designed to incorporate within-firm diversity.

Starting from this simple premise we build a general framework for studying the allocation of skills within the firm, yet where firms coexist while competing in the labor market. We provide a tractable model that is closely related to models commonly used in many applications in macro, labor, and public economics. It incorporates the notion of diverse organizations, where *internally* the demand for skills trades off higher skill levels against more diversity. As a result, in equilibrium firms will consist of a non-degenerate distribution of skills. And since firms hire in a competitive labor market, *external* market wages determine the relative cost of skills economy wide. If firms differ in their firm-specific Total Factor Productivity (TFP), then those internal and external trade-offs typically do not line up such that the demand for skills of different firms is the same. As a result, firms will have different distributions of skills. Our theory thus predicts diversity between organizations as well as within organizations.

Diversity matters within the group. Recent work establishes that diversity is particularly important in collaborative groups and formal organizations. Hong and Page (2001 and 2004) propose a general framework for the collaboration of problem solving agents that is based on internal languages in which they encode solutions. They show that groups of diverse problem solvers, randomly selected from the population outperform groups of high-ability problem solvers. When there are already several experts with similar skills, adding a differently skilled worker will contribute to the resolution by adding a different perspective. Even if the added worker is of lower skill, eventually her individual contribution to the problem will be superior to that of one more of already numerous higher skilled expert. Having a seventh logistics expert streamlining the baggage loading process in for an airline may lead to a minimal improvement if only the ground personnel know that most of the delays are caused by the wheels of one particular type Samsonite suitcase that get stuck in the conveyor belt.

For the purpose of embedding diverse organizations in a general equilibrium framework, our approach is much less sophisticated than Hong and Page. Ours consists of a simple problem solving process in which there is a Poisson arrival rate of a solution within the pool of identical agents. The more agents work on the problem, the higher the expected contribution to the value of the problem at hand. However, because of their common perspective on the same problem, the arrival rate is decreasing in the size of the group. Larger homogeneous groups add more value, but the marginal contribution is decreasing. This then naturally generates a complementarity with other skill levels. The firm level technology aggregates the contribution of the different skill categories, and given the decreasing marginal contribution of each skill level, the lower skilled workers will eventually have a higher marginal productivity which makes it profitable to hire them.

This firm level technology is embedded in a competitive labor market. In assigning skills within the firm, there is not only a trade-off between the marginal productivity at different skill levels, but also with the market wage. The key insight of this paper is that diversity within organizations leads to diversity between organizations. Firms that hire their skilled workers in a competitive labor market and pay common competitive wages will choose a different distribution of skills.

The main implications of this technology is that the firm size is endogenous and consists of a non-degenerate distribution of skills. The imperfect substitutability of workers as inputs in production implies that the size of the firm is endogenous. For reasons of comparative advantage in different jobs, firms in equilibrium decide to hire workers with different talent. Of course, quite a lot is known about the size distribution of firms (for recent examples, see Luttmer (2007) and Rossi-Hansberg and Wright (2007)). The interest here is how firm size relates to the internal distribution of skills within the firm. In particular, firms of different sizes will pay different average wages. This is due to the fact that firms differ in their composition of talent. Firms with higher firm-specific TFP will on average more skilled workers, and as a result, the average wage paid is higher even though workers of the same skill get paid identical wages. We show that only if the elasticity of substitution between different skills is constant and there are no indivisibilities or bounds on the marginal productivity of skills, will the distribution of skills within different firms be identical. But this requires that the arrival rate of solutions to problems is infinite when any workers embarks on a new problem.

In general, the skill distribution of larger firms stochastically dominates the distribution of smaller firms. This is illustrated in Figure 1. Larger firms hire over a wider range of skills and as a result they have more levels in their hierarchy. The immediate implication is that their CEO is more skilled. It also follows that the top of the firm is "leaner", i.e., 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 therefore hire on average more skilled workers and therefore pay on average higher wages. This can explain a well-documented fact in the empirical labor literature, that there is an *employer-size wage premium.*

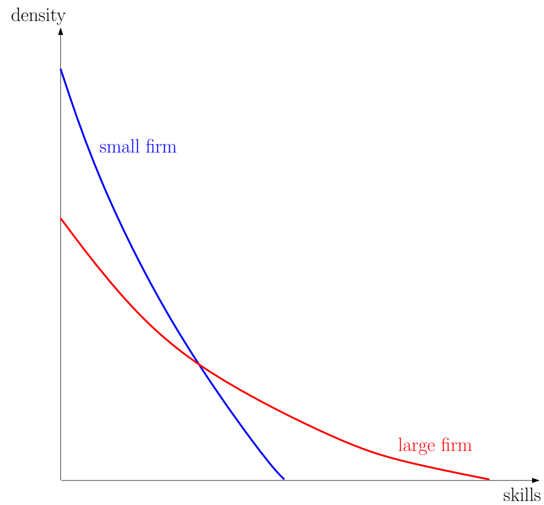

Figure 1: Stochastic Dominance of Skill Distribution in Large Firms

One important finding relative to the existing literature, including in macro-economic applications, is that the CES firm production technology is a knife-edge result. Under CES, all firms are identical in the sense that they have the same distribution of skills. While they differ in size – more productive, higher TFP firms are larger – they hire the same mix of skills, and the skill distribution within any firm is the exact mirror image of the economy skill distribution. We show that the CES property is a necessary and sufficient condition for identical organizations, and as a result, for any non-CES technology organizations will be diverse. We are unaware of such a result to date.

We use our model to analyze how organizations evolve in an environment that is changing. 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 mergers and acquisitions. Since these tend to occur in waves, we model those as stochastic dominance changes in the distribution of TFP. Merger waves lead to a higher concentration of high TFP firms. Our results show that in merger waves, 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 it's TFP will decrease due to increased competition.

The decrease in the employment size at all skill levels provides a market driven explanation for the fact that mergers lead to downsizing. Because the change in firm size is mediated by equilibrium prices in this frictionless model without unemployment, it is not hard to see that downsizing is beneficial for workers. Wages are higher after the merger wave.

We consider extensions of our model. In particular, we consider the case as in Lucas (1978), where a minimum scale of output is needed. This is the case where the return on hiring few workers of a given skill is negative. We find that in equilibrium, firms with larger capital stocks will be larger and will find it profitable to hire proportionally more high skilled workers. This implies that the skill distribution in large firms is skewed to the right compared to the distribution in small firms. In addition, the highest skilled manager in the large firm will be more skilled than the CEO in a small firm. The key difference with the general case is that now there are many highest skilled types in the firm (a board instead of one CEO). As a result, the support of skills of the small firm is included in the support of skills of the large firm. 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 no or small 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, 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.

It is worth pointing out that we do not necessarily see the problem-solving process as a sequential process in which higher skilled agents coordinate problems and attempt to solve problems for which lower skilled workers failed to find a solution. Our production technology can therefore be interpreted as a *polyarchy* instead of a *hierarchy*. Sah and Stiglitz (1986) define a polyarchy as a system where there are several decision makers who can undertake projects (or ideas) independently of one another; in a hierarchy only a few individuals can undertake projects while others provide support in decision making.

There is a large body of work studying the firm as an information processing organization. A recent revival pioneered amongst others by Garicano (2000), Antràs, Garicano and Rossi-Hansberg (2006), Garicano and Rossi-Hansberg (2006)) analyzes the organization of knowledge in hierarchies. Differently skilled agents specialize in the type of problems to solve, with lower skilled agents passing on those problems they have not been able to solve. A cost of communication partitions the skill distribution in a a finite number of hierarchical levels. Our work builds on this literature. Like those models, our model is a version of a many-to-one matching model of workers to an individual firm. Crawford and Knoer (1981) and Crawford and Kelso (1982) provide a general framework and conditions (notably, gross substitutes) under which a market mechanism will lead to an equilibrium allocation and wage vector.[1] We impose more structure by assuming a particular functional form, and we allow for the possibility of a continuum of skill levels. We represent the firm level production processes that embody heterogeneity using the tools of the standard Arrow-Debreu framework. The ultimate objective is to propose a theory that uses a set of equilibrium tools that are common in applied work and macro economic models, yet without assuming an aggregate production function. We provide a tractable framework for skill

[1]See also Kremer (1993), Kremer and Maskin (2004). These O-ring type production technologies can be interpreted as problem-solving technologies. A small mistake by one worker in the production chain can have implications of unprecedented dimensions. One bug in the software may lead to the malfunctioning of millions of electronic devices, or the inadequate quality control for lead in paint can lead to a worldwide recall of a toy.

heterogeneity within the firm based on a commonly used CES production technology with a seemingly minor twist. The advantage is that the model can be analyzed like the standard CES model but the economic implications both at the firm level and in the aggregate economy are substantially different from the CES model.

2 The Model

Population. Consider a population consisting of agents endowed with talent x, a one-dimensional skill characteristic. There are N different types of skills x_i $(x_1, x_2, ..., x_N)$ increasing in i and with a measure $m(x_i)$ of each type. The total measure of agents is normalized to one. 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 total factor productivity. Let $\mu(A)$ denote the measure of each type A.

The problem-solving technology. In our model firm, of each skill x there is a contiuum of agents with measure $n(x)$ employed by the firm. Consider each skill group in isolation and denote by $h(n)\,x$ the expected value of problems solved after hiring a measure n of type x workers. We think of these workers continuously attempting new problems. Any new problem is circulated amongst all n workers of the same type, and a solution is found randomly. If a solution is not found by any type x worker, the problem is abandoned. When more agents of skill x are employed, a new problem is attempted by more workers. Because the workers have common skill sets, an attempt after other agents have tried and failed to find the solution leads to a lower solution probability. We therefore represent the arrival of a solution probability by a non-homogeneous Poisson process with arrival rate $\lambda(n)$.[2] This is a Poisson process with a non-constant arrival rate, which in our case is assumed to be decreasing: $\lambda' < 0$. The firm is concerned with the expected number of problems solved, which is given by $h(n)x$ where

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

Adding more agents increases the number of problems solved, i.e., $h'(n) > 0$, but it does so at a decreasing rate, i.e., $h''(n) < 0$ (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. This is illustrated in Figure 2.

Across different skill types, skill sets are different, and we assume that the problem-solving capacity of different x types is as if they consider different new problems. We assume that the total value of problems solved is determined by a standard additive aggregator:

$$L(\mathbf{n}) = \left[\sum_{i=1}^{N} h(n_i)x_i\right]^{\beta},$$

[2] A formal definition of a non-homogeneous Poisson process is provided in the Appendix.

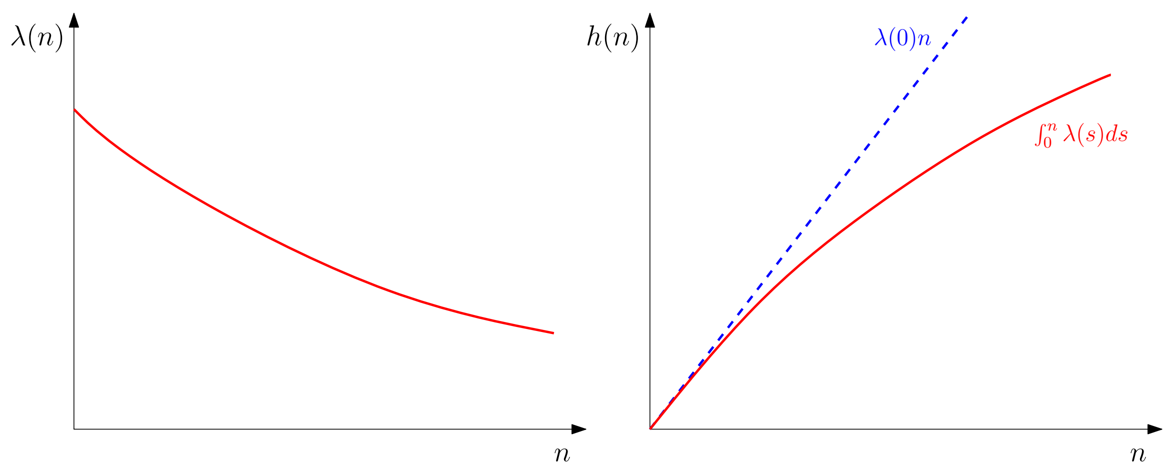

Figure 2: A. The non-homogeneous Poisson arrival rate — B. The expected number of problems solved.

where $\mathbf{n}$ is the vector of quantities $n_i = n(x_i)$ of each skill x_i, and $\beta > 0$. Here we can interpret the role of the higher type x either as a higher probability of solving the problem or as a higher return say on a harder problem.

Observe that even with $\beta = 1$ the different skilled inputs are not perfect substitutes due to the concavity in h.[3] In particular, lower skilled types x_i may generate higher marginal value than higher types $x_j > x_i$ as long as n_i is sufficiently smaller than n_j : $h'(n_i)x_i > h'(n_j)x_j$ provided $n_i \ll n_j$. The complementarity in this aggregator captures the notion that diverse skill types are valuable for problem solving. Observe also that while there is a rank of skills, with higher x types being more successful at solving problems, the problem-solving must not necessarily be interpreted as a hierarchical process. The organizational form can equally well be understood as a polyarchy of different-but-equal problem solving units (see Sah and Stiglitz (1986)).

Finally, solved problems translate in output. Firms produce output y using their given TFP A and the aggregate value of problems solved L given their skill vector $\mathbf{n}$. The production function is given by

$$y = AL(\mathbf{n}).$$

It is important to note at this stage that y is a firm-level production function and that in general it is *not equal to the aggregate production function*. For most of the paper, we consider a discrete distribution of types x. A continuous distribution of types is analogously represented by

$$L(\mathbf{n}) = \left[\int h(n(x))x dF_A(x)\right]^{\beta}$$

where $F_A(x)$ denotes the distribution of skills in firm A. Below we derive that this is the continuous limit of the production technology with finite skill types.

[3] To see the complementarity in inputs, consider for example a Cobb-Douglas like formulation where h is logarithmic and x_i are the weights: $L = \sum_{i=1}^{N} \ln(n_i)x_i$.

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:

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

where n_i is short for $n_i(A)$. A competitive equilibrium of the economy can be defined as follows:

Definition 1 *In a competitive equilibrium in this economy: 1. Firms maximize profits π_A; 2. workers choose the job with the highest wage offered $w(x)$ for a type x; 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

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

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, i.e., 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

We now proceed with solving the equilibrium allocation in the model. First, we state the firm's optimization problem.

3.1 Identically Distributed Organizations

Suppose we were to start out with a Constant Elasticity of Subsitution (CES – σ is constant) 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 $L = \left[\sum_{i=1}^{N} bn_i^{\gamma}x_i\right]^{1/\gamma}$. It turns out – as we will establish in the next Lemma – that the most general form is

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

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

Lemma 1 *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 subsitution 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 Appendix. ■

The immediate implication of the firm-level CES technology is that all firms have an identical skill composition.[4]

Proposition 1 *In equilibrium all firms have the same skill distribution $F_A(x)$ equal to the economy's skill distribution $F(x)$ if and only if the production technology is CES with $a \geq 0$.*

Proof. In Appendix. ■

The proof documents in detail that this follows from the homotheticity property of the CES technology where the equilibrium allocation depends on the ratio of the inputs, not their value. From the First-Order Condtions of the firm's problem, it follows that

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

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

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

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

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

for every A. As a result, the distribution of skills within the firms is identical to the distribution of skills in the market. [5]

Then given identical distributions, the following result immediately follows.

[4] Below, in section 6 we analyze the case where $a < 0$.

[5] One qualifying comment is due here. In the Appendix (Lemma 2) we show that the production technology is always quasi-concave, and strictly concave when $\beta < \frac{1}{\gamma}$. In the latter case the firm's problem is also quasi-concave, but that is not guaranteed when $\beta > \frac{1}{\gamma}$. In fact, when β is sufficiently large, the degree of gross complementarity may be so strong that firms have some monopoly power. With extreme complementarities, ultimately all workers will be employed by the most productive firm. In the context of an exchange economy, Ostroy and Zame (1994) derive conditions on a general class of utility functions under which nonatomic markets are non-competitive. A market equilibrium is guaranteed to be competitive provided there is sufficient substitutability. Throughout this paper (with and without CES technologies) we will implicitly maintain the assumption that β is not too large, thus guaranteeing sufficient substitutability and an equilibrium that is competitive.

Proposition 2 *Under CES with $a \geq 0$:*

1. *There is full support of the distribution of all firms; and*

2. *There is no firm size-wage premium (firms of different sizes pay identical average wages)*

The key insight here is that for the CES technology $h(n) = a + bn^{\gamma}$, at zero the derivative is infinite $h'(0) = \infty$. The immediate implication of this is that no matter what the equilibrium wage is, 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, that will imply they hire only a tiny fraction of her time. Observe that Lemma 1 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 skilled as the CEO of large companies like General Electric, even if only a tiny fraction of their time. In terms of our problem-solving technology, this comes down to the properties of the stochastic process that governs the Poisson arrival rate of solutions. Imposing the CES structure on h implies that the non-homogeneous Poisson process has an infinite arrival rate as the number of workers in a skill category becomes small. To see this, consider for example, $h(n) = bn^{\gamma}$, i.e. with $b > 0, \gamma < 1$, and $a = 0$. In order to obtain this h function with our non-homogeneous 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.[6] As a result, there is no longer any imperfection in the problem-solving technology.[7] In other words, when no-one else is around, the skilled worker is suddenly endowed with an infallible ability to solve any problem immediately. It seems therefore reasonable to focus on processes where the arrival rate is bounded, thus moving beyond the CES assumption on h. In the next section we will find that this also implies more realistic properties of the firm distribution across firms.

3.2 Diverse Organizations

We now proceed by analyzing the problem in which it is explicitly assumed that under no circumstance agents have an infinite problem-solving ability and that as a result there is always a residual amount

[6] Considering the case in which we have a positive sunk cost ($a < 0$), then the arrival rate is exactly the same (the constant doesn't matter here). However, the presence of a puts a bound on the minimal n that would be chosen, avoiding the n in which the arrival rate is arbitrarily large. We address the case of CES with a sunk cost in section 6

[7] The CES example suffers from the problem pointed out by Faingold (2005) in continuous time games with imperfect monitoring. As the period length shrinks to zero, the number of signals observed in any given interval of real time increases without bound, and as a result, the monitoring imperfection vanishes.

of unresolved problems. This is equivalent to assuming that the solution to problems does not arrive at an infinite rate. Given concavity of h, this is equivalent to assuming that the marginal product of the first worker any skill level is bounded

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

Examples that satisfy this condition include 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 $h'(0) \to \infty$.

Further, 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. The implication of this assumption is that for any changes in prices (wages), changes in allocations are completely determined by the substitution effect. Moreover, for values of β in the neighborhood of 1, the results do not change qualitatively either because of the generic local uniqueness of equilibrium in Arrow-Debreu economies.

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 within organization diversity of skills. To see this, observe that the key implication of the bounded arrival of new solutions is that firms with different levels of TFP will have different marginal returns from hiring any given worker. This can be seen from inspection of the firm's profits where output is multiplied by A and therefore at each skill level, output is more valuable in higher A firms given identical n_i:

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

There are non-negativity constraints on n_i that will now be binding whenever the marginal product $Ah'(n_i)x_i$ is below the wage rate $w(x_i)$. This is immediately evident from the First-Order Conditions which imply:

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

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 x_i.

Given continuity of the distribution of TFP, for each skill level x_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 x_i. Thus, the critical TFP firm satisfies

$\underline{A}(x_i) = \frac{w(x_i)}{\bar{h}x_i}$. The demand of a firm A for skill x_i therefore satsifies:

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

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

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

For the case of a continuum of skills we take $m(x_i) = F(x_i) - F(x_i - \Delta)$ 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:

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

We can show that firms now will not hire on the entire support of skills, but will have a cut off rule for the highest skill a firm hires given its amount of TFP A. Call the cut-off rule x_{CEO}.

Proposition 3 *Firms with higher A have a larger labor force, and they hire more of all skill types.*

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 CES technology is a special case here and this result therefore also holds. The question remains how the skill distributions within the different firms compare.

We will denote the highest skill type x that a firm with TFP A hires by $x_{CEO}(A)$. It corresponds to the highest skill level the cutoff firm $A(x_i)$ is willing to hire, i.e., $x_{CEO}(A) = \frac{w(x_i)}{\bar{h}A}$. We can now establish the following proposition:

Proposition 4 *If $f'(x_i) < 0$, the highest skilled worker $x_{CEO}(A)$ is increasing in A and therefore in the size of the firm.*

Proof. In Appendix. ■

The next result follows immediately from the proof of Proposition 4.

Corollary 1 *Smaller firms hire from a smaller range of skills than larger firms: supp $f_{\underline{A}} \subset$ supp $f_{\overline{A}}$ for all $\underline{A} < \overline{A}$.*

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, i.e. λ is not too convex. That means that there is no sudden drop in the arrival rate followed by a constant arrival. For the 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.

Proposition 5 *There is single-crossing of the densities:* $\frac{d^2\left(\frac{n_i(A)}{n(A)}\right)}{dAdx_i} > 0$

Proof. In Appendix. ■

Then 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 6 *(Stochastic Dominance). The skill distribution of larger firms stochastically dominates that of smaller firms.*

Example. To see how different TFP firms design their organizations with different skill distributions, consider the following example with the skill distribution Pareto (on support $x > 1$ and with coefficient 1), the firm TFP distribution uniform on $[0, 1]$ and where there is 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.

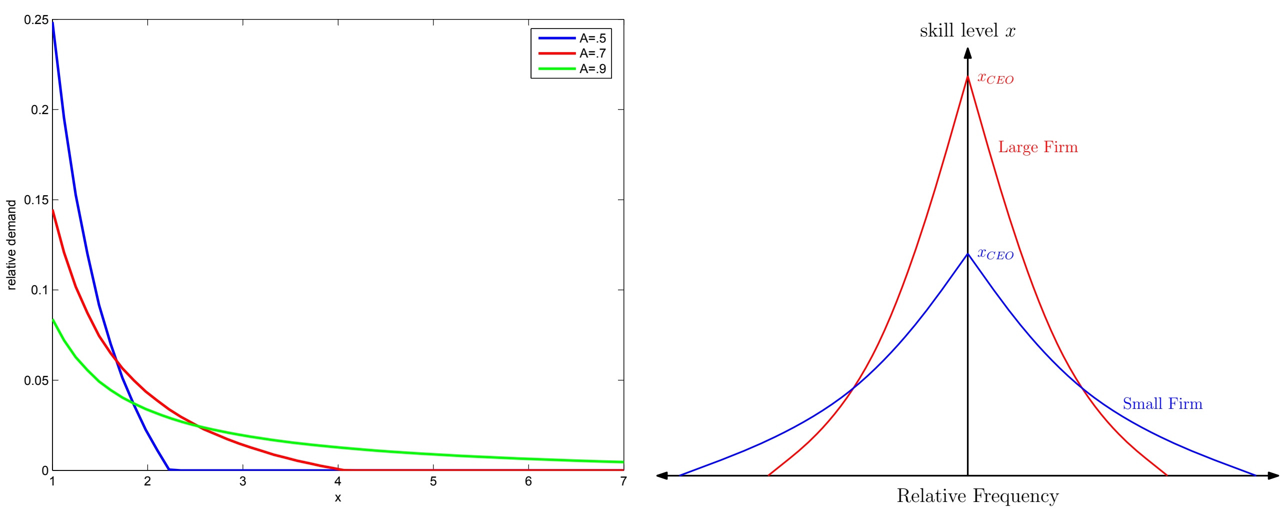

Figure 3: A. The pdf of skilled workers. – B. Organigram in different firms.

If the skill distribution is taken as the guidance for the organigram of a firm, this has immediate implications for what it will look like for firms of different sizes. Larger firm will not only hire more skilled workers for their top position of CEO, 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, i.e. equal skills earn equal wages.

Proposition 7 *(Firm Size – Wage Premium). Larger firms pay higher wages than smaller firms.*

These findings on the organizational design of firms is consistent with empirical evidence. There is ample evidence of an Employer Size – Wage Effect. Large employers pay more than small employers (early evidence goes back to Lester (1967)). Several explanations have been proposed ranging from working conditions and compensating differentials to union avoidance and product market power. Brown and Medoff (1989) find that these explain little of the size–wage effect, and that the largest share is explained by differences in the labor quality. Larger firms hire more skilled labor. This is consistent with our finding in Proposition 7. Also, the effect is sizable: around the average, a one standard deviation increase in firm size (7 workers more) leads to a $6-15\%$ increase relative to the earnings one standard deviation below (Brown and Medoff (1989)). Of course, a substantial component in the empirical debate is about unobservable heterogeneity. Here we consider the type x to represent the true type, including if it is unobservable.

Skill levels and salaries of the CEOs are higher in larger firms. Since Roberts (1956), it has repeatedly been confirmed that CEO compensation is increasing in 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 Tervio (2008)) confirm this finding and refer to it as Roberts' Law. They find an estimate for $\widehat{\kappa} \simeq 1/3$ where $w \sim S^{\kappa}$. Below in section 7 we use CEO compensation to back out the distribution of TFP across different firms.

More productive firms pay on average higher wages. For evidence, see amongst others Krueger and Summers (1988). This is consistent with the predictions of the model. Even though in this competitive market with complete information, workers of the same skill get the same wages, the first order stochastic dominance of the skill distribution of larger firms implies that the more productive firms pay higher wages on average.

4 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.

First we consider changes in the concentration of productive resources of firms. In waves of mergers and aquisitions the concentration will increase, whereas in periods of downsizing, spin-offs and out-sourcing, the concentration of TFP will decrease. We consider a simple experiment of a change in the distribution of A. In particular, we compare a given economy with distributions of TFP $G(A)$ to an otherwise identical economy with distribution $G_1(A)$ and where $G_1(\cdot)$ First-Order Stochastically Dominates (FOSD) $G(\cdot)$, i.e. $G(\cdot) > G_1(\cdot)$ for all A.

Proposition 8 *As the distribution of TFP becomes more concentrated (in the sense of FOSD) firms become smaller: the demand at a firm with constant A for each skill type n_i decreases, wages increase, and the skill type of the CEO x_{CEO} decreases.*

Proof. In 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, that also implies that the skill level of the CEO, x_{CEO}, decreases. In waves of increased mergers and aquisitions for example, the model predicts that wages of all skill levels increase and that the skill level of the CEO in a given firm A decreases.

This may provide an explanation for the fact that mergers lead to downsizing. The explanation here is market driven: mergers tend to occur in waves, thus increasing the demand for skills, pushing up wages and resulting in lower equilibrium quantities employed. In this frictionless model without unemployment, downsizing is beneficial for workers since wages are higher after the merger wave. It is precisely the wage increase that lead firms to downsize.

In addition to changes in the concentration of productive resources, an obvious change is the impact of technological change in the problem-solving production function. Over recent decades, there have been enormous advances in technology. 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 $\lambda(n; a)$ parameterized by a and assume that λ is everywhere increasing in a. The immediate implication is that the marginal productivity of every skill group increases since $h'(n; a) = \lambda(n; a)$. Because λ is increasing in a 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* $\frac{dh'(n;a)}{da} > 0$*, all wages increase.*

Proof. In 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 not obvious whether the skill level of the CEO will increase or decrease. Consistent with Gabaix and Landier (2008) though, wages of the CEOs go up unambiguously.

5 Investment in Skills: Endogenous Heterogeneity

Consider an economy with ex ante identical agents and a technology where each can choose to invest in education to obtain a level of skills x_i. The cost of education is given by

$$C(x_i) = a + c(x_i)$$

consisting of a fixed cost $a \geq 0$ 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.[8] Considering that in equilibrium all skill are supplied, we have that:

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

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:

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

where:

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

In the previous sections, equilibrium was determined by an exogenous distribution of skills $f(x_i)$ and an endogenous wage schedule. Now, $w(x_i)$ is exogenously pinned down by the cost function $C(x_i)$, while $f(x_i)$ is determined endogenously in $(\star)$.

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

[8] Alternatively it is equal to some constant

equilibrium agents will choose invest levels that are different, yet obtaining 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 10 *The equilibrium distribution of skills is always uni-modal and has a long right tail. When there is no fixed cost of investment ($a = 0$), the density is everywhere downward sloping.*

Proof. In 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 non-degenerate distribution of firms.[9]

Example. 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 $x^* = \sqrt{\frac{3}{2}}$. Notice that every TFP above a given threshold will hire a given skill, but now each TFP level has a minimum and a maximum skill thresholds.

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

We now consider the parameter values $a = \frac{1}{2}$, $c = \frac{1}{3}$ and $\gamma = 2$. The graph of $A(x_i)$ is given in Figure 4.

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:

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

Using the expression we obtained for $A(x_i)$ above, $A = \frac{a+cx_i^2}{\gamma x_i}$ and $-cx_i^2 + \gamma A x_i - a = 0$, we can get the minimum and the maximum skill thresholds for a given A: $x_{CEO}(A)$ and $x_{Janitor}(A)$. From the case in which $\Delta = 0$, we obtain the minimum company in activity, the one in which $\gamma^2 A^2 - 4ac = 0$. For our parameters $\underline{A} = \sqrt{\frac{1}{6}} \approx 0.40825$. Then, solving the equation above, we have:

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

In our example $x(\underline{A}) = \sqrt{\frac{3}{2}} = x^*$, as we should expect. Graphically, the demand for 4 different type TFP firms in this example is given in Figure 4.B.

[9] For a framework with spillovers from technology adoption and the ensuing endogenous heterogeneity of ex ante identical agents, see for example Eeckhout and Jovanovic (2002).

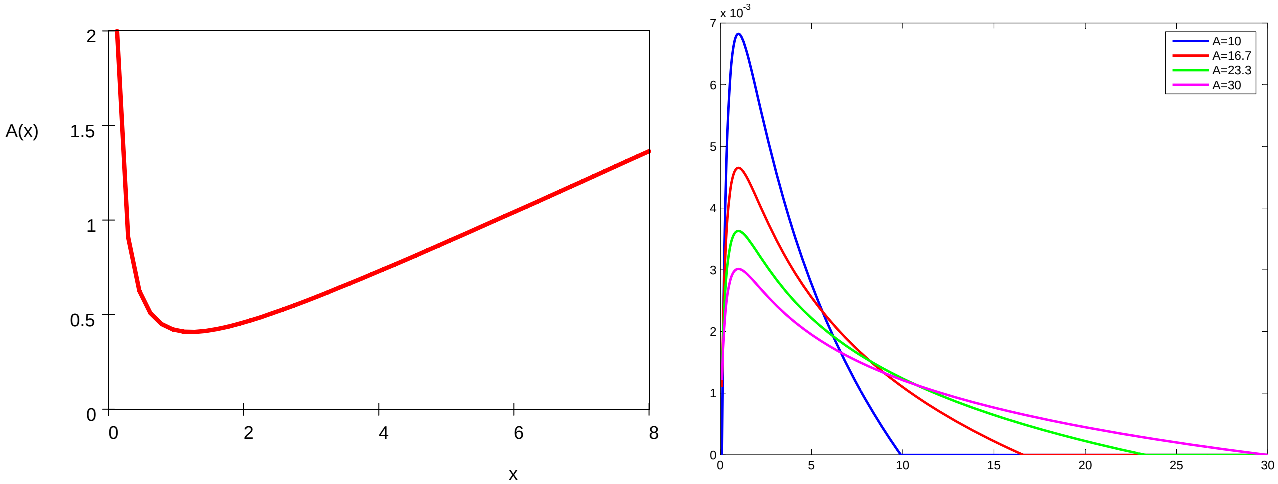

Figure 4: **A.** The equilibrium function $A(x_i)$. — **B.** The endogenous demand for skills: $n(x)$.

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

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

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

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

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

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

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:

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

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 5. The Right panel of the Figure plots the density for the same example when there is no sunk cost of investment $(a = 0)$. In that case, the density is everywhere downwards sloping.

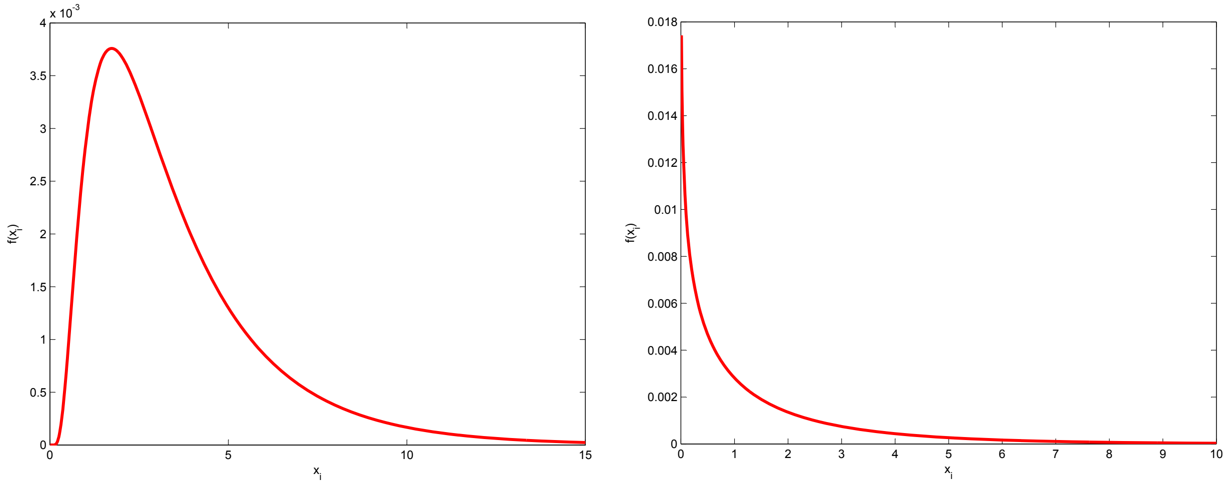

Figure 5: Investment leads to endogenous distribution of skills. The distribution is unimodal when there is a fixed cost of investment (Left) and otherwise everywhere downward sloping (Right).

6 Discussion and Extensions

6.1 Relation to Lucas' Span of Control Model

One obvious way to relax 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 non-convex. The production technology then is:

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

This formulation is remeniscent of Lucas' (1978) span of control technology of the manager (though not of the workers): a firm needs exactly on 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 $\left(-\frac{a}{b}\right)^{1/\gamma}$. There is however no maximum scale. This is illustrated in Figure 6.A.

In equilibrium, firms will now differ but due to the non-convexity 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

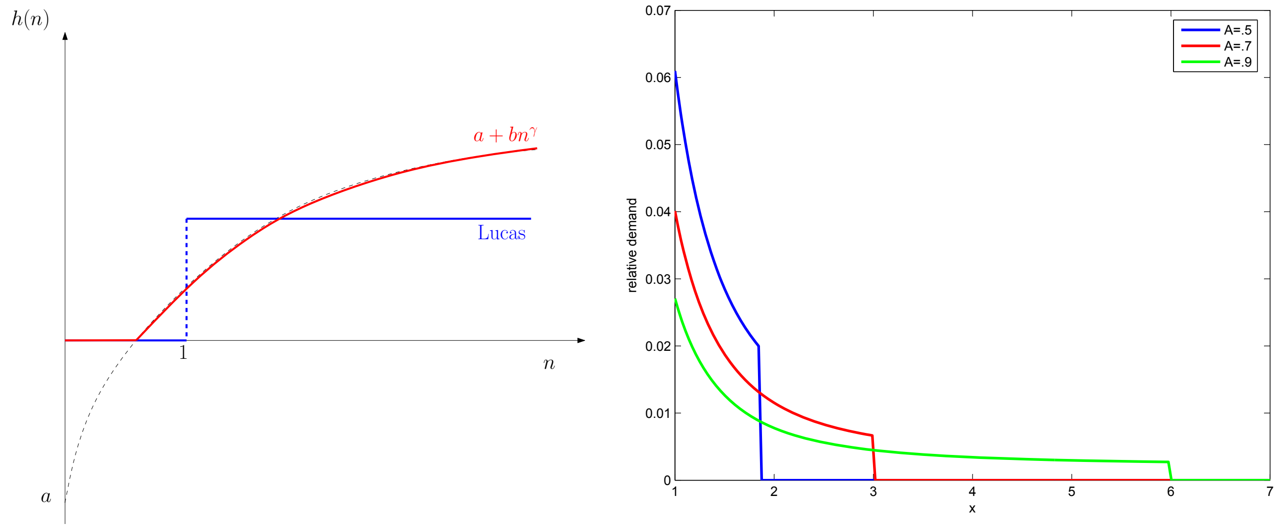

Figure 6: **A.** Span of Control: $h(n)$ versus Lucas (1978). — **B.** Skill Densities, CES and Sunk Costs.

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 $A(x_i)$ will therefore be indifferent between hiring and not hiring provided

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

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

Proposition 11 *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 non-convexity which 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 is in Figure 6.B. At the highest level, there is a sharp drop in the density due to the minimum size employed.

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:

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

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 $A(x_i) = \left\{A \in \overline{\mathcal{A}} \,|\, h(n_x(A)) = 0\right\}$. Therefore, there exists a threshold such that if $A < A(x_i)$, $\max\left\{0, a + n_x(A)^{\frac{1}{2}}\right\} = 0$. 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:

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

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

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

Claim 1 *$x_i \to \infty$ as $A(x_i) \to 1$.*

Proof. In Appendix. ■

Claim 2 *$A(1) > 0$, i.e., some firms shut down in equilibrium.*

Proof. In Appendix. ■

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

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

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

$$\begin{aligned} x &= \frac{\sqrt{3}A}{0.5\left(1-A^3\right)^{\frac{1}{2}}} \\ A &= 0.25 \end{aligned}$$

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

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

Substituting $A\left(x\right)$ and rearranging, we have:

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

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

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

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

where $A\left(x\right)$ solves (2). 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.

6.2 Decreasing Elasticity of Substitution

The fact that $h'(0) = \infty$ does nonetheless not necessarily imply that all firms are identical. While there is full support for all firms, in the absence of CES firm size distributions differ. In particular, when there is a Decreasing Elasticity of Substitution (DES), $\frac{\partial \sigma}{\partial n} < 0$, larger firms will hire more skilled workers.

Proposition 12 *Let $\sigma' < 0$. If the density of x is decreasing then:*

1. *All firms hire workers of all types (full support distributions);*
2. *Average skills and average wages are higher in larger firms than in smaller firms;*
3. *The skill and wage distribution in larger firms First-Order Stochastically dominates those in small firms.*

Proof. In Appendix. ■

Observe that the DES case is the most logical case when the $h(\cdot)$ function is bounded above. As a result, σ will eventually be decreasing in n.

7 Deriving the Aggregate Distribution of TFP across Firms

In recent years there has been an sharp increase in the attempts to measure productivity. Observing the productivity of an economy or a firm is central in understanding the firm's individual profitability and the aggregate state of the economy. Bartelsman and Doms (2000) review the empirical literature that uses longitudinal micro-level 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.

Direct and precise observation of TFP is virtually impossible, all estimation procedures build, implicitly or explicitly, on a theoretical model in the background. By construction, 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. 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.

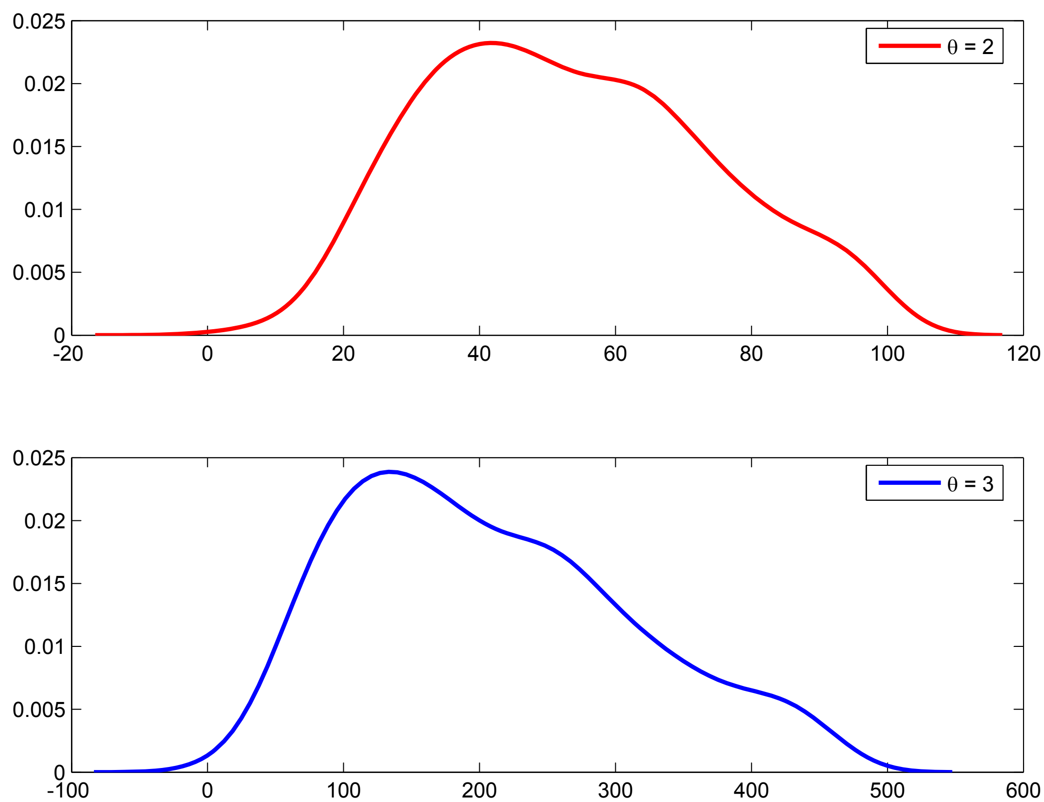

Figure 7: Implied TFP distribution (A) across different firms for different investment costs (θ).

Suppose the underlying model of the economy is ours, where organizations differ in the distribution of skills of their workers. We know then that the production technology is not CES, and we can use information on the distribution of wages 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 x_i the first-order condition holds, including at x_{CEO}. By construction, the CEO is the type for whom $n(x_{CEO}) = 0$, $h'(n)$ is evaluated at zero, which is common to all firms. This allows us to identify A from CEO characteristics only:

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

Instead of using the CEO skill level x_{CEO}, we can also use the investment decision as in section 5 above. 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

$$A = Kw(x_{CEO})^{1-1/\theta},$$

where $K = \frac{b^{1/\theta}}{h'(0)}$ is a constant.

Using Compustat Executive Compensation Data, we obtain the distribution of $w(x_{CEO})$, up to a constant K, provided we know the θ. Recall that θ measures the curvature of the investment cost. In Figure 7 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 K, and even for different θ's, what 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.

8 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 paper we embed a stylized notion of diversity in an equilibrium framework and show that within-group diversity in general induces between-group diversity. Firms will in general differ in their skill distribution and more productive firms will have skill distributions that stochastically dominate. This implies also that they will hire CEOs that are more skilled. We highlight the fact that assuming CES production technologies for firms is convenient but highly specific: CES is necessary and sufficient for having identical firms.

We have analyzed the impact of changes in the environment and have shown how mergers and acquisitions in this framework lead to downsizing. We also derived the induced TFP distribution within the economy using the compensation of CEOs. We argue that this is a tractable model for analyzing the aggregate macro implications of organizational diversity in an equilibrium framework.

9 Appendix

Definition of a non-homogeneous 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. $\Pr[\text{exactly } 1 \text{ event in } (t, t+h)] = \lambda(t) h + o(h)$.

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 non-homogeneous 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. $P(N(v) - N(u) = n) = \frac{\left(\int_u^v \lambda(t) dt\right)^n}{n!} e^{-\int_u^v \lambda(t) dt}$.

Derivation of the continuous case. We need to be careful about which assumptions we impose on $n(x)$ for writing down the continuous case. If we rewrite the model with Δs, we are using a partition/refinement argument, which delivers a Riemann integral[10]. Based on this, we must have a

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

piecewise continuous $n(x)$. Consider 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 $(\mathcal{P},\mathcal{X})$. We can show that any piecewise continuous function satisfies integrability. The continuous case can derived from taking the appropriate limit for $\Delta \to 0$

$$L(\mathbf{n},\mathbf{x}) = \left[\int h(n)xdx\right]^{\beta}$$

A special case with h CES:

$$L(\mathbf{n},\mathbf{x}) = \left[\sum_{i=1}^{N} n_i^{\gamma} x_i^{\alpha}\right]^{\beta}$$

then becomes in the continuous case:

$$L(\mathbf{n},\mathbf{x}) = \left[\int n_i^{\gamma} x_i^{\alpha} dn_i\right]^{\beta}.$$

Gross Complements and Gross Substitutes. From the firm's objective function, we derive

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

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 3 *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 is the range of parameters where inputs in production are complements, and the green area where they are substitutes.

Proof of Lemma 1

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

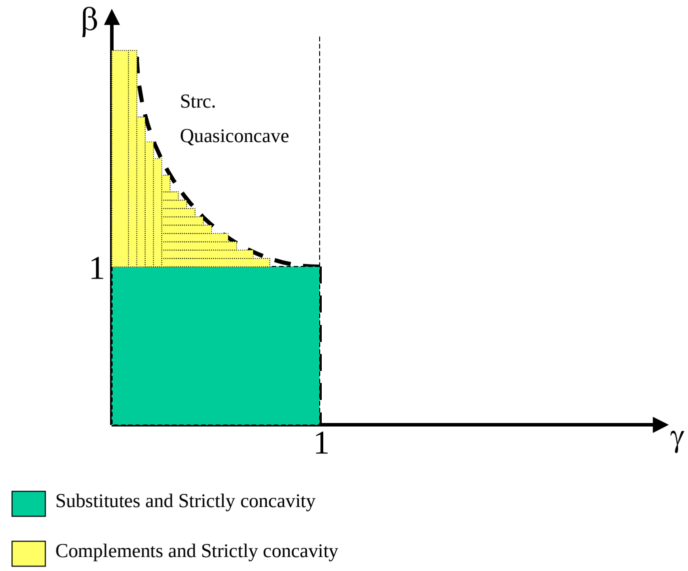

Figure 8: Complements and Substitutes

1. *The elasticity of subsitution 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. Part 1. Since σ is a constant, 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:

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

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:

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

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

$$h'(tn_i) = c h'(n_i)$$

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 bd 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:

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

■

Proof of Proposition 1.

Proposition 1 *In equilibrium all firms have the same skill distribution $F_A(x)$ equal to the economy's skill distribution $F(x)$ if and only if the production technology is CES.*

Proof. From the First-Order Condtions 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. ■

Concavity of the firm's objective function.

Lemma 2 *If $\gamma\beta < 1$, then the firm's objective function as defined generally above is strictly concave, whenever $a \geq 0$.*

Proof. Notice that:

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

Rearranging:

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

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

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

Which implies:

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

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 don't have any assumptions on a. If we consider $a \geq 0$, we notice that a sufficient condition would be $\gamma\beta < 1$ (I'm already assuming by concavity of $h(\cdot)$ that $\gamma < 1$). To get Inada conditions, we necessarily have $a = 0$. If $a < 0$, then we wouldn't have strict concavity holding for all $\mathbf{n}$.

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

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

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

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

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

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 $a \geq 0$. ■

Proof of Proposition 4.

Proposition 4 *If $f'(x_i) < 0$, the highest skilled worker $x_{CEO}(A)$ is increasing in A and therefore in the size of the firm.*

Proof. Taking the total derivative of (1) with respect to x_i, we have:

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

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

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

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

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 $\frac{dA(x_i)}{dx_i} > 0$, i.e., 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 $x_{CEO}(A)$ is increasing in A. ■

Proof of Proposition 5.

Proposition 5 *There is single-crossing of the densities:* $\frac{d^2\left(\frac{n_i(A)}{n(A)}\right)}{dAdx_i} > 0$

Proof. Observe that:

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

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

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

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

Now consider the distribution of skills. Define:

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

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

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

Then:

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

Taking the derivative with respect to x_i:

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

And the cross-derivative:

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

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

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. ■

Proof of Proposition 8

Proposition 13 *As the distribution of TFP becomes more concentrated (in the sense of FOSD) firms become smaller: the demand at a firm with constant A for each skill type n_i decreases, wages increase, and the skill type of the CEO x_{CEO} decreases.*

Proof. Recall that the market clearing condition for skill type i is:

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

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

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

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}; 3. lower demand n_i in all firms. To see the effect on demand, observe that from F.O.C.s we have:

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

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. ■

Proof of Proposition 9

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

Proof. Consider the market clearing condition

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

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

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

$$= \int_{\frac{w(x_i)}{h'(0;a)x_i}}^{\overline{A}} \frac{dh'^{-1}\left(\frac{w(x_i)}{Ax_i};a\right)}{da} dG(A) + \underbrace{h'^{-1}\left(h'(0)\right)}_{=0} g\left(\frac{w(x_i)}{h'(0;a)x_i}\right) * \frac{w(x_i)}{[h'(0;a)]^2 x_i} \frac{dh'(0;a)}{da}$$

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

In Lemma 3 we establish that h'^{-1} is increasing in a. As a result, an increase in a raises the LHS, while keeping the RHS constant. So we need to change wages to preserve the equality. Notice that:

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

Therefore, an increase in a generates a decrease in wages for all skills. From implicit function theorem, we have:

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

■

Lemma 3 *If* $h(n;a)$ *is strictly concave and twice differentiable and* $h'(n;a)$ *is monotonic in the parameter* a*, then* $h'^{-1}(n,a)$ *is also monotone in* a *in the same direction.*

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

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

Applying total derivative:

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

Rearranging:

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

Looking at the RHS 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 $h'^{-1}(n;a)$. Therefore, inversion preserves the monotonic increase or decrease in a. ■

Proof of Proposition 10.

Proposition 14 *The equilibrium distribution of skills is always uni-modal and has a long right tail. When there is no fixed cost of investment ($a = 0$), the density is everywhere downward sloping.*

Proof. Given the equilibrium condition ($\star$), the derivate of $A(x_i)$ with respect to x_i is:

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

Therefore, we have that $A'(x_i) > 0$ if and only if $w'(x_i)x_i - w(x_i) > 0$. Since $w : (0,\overline{x}) \to \mathbb{R}$ 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 $y = x_i$,[11] 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 $w'(x_i)x_i - w(x_i) > 0 \Rightarrow A'(x_i) > 0$. 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:

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

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

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

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

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

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. ■

Proof of Claim 1.

Claim 4 *$x_i \to \infty$ as $A(x_i) \to 1$.*

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

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

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

Proof of Claim 2.

Claim 5 *$A(1) > 0$, i.e., some firms shut down in equilibrium.*

Proof. From (2), we have:

$$3A(1)^2 = (-a)^2\left[1 - A(1)^3\right]$$

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

Proof of Proposition 12.

Proposition 12 *Let $\sigma' < 0$ and for any β not too large. If the density of x is decreasing then:*

1. *All firms hire workers of all types (full support distributions);*

2. *Average skills and average wages are higher in larger firms than in smaller firms;*

3. *The skill and wage distribution in larger firms First-Order Stochastically dominates those in small firms.*

Proof. The proof considers a local change around the equilibrium allocation for a two firm, two skill economy. We start from the equilibrium conditions, with endogenous variables: $n_1^1,\ n_2^1,\ n_1^2,\ n_2^2,\ w_1,\ w_2$.

$$\begin{cases} A_1\beta\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h'\left(n_1^1\right)x_1=w_1 & (1)\\ A_1\beta\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h'\left(n_2^1\right)x_2=w_2 & (2)\\ A_2\beta\left[h\left(n_1^2\right)x_1+h\left(n_2^2\right)x_2\right]^{\beta-1}h'\left(n_1^2\right)x_1=w_1 & (3)\\ A_2\beta\left[h\left(n_1^2\right)x_1+h\left(n_2^2\right)x_2\right]^{\beta-1}h'\left(n_2^2\right)x_2=w_2 & (4)\\ n_1^1+n_1^2=m\left(x_1\right) & (5)\\ n_2^1+n_2^2=m\left(x_2\right) & (6)\end{cases}$$

For the general case, when $\beta\neq 1$, we can reduce the system to:

$$\begin{cases} A_1\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h'\left(n_1^1\right)-A_2\begin{bmatrix} h\left(m\left(x_1\right)-n_1^1\right)x_1\\ +h\left(m\left(x_2\right)-n_2^1\right)x_2\end{bmatrix}^{\beta-1}h'\left(m\left(x_1\right)-n_1^1\right)=0 & (F_1)\\ A_1\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h'\left(n_2^1\right)-A_2\begin{bmatrix} h\left(m\left(x_1\right)-n_1^1\right)x_1\\ +h\left(m\left(x_2\right)-n_2^1\right)x_2\end{bmatrix}^{\beta-1}h'\left(m\left(x_2\right)-n_2^1\right)=0 & (F_2)\end{cases}$$

The main problem is that this is a non-linear non-separable system.

From $\frac{(F_1)}{(F_2)}$, we have:

$$\frac{h'\left(n_1^1\right)}{h'\left(n_2^1\right)}=\frac{h'\left(m\left(x_1\right)-n_1^1\right)}{h'\left(m\left(x_2\right)-n_2^1\right)}$$

Then, let's prepare ourselves for the IFT:

$$D_AF=\begin{bmatrix}\frac{\partial F_1}{\partial A_1} & \frac{\partial F_1}{\partial A_2}\\ \frac{\partial F_2}{\partial A_1} & \frac{\partial F_2}{\partial A_2}\end{bmatrix}$$

where:

$$\frac{\partial F_1}{\partial A_1}=\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h'\left(n_1^1\right)$$

$$\frac{\partial F_1}{\partial A_2}=-\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-1}h'\left(m\left(x_1\right)-n_1^1\right)$$

$$\frac{\partial F_2}{\partial A_1}=\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h'\left(n_2^1\right)$$

$$\frac{\partial F_2}{\partial A_2}=-\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-1}h'\left(m\left(x_2\right)-n_2^1\right)$$

And,

$$D_nF=\begin{bmatrix}\frac{\partial F_1}{\partial n_1^1} & \frac{\partial F_1}{\partial n_2^1}\\ \frac{\partial F_2}{\partial n_1^1} & \frac{\partial F_2}{\partial n_2^1}\end{bmatrix}$$

where:

$$\frac{\partial F_1}{\partial n_1^1} = A_1\left\{(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(n_1^1\right)\right]^2 x_1+\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h''\left(n_1^1\right)\right\}$$
$$-A_2\left\{\begin{array}{c} -(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(m\left(x_1\right)-n_1^1\right)\right]^2 x_1 \\ -\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-1}h''\left(m\left(x_1\right)-n_1^1\right) \end{array}\right\}$$

$$\frac{\partial F_1}{\partial n_2^1} = \left\{\begin{array}{c} A_1(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}h'\left(n_1^1\right)h'\left(n_2^1\right)x_2+ \\ A_2(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2}h'\left(m\left(x_1\right)-n_1^1\right)h'\left(m\left(x_2\right)-n_2^1\right)x_2 \end{array}\right\}$$

$$\frac{\partial F_2}{\partial n_1^1} = \left\{\begin{array}{c} A_1(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}h'\left(n_2^1\right)h'\left(n_1^1\right)x_1+ \\ A_2(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2} \\ *h'\left(m\left(x_2\right)-n_2^1\right)h'\left(m\left(x_1\right)-n_1^1\right)x_1 \end{array}\right\}$$

$$\frac{\partial F_2}{\partial n_2^1} = A_1\left\{(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(n_2^1\right)\right]^2 x_2+\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h''\left(n_2^1\right)\right\}$$
$$-A_2\left\{\begin{array}{c} -(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(m\left(x_2\right)-n_2^1\right)\right]^2 x_2 \\ -\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-1}h''\left(m\left(x_2\right)-n_2^1\right) \end{array}\right\}$$

Then, we have:

$$\det D_n F = \frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial n_2^1} - \frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial n_2^1}$$

So:

$$\frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial n_2^1} =$$
$$\left[\begin{array}{c} A_1\left\{(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(n_1^1\right)\right]^2 x_1+\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h''\left(n_1^1\right)\right\} \\ -A_2\left\{\begin{array}{c} -(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(m\left(x_1\right)-n_1^1\right)\right]^2 x_1 \\ -\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-1}h''\left(m\left(x_1\right)-n_1^1\right) \end{array}\right\} \end{array}\right]$$
$$*\left[\begin{array}{c} A_1\left\{(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(n_2^1\right)\right]^2 x_2+\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-1}h''\left(n_2^1\right)\right\} \\ -A_2\left\{\begin{array}{c} -(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(m\left(x_2\right)-n_2^1\right)\right]^2 x_2 \\ -\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-1}h''\left(m\left(x_2\right)-n_2^1\right) \end{array}\right\} \end{array}\right]$$

Rearranging:

$$\frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial n_2^1} =$$

$$\left[\begin{array}{c} A_1\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}\left\{(\beta-1)\left[h'\left(n_1^1\right)\right]^2 x_1+\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]h''\left(n_1^1\right)\right\} \\ -A_2\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2} * \\ \left\{\begin{array}{c} -(\beta-1)\left[h'\left(m\left(x_1\right)-n_1^1\right)\right]^2 x_1 \\ -\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]h''\left(m\left(x_1\right)-n_1^1\right) \end{array}\right\} \end{array}\right]$$

$$* \left[\begin{array}{c} A_1\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}\left\{(\beta-1)\left[h'\left(n_2^1\right)\right]^2 x_2+\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]h''\left(n_2^1\right)\right\} \\ -A_2\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2} * \\ \left\{\begin{array}{c} -(\beta-1)\left[h'\left(m\left(x_2\right)-n_2^1\right)\right]^2 x_2 \\ -\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]h''\left(m\left(x_2\right)-n_2^1\right) \end{array}\right\} \end{array}\right]$$

and

$$\frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial n_2^1} = \left\{\begin{array}{c} A_1(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}h'\left(n_1^1\right)h'\left(n_2^1\right)x_2+ \\ A_2(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2} \\ *h'\left(m\left(x_1\right)-n_1^1\right)h'\left(m\left(x_2\right)-n_2^1\right)x_2 \end{array}\right\}$$

$$* \left\{\begin{array}{c} A_1(\beta-1)\left[h\left(n_1^1\right)x_1+h\left(n_2^1\right)x_2\right]^{\beta-2}h'\left(n_2^1\right)h'\left(n_1^1\right)x_1+ \\ A_2(\beta-1)\left[h\left(m\left(x_1\right)-n_1^1\right)x_1+h\left(m\left(x_2\right)-n_2^1\right)x_2\right]^{\beta-2} \\ *h'\left(m\left(x_2\right)-n_2^1\right)h'\left(m\left(x_1\right)-n_1^1\right)x_1 \end{array}\right\}$$

Now, consider the symmetric equilibrium in which $A_1 = A_2$, $n_1^1 = \frac{m(x_1)}{2}$ and $n_2^1 = \frac{m(x_2)}{2}$. Then, we have:

$$\left.\frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial n_2^1}\right|_{A_1=A_2=A} =$$

$$\left(2A\left[h\left(\frac{m(x_1)}{2}\right)x_1+h\left(\frac{m(x_2)}{2}\right)x_2\right]^{\beta-2}\right)^2\left\{\begin{array}{c} (\beta-1)\left[h'\left(\frac{m(x_1)}{2}\right)\right]^2 x_1 \\ +\left[h\left(\frac{m(x_1)}{2}\right)x_1+h\left(\frac{m(x_2)}{2}\right)x_2\right]h''\left(\frac{m(x_1)}{2}\right) \end{array}\right\}$$

$$* \left[\left\{\begin{array}{c} (\beta-1)\left[h'\left(\frac{m(x_2)}{2}\right)\right]^2 x_2 \\ +\left[h\left(\frac{m(x_1)}{2}\right)x_1+h\left(\frac{m(x_2)}{2}\right)x_2\right]h''\left(\frac{m(x_2)}{2}\right) \end{array}\right\}\right]$$

and

$$\frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial n_2^1} = \left[2A(\beta-1)\left[h\left(\frac{m(x_1)}{2}\right)x_1+h\left(\frac{m(x_2)}{2}\right)x_2\right]^{\beta-2}h'\left(\frac{m(x_1)}{2}\right)h'\left(\frac{m(x_2)}{2}\right)\right]^2 x_2 x_1$$

Then, $\det D_n F$ becomes:

$$\det D_n F = 4A^2\left[h\left(\frac{m(x_1)}{2}\right)x_1+h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-4} *$$

$$\left\{\begin{array}{c} (\beta-1)\left[h''\left(\frac{m(x_2)}{2}\right)\left[h'\left(\frac{m(x_1)}{2}\right)\right]^2 x_1+h''\left(\frac{m(x_1)}{2}\right)\left[h'\left(\frac{m(x_2)}{2}\right)\right]^2 x_2\right] \\ +\left[h\left(\frac{m(x_1)}{2}\right)x_1+h\left(\frac{m(x_2)}{2}\right)x_2\right]^2 h''\left(\frac{m(x_1)}{2}\right)h''\left(\frac{m(x_2)}{2}\right) \end{array}\right\}$$

If $\beta < 1$, this is necessarily different than zero. Otherwise, this could be zero but the set of parameters in which this occurs has mean zero. Then:

$$D_n^{-1}F = \frac{1}{|\det D_n F|}\begin{bmatrix} \frac{\partial F_2}{\partial n_2^1} & -\frac{\partial F_1}{\partial n_2^1} \\ -\frac{\partial F_2}{\partial n_1^1} & \frac{\partial F_1}{\partial n_1^1} \end{bmatrix}$$

Then:

$$\begin{bmatrix} \frac{\partial n_1^1}{\partial A_1} & \frac{\partial n_1^1}{\partial A_2} \\ \frac{\partial n_2^1}{\partial A_1} & \frac{\partial n_2^1}{\partial A_2} \end{bmatrix} = -D_n^{-1}F * D_A F$$

Substituting, we have:

$$\begin{bmatrix} \frac{\partial n_1^1}{\partial A_1} & \frac{\partial n_1^1}{\partial A_2} \\ \frac{\partial n_2^1}{\partial A_1} & \frac{\partial n_2^1}{\partial A_2} \end{bmatrix} = -\frac{1}{|\det D_n F|}\begin{bmatrix} \frac{\partial F_2}{\partial n_2^1} & -\frac{\partial F_1}{\partial n_2^1} \\ -\frac{\partial F_2}{\partial n_1^1} & \frac{\partial F_1}{\partial n_1^1} \end{bmatrix} * \begin{bmatrix} \frac{\partial F_1}{\partial A_1} & \frac{\partial F_1}{\partial A_2} \\ \frac{\partial F_2}{\partial A_1} & \frac{\partial F_2}{\partial A_2} \end{bmatrix}$$

Then:

$$\frac{\partial n_1^1}{\partial A_1} = -\frac{1}{|\det D_n F|}\left(\frac{\partial F_2}{\partial n_2^1} * \frac{\partial F_1}{\partial A_1} - \frac{\partial F_1}{\partial n_2^1} * \frac{\partial F_2}{\partial A_1}\right)$$

Then:

$$\frac{\partial F_2}{\partial n_2^1} * \frac{\partial F_1}{\partial A_1} =$$

$$\begin{bmatrix} A_1\left\{(\beta-1)\left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(n_2^1\right)\right]^2 x_2 + \left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-1} h''\left(n_2^1\right)\right\} \\ -A_2\left\{\begin{matrix} -(\beta-1)\left[h\left(m(x_1) - n_1^1\right)x_1 + h\left(m(x_2) - n_2^1\right)x_2\right]^{\beta-2}\left[h'\left(m(x_2) - n_2^1\right)\right]^2 x_2 \\ -\left[h\left(m(x_1) - n_1^1\right)x_1 + h\left(m(x_2) - n_2^1\right)x_2\right]^{\beta-1} h''\left(m(x_2) - n_2^1\right) \end{matrix}\right\} \end{bmatrix}$$
$$*\left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-1} h'\left(n_1^1\right)$$

at $A_1 = A_2 = A$ and symmetric equilibrium, we have:

$$\left.\frac{\partial F_2}{\partial n_2^1} * \frac{\partial F_1}{\partial A_1}\right|_{A_1=A_2=A} =$$

$$2A\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-3}\begin{bmatrix} (\beta-1)\left[h'\left(\frac{m(x_2)}{2}\right)\right]^2 x_2 + \\ \left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right] h''\left(\frac{m(x_2)}{2}\right) \end{bmatrix} h'\left(\frac{m(x_1)}{2}\right)$$

and

$$\frac{\partial F_1}{\partial n_2^1} * \frac{\partial F_2}{\partial A_1} =$$

$$\left\{\begin{matrix} A_1(\beta-1)\left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-2} h'\left(n_1^1\right)h'\left(n_2^1\right)x_2 + \\ A_2(\beta-1)\left[h\left(m(x_1) - n_1^1\right)x_1 + h\left(m(x_2) - n_2^1\right)x_2\right]^{\beta-2} \\ *h'\left(m(x_1) - n_1^1\right)h'\left(m(x_2) - n_2^1\right)x_2 \end{matrix}\right\}$$
$$*\left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-1} h'\left(n_2^1\right)$$

again, at $A_1 = A_2 = A$, we have:

$$\left.\frac{\partial F_1}{\partial n_2^1} * \frac{\partial F_2}{\partial A_1}\right|_{A_1=A_2=A} =$$

$$2A(\beta-1)\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-3} h'\left(\frac{m(x_2)}{2}\right)^2 h'\left(\frac{m(x_1)}{2}\right)x_2$$

Putting everything together at $A_1 = A_2 = A$, we have:

$$\left.\frac{\partial F_2}{\partial n_2^1} * \frac{\partial F_1}{\partial A_1}\right|_{A_1=A_2=A} - \left.\frac{\partial F_1}{\partial n_2^1} * \frac{\partial F_2}{\partial A_1}\right|_{A_1=A_2=A} =$$

$$2A\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-3} \left[\begin{array}{c} (\beta-1)\left[h'\left(\frac{m(x_2)}{2}\right)\right]^2 x_2 + \\ \left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]h''\left(\frac{m(x_2)}{2}\right) \end{array}\right] h'\left(\frac{m(x_1)}{2}\right)$$

$$-2A(\beta-1)\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-3} h'\left(\frac{m(x_2)}{2}\right)^2 h'\left(\frac{m(x_1)}{2}\right)x_2$$

$$= 2A\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-2} h''\left(\frac{m(x_2)}{2}\right)h'\left(\frac{m(x_1)}{2}\right)$$

Therefore:

$$\left.\frac{\partial n_1^1}{\partial A_1}\right|_{A_1=A_2=A} = -\frac{\left\{2A\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-2} h''\left(\frac{m(x_2)}{2}\right)h'\left(\frac{m(x_1)}{2}\right)\right\}}{|\det D_n F|} > 0$$

Now, let's calculate $\frac{\partial n_2^1}{\partial A_1}$. Then, we have:

$$\left.\frac{\partial n_2^1}{\partial A_1}\right|_{A_1=A_2=A} = -\frac{1}{|\det D_n F|}\left(\frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial A_1} - \frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial A_1}\right)$$

Then, let's substitute this step by step:

$$\frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial A_1} =$$

$$\left[\begin{array}{c} A_1\left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-2}\left\{(\beta-1)\left[h'\left(n_1^1\right)\right]^2 x_1 + \left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]h''\left(n_1^1\right)\right\} \\ -A_2\left[h\left(m(x_1)-n_1^1\right)x_1 + h\left(m(x_2)-n_2^1\right)x_2\right]^{\beta-2} * \\ \left\{\begin{array}{c} -(\beta-1)\left[h'\left(m(x_1)-n_1^1\right)\right]^2 x_1 \\ -\left[h\left(m(x_1)-n_1^1\right)x_1 + h\left(m(x_2)-n_2^1\right)x_2\right]h''\left(m(x_1)-n_1^1\right) \end{array}\right\} \end{array}\right]$$

$$* \left[h\left(n_1^1\right)x_1 + h\left(n_2^1\right)x_2\right]^{\beta-1} h'\left(n_2^1\right)$$

at $A_1 = A_2 = A$, we have:

$$\left.\frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial A_1}\right|_{A_1=A_2=A} =$$

$$2A\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(\frac{m(x_2)}{2}\right)x_2\right]^{2\beta-3}\left[\left\{\begin{array}{c} (\beta-1)\left[h'\left(\frac{m(x_1)}{2}\right)\right]^2 x_1 \\ +\left[h\left(\frac{m(x_1)}{2}\right)x_1 + h\left(n\frac{m(x_2)}{2}\right)x_2\right]h''\left(\frac{m(x_1)}{2}\right) \end{array}\right\}\right] h'\left(\frac{m(x_2)}{2}\right)$$

and

$$\frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial A_1} =$$

$$\left\{ \begin{array}{c} A_1 (\beta - 1) \left[h\left(n_1^1\right) x_1 + h\left(n_2^1\right) x_2 \right]^{\beta - 2} h'\left(n_2^1\right) h'\left(n_1^1\right) x_1 + \\ A_2 (\beta - 1) \left[h\left(m(x_1) - n_1^1\right) x_1 + h\left(m(x_2) - n_2^1\right) x_2 \right]^{\beta - 2} \\ * h'\left(m(x_2) - n_2^1\right) h'\left(m(x_1) - n_1^1\right) x_1 \end{array} \right\}$$

$$* \left[h\left(n_1^1\right) x_1 + h\left(n_2^1\right) x_2 \right]^{\beta - 1} h'\left(n_1^1\right)$$

Then, at $A_1 = A_2 = A$, we have:

$$\left. \frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial A_1} \right|_{A_1 = A_2 = A} =$$

$$2A(\beta - 1) \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(\frac{m(x_2)}{2}\right) x_2 \right]^{2\beta - 3} h'\left(\frac{m(x_2)}{2}\right) \left[h'\left(\frac{m(x_1)}{2}\right) \right]^2 x_1$$

Then, we have:

$$\left. \frac{\partial F_1}{\partial n_1^1} * \frac{\partial F_2}{\partial A_1} \right|_{A_1 = A_2 = A} - \left. \frac{\partial F_2}{\partial n_1^1} * \frac{\partial F_1}{\partial A_1} \right|_{A_1 = A_2 = A} =$$

$$2A \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(\frac{m(x_2)}{2}\right) x_2 \right]^{2\beta - 3} \left[\left\{ \begin{array}{c} (\beta - 1) \left[h'\left(\frac{m(x_1)}{2}\right) \right]^2 x_1 \\ + \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(n\frac{m(x_2)}{2}\right) x_2 \right] h''\left(\frac{m(x_1)}{2}\right) \end{array} \right\} \right] h'\left(\frac{m(x_2)}{2}\right)$$

$$-2A(\beta - 1) \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(\frac{m(x_2)}{2}\right) x_2 \right]^{2\beta - 3} h'\left(\frac{m(x_2)}{2}\right) \left[h'\left(\frac{m(x_1)}{2}\right) \right]^2 x_1$$

$$= 2A \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(\frac{m(x_2)}{2}\right) x_2 \right]^{2\beta - 2} h''\left(\frac{m(x_1)}{2}\right) h'\left(\frac{m(x_2)}{2}\right)$$

Then:

$$\frac{\partial n_2^1}{\partial A_1} = - \frac{\left\{ 2A \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(\frac{m(x_2)}{2}\right) x_2 \right]^{2\beta - 2} h''\left(\frac{m(x_1)}{2}\right) h'\left(\frac{m(x_2)}{2}\right) \right\}}{|\det D_n F|} > 0$$

Then:

$$\frac{\partial \left(\frac{n_1^1}{n_2^1}\right)}{\partial A_1} = \frac{\frac{\partial n_1^1}{\partial A_1} * n_2^1 - \frac{\partial n_2^1}{\partial A_1} * n_1^1}{\left(n_2^1\right)^2}$$

$$\left. \frac{\partial \left(\frac{n_1^1}{n_2^1}\right)}{\partial A_1} \right|_{A_1 = A_2} = \frac{2A \left[h\left(\frac{m(x_1)}{2}\right) x_1 + h\left(\frac{m(x_2)}{2}\right) x_2 \right]^{2\beta - 2}}{|\det D_n F| \left(\frac{m(x_2)}{2}\right)^2} \left[\begin{array}{c} -h''\left(\frac{m(x_2)}{2}\right) h'\left(\frac{m(x_1)}{2}\right) \frac{m(x_2)}{2} \\ +h''\left(\frac{m(x_1)}{2}\right) h'\left(\frac{m(x_2)}{2}\right) \frac{m(x_1)}{2} \end{array} \right]$$

Therefore, we have:

$$\left. \frac{\partial \left(\frac{n_1^1}{n_2^1}\right)}{\partial A_1} \right|_{A_1 = A_2} > 0 \text{ if}$$

$$-h''\left(\frac{m(x_2)}{2}\right) h'\left(\frac{m(x_1)}{2}\right) \frac{m(x_2)}{2} > -h''\left(\frac{m(x_1)}{2}\right) h'\left(\frac{m(x_2)}{2}\right) \frac{m(x_1)}{2}$$

Rearranging, we have:

$$-\frac{h''\left(\frac{m(x_2)}{2}\right)}{h'\left(\frac{m(x_2)}{2}\right)}\frac{m\left(x_2\right)}{2}>-\frac{h''\left(\frac{m(x_1)}{2}\right)}{h'\left(\frac{m(x_1)}{2}\right)}\frac{m\left(x_1\right)}{2}$$

■

References

Antràs, Pol, Luis Garicano and Esteban Rossi-Hansberg, "Offshoring in a Knowledge Economy", *Quarterly Journal of Economics* **121(1)**, 2006, 31-77.

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

Brown, Charles, and James Medoff, "The Employer Size–Wage Effect", *Journal of Political Economy* **97(5)**, 1989, 1027-1059.

Crawford, Vincent, and Alexander Kelso, "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica* **50**, 1982.

Eeckhout, Jan, and Boyan Jovanovic, "Occupational Sorting and Development", NBER working paper **w13686**, 2007.

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

Gabaix, Xavier, and Augustin Landier "Why has CEO Pay Increased so Much?", *Quarterly Journal of Economics*, forthcoming 2008.

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

Hong, Lu, and Scott Page, "Problem Solving by Heterogeneous Agents," *Journal of Economic Theory* **97**, 2001, 123-163.

Hong, Lu, and Scott Page, "Groups of diverse problem solvers can outperform groups of high-ability problem solvers," *Proceedings of the National Academy of Sciences*, November 16, 2004, 16385-16389.

Krueger, Alan, and Lawrence Summers, "Efficiency Wages and the Inter-Industry Wage Structure" *Econometrica* **56(2)**, 1988, 259-293.

Lester, Richard A., "Pay Differentials by Size of Establishment", *Industrial Relations* **7**, 1967, 57-67.

Lucas, Robert, "On the Size Distribution of Business Firms." *Bell Journal* 1978.

Luttmer, Erzo, "Selection, Growth, and the Size Distribution of Firms," *Quarterly Journal of Economics* **122(3)**, 2007, 1103-1144.

Ostroy, Joseph M., and William R. Zame, "Nonatomic Economies and the Boundaries of Perfect Competition," *Econometrica* **62(3)**, 1994, 593-633.

Page, Scott E., *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton University Press, 2007.

Rajan, Raghuram, and Julie Wulf, "The Flattening Firm: Evidence from Panel Data on the Changing Nature of Corporate Hierarchies," *Review of Economics and Statistics*, **88(4)**, 2006, 759-773.

Rossi-Hansberg, Esteban, and Mark Wright, "Establishment Size Dynamics in the Aggregate Economy", *American Economic Review* **97(5)**, 2007, 1639-1666.

Sah, Raaj Kumar, and Joseph E. Stiglitz, "The Architecture of Economic Systems: Hierarchies and Polyarchies,"*American Economic Review* **76(4)**, 1986, 716-727.

Tervio, Marko, "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review* **98(3)**, 2008, 642-668.