Zipf's Law for Firms: A Static and Microfounded Theory*

François Geerolf †
Toulouse School of Economics

VERY PRELIMINARY AND INCOMPLETE
[Last Version]

This version: July 2014
First version: April 2014

Abstract

A model of knowledge-based production hierarchies based on Garicano (2000) generates Zipf's Law for firm sizes in the upper tail, with very minimal assumptions on the underlying distribution of agents' skills. When the density distribution of higher skills is bounded away from zero, the span of control of managers follows a Pareto distribution of endogenous tail coefficient equal to two in the upper tail between adjacent levels of hierarchical organization, four thirds between two of these levels, and in general $2^L/(2^L - 1)$ between L levels, converging to one, or Zipf's law, when the number of levels increases. These ancillary predictions for span of control of intermediary managers are verified in the French matched employer-employee data, with a very good degree of precision. The model attributes firms' size unbounded heterogeneity as well as high labor income inequality to a potentially arbitrarily small heterogeneity in agents' skills, without any functional form assumption. The microfoundations of this model suggest that firm sizes' distributions have no direct implications for the productivity distribution of firms, and a well-defined upper tail behavior for the labor income distribution consistent with preliminary evidence.

Keywords: Zipf's Law, Pareto distributions, Hierarchies
JEL classification: D2, L2

*I thank Ariel Burstein, Jacques Crémer, Jan Eeckhout, Xavier Gabaix, Christian Hellwig, Martí Mestieri, François de Soyres, Robert Ulbricht and seminar participants at Toulouse School of Economics for useful comments. I am especially indebted to Maxime Liegey for presenting the data on French team and plant sizes during a student workshop at Toulouse School of Economics.

†E-mail: francois.geerolf@polytechnique.org. This is very preliminary, all comments are welcomed. In particular, due to short-run data constraints, the preliminary evidence consists of a patchwork of different sources. An Online Appendix to this paper is available here.

Introduction

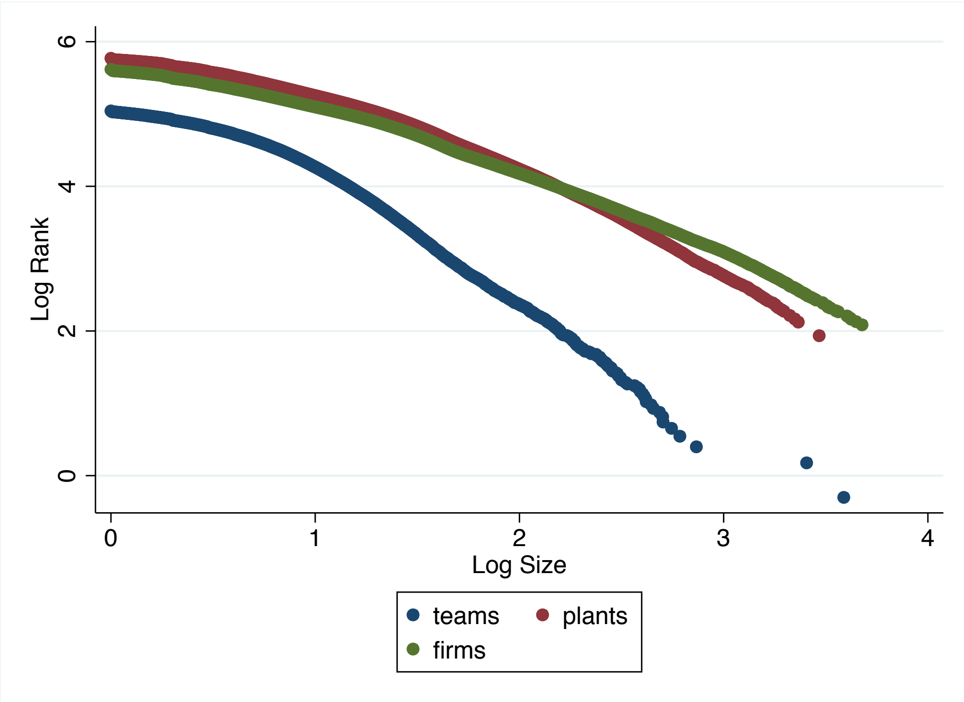
It is well known that firm size's distribution approximately follows Zipf's law in the upper tail. Models explaining this Pareto distribution all at some level rely on a stochastic random growth process with statistical frictions, following Gibrat (1931)'s law, or on some other Pareto distribution underlying the economy's primitives, whether it be entrepreneurial skill or firm's productivity. In this paper, I propose a static mechanism, able to microfound the existence of Zipf's law, based on a static production hierarchies model a la Garicano (2000), and which makes no functional form on the underlying distribution of primitives, in this model agents' skills. The model makes ancillary predictions about the distribution of span of controls between intermediary levels of management, which are borne out by the French data, as Figure 1 below shows. In theory, the span of control of managers over employees down one level of hierarchical organization should follow a Pareto distribution in the upper tail with a coefficient equal to two, $4/3 \approx 1.33$ down two levels of management, $8/7 \approx 1.14$ down three levels, getting closer to a Pareto with coefficient one, or Zipf's law, as the number of levels of management increases. In the French data, the point estimates for the corresponding numbers are 1.96, 1.30 and 1.14 respectively.

It is also well known that Zipf's law holds only approximately in the data. In particular, the coefficient on the Pareto is slightly higher than one, the Pareto holds only for the upper tail of the distribution, while the lognormal is a better fit for the bulk of the distribution. And very big firms are too small relative to what Zipf's law would suggest, which is usually interpreted as a consequence of antitrust regulation, in the random growth literature.[1] Furthermore, Gibrat's law does not hold so well in the data, as the variance of growth rates decreases with size. Moreover, establishment size also is Pareto distributed in the upper tail with a higher coefficient equal to approximately 1.30, not just in the French data but also in the US Data (Figure 2), something the random growth model cannot really account for, unless statistical frictions (the height of the "reflecting barrier") are higher for establishments than for firms. In contrast to random growth models, the static hierarchies model I will present can make sense of all these empirical facts.

Microfounding Zipf's law is not just important for its own sake but because many issues crucially rely on understanding why firm sizes are so heterogeneous in the economy. Trade models often assume that productivities are distributed Pareto, while span of control models sometimes assume that managers' talents are. The fact that heterogeneity in productivities across firms is very large is crucial to finding large effects of reallocation across firms after trade openness as in Melitz (2003) type models, or large potential costs of misallocation of capital across firms and of size dependent regulations. This paper, in contrast, would suggest that the heterogeneity underlying the economy can very well be bounded, and even infinitesimal, and nonetheless generate Zipf's law for firms in the upper tail. Moreover, if the reason why Zipf's law holds in the data is static, then the dynamics of firm growth need no longer be constrained by the overall

[1] Note that one cannot see this on the graph presented below for confidentiality reasons: the biggest of the firms are not shown.

Figure 1: French Production Hierarchies



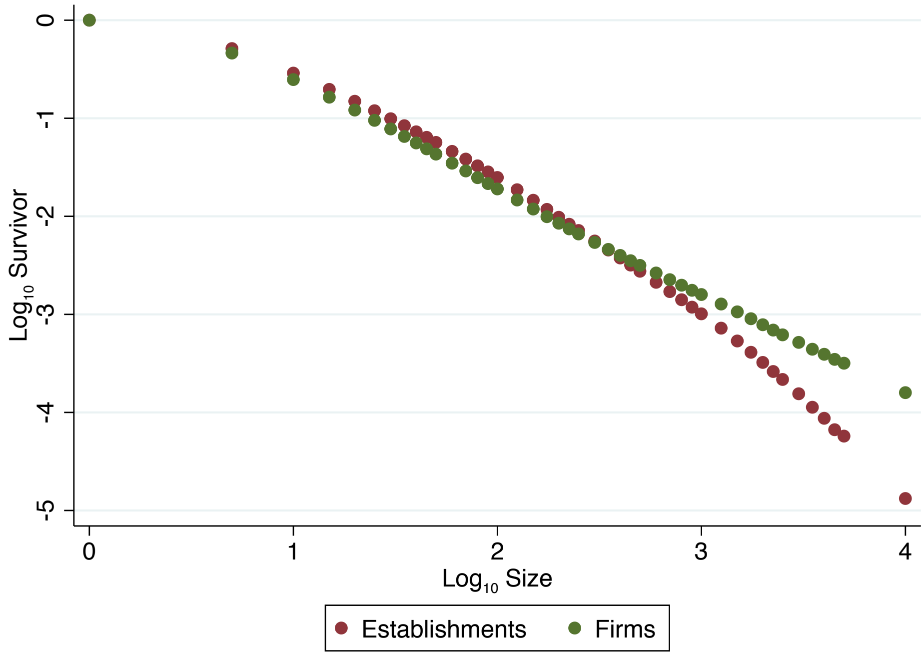
Pareto coeff., **DADS Data**: 1 level: $\boxed{\mathbf{1.96}}$ 2 levels: $\boxed{\mathbf{1.30}}$ 3 levels: $\boxed{\mathbf{1.14}}$

Pareto coeff., **Theory**: 1 level: $\boxed{\mathbf{2}}$ 2 levels: $\boxed{\mathbf{4/3}}$ 3 levels: $\boxed{\mathbf{8/7}}$

Note: Source: French matched employer-employee Data (DADS) reproduced in Liegey (2014). "Size" is defined in terms of employment. The theoretical Pareto coefficients are given by $2^L/(2^L-1)$ for span of control between L levels of hierarchical organization, as section 3 will show. "teams" ($L=1$), "plants" ($L=2$), and "firms" ($L=3$) follow Caliendo et al. (2014)'s terminology.

distribution of firm sizes we observe. In particular, the model can reconcile the deviations from Gibrat's law that economists have long been observing in the data with the overall firm size distribution. Furthermore, the mechanism put forward in this model can generate a very large amount of labor income inequality, even resulting from an initially very homogenous skill distribution. This is because, to paraphrase Rosen (1982), slight improvements in upper level decisions have an enormous influence as a whole by affecting the productivity of lesser ranking workers, as in the competitive equilibrium even slightly more talented individuals command many more employees. This contrasts greatly with existing theories of inequalities, which attribute Pareto distributions in incomes to very heterogeneous skills, or equivalently to an unlikely very long sequence of good idiosynchratic shocks to human capital at the individual level, following the random growth tradition applied to income by Champernowne (1953). The contribution of this paper to the "superstars" literature is to derive the distribution of firm sizes and of top managers' income in the upper tail jointly from a general underlying talent distribution; it shows for example that decreasing communication costs lead both to an increase in the size of firms and to growing labor income inequality. Finally, understanding what determines firm sizes is important

Figure 2: US Firm and Establishment Sizes



Pareto coeff., **Census Data**:	**Establishments:** $\boxed{\mathbf{1.33}}$	**Firms:** $\boxed{\mathbf{1.01}}$
Pareto coeff., **Theory**:	2 **levels:** $\boxed{\mathbf{4/3}}$	7 **levels:** $\boxed{\mathbf{128/127}}$

Note: Source: Census Bureau, Statistics of US Businesses, 1990. I show the cross-sectional distribution of US firm and establishment sizes in 1990, but the Pareto coefficients do not vary across years. In the US case, I do not have data on hierarchies directly, but the model allows to interpret the size distribution of establishments as characteristic of two levels of hierarchy ($4/3 \approx 1.33$). The number of hierarchical organization is given as a function of the measured Pareto coefficient α by: $\log_2\left(\frac{\alpha}{\alpha-1}\right) \approx 7$.

for industrial organization theory at large and for antitrust policy in particular. The boundaries of the firm in this model are not determined by economies of scale or of scope or otherwise, but by the indivisibility of a top manager's time. The model and its empirical success seem to lend support to a knowledge-based hierarchy view of the firm, with labor markets organized internally.

The static model which forms the basis of this paper and generates a Zipf's law for firm sizes is a well-known Garicano (2000) production hierarchies model, where time can be used for production as well as for communicating one's knowledge. In such a model, production workers encounter problems when producing. When they do not know the solution to this problem, they can potentially ask someone more knowledgeable than them for a solution. If the time of communicating knowledge, called "helping time", is lower than the time of producing by oneself, higher production can sometimes be achieved by having more knowledgeable agents specialize in communicating solutions to problems. Because workers ask these agents only when they do not know the solution themselves, these agents are called managers. The theoretical contribution of this paper is to show that in general, the distribution of span of control for managers over agents down L levels is a Pareto distribution of coefficient $2^L/(2^L-1)$ in the upper tail, under minimal assumptions for the distribution of problems as well as the distribution of workers' skills. Hence, Zipf's law obtains in the upper tail with a sufficiently large number of hierarchical levels. While Garicano (2000) and subsequent papers like Garicano and Rossi-Hansberg (2006) have looked mostly at the impact of organization and decreases in costs of communication on wage inequality, in models either with learning or with an exogenouly low number of hierarchical organization, I show that taking the distribution of skills as exogenous as well as allowing for multiple layers of management naturally gives rise to the Pareto properties described above. Incidentally, the distribution of span of controls across firms will also have some implications for the upper tail of the wage distribution, albeit in a less direct way than in Gabaix and Landier (2008). In particular, because the level of complementarities is endogenous and varying across agents in a Garicano (2000) model, the resulting wage distribution will not necessarily be Pareto, even though it will potentially be very skewed.[2]

The intuition for why a Pareto distribution for span of control is obtained in the upper tail is that the most knowledgeable managers work with the most knowledgeable workers in the competitive equilibrium of this model. This is because of a complementarity between workers' and managers' skills: at an optimal allocation, the most skilled managers must be shielded from answering easy questions, as they can make a more productive use of their time. To the limit, when heterogeneity goes to zero, or when the number of layers of hierarchical organization increases, the most productive workers are able to solve almost all the problems by themselves, so that they almost never ask managers. Those very productive managers can therefore handle many workers, who could almost produce by themselves, and only ask for their help only when the question is very difficult. However, not all managers can work with those very productive, almost self-sufficient workers, for there is only a limited supply of very skilled workers. Less knowledgeable managers therefore have to "tap" progressively into the knowledge of less knowledgeable workers. The way in which they have to "tap" into this knowledge is given by a Pareto distribution of tail coefficient equal to two, for the same mathematical reason as in Geerolf (2013).[3] Just as in Geerolf (2013), the distribution of the largest firms' sizes corresponding to the most productive workers and the most productive managers does not depend on the functional forms underlying the primitives - distribution of arriving problems, and of agents' skills - because relatively few managers and few workers underly this distribution, so that to the limit, the distribution can be approximated by a uniform distribution. When there are not just one, but many levels of hierarchical organization, the formula for the Pareto's endogenous tail coefficient is $2^L/(2^L-1)$ in general. The intuition for that is that managers at the top of the firm manage other intermediary

[2]In contrast, Gabaix and Landier (2008) assume that firms' sizes and managers' skills interact in a linear way. Their empirical strategy indeed does not allow to identify separately managers' skills from complementarities. Moreover, the size of firms is exogenous, while in this model it is endogenous.

[3]There is a clear analogy between lenders and borrowers' leverage ratios in Geerolf (2013) and workers and managers' span of control in this paper.

managers, who themselves manage other agents. The tail for span of control is thus thicker with multiple levels of intermediary management than with just one of them. For example, in the case of two layers of management, I show that the number of workers that the intermediary managers supervise is given by a Pareto distribution of coefficient four in the space of top managers: there are fewer managers than intermediary managers. Since their own span of control over intermediary management is given by a Pareto with coefficient two, total span of control therefore has a tail Pareto exponent given by $1/(1/2+1/4)=4/3$, since it is just a multiplication of the two above distributions. This reasoning generalizes straightforwardly to a case with L layers, with an exponent of $2^L/(2^L-1)$ in general.

One thing that crucially distinguishes this theory from a random growth mechanism is therefore that it not only predicts a Zipf's law for firm sizes - the data lends support to both theories in this respect - but that it also predicts Pareto distributions with endogenous tail coefficients equal to $2^L/(2^L-1)$ for $L=1,2,...$ between intermediary levels of management, something that random growth theory does not. This is a very clear testable implication of the model. The fact that these Pareto distributions are indeed observed in the French data for firms, with these exact tail coefficients, as well as in the Census Data for establishments, seems to lend support to the theory presented in this paper. Another empirical success of this theory is that the way in which the distribution of spans of control with non zero heterogeneity and helping time differ from the exact Pareto distribution is similar to what is actually found in the data.

An important take from the model is that heterogeneity in the primitives, here agents' skills, can very well have finite support, even a support with infinitesimally small measure, and yet heterogeneity in outcomes appear unbounded. To the best of my knowledge, this contrasts with all existing literature which attributes outcomes with infinite support to very heterogeneous causes, ones with infinite support as well. Among many examples, managers' skills in Lucas (1978) have unbounded support - they are even distributed Pareto; or productivities are Pareto distributed in some models of heterogeneous firms with decreasing returns to scale and fixed costs. In contrast, the model will allow to map an empirically very unequal labor income distribution to an underlying potentially very equal distribution in skills. Even on a purely methodological ground, whether the "true model" of the economy has an unbounded dispersion of primitives (productivity, demand, skills), or a bounded one, as this model would suggest, is likely to matter not just quantitatively, but also qualitatively. I will come back to this point in the literature review below when referring to the relevant literature on models with heterogenous firms.

The rest of the paper proceeds as follows. Section 1 presents eight stylized facts on the firm size distribution and firm dynamics, seven of which the random growth literature has a hard time explaining. Section 2 develops a Garicano (2000) model of production hierarchies. Section 3 focuses on the properties of the equilibrium span of control distribution, without reference to equilibrium prices; and shows that the model can make sense of all eight stylized facts. Section 4 calculates the labor income distribution sustaining these allocations. Section 5 concludes.

Literature. This model speaks to the random growth literature generating Zipf (1949)'s Law, a survey of which is given for example in Gabaix (2009). The intuition in these models for why heterogeneity in firm sizes, city sizes or incomes are very large is that firms, cities or individuals in the tail of the Pareto distribution had a particularly long and unlikely continued sequence of good idiosynchratic shocks. Models along these lines crucially assume Gibrat (1931)'s law holds[4]; empirically, this law is not exactly verified, as the variance of firm growth rates decreases a bit with size: Mansfield (1962) presents early evidence along these lines. Models generating Pareto distributions from a random growth models comprise for example Champernowne (1953), Simon (1955), Simon and Bonini (1958), Kesten (1973), Sutton (1997), Gabaix (1999), Axtell (2001), Luttmer (2007), Rossi-Hansberg and Wright (2007). Importantly, these random growth models are not usually considered as being microfounded, as the source of idiosynchratic random shocks is not understood and more importantly, these shocks are assumed to be uninsured, even though they come repeatedly. The paper also speaks to the literature on firm dynamics, among many examples Jovanovic (1982), Hopenhayn (1992), Cooley and Quadrini (2001), Klette and Kortum (2004), none of which however generates Pareto distributions.

The paper most chiefly belongs to the span of control literature, developed since Lucas (1978), who assumes Pareto distribution in skills of managers together with homogeneity of workers' skills to explain Pareto distributions in span of control; and to the literature on the "economics of superstars", initiated by Rosen (1981). It also speaks to the literature on organizational structure, a survey of which is given in Radner (1992) for the older literature. The paper builds on Garicano (2000)'s production hierarchies model to investigate the distribution of firm sizes in the economy. More precisely, it draws heavily from the applied production hierarchies models developed in Garicano and Rossi-Hansberg (2004), Garicano and Rossi-Hansberg (2006) and Antràs et al. (2006), and Caliendo and Rossi-Hansberg (2012). It uses the same intuition, and a similar methodology, as Geerolf (2013) to generate a Pareto distribution of tail coefficient equal to two. However, the mechanism to generate more skewed Pareto distributions (with lower tail coefficients) is different, and here relies on a static mechanism, and not from a dynamic one as in Geerolf (2013). The use of French data for production hierarchies, and in particular the occupation code as an indication of workers' position in a firm's hierarchical structure follows Caliendo et al. (2014) and Liegey (2014).

The model is part of the competitive assignment literature, comprising Roy (1950), Rosen (1974), Sattinger (1975), Rosen (1981), Teulings (1995), Terviö (2008) and Gabaix and Landier (2008). In particular, a crucial assumption underlying Garicano (2000) as well as in similar models of sorting is that quality and quality of workers are imperfect substitutes, and that there is no such thing as a concept of "efficiency units of labor" (see Eeckhout and Kircher (2012) for a unified treatment). Note that in contrast to Lucas (1978), Gabaix and Landier (2008) take firm sizes as exogenous. Because the distribution of CEOs income is less skewed than Zipf, they

[4] Gibrat (1931) wanted to explain why the distribution of firm sizes was approximately log-normal, so he did not need statistical frictions. We now know that Pareto fits the distribution much better in the upper tail, although log-normal gives a better fit for the rest of the distribution.

interpret it as evidence that the distribution of talents goes to zero relatively rapidly for top talents, which follows from the assignment equation. Terviö (2008) develops a similar line of reasoning, without being explicit on the functional form for talent. More generally, it has been known for a long time that competitive assignment models were able to create some inequality in earnings, which were convex in the underlying heterogeneity in agents, because of the amplifying nature of the assignment process. What this paper shows, is that under minimal assumptions, higher skills can command a lot more ressources than very close skills. Unlike in Gabaix and Landier (2008), the levels of complementarities is however not assumed to be fixed and output linear in managers' talents, but comes from the fact that managers must "complement" workers' production by solving the problems they pass on to them. The link between span of control and equilibrium wages therefore is not mechanic, and in particular Pareto distributions in span of control will not mechanically result in Pareto distributions for labor incomes.

The paper is an alternative to random growth theory to microfound with a very limited number of assumptions on primitives why firms' sizes tend to be distributed Pareto. It is therefore potentially relevant to far more distant pieces of the economics literature, since research on firm heterogeneity takes place at the intersection of macroeconomics, labor economics, trade and industrial organization. For example a very large literature in trade with heterogenous firms following Melitz (2003) uses evidence of Pareto firm sizes to justify Pareto distributions in productivity: Antràs and Helpman (2004), Helpman et al. (2004), Ghironi and Melitz (2005), Bernard, Redding and Schott (2007), Melitz and Ottaviano (2008), Chaney (2008), Atkeson and Burstein (2010), Arkolakis (2010), Helpman et al. (2010), Eaton et al. (2011), among many other examples - see Bernard, Jensen, Redding and Schott (2007) and Melitz and Trefler (2012) for surveys of the trade literature.[5] The rather general results in Arkolakis et al. (2012) on the welfare consequences of trade are valid for heterogenous firms models only under the Pareto functional form assumption on productivity.[6] There is also a large literature in macroeconomics on misallocation, speaking to the welfare consequences of microeconomic and macroeconomic distortions, using abundantly the firm size distribution to back out productivity differences, among which Hopenhayn and Rogerson (1993), Restuccia and Rogerson (2008), Hsieh and Klenow (2007).[7] In the light of this paper, it seems on the contrary that the shape of firm size's distribution says very little on the underlying distribution of productivity across them, even on the relative importance of their dispersion. More generally, the fact that primitives can exhibit only a bounded distribution, even one with an arbitrarily small dispersion, and yet generate very heterogenous outcomes, is a potentially useful insight for many strands of the literature.

[5] More rarely, a Pareto distribution in some other primitive is used: for example, a Pareto distribution in demand shifters in Caliendo and Rossi-Hansberg (2012).

[6] In Arkolakis et al. (2012), assumption $R3$, see p. 103.

[7] For example, in Restuccia and Rogerson (2008), the productivity differences between the most and the least productive firms, which are crucial for the welfare results, are calibrated to be in a ratio of 1 to 3.98 from the 1 over $100,000$ ratio observed on the firm size distribution. Similarly, Hopenhayn and Rogerson (1993) use the size distribution of firms aged 0-6 years to back out productivity differences between entering firms.

1 Evidence on the Firm Size Distribution and Firm Dynamics

I first lay out eight stylized facts regarding the firm size distribution and the distribution of firms' growth rates. All of them but one (Stylized Facts 1 to 7) are already well known in the literature. Stylized Fact 8 is new and concerns French production hierarchies mentioned in the introduction. Only Stylized Fact 1 is consistent with random growth theory. All the others are at some level inconsistent with the theory of random growth with statistical frictions, but will be consistent with the model developed in the rest of the paper.

1.1 Overall Firm Size Distribution

Stylized facts 1-4 concern the overall firm size distribution and are illustrated on Figure 13 taken from Axtell (2001). Stylized fact 1 states the Zipf's law character of the firm size distribution, that random growth theory seeks to explain.

Stylized Fact 1 (Overall Firm Size Distribution)**.** *The distribution of firm sizes is approximately a Pareto distribution of coefficient equal to one (Zipf's law).*

The overall firm size distribution however exhibits three deviations from Zipf's law: the fitted Pareto coefficient is not exactly one, Zipf's does not apply for the very upper nor for the lower tail of the distribution, but only for firms with a number of employees between 50 and $100,000$.

Stylized Fact 2 (Deviation: Pareto Coefficient)**.** *The best fit of the Pareto distribution to the distribution of firm sizes is a Pareto with a coefficient equal to* 1.059 *for the United States.*

Stylized Fact 3 (Deviation: Large Firms)**.** *The firm size distribution is bounded unlike the Pareto distribution, and has a thinner tail than the Pareto distribution.*

Stylized Fact 4 (Deviation: Small Firms)**.** *Pareto is a good approximation only for the upper tail of the firm size distribution. For the bulk of the distribution, log normal is a much better fit.*

1.2 Firm Dynamics

Ever since the random growth literature, Zipf (1949)'s law has been seen as indirect evidence for scale independance in the growth of firms. However, it has been known for a long time, at least since Mansfield (1962), that Gibrat's law for firm growth does not hold so well in the data. For example, Figures 15a and 15b are taken from Rossi-Hansberg and Wright (2007).

Stylized Fact 5 (Deviation from Gibrat's Law)**.** *Firm growth does not follow Gibrat's law. In particular, the variance of growth rates decreases with firm's size.*

Cabral and Mata (2003) present even more direct evidence which tends to cast some doubt on the random growth model as an explanation for the firm size distribution. They show on Portuguese micro level data that the distribution of firm sizes is already very skewed to the left at the time of birth. More importantly, the fact that small firms conditional on survival grow at a faster rate than large firms has been argued to be consistent with Gibrat's law if there is a lot of firm exit (for example, because firms gradually learn how productive they are, as in Jovanovic (1982)). However, Cabral and Mata (2003) show that selection only accounts for a very small fraction of the evolution of firm sizes.

Stylized Fact 6 (Size Distribution by age)**.** *The distribution of firm sizes is already very skewed to the left at the time of birth.*

1.3 Disagreggated Size of Production Units

The perhaps most important stylized facts that this paper is after concern the disaggregated size of production units. Stylized fact 7 concerns the distribution of plant sizes in the US data, highlighted in particular in Rossi-Hansberg and Wright (2007) (pp 1948), and illustrated on Figure 2. To be precise, Rossi-Hansberg and Wright (2007) do not make explicit the exact Pareto coefficient on the distribution of establishments. Only do they note that the size distribution of establishments in general has a thinner tail than the Pareto distribution of firms.[8]

Stylized Fact 7 (Establishments)**.** *The size distribution of establishments has a thinner tail than the size distribution of firms. In the Census US data, the upper tail of the establishment size follows a Pareto with a coefficient close to* 1.33*.*

Stylized fact 8 is based on the use of French micro-level data, and the methodology of Caliendo et al. (2014) to delimit production hierarchies, and is illustrated on Figure 1.

Stylized Fact 8 (Hierarchies)**.** *In the French data, the distribution of span of control down one level of hierarchy follows a Pareto distribution of coefficient* 1.96*. The distribution of span of control down two levels of hierarchy follows a Pareto distribution of coefficient* 1.33*. The distribution of span of control down three levels of hierarchy follows a Pareto distribution with coefficient* 1.14*.*

The model developed in the next section will be able to make sense of all the above stylized facts, not just of Stylized Fact 1.

2 Model

In this section and the following, I develop and study the upper tail properties of a Garicano (2000) model of production hierarchies. In order to get quickly at the main results of the model, I will assume a fixed, exogenous distribution of skills - although endogenous skill acquisition is potentially a fruitful extension. Unlike Antràs et al. (2006) however, I will not here "force" agents

[8]Rossi-Hansberg and Wright (2007) write: "It is worth noting that the size distribution of enterprises is much closer to the Pareto, especially if we focus attention on enterprises with between 50 and 10,000 employees. The differences between the size distributions for establishments and enterprises may shed light on the forces that determine the boundaries of the firm. Our theory focuses, however, on the technology of a single production unit and does not address questions of ownership or control."

to join teams. They will work in cooperation with others only when they have an interest in doing so, that is when they gain more in team than by being self-employed: I will show this is the case when heterogeneity in skills, or the cost of communicating solutions to problems, is sufficiently low. Since occupational choice is a crucial element for the formation of Pareto distributions for span of control, I don't want to restrict the choice set exogenously in this dimension. Again, unlike in Antràs et al. (2006), I do not here restrict the number of layers of hierarchical organization to be one in the maximum, but the number of layers L will be determined endogenously. This higher number of layers is crucial to getting the Pareto results.

I consider a static economy, with a continuum of agents indexed by i. All agents value consumption in the same way according to a linear utility function. Agents are endowed with one unit of time, which they supply for production or for helping others solve problems. When producing, they encounter problems, whose set is indexed on $[0, 1]$. The arrival of these problems is uniformly distributed on this interval, and the problems' indexes are higher when these problems are harder to solve - that is, less agents know how to solve them. This is without loss of generality, as from any initial draw of problems and their probability distributions, one can just relabel them such that this is verified. As in this literature, I shall denote the cumulative distribution function for the arrival of problems by $F(x) = x$ on $[0, 1]$.

Agents' heterogeneity. Agents i are assumed to be heterogeneous in terms of how many problems they can solve by themselves: they generally do not know everything. It is assumed that knowledge is cumulative, as in Garicano and Rossi-Hansberg (2004) or Antràs et al. (2006), so that an agent with a higher skill knows everything that an agent with lower skill knows - see Garicano (2000) for a discussion.

An agent i has a skill which is indexed by the hardest problem he can solve: an agent i with skill z^i can solve all problems contained in set $[0, z^i]$. The distribution of agents' skills over problems is given by a cumulative distribution function $G(.)$, with density $g(.)$ over $[0, 1]$. It is assumed that the complexity of producing the good in the economy is such that only the most skilled of them would know how to produce the good by themselves. In other words that the density function has support $[1 - \Delta, 1]$, where Δ indexes, without summarizing in general, the level of skill heterogeneity in the economy.[9]

Production. The organization of production closely follows Garicano (2000): apart from producing with their time as a factor of production, agents can also transmit solutions to other agents. When production workers know the solution to these problems, they can produce one unit of output per unit of time. But if they do not, they can ask their managers. Regardless of

[9]One could generalize this to having a non-trivial measure of agents able to solve all problems, but this would only add subcases, without adding much intuition to the model. On the other hand, the results rely on some agents being able to solve all problems arising in production, at least to the limit. It is doubtful that agents would in equilibrium agree to produce a good which would pose unsolvable problems in production even to the most skilled of them. Another option would be to endogenize the choice of the good being produced.

whether the manager knows the solution or not, it takes $h < 1$ units of time for a manager to communicate a solution to the worker. $h < 1$ is assumed for agents to find it (sometimes) optimal to form teams: if the time of communicating knowledge is greater than the time of producing, then it is always better to engage in self-production. h is called the helping time. Potentially, managers can also ask other agents to solve the problems if they do not know the solutions to them, so that hierarchies can form.

A crucial assumption is that workers cannot diagnose or "label" problems when they cannot solve them and therefore do not know who may know the solution to problems they cannot solve, otherwise the organization would never have a pyramidal structure, and agents would directly go to the agents who know. I call manager of type l a manager who answers questions which have been transmitted l times: managers of type 1 directly supervise production workers, managers of type 2 answer questions of managers of type 2, and so on.

Equilibrium. In equilibrium, agent i with skill z^i chooses to allocate his time being a production worker, a manager of type l (with potentially any $l \in \{1, 2, ...\}$), or self-employed in order to maximize his expected utility - that is, his income since utility is linear. Denote by L the maximum number of layers of hierarchical organization in this economy - I will show later than under an indivisibility assumption on managers' time leads to L to be a finite number.

Formally, an agent is denoted as having a negative position in production worker time $t^i_W(.)$ when he hires workers, and a positive one $t^i_W(z^i)$ when producing as a production worker. Symmetrically, an agent can hire managers of lower levels who pass on problems they cannot solve, so that the manager of type l time $t^i_{M_l}(z^i)$ can be positive or negative for any $l \in \{1, ..., L-1\}$. In contrast, an upper level manager by definition cannot be hired by anyone. An agent can also spend time t^i_S being self-employed.

Denoting by $w_0(z)$ the wage from being a production worker when of skill z (or the price of production time at that skill level), and similarly by $w_l(z)$ the wage from being a manager of type $l \in \{1, ..., L-1\}$, and given that self-employment gives z as income, the agent chooses to occupy his time so as to maximize his income (I) subject to his total time constraint of one unit (TC) and his communication time constraints (CTC_0) and (CTC_l) given by the unit time of communicating h multiplied by the number of time he will have to answer his workers' questions:

$$\max_{\left(t^i_W(.), \{t^i_{M_l}(.)\}_{l=1}^{L-1}, t^i_S\right)} \quad z^i t^i_M + z^i t^i_S + \int_z w_0(z) t^i_W(z) dz + \sum_{l=1}^{L-1} \int_z w_l(z) t^i_{M_l}(z) dz \qquad (I)$$

$$\text{s.t.} \quad t^i_M + t^i_S + \int_z \max\{t^i_W(z), 0\} dz + \sum_{l=1}^{L-1} \int_z \max\{t^i_{M_l}(z), 0\} dz \leq 1 \qquad (TC)$$

$$\text{s.t.} \quad \int_{z \neq z^i} h(1-z) t^i_W(z) dz \leq t^i_{M_1} \qquad (CTC_0)$$

$$\text{s.t.} \quad \forall l \in \{1, ..., L-1\}, \quad \int_{z \neq z^i} h(1-z) t^i_{M_l}(z) dz \leq t^i_{M_{l+1}} \qquad (CTC_l)$$

$$\text{s.t.} \quad \forall z \neq z^i, \quad t^i_W(z) \leq 0$$

$$\text{s.t.} \quad \forall l \in \{1, ..., L-1\}, \forall z \neq z^i, \quad t^i_{M_l}(z) \leq 0$$

Note that if $t^i_W(z) > 0$, then agent i is a production worker, if $t^i_{M_l} > 0$, then agent i is a manager of type l, and if $t^i_S > 0$, then agent i is self-employed. The before last constraint states that an agent can hire any type of production worker, but can only supply a positive amount of time $t^i_W(z^i)$ with his skill. This is why all other $t^i_W(z)$ for $z \neq z^i$ must be non positive. Similarly, an agent can hire any type of intermediary manager in principle, but can only supply a positive amount of intermediary time $t^i_{M_l}(z^i)$ for $l \in \{1, ..., L-1\}$ with his skill. A Competitive Equilibrium of this production economy $\mathcal{E}^1$ is then defined as follows.

Definition 1 (Competitive Equilibrium of $\mathcal{E}^1$). *A* Competitive Equilibrium for Economy $\mathcal{E}^1$ *is a wage function for production workers $w_0(.)$ and allocations of time $\left(t^i_W(.), \{t^i_{M_l}(.)\}_{l=1}^{L-1}, t^i_S\right)$ for all agents i such that agents maximize their income (I) under the time constraint (TC), the communication time constraints (CTC_l) for $l \in \{0, ..., L-1\}$, taking the wage functions $w_l(.)$ for $l \in \{0, ..., L-1\}$ as given, and the market for work time, and lower-level managers' time clears for all skill levels, that is:*

$$\forall z, \quad \int_i t^i_W(z) di = 0. \tag{MC_0}$$

$$\forall l \in \{1, ..., L-1\}, \quad \forall z, \quad \int_i t^i_{M_l}(z) di = 0. \tag{MC_l}$$

Because the program is linear in $\left(t^i_W(.), \{t^i_{M_l}(.)\}_{l=1}^{L-1}, t^i_S\right)$, it will be optimal in equilibrium for agents to do one of three things: use their whole time producing in a team, or managing and communicating solutions at a given level of management, or engage in self-employment. In other words, the model I am working with is very similar to that in the seminal Garicano (2000), or Garicano and Rossi-Hansberg (2006). I will now solve for the equilibrium of this model and state the main result of the paper about the distribution for spans of control of managers. Following Caliendo and Rossi-Hansberg (2012), I will assume that a top manager can only work in one firm, and not manage multiple firms.[10]

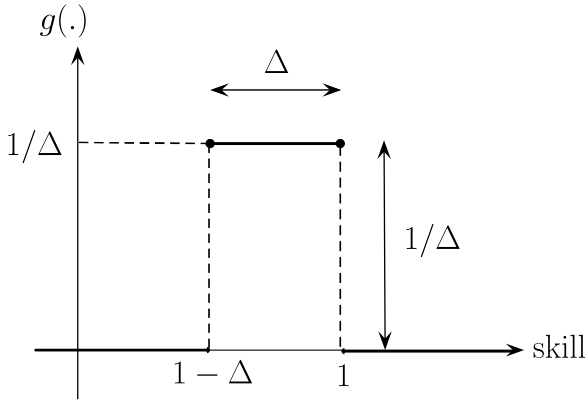
Assumption 1 (Indivisibility). *A firm is run by a manager working full time at the top of his organization.*

Finally, a lot in this paper will revolve around the uniform distribution of skills. In the paper, I will refer to the uniform distribution of skills with heterogeneity Δ which corresponds to the distribution in Definition 2.

[10] I could as well assume that firms can be run by two or three agents, or that agents can run at most an integer number of firms, as long as it is a fixed number. In practice, conflict of interest or management issue would certainly arise in that case, so that the constraint can be thought of as a reduced form of an upper level model with an explicit decision-making process.

Definition 2 (Uniform Distribution)**.** *The* uniform distribution of skills with heterogeneity parameter Δ *is defined as a function of* Δ *by:*

$$g(x) = \begin{cases} 0 & \textit{if } x < 1 - \Delta \\ 1/\Delta & \textit{if } x \in [1 - \Delta, 1] \\ 0 & \textit{if } x > 1. \end{cases}$$



3 Span of Control

As can be inferred from Definition 1, a Competitive Equilibrium of the model defined above is potentially a high dimensional object. In this section, I will study independantly allocations in this model with no mention of the equilibrium prices sustaining these allocations. As in Garicano (2000), all of these allocations are Pareto-optimal and can be decentralized using labor markets inside the firm something I turn to in Section 4.

3.1 Ruling out Self-Employment

The following lemma states the first result: for helping time h and hetereogeneity Δ sufficiently low, there is no self-employment in the equilibrium of this model.

Lemma 1 (No self-employment Condition)**.** *For a sufficiently low value of skill heterogeneity* Δ *and of communication time* h*, there is no self-employment in equilibrium.*

Proof. See Appendix C.1. □

In the rest of the paper, I assume that Δ and h are indeed sufficiently low. An alternative would be to assume as in Antràs et al. (2006) that self-employment is not an option from the outset.

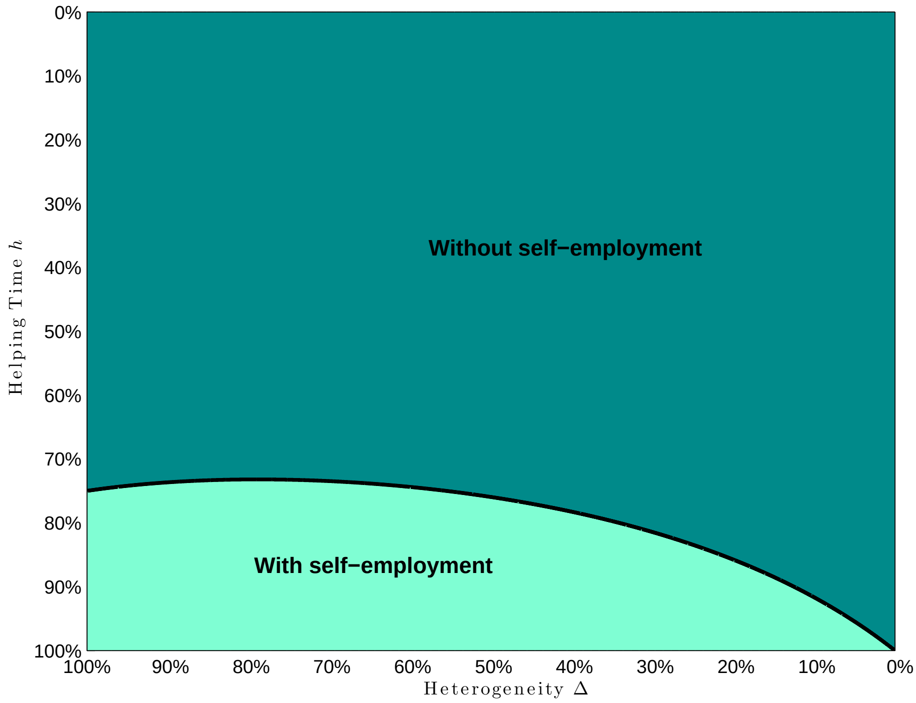
Assumption 2 (No self-employment)**.** Δ *and/or* h *are low enough, so that there is no self-employment in equilibrium.*

To get an idea of how restrictive this assumption might be, one can look at the shape of Assumption 2 in the case of a uniform distribution of skills with disagreement Δ (see Definition

2). The set $\mathcal{A}_2$ of (Δ, h) values such that the assumption is verified can then be expressed in closed form:

$$\mathcal{A}_2 = \left\{ (\Delta, h); \quad (\Delta, h) \in [0,1]^2; \quad h > \frac{2\sqrt{1+2\Delta-2\Delta^2}-1-\Delta}{1+2\Delta-3\Delta^2} \right\}$$

Figure 3: Validity of Assumption 2 in the Uniform Case



Note: The parameter space for which there is no self-employment is the upper-right panel of this Figure. Note that for any value of helping time h, even one very close to production time ($h = 100\%$), there exists some upper bound on heterogeneity so that if heterogeneity is lower than this threshold, there is no self-employment in equilibrium.

This particular expression is derived in Appendix C.4. The self-employment and no self-employment regions are drawn on Figure 3 as a function of the heterogeneity parameter Δ. Assumption 2 is valid across most of the parameter space. In particular, for any $h < 1$, there exists $\Delta > 0$ such that for any heterogeneity Δ lower than this threshold, the assumption is verified.

3.2 Allocations

Proposition 1 describes, given a maximum number of layers of management, the allocations in this model: the occupational choice of agents, and who works for whom (defining the boundaries of the firm). Proposition 2 gives the equilibrium number of layers, for a finite number of workers - and when the skill distribution with probability distribution function given by $G(.)$ is the continuous limit of the underlying discrete skill distribution. Together, Propositions 1 and 2 uniquely define the allocations of a Competitive Equilibrium defined in Definition 1.

Before stating Proposition 1, it is convenient to denote the limiting bounds for the support of beliefs for any number of layers L as $[z_0^L, z_{L+1}^L] = [1-\Delta, 1]$.

Definition 3 (z_0^L, z_{L+1}^L)**.** *Let z_0^L and z_{L+1}^L denote the bounds of the support of skills:*

$$z_0^L = 1 - \Delta \qquad z_{L+1}^L = 1.$$

Proposition 1 (Allocations, given L)**.** *Assume a maximum number of layers L of hierarchical organization. There exists L endogeneous cutoffs $\{z_l^L\}_{l=1}^L$ uniquely given by a set of L equations:*

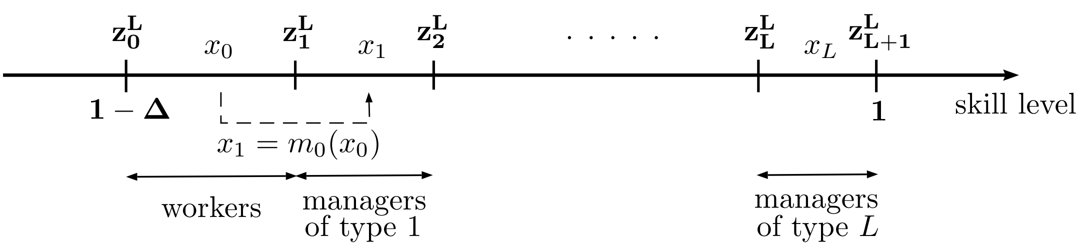
$$\forall l \in \{0, ..., L-1\}, \quad G(z_{l+2}^L) - G(z_{l+1}^L) = h \int_{z_l^L}^{z_{l+1}^L} (1-u)g(u)du,$$

such that agents with skills in $[z_0^L, z_1^L]$ are production workers, and pass on problems to managers of type 1 with skills in $[z_1^L, z_2^L]$ and so on, until managers of type L in $[z_L^L, z_{L+1}^L]$, the top managers of the firm. Workers and different levels of managers pass on problems to each other according to increasing matching functions $\{m_l(.)\}_{l=0}^{L-1}$, respectively defined on $[z_l^L, z_{l+1}^L]$ with values in $[z_{l+1}^L, z_{l+2}^L]$ through:

$$\forall l \in \{0, ..., L-1\}, \quad \forall x_l \in [z_l^L, z_{l+1}^L], \quad m_l'(x_l)g(m_l(x_l)) = h(1-x_l)g(x_l) \quad \textit{and} \quad m_l(z_l^L) = z_{l+1}^L.$$

Proof. See Appendix C.2. □

Figure 4: Equilibrium Allocations: Notations



Note that the matching functions depend on the assumed maximum number of layers L, just as the cutoffs, and so should normally be denoted by $\{m_l^L(.)\}_{l=0}^{L-1}$, but this is omitted for conciseness (see Figure 4 above for the detail of the notations).

A second result from the paper is that the equilibrium number of layers is given in this Garicano (2000) model endogenously by equation (L) in Proposition 2 below. This result is intuitive: when the number of layers increases, the number of agents in the upper layer becomes smaller and smaller, and tends to zero, as stated in the following Lemma 2.

Lemma 2 (Size of Top Layer)**.** $z_L^{L+1} - z_L^L \to 0$ *when* $L \to \infty$.

Proof. This results straightforwardly from Proposition 1. See Appendix C.3. $\square$

When the number of agents in this upper-level layer becomes lower than one, this add-in of a new layer becomes irrelevant, under the indivisibility assumption 1. Lemma 2 thus leads to Proposition 2.

Proposition 2 (Equilibrium L)**.** *For any finite number of workers N_w, the maximum levels of hierarchical organization is given by:*

$$L = \max_{L \in \mathbb{N}} \left\{ L \quad s.t. \quad 1 - z_L^L \geq \frac{1}{N_w} \right\} \tag{L}$$

It is useful to look at some numbers to get an idea of the speed of convergence. For example, the case of a uniform distribution over the maximum support of skill hetereogeneity $[0, 1]$ (that is, hetereogeneity is $\Delta = 1$) is investigated in Table 1. From this table, one notes that if the number of workers is one million, that is $N_w = 1,000,000$ then the integer constraint becomes binding for $L = 4$, so that the maximum levels of hierarchical organization is given by $L = 3$. In this model, the limited amount of time that a top manager is what determines the boundaries of the firm; and even the industrial organization of the economy. This is potentially interesting because scholars have long been interested in the distribution of firm sizes to know for example whether, if anything, something ought to be done about the high number of big firms. In this theory, both the number of firms and the number of employees in each firm are endogenously determined by the indivisibility of managers' time; and so is the industrial organization structure of a sector, even if the underlying technology has constant returns.

Proposition 1 and Proposition 2 together fully characterize the equilibrium allocations in the model. The main results from the paper then pertain to the endogenous span of control distributions that result from these allocations, which I turn to in the next section.

3.3 Span of Control Distributions

Denote by $N_{l_1 \to l_2}^{l_3}(x_{l_3})$ the equilibrium span of control of layer l_1 over layer $l_2 < l_1$, expressed as a function of workers of layer l_3 skills, or in the "space" of workers of layer l_3 skills. For example, the time constraint of a team manager in layer $l+1$ with skill x_{l+1} helping a number $N_{l+1 \to l}^{l}(x_l)$ of workers of skill x_l gives his equilibrium span of control $N_{l+1 \to l}^{l}(x_l)$ through his time communicating constraint (CTC_l), which holds with equality at the optimum:

$$h\,(1 - x_l)\, N_{l+1 \to l}^{l}(x_l) = 1 \quad \Rightarrow \quad N_{l+1 \to l}^{l+1}(x_{l+1}) = \frac{1}{h\,\left(1 - m_l^{-1}(x_{l+1})\right)}.$$

All the other span of control functions $N_{l_1 \to l_2}^{l_3}(x_{l_3})$ result from these fundamental span of control functions between intermediary levels of management, up sometimes to a small change of

Table 1: UNIFORM SKILL DISTRIBUTIONS: CUTOFFS FOR DIFFERENT VALUES OF L, WITH $\Delta = 100\%$ AND $h = 75\%$

	$\mathbf{z_1^L}$	$\mathbf{z_2^L}$	$\mathbf{z_3^L}$	$\mathbf{z_4^L}$	$\mathbf{z_5^L}$	$\mathbf{z_6^L}$
$\mathbf{L = 1}$	0.666667					
$\mathbf{L = 2}$	0.625993	0.948538				
$\mathbf{L = 3}$	0.625000	0.947266	0.998958			
$\mathbf{L = 4}$	0.625000	0.947266	0.998957	1.000000		
$\mathbf{L = 5}$	0.625000	0.947266	0.998957	1.000000	1.000000	
$\mathbf{L = 6}$	0.625000	0.947266	0.998957	1.000000	1.000000	1.000000

Note: In this table, the cutoffs are calculated with distribution of skills in the population given by $G(z) = \frac{z-(1-\Delta)}{\Delta} \mathbb{1}\,(1-\Delta, 1)\,(z) + \mathbb{1}\,(1, +\infty)\,(z)$. Precision up to 6 digits is displayed. Successively, the maximum number of levels is restricted to being $L = 1$, $L = 2$, $L = 3$, $L = 4$, $L = 5$, $L = 6$. Heterogeneity in skills is taken to be maximal $\Delta = 100\%$ and so is helping time $h = 75\%$. For comparative statics on h and Δ, and the resulting cutoffs, refer to Table 2 in Appendix B for numerical values and Figures 7 and 20 for a graphical illustration.

variable through matching functions $m_l(.)$ for some $l \in \{1, ..., L-1\}$. The notations for different span of control distributions are illustrated in the case of a firm with $L = 2$ layers on Figure 4, inspired by Figure 1 in Rosen (1982)'s seminal contribution.

Note that we do not need the wage function to solve for the span of control distribution in this economy, unlike in Geerolf (2013).[11] This is why the order of differential equations in this paper never is greater than one, which simplifies the problem considerably.

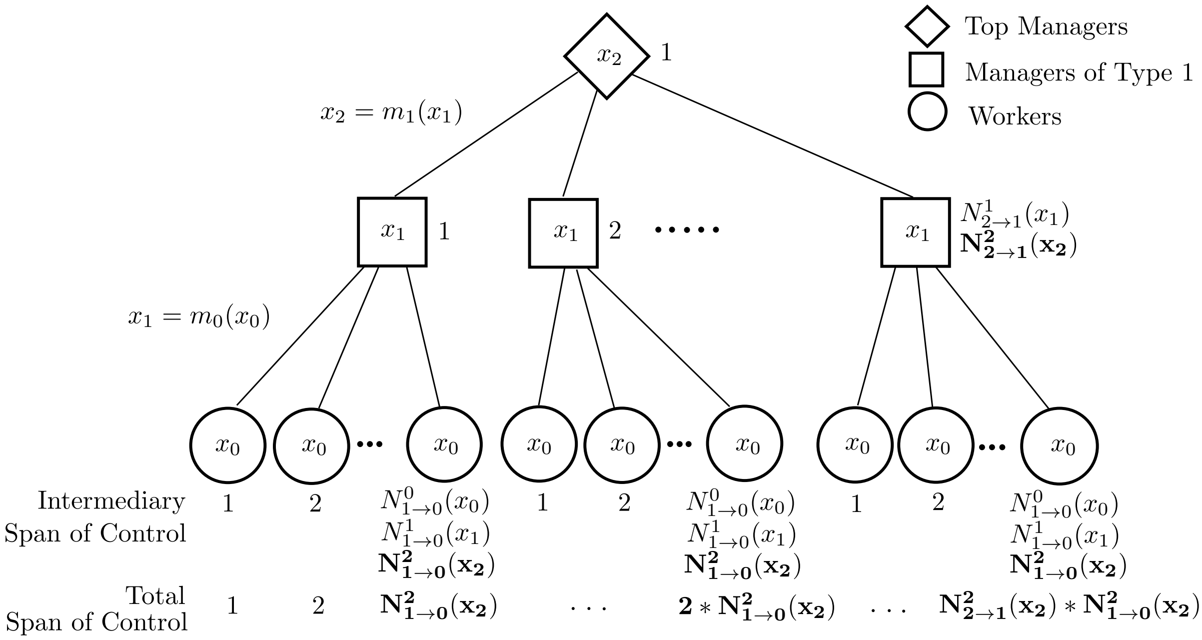
As can be seen in Table 1, the measure of top managers becomes very small when the number of layers increases. The measure of managers of type $L-1$ also becomes very small. Intuitively, this is the reason why the model's predictions on Pareto distributions do not depend on the underlying distribution of skills: the support of the distribution of skills which really matters for values of high span of controls is very small, and any smooth bounded away from zero, distribution, in such a region can as a first approximation be approximated by a uniform. Given that all the limit results shown in the following will always be given with respect to the uniform distributions, it is therefore useful first to solve for the case with a uniform distribution of skills, where all allocations and span of control distributions obtain in closed form.

3.3.1 Special Case: Uniform Distribution of Skills

Let us look first at the span of control distribution of managers down one level of hierarchical organization, and then generalize to multiple levels.

[11] This is because in Geerolf (2013), the leverage factor depended on the price of bond contracts sold, which were a equilibrium outcome of the assignment process.

Figure 5: Span of Control: Notations - Example with one Firm, $L = 2$ Layers



Note: This chart is inspired from Rosen (1982)'s Figure 1. It represents a firm with $L = 2$ layers, whose top manager (or manager of type-2) has skill x_2, supervises intermediary managers (managers of type-2) of skill $x_1 = m_1^{-1}(x_2)$, who themselves supervise workers of skill $x_0 = m_0^{-1}(x_1)$. The span of control of the top managers over intermediary managers is denoted by $N^1_{2\to1}(x_1)$ as a function of intermediary managers' skills, and $N^2_{2\to1}(x_2)$ as a function of their own skill (so that $N^2_{2\to1} \circ m_1 = N^1_{2\to1}$). Total Span of Control of a top manager over both intermediary managers and workers is given by $N^2_{2\to1}(x_2) * (N^2_{1\to0}(x_2) + 1)$ which is behaves like $N^2_{2\to1}(x_2) * N^2_{1\to0}(x_2)$ when span of control $N^2_{1\to0}(x_2)$ is large.

One Level of Hierarchical Organization. In the uniform case, the primitives of the model are summarized by the heterogeneity in skills Δ and the helping time h (see Definition 2). In particular, the inverse of the equilibrium matching functions is given in closed form through:

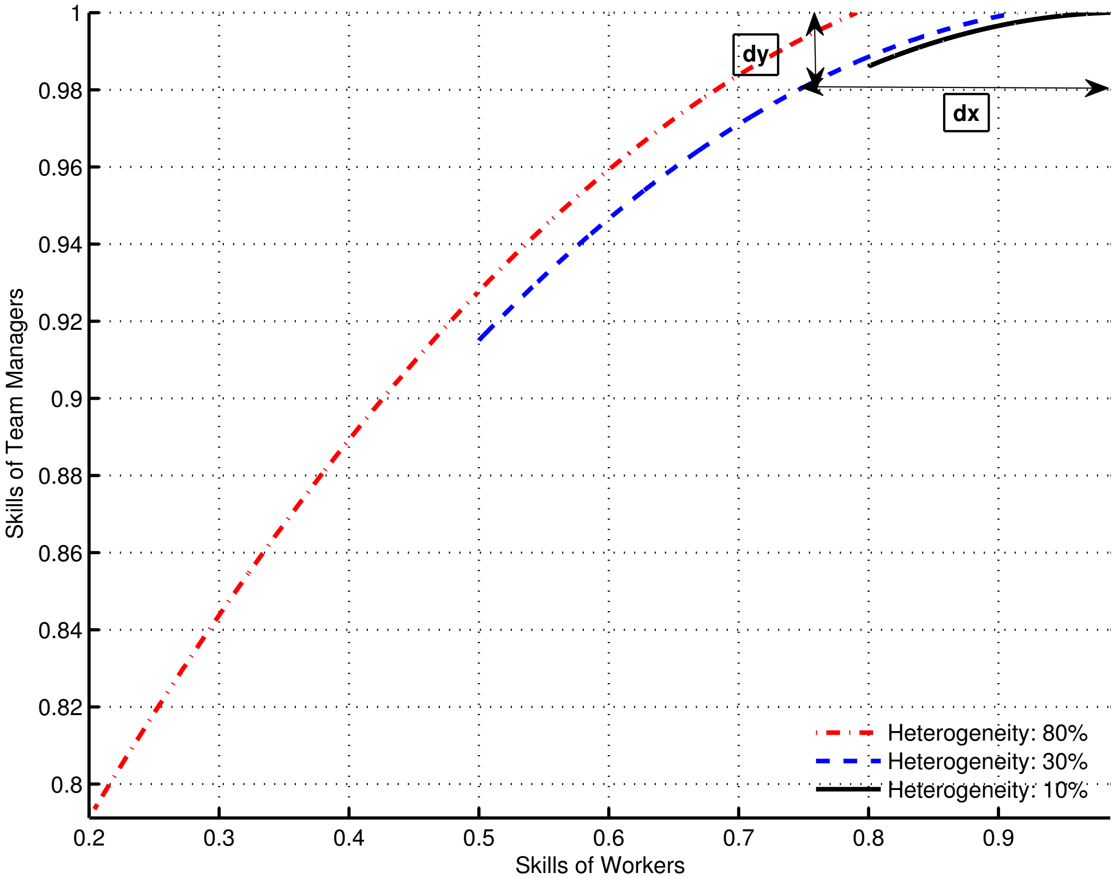
$$z^L_{l+2} - m_l(x_l) = h \int_{x_l}^{z^L_{l+1}} (1-u)du \quad \Rightarrow \quad 1 - m_l^{-1}(x_{l+1}) = \sqrt{(1 - z^L_{l+1})^2 - \frac{2}{h}(z^L_{l+2} - x_{l+1})}.$$

This allows to express the span of control between adjacent levels of hierarchical organization in the following Lemma 3.

Lemma 3 (One Level, Uniform Case)**.** *Let the density of skills $g(.)$ be uniform with heterogeneity Δ. The span of control between adjacent levels of hierarchical organization is then given by a shifted Pareto distribution of coefficient equal to two:*

$$N^{l+1}_{l+1\to l}(x_{l+1}) = \frac{1}{h\sqrt{(1 - z^L_{l+1})^2 + \frac{2}{h}(z^L_{l+2} - x_{l+1})}} \sim \text{Pareto}(2).$$

Figure 6: MATCHING FUNCTION FROM WORKERS TO TEAM MANAGERS, $h = 70\%$ - UNIFORM DISTRIBUTION



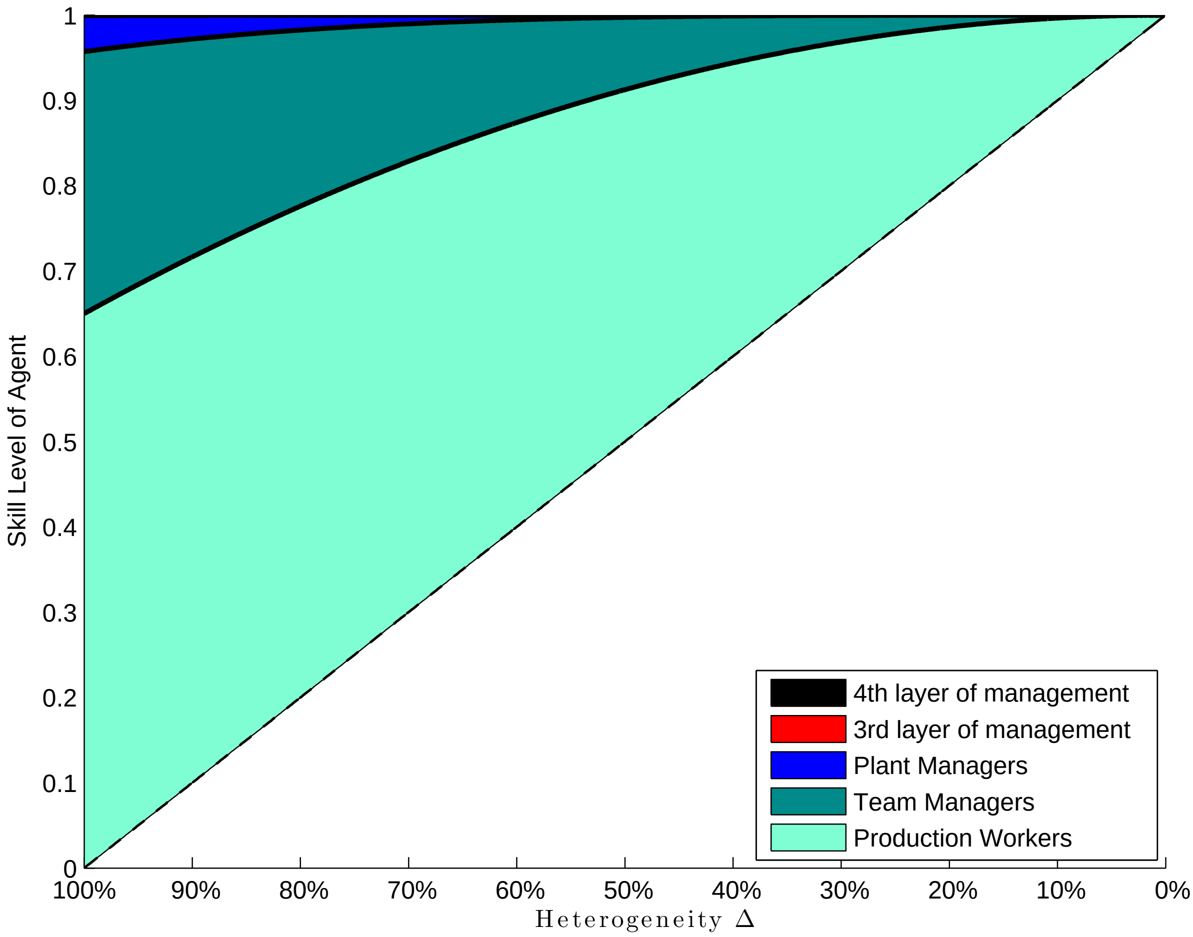
Note: This figure gives the matching function from workers to team managers derived in Section 2, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta} \mathbb{1}\,(1-\Delta, 1)\,(z) + \mathbb{1}\,(1, +\infty)\,(z)$. As heterogeneity Δ becomes small, this assignment function $m_l(x_l)$ becomes flat for $x_l \sim z^L_{l+1}$. This is because to the limit, the measure of managers $dy = dm_l(x) = m'_l(x_l)dx$ corresponding to a given measure dx of workers becomes small, as the most skilled managers work with workers who almost never ask for their help, which economizes on their valuable time.

When Δ or h become small, or when the number of layers increases, this becomes arbitrarily close to a Pareto distribution of tail coefficient equal to two, because in that case the cutoffs z^L_{l+1} and z^L_{l+2} approach one, and so the span of control distribution becomes arbitrarily close to a Pareto with a tail coefficient equal to 2, as for $x_{l+1} \in [z^L_{l+1}, z^L_{l+2}]$:

$$N^{l+1}_{l+1 \to l}(x_{l+1}) \sim \frac{1}{\sqrt{2h}} \frac{1}{\sqrt{z^L_{l+2} - x_{l+1}}}.$$

More precisely, it is a "shifted" Pareto in the sense that there is a maximum value for span of control, since the denominator does not quite attain zero even for extreme values of skills.

Figure 7: Occupational Choices for a Uniform Distribution of Skills, and $h = 70\%$

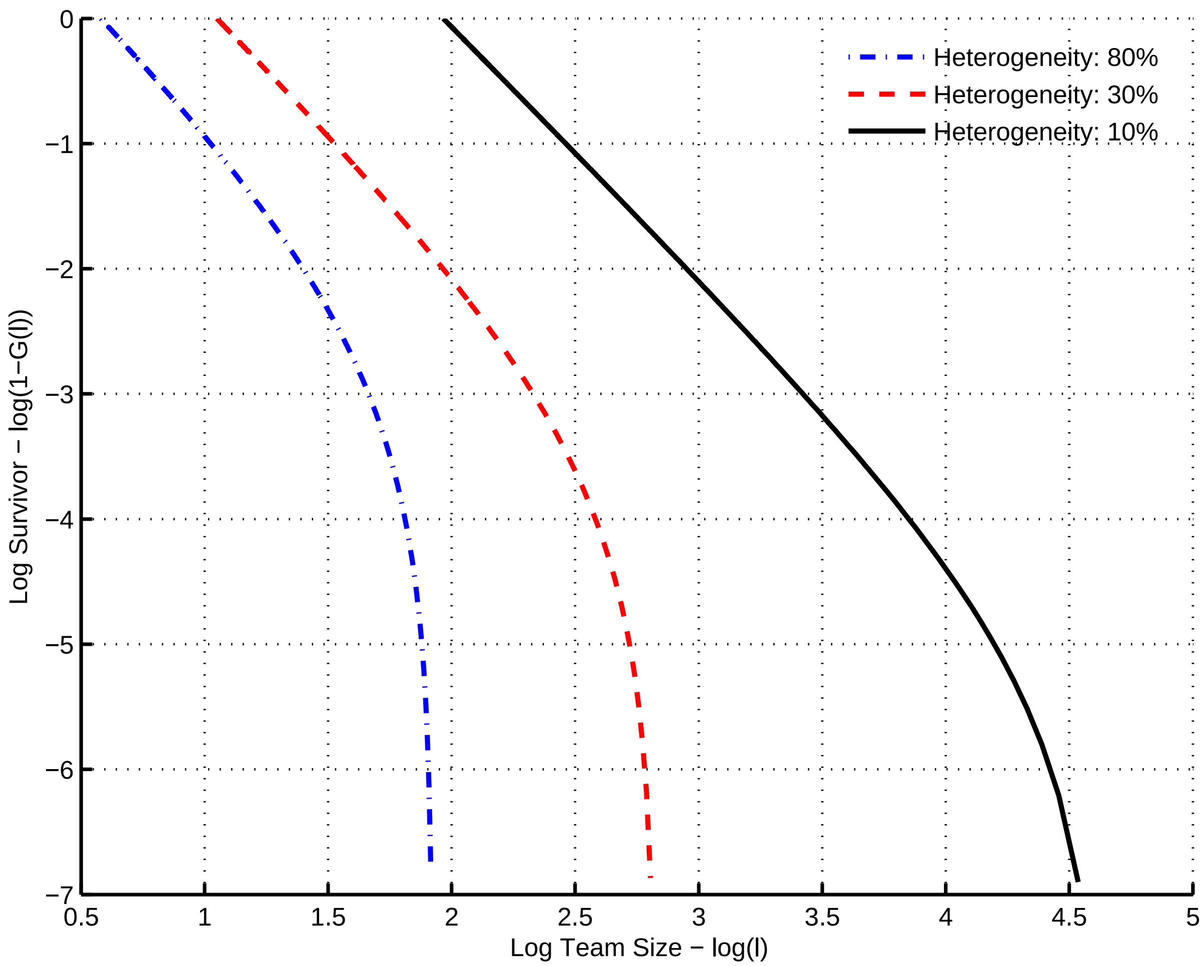


Note: This figure gives the occupational choices of agents when $h = 70\%$, and the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta}\mathbb{1}(1-\Delta,1)(z) + \mathbb{1}(1,+\infty)(z)$. Note that the 3rd layer and the 4th layer of management can barely be seen on the graph, because their measures are very small, even for high heterogeneity. This gives an intuition for why the Pareto distribution for span of control obtains for any value of overall skill heterogeneity: managers in the upper levels of management endogenously are little different in terms of skills. Note that the decrease in the measure of team managers, plant managers, is not an artefact due to the decrease in hetereogeneity Δ (see the Online Appendix): the measure of managers decreases faster than heterogeneity. In other work, the fraction of managers relative to that of workers goes to zero: this is because production workers are now able to solve almost all problems by themselves.

Note also that in that case, both z^L_{l+1} and z^L_{l+2} both go to one. One might of course worry that this finding is very specific to having communication much more efficient than production (h low) or that the finding does not hold in economies with some non negligible amount of skill heterogeneity (Δ low). However, we shall see next that it is not the case. In fact, because when new layers of management are added, the size of the new layers endogenously become small, this in effect makes heterogeneity go to zero for the span of control of top managers over their direct subordinates, without any assumption on the total distribution of skills. Moreover, convergence is typically very fast (see Appendix C.6). I will come back to this point later.

The intuition for why the span of control of top managers becomes very high, even more so when there is an arbitrarily small heterogeneity in skills in the economy, comes from the fact that all workers ask relatively few questions when heterogeneity goes to zero: they are then able to solve almost all problems by themselves. The fraction of managers required to answer these questions also goes to zero, in the limit, as can be seen on Figure 7. The marginal worker, who is just indifferent between being a worker and a manager, then is really able to solve almost all problems by himself, and the corresponding manager has a very high span of control. This reasoning actually also holds for relatively high values of heterogeneity: again as in Figure 7, adding new layers of management has the same effect as reducing heterogeneity in skills between the upper layer.

Figure 8: Distribution of Span of Control Down One Level of Hierarchical Organization ("Teams"), Log-Log Scale, $h = 70\%$ - Uniform Skill Distribution



Note: This figure gives the theoretical distribution of team sizes as derived in Section 2, for different values of Δ. As heterogeneity Δ goes to zero, the distribution approaches the full Pareto distribution in the upper tail. When the number of layers of hierarchical organization is endogenous, the number of agents in the last and before last layers endogenously go to zero, so that the full Pareto is obtained.

The fact that span of control is the distributed like a shifted Pareto distribution in the upper tail comes from the mathematical reasoning above: the span of control is an inversely proportional function of the skill of the corresponding worker. This Pareto distribution has an endogenous tail coefficient equal to two, which can be seen on Figure 8, even for relatively high values of heterogeneity: on this Figure the slope of the linear part is -2 on a log-log scale when the survivor function is plotted against the value for span of control. Intuitively, this coefficient two comes from the shape of the matching function for the most skilled managers and workers, which becomes flat for those agents when heterogeneity goes to zero, as can be seen on Figure 6 above. The intuition for why this function is flat is that managers with high skills, close to one, are matched with workers who also have relatively high skills, and can in effect solve almost all problems by themselves. It is all the more true that heterogeneity is relatively low. Again, for non zero helping time or heterogeneity, the distribution of span of control we obtain could reasonably be called a "shifted Pareto": in particular, it has an upper bound (though it is very high for most values of the parameters) and does not quite have the Pareto property in the very upper tail. This fact can also be seen on Figure 6, the matching function is not exactly flat for non-zero heterogeneity. Note that this behavior is exactly what is observed in the data, where real economic variables are bounded, and the distribution of firms indeed has fewer firms in the upper tail than the Pareto benchmark would suggest (Stylized Fact 3).[12] The way in which the distribution of span of control differs from the full Pareto distribution is illustrated through comparative statics exercices in Figures 21, 22, and 23 in Appendix B.

Two Levels of Hierarchical Organization. Now that we have established that span of control down one level of hierarchical organization is a shifted Pareto of coefficient two in the uniform case, let us generalize this for the case of two layers of hierarchical organization. In particular, a case of interest is for the span of control of top managers, which pins down the equilibrium size distribution of firms, since each firm is run by a top manager:

$$N^L_{L \to L-1}(x_L) = \frac{1}{h\sqrt{(1-z^L_L)^2 + \frac{2}{h}(1-x_L)}} \sim \frac{1}{\sqrt{2h}} \frac{1}{\sqrt{1-x_L}}.$$

Of course, when $L > 1$, the agents below the top managers themselves manage other workers. Their span of control in turn is given by:

$$N^{L-1}_{L-1 \to L-2}(x_{L-1}) = \frac{1}{h\sqrt{(1-z^{L-1}_{L-1})^2 + \frac{2}{h}(z^{L-1}_L - x_{L-1})}}$$

[12] There is an ongoing fierce debate in the literature about whether the distribution of firm sizes is Pareto or log-normal. The same debate rages about the distribution of city sizes (see Eeckhout (2004), Levy (2009), Eeckhout (2009)), as well as about that of trade flows (Head et al. (2014)). The model presented here generates a shifted Pareto only in the upper tail: in particular, the distribution of small firms depends on the distribution of skills; and the way in which the Pareto is shifted in the upper tail depends on the level of skill heterogeneity as well as on helping time.

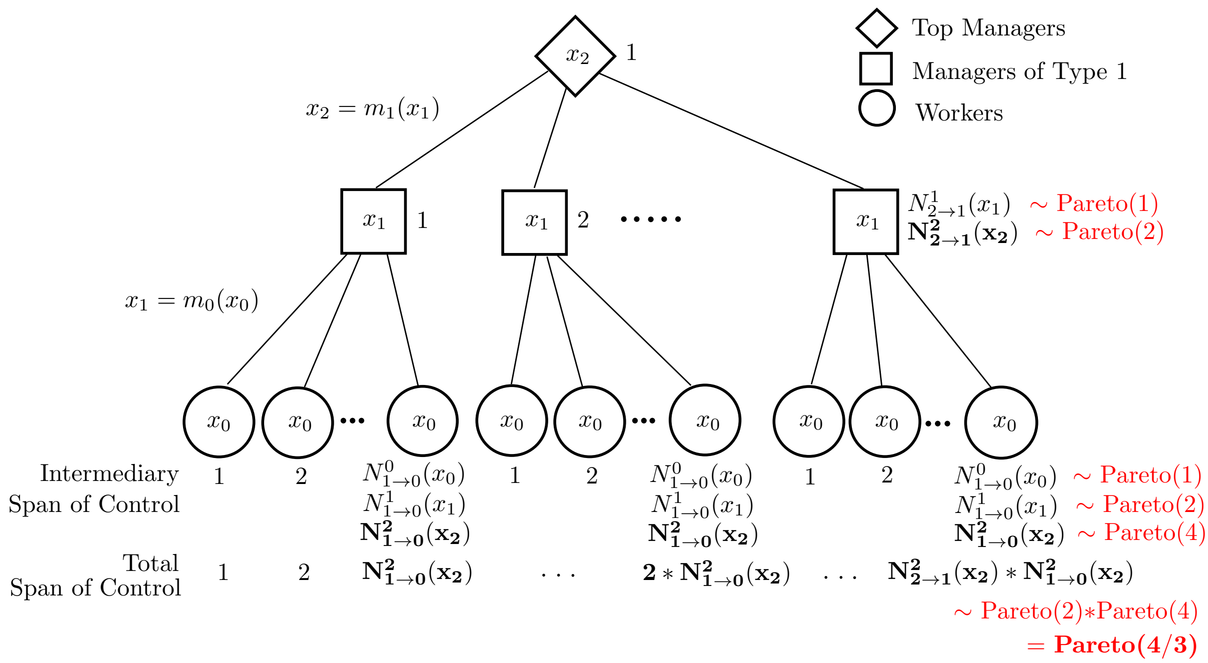
Expressed as a function of managers beliefs, this span of control function is given by:

$$N^L_{L-1\to L-2}(x_L) = N^{L-1}_{L-1\to L-2}\left(m_L^{-1}(x_L)\right)$$
$$N^L_{L-1\to L-2}(x_L) = \frac{1}{h\sqrt{(1-z^L_{L-1})^2 - \frac{2}{h}(1-z^L_L) + \frac{2}{h}\sqrt{(1-z^L_L)^2 + \frac{2}{h}(1-x_L)}}}.$$

Similarly, this is becomes arbitrarily close to a Pareto distribution of tail coefficient equal to four, when Δ or h become small, or when the number of layers increases, as for $x_L \in [z^L_L, 1]$:

$$N^L_{L-1\to L-2}(x_L) \sim \frac{1}{\sqrt[4]{8h}}\frac{1}{\sqrt[4]{1-x_L}}.$$

Figure 9: Span of Control: Pareto Coefficients, $L = 2$ Layers



Note: This chart complements Figure 5 in giving an intuition for the formula $2^L/(2^L-1)$ in the case $L = 2$. The Pareto distribution for total span of control has a coefficient equal to $4/3 = 2^2/(2^2-1)$, because of the multiplication of a Pareto with coefficient 2, that coming from the span of control of intermediary managers $N^2_{2\to 1}$, and one with coefficient 4, that originating from the span of control of intermediary managers on workers, in the space of top managers' skills $N^2_{1\to 0}$.

Importantly, the total span of control of an upper level managers over workers down two levels of hierarchical organization is given by the product of the two spans of control, since agents of level $L-1$ themselves manage agents of level of $L-2$. The total span of control of a top level manager down two levels of hierarchical organization is therefore given by:

$$N^L_{L\to L-2}(x_L) = N^L_{L\to L-1}(x_L) N^L_{L-1\to L-2}(x_L)$$
$$N^L_{L\to L-2}(x_L) = \frac{1}{h^2\sqrt{(1-z^L_L)^2 + \frac{2}{h}(1-x_L)}} \frac{1}{\sqrt{(1-z^L_{L-1})^2 - \frac{2}{h}(1-z^L_L) + \frac{2}{h}\sqrt{(1-z^L_L)^2 + \frac{2}{h}(1-x_L)}}}.$$

Again, one can approximate this as Δ or h become small, or when the number of layers increases by:

$$N^L_{L\to L-2}(x_L) \sim \frac{1}{\sqrt{2h}}\frac{1}{\sqrt{1-x_L}}\frac{1}{\sqrt[4]{8h}}\frac{1}{\sqrt[4]{1-x_L}} \sim \frac{1}{\sqrt{2h}\sqrt[4]{8h}}\frac{1}{{}^{3/4}\!\sqrt{1-x_L}}.$$

This is a Pareto distribution with coefficient equal to 4/3. The steps of the reasoning for $L=2$ levels of Hierarchical Organization are illustrated on Figure 9, where the firm has only two levels of hierarchical organization, in total. The total span of control of the top manager in that case is given by the multiplication of his span of control of managers of type 1 and of each one of these managers of type 1 on workers. The first distribution of span of control is given by a Pareto of coefficient two, as was explained before, and the second one if given by a Pareto of coefficient four.

Multiple Levels of Hierarchical Organization. More generally, what we have shown in the case $L=2$ generalizes to any L levels of hierarchical organization. By recursion, the span of control distribution down one level of hierarchical organization, seen in the space of managers' of type L's skills, of $N^L_{n+1\to n}(x_L)$ is a shifted Pareto distribution of coefficient equal to 2^{L-n}, and allows to state the following lemma.

Lemma 4 (One Level, Different Space, Uniform Case)**.** *Let the density of skills $g(.)$ be uniform with heterogeneity Δ. The span of control between adjacent levels of hierarchical organization in the space of higher managers' skills, denoted by $N^L_{n+1\to n}(x_L)$ is a shifted Pareto distribution of coefficient equal to 2^{L-n}:*

$$N^L_{n+1\to n}(x_L) \sim \text{Pareto}(2^{L-n}).$$

For conciseness, I do not express these shifted Paretos explicitly. For example, the case with $L=n+2$ was solved previously:

$$N^L_{L-1\to L-2}(x_L) = \frac{1}{h\sqrt{(1-z^L_{L-1})^2 - \frac{2}{h}(1-z^L_L) + \frac{2}{h}\sqrt{(1-z^L_L)^2+\frac{2}{h}(1-x_L)}}}.$$

The shifted Paretos are given explicitly by the formula:

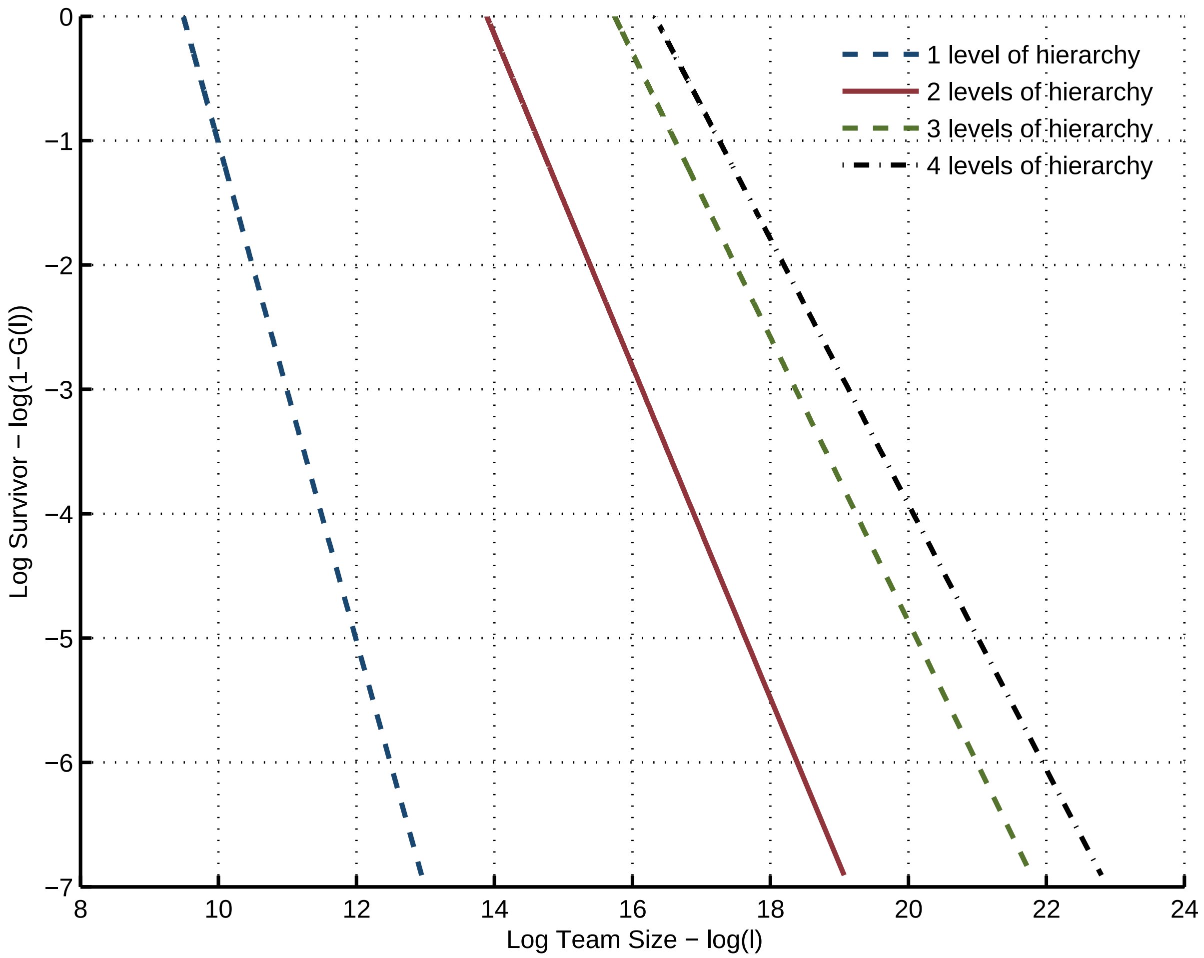
$$N^L_{n+1\to n}(x_L) = \frac{1}{h\left(1-m_n^{-1}\circ ...\circ m_{L-1}^{-1}(x_L)\right)} = \frac{1}{h\left[1-\left(\bigotimes_{l=n}^{L-1} m_l^{-1}\right)(x_L)\right]},$$

where each one of the assignment functions are given as a function of the endogenous cutoffs $\{z^L_l\}^L_{l=1}$ in Proposition 1, namely:

$$\forall l\in\{0,...,L-1\},\quad \forall x_{l+1}\in[z^L_{l+1}, z^L_{l+2}],\quad m_l^{-1}(x_{l+1}) = 1-\sqrt{(1-z^L_{l+1})^2-\frac{2}{h}(z^L_{l+2}-x_{l+1})}.$$

Proof. The proof for the tail coefficient of the Pareto proceeds by induction. The case for $L=1$ was shown in Lemma 3. Assume that Lemma 4 is true for some $L-1$ with $L\geq 2$, let us show

Figure 10: Distribution of Firm, Plant, Team Sizes, Log-Log Scale, $\Delta = 7/10$, $h = 8/10$, Uniform Distribution, with $L = 4$



Note: This Figure represents the theoretical distribution of span of control for top managers over employees down one to four levels or hierarchical organization, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta} \mathbb{1}\,(1-\Delta, 1)\,(z) + \mathbb{1}\,(1, +\infty)\,(z)$. The slopes on a log-log scale are -2, $-4/3$, $-8/7$ and $-16/15$ respectively. It can be compared with Figure 1 corresponding to the data counterpart.

it is true for L. Noting that:

$$N^L_{n+1\to n}(x_L) = \frac{1}{h\left(1 - m_n^{-1} \circ \ldots \circ m_{L-1}^{-1}(x_L)\right)} = \frac{1}{h\left[1 - \left(\bigotimes_{l=n}^{L-1} m_l^{-1}\right)(x_L)\right]}$$

$$\sim \frac{1}{h\sqrt{1 - \bigotimes_{l=n+1}^{L-1}(m_l^{-1})(x_L)}} \quad \text{from Lemma 3.}$$

By induction hypothesis:

$$\frac{1}{1 - \bigotimes_{l=n+1}^{L-1}(m_l^{-1})(x_L)} \sim \text{Pareto}(2^{L-n-1}),$$

where Pareto(2^{L-n-1}) denotes a shifted Pareto distribution with a tail coefficient equal to 2^{L-n-1}. Then:

$$N^L_{n+1\to n}(x_L) \sim \text{Pareto}(2^{L-n}),$$

which proves the proposition for L and concludes the proof by induction. $\square$

From Lemma 4, we are able to state the main result of the paper in the case of a uniform distribution in Lemma 5.

Lemma 5 (L Levels, Uniform Case). *Let the density of skills $g(.)$ be uniform with heterogeneity Δ. Total span of control of managers of type n on levels of hierarchical organization down L levels, denoted by $N^n_{n\to n-L}(x_n)$ is a shifted Pareto distribution of coefficient equal to $2^L/(2^L-1)$:*

$$N^{n+L}_{n+L\to n}(x_{n+L}) \sim \text{Pareto}\left(\frac{2^L}{2^L-1}\right).$$

Proof. This proposition, which is the main result of the paper, is just a corollary of the previous lemma since:

$$N^L_{L\to 0}(x_L) = N^L_{L\to L-1}(x_L) * N^L_{L-1\to L-2}(x_L) * ... * N^L_{1\to 0}(x_L) = \prod_{l=0}^{L-1} N^L_{L-l\to L-l-1}(x_L).$$

Applying the previous Lemma 4 to each one of the terms in the product $N^L_{L-l\to L-l-1}(.)$, which are therefore distributed according to shifted Paretos with a coefficient $2^{L-(L-l-1)} = 2^{l+1}$, allows to conclude by noting that the Pareto coefficient α_L is solution of:

$$\frac{1}{\alpha_L} = \sum_{l=0}^{L-1} \frac{1}{2^{l+1}} = \frac{1}{2}\frac{1-\frac{1}{2^L}}{1-\frac{1}{2}} = \frac{2^L-1}{2^L} \quad \Rightarrow \quad \alpha_L = \frac{2^L}{2^L-1}.$$

$\square$

Let us denote by α_L the tail coefficient on the Pareto distribution for span of control given as $\alpha_L = 2^L/(2^L-1)$. The following Table gives the values of α_L for $L = 1, 2, 3, 4, 5, ..., +\infty$.

L	1	2	3	4	5	...	$+\infty$
α_L (exact)	2	4/3	8/7	16/15	32/31	...	1
α_L (approx.)	2.00	1.33	1.17	1.07	1.03	...	1.00

These different coefficients are illustrated on Figure 10. In particular, the theoretical coefficients are close to those observed in the data (see Figure 1), and converge to 1 as the number of layers increases. Of course, the uniform distribution of skills is a very particular one. However I show in the next section that what was shown here is more generally true in the upper tail of the span of control distributions for any smooth distribution function, provided its density is continuous ad bounded away from zero for higher skills. This is an arguably rather limited set of assumptions.

3.3.2 General Case

I next show that no matter what the distribution of skills in the population is, provided that it is bounded away from zero for high skills, and continuous, the Pareto distributions for span of control obtain in the upper tail, exactly in the same way as in the uniform case. Let me first formalize these assumptions in Assumption 3.

Assumption 3 (Density Function of Skills). *The density function of skills $g(.)$ is continuous on $[1-\Delta, 1]$, and bounded away from zero near one:*

$$\exists m > 0, \quad \exists \eta > 0, \quad \forall x \in [1-\eta, 1], \quad g(x) \geq m > 0.$$

The fact that the conclusions on Pareto distributions, and in particular on their tail coefficients, do not depend on the underlying distribution of skills (apart from them satisfying Assumption 3) may seem as a perhaps surprising result at first sight. The intuition is that as the number of layers increases (or heterogeneity goes to zero), the distribution of skills which matters is drawn for a smaller segment of the total skill distribution. To the limit, any function can be approximated by a uniform distribution under minimal regularity assumptions. In particular, the very large heterogeneity underlying the upper tail of the span of control distribution actually hinges on the behavior of a relatively small segment of managers' and workers' skill distribution. Proposition 3 only generalizes Lemma 3 in the case of a uniform distribution to any distribution of skills satisfying Assumption 3.

Proposition 3 (One Level, General Case). *The span of control of a manager of type $l+1$ over employees down one level of hierarchical organization $N_{l+1\to l}^{l+1}(.)$ is arbitrarily close in the upper tail to the shifted Pareto distribution of coefficient two obtained in the uniform case:*

$$\forall \epsilon > 0, \quad \exists \eta > 0, \quad \forall x_{l+1} \in [z_{l+2}^L - \eta, z_{l+2}^L],$$
$$\frac{1}{h\sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1-\epsilon)}(z_{l+2}^L - x_{l+1})}} \leq N_{l+1\to l}^{l+1}(x_{l+1}) \leq \frac{1}{h\sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1+\epsilon)}(z_{l+2}^L - x_{l+1})}}.$$

This will be denoted as:

$$N_{l+1\to l}^{l+1}(x_{l+1}) \simeq \text{Pareto}(2).$$

The shifted Pareto on the left-hand side is a lower bound in the upper tail:

$$\frac{1}{h\sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1-\epsilon)}(z_{l+2}^L - x_{l+1})}} \sim \frac{\sqrt{A(1-\epsilon)}}{\sqrt{2h}} \frac{1}{\sqrt{z_{l+2}^L - x_{l+1}}}.$$

Symmetrically, the shifted Pareto on the right-hand side is an upper bound:

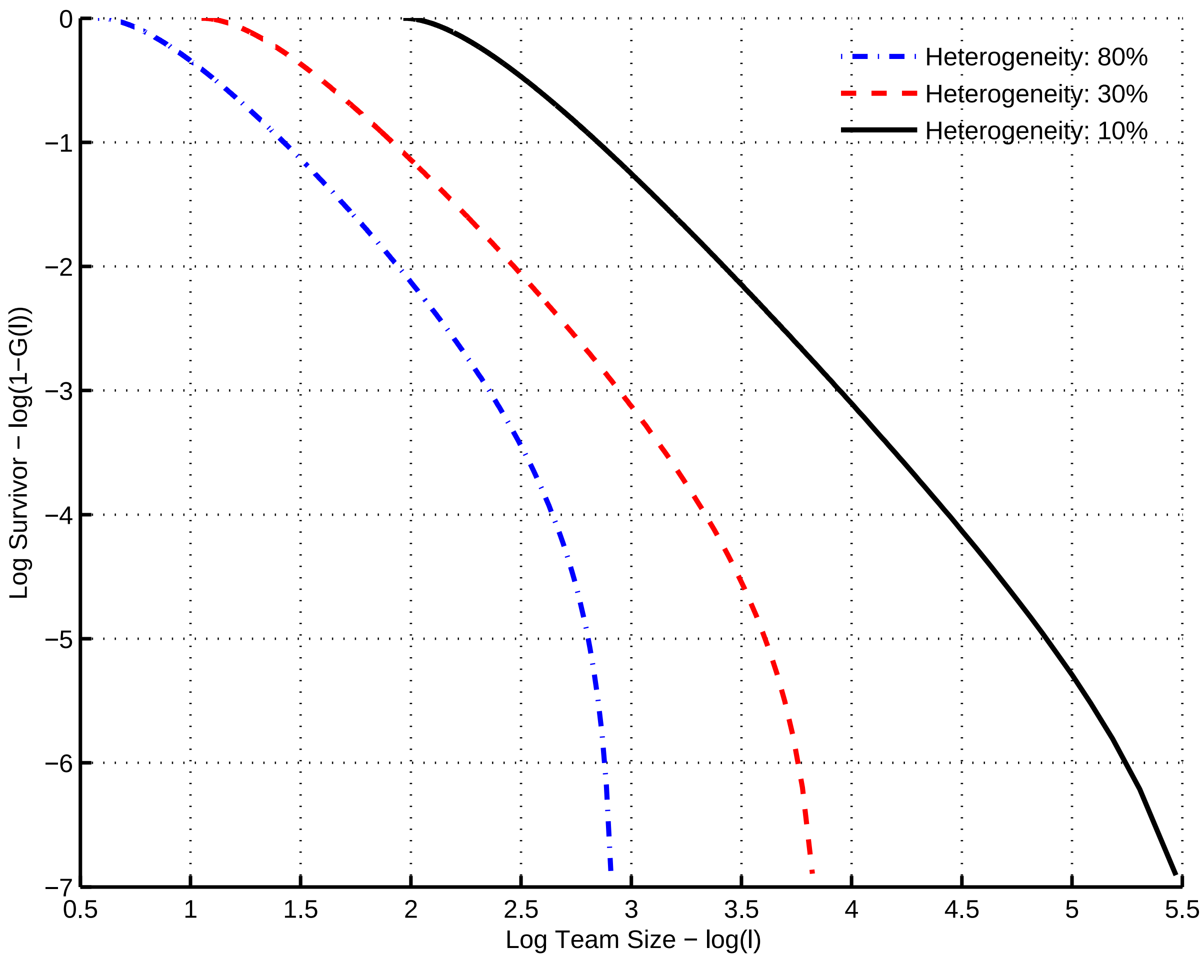
$$\frac{1}{h\sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1+\epsilon)}(z_{l+2}^L - x_{l+1})}} \sim \frac{\sqrt{A(1+\epsilon)}}{\sqrt{2h}} \frac{1}{\sqrt{z_{l+2}^L - x_{l+1}}}.$$

As ϵ goes to zero, these shifted Paretos become infinitesimally close to each other, so that $N_{l+1\to l}^{l+1}(x_{l+1})$ is also arbitrarily close a shifted Pareto in the upper tail, which I will denote in the following as:

$$N_{l+1\to l}^{l+1}(x_{l+1}) \simeq \text{Pareto}(2).$$

As an illustration, Figure 11 plots the equivalent of Figure 8 in the case where the distribution of skills is given by an increasing function. One can see how the behavior in the lower tail of the distribution differs a bit from the uniform case, but that the upper tail behavior, and in particular the slope -2 for the survivor function on a log-log scale, remains unchanged. (see the Online Appendix, for more Figures in the increasing case)

Figure 11: Distribution of Span of Control Down One Level of Hierarchical Organization ("Teams"), Log-Log Scale, $h = 70\%$ - Increasing Skill Distribution



Note: This figure gives the theoretical distribution of team sizes, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{(z-(1-\Delta))^2}{\Delta^2}\mathbb{1}\left[1-\Delta,1\right](z) + \mathbb{1}\left[1,+\infty\right](z)$. One can compare to Figure 8 for an equivalent in the uniform case. Note that the lower tail behavior differs a bit with this increasing distribution. However, similarly as in the uniform case, as heterogeneity Δ goes to zero, the distribution approaches the full Pareto distribution with coefficient two in the upper tail.

Proof. From the market clearing equation for skills:

$$m_l'(x_l) = h(1-x_l)\frac{g(x_l)}{g(m_l(x_l))}$$

Denote by A the limit of $\frac{g(x_l)}{g(m_l(x_l))}$ when $x_l \to z_{l+1}^L$. Fix $\epsilon > 0$ as small as one wants. Then there exists η such that for all $x_l \in [z_{l+1}^L - \eta, z_{l+1}^L]$, we have:

$$Ah(1-x_l)(1-\epsilon) \leq m_l'(x_l) \leq Ah(1-x_l)(1+\epsilon)$$

Integrating between $[x_l, z_{l+1}^L]$, this gives:

$$\begin{aligned}
&Ah(1-\epsilon)\left(z_{l+1}^L - \frac{(z_{l+1}^L)^2}{2} - x_l + \frac{x_l^2}{2}\right) \leq z_{l+2}^L - m_l(x_l) \leq Ah(1+\epsilon)\left(z_{l+1}^L - \frac{(z_{l+1}^L)^2}{2} - x_l + \frac{x_l^2}{2}\right) \\
&\Rightarrow \quad 2\frac{z_{l+2}^L - m_l(x_l)}{Ah(1+\epsilon)} - 2z_{l+1}^L + (z_{l+1}^L)^2 \leq x_l^2 - 2x_l \leq 2\frac{z_{l+2}^L - m_l(x_l)}{Ah(1-\epsilon)} - 2z_{l+1}^L + (z_{l+1}^L)^2 \\
&\Rightarrow \quad 2\frac{z_{l+2}^L - m_l(x_l)}{Ah(1+\epsilon)} + \left(1 - z_{l+1}^L\right)^2 \leq (1-x_l)^2 \leq 2\frac{z_{l+2}^L - m_l(x_l)}{Ah(1-\epsilon)} + \left(1 - z_{l+1}^L\right)^2 \\
&\Rightarrow \sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1+\epsilon)}(z_{l+2}^L - x_{l+1})} \leq 1 - m_l^{-1}(x_{l+1}) \leq \sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1-\epsilon)}(z_{l+2}^L - x_{l+1})} \\
&\Rightarrow \frac{1}{h\sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1-\epsilon)}(z_{l+2}^L - x_{l+1})}} \leq N_{l+1\to l}^{l+1}(x_{l+1}) \leq \frac{1}{h\sqrt{(1-z_{l+1}^L)^2 + \frac{2}{Ah(1+\epsilon)}(z_{l+2}^L - x_{l+1})}}.
\end{aligned}$$

□

Similarly, Proposition 4 generalizes Lemma 4 and Proposition 5 generalizes Lemma 5, to a case where Assumption 3 holds.

Proposition 4 (One Level, Different Space, General Case)**.** *The span of control between adjacent levels of hierarchical organization in the space of higher managers' skills, denoted by* $N_{n+1\to n}^L(x_L)$ *is arbitrarily close to the shifted Pareto distribution of coefficient equal to* 2^{L-n} *obtained in the uniform case:*

$$N_{n+1\to n}^L(x_L) \simeq \text{Pareto}(2^{L-n}).$$

Proposition 5 (L Levels, General Case)**.** *Total span of control of managers of type n on levels of hierarchical organization down L levels, denoted by* $N_{n\to n-L}^n(x_n)$ *is arbitrarily close to the shifted Pareto distribution of coefficient equal to* $2^L/(2^L-1)$ *obtained in the uniform case:*

$$N_{n+L\to n}^{n+L}(x_{n+L}) \simeq \text{Pareto}\left(\frac{2^L}{2^L-1}\right).$$

Proof. These two Propositions result directly from Proposition 3 in the same way as Lemmas 4 and 5 result from Lemma 3. The proof for the tail coefficient of the approximating shifted

Pareto proceeds by induction, using the case for $L = 1$ as a starting statement, assuming that the statement is true for some $L-1$ with $L \geq 2$ and showing it is true for L. Similarly, the proof of Proposition 5 follows from multiplying individual span of control distributions together, which gives the same formula for the total span of control distributions' tail coefficient $\alpha_L = 2^L/(2^L - 1)$. Refer to the proofs of Lemmas 4 and 5 for details, replacing $\sim$ by $\simeq$ equivalence relations. $\square$

4 Labor Income Distribution [Very Preliminary and Incomplete]

The equilibrium spans of control for managers was derived without any reference to equilibrium prices. This is because the solution is more easily found through the planner's problem. However one can calculate how this competitive equilibrium arises in practice, and in particular use this to calculate the utility that every agent can achieve in the equilibrium of this model, which is decentralized through internal labor markets. Interestingly, the income distribution in the upper tail looks similar to the one that is observed in the data. Due to short-run data constraints, this part is very preliminary.

4.1 Fixed Component

As is well known in these hierarchies models, the competitive equilibrium can be decentralized with a set of transfers inside the firm, which can be viewed as internal labor markets. And just as the span of control distribution in the upper tail does not depend on the distribution of skills in the population, the upper tail of the labor income distribution will be essentially invariant to the shape of the skill distribution.

Denote by $w_0(.)$ the wage that production workers get, and by convention $w_L(.)$ the wage that the top level of the hierarchy "implicitly" gets through production, given therefore by:

$$\forall x_L \in [z_L^L, z_{L+1}^L], \quad w_L(x_L) = x_L.$$

The return for a manager of level l is then defined by what he gets as a wage minus what he pays as wages to the managers or workers down one level of hierarchical organization (see Garicano and Rossi-Hansberg (2006) for details):

$$R_{l+1}(x_{l+1}) \equiv \frac{w_{l+1}(x_{l+1}) - w_l(x_l)}{h(1 - x_l)}$$

By convention, let us similarly define the return function for a production worker as:

$$\forall x_0 \in [z_0^L, z_1^L], \quad R_0(x_0) = w_0(x_0).$$

All the income a production worker gets is indeed its wage: he is not able to leverage his skills, and neither does he need to pay any employee.

Proposition 6 (Internal Labor Markets)**.** *The implicit wages received by each layer of the hierarchy are given through inverse recursion, starting from $w_L(x_L) = x_L$ by:*

$$\forall l \in \{0, ..., L-1\}, \quad w_l(x_l) = (1-x_l) \int_{z_l^L}^{x_l} \frac{w_{l+1}(m_l(u))}{(1-u)^2} du + (1-x_l) \frac{w_l(z_l^L)}{1-z_l^L}$$

The initial conditions for these differential equations in integral form are solution of:

$$\forall l \in \{0, ..., L-1\}, \quad R_{l+1}(z_{l+1}^L) = R_l(z_{l+1}^L).$$

Proof. By optimization of a manager in level $l+1$, who chooses the skill x_l of his workers or intermediary managers to maximize his expected income, it must be that, for all $l \in \{0, ..., L-1\}$:

$$m_l^{-1}(x_{l+1}) = \arg\max_{x_l} \frac{w_{l+1}(x_{l+1}) - w_l(x_l)}{h\,(1-x_l)} \quad \Rightarrow \quad (1-x_l)\, w_l'(x_l) + w_l(x_l) = w_{l+1}(m_l(x_l))$$

$$\Rightarrow \quad \frac{d}{dx_l}\left(\frac{w_l(x_l)}{1-x_l}\right) = \frac{w_{l+1}(m_l(x_l))}{(1-x_l)^2} \quad \Rightarrow \quad \frac{w_l(x_l)}{1-x_l} - \frac{w_l(z_l^L)}{1-z_l^L} = \int_{z_l^L}^{x_l} \frac{w_{l+1}(m_l(u))}{(1-u)^2} du.$$

The L indifference conditions for agents with skills z_l^L with $l \in \{0, ..., L-1\}$, which are initial conditions for the wage functions, are then written as follows:

$$\forall l \in \{0, ..., L-1\}, \quad R_{l+1}(z_{l+1}^L) = R_l(z_{l+1}^L)$$

By reverse induction, this defined the whole sequence of wage functions $w_l(.)$ for all $l \in \{0, ..., L-1\}$, starting from the known $w_L(x_L) = x_L$. □

Of particular interest is the theoretical distribution of top managers' wages, both because it does not depend on the underlying distribution of skills as for the distribution of span of control, and because inequality issues have gained some prominence in the public debate recently. We are therefore interested in the distribution of $R_L(x_L)$, which obtains rather straighforwardly from Proposition 6.

Proposition 7 (Top Labor Income Distribution)**.** *The incomes at the top are given as a function of the skill of top managers x_L by:*

$$R_L(x_L) = R_L\left(z_L^L\right) + \int_{z_{L-1}^L}^{m_{L-1}^{-1}(x_L)} \frac{g(u)}{g\left(m_{L-1}(u)\right)} du.$$

More generally, total returns of managers are given as:

$$R_{l+1}(x_{l+1}) = R_{l+1}\left(z_{l+1}^L\right) + \int_{z_l^L}^{m_l^{-1}(x_{l+1})} w_{l+1}'\left(m_l(u)\right) \frac{g(u)}{g\left(m_l(u)\right)} du.$$

Proof. The first result follows straightforwardly from the second, which is more general: set $l = L-1$ and use $w_L(x_L) = x_L$ so that $w_L'(x_L) = 1$ for all $x_L \in [z_L^L, 1]$. We need to integrate result in Proposition 6 by parts:

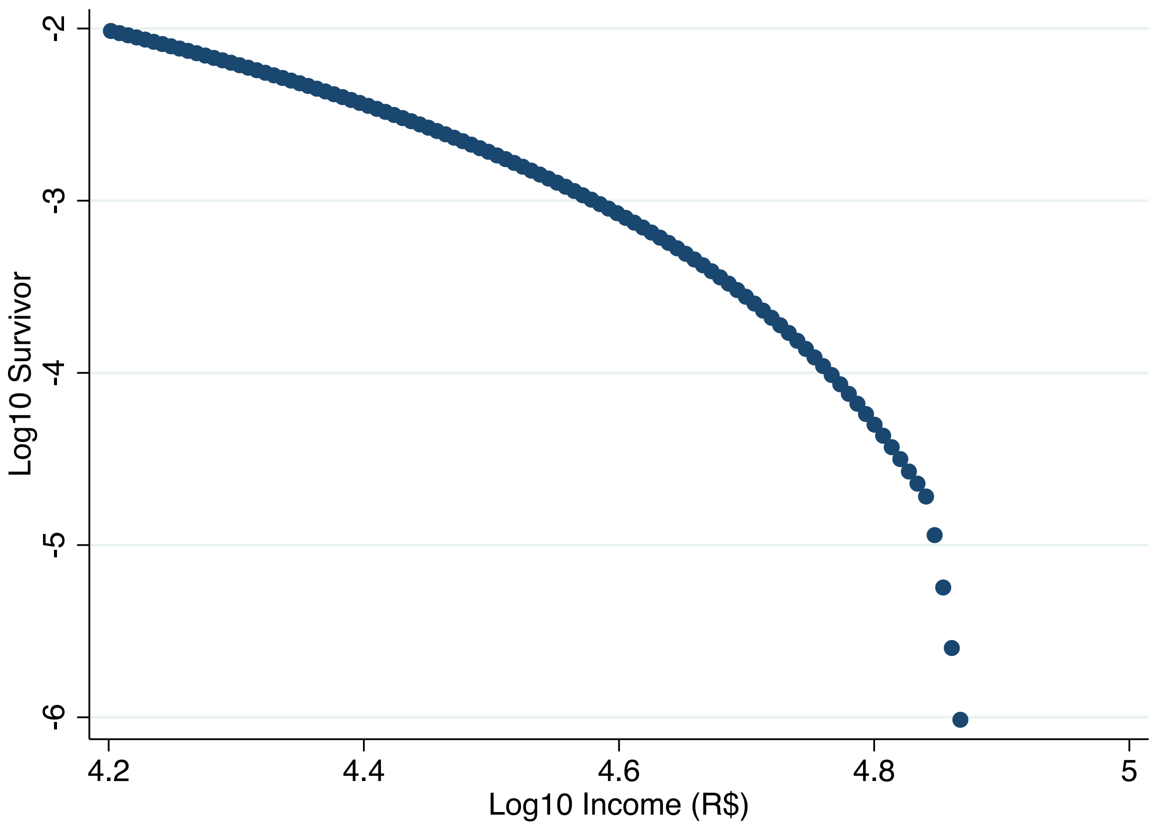
$$\int_{z_l^L}^{x_l} \frac{w_{l+1}\left(m_l(u)\right)}{(1-u)^2} du = \left[\frac{w_{l+1}\left(m_l(u)\right)}{1-u}\right]_{z_l^L}^{x_l} - \int_{z_l^L}^{x_l} \frac{m_l'(u) w_{l+1}'\left(m_l(u)\right)}{1-u} du.$$

Using then Proposition 1 to replace m_l' through $m_l'(u)g\left(m_l(u)\right) = h(1-u)g(u)$, and replacing $w_l(x_l)$ given by Proposition 6:

$$\begin{aligned} R_{l+1}\left(x_{l+1}\right) &= \frac{w_{l+1}\left(x_{l+1}\right) - w_l\left(x_l\right)}{h\left(1-x_l\right)} \\ &= \frac{w_{l+1}\left(x_{l+1}\right)}{h\left(1-x_l\right)} - \frac{1}{h}\left[\frac{w_{l+1}\left(m_l(u)\right)}{1-u}\right]_{z_l^L}^{x_l} + \int_{z_l^L}^{x_l} w_{l+1}'\left(m_l(u)\right) \frac{g(u)}{g\left(m_l(u)\right)} du - \frac{w_l\left(z_l^L\right)}{h\left(1-z_l^L\right)} \\ R_{l+1}\left(x_{l+1}\right) &= \frac{w_{l+1}\left(z_{l+1}^L\right) - w_l\left(z_l^L\right)}{h\left(1-z_l^L\right)} + \int_{z_l^L}^{m_l^{-1}\left(x_{l+1}\right)} w_{l+1}'\left(m_l(u)\right) \frac{g(u)}{g\left(m_l(u)\right)} du. \end{aligned}$$

$\square$

Figure 12: Example of Top 1% Theoretical Distribution



Note: In this figure, the theoretical countercumulative distribution function of Top 1% wages is represented, with the following parameters: $L = 4$, $s = 1$, $h = 70\%$ and $\Delta = 70\%$. It should be compared to the empirical wage distribution function in the Brazilian data on Figure 16.

4.2 Variable Component

As was seen in the previous section, the Pareto distribution for span of control of managers does not translate mechanically into a Pareto distribution for their *fixed* wages, as talented managers give an endogenous and increasing part of their wages to more talented lower-level managers. However, the empirical wage distribution resembles a lot this distribution, at least in the Brazilian data on Figure 16.

However, it is useful to note that since the distribution of span of controls of managers is distributed Pareto, any unexpected shock on (for example) the price of the good sold, which was until now taken to be one, will potentially translate into Pareto distributions for the unexpected income of managers with the same coefficient, depending on how this unexpected windfall is shared between different levels of the hierarchy.

In line with the theory, the variable component is given by a Pareto distribution in the upper tail, at least for US Chief Executive Officers in the ExecuComp Data, as can be seen on preliminary evidence on Figure 18. However, more remains to be investigated.

5 Conclusion

In this paper, I have developed a new static, microfounded, hierarchy-based model of Zipf's law for firm size's distribution. Functional form assumptions are very commonly based on the observed Zipf's law for firm sizes, such as the distribution of productivities in the heterogenous firms literature. My model suggests on the contrary that no mechanic link between the two exists, but that Zipf's law arises regardless of the underlying distribution of productivities. As already emphasized in the literature review, the implications of the finding therefore span many different parts of the economics literature, from the magnitude of misallocations stemming from size-dependent regulations to the sources of high labor income inequality. In particular, the fact that heterogeneity can very well be bounded, even infinitesimal, and yet yield what looks like an unbounded heterogeneity in outcomes, suggest that more parsimonious structural models, in the sense of incorporating less ex-ante hetereogeneity on primitives, can be written to explain the same observed phenomena.

Unfortunately, the model is very much a first pass in many respects. Its most important limitation is perhaps that skills are assumed to be a given in the model, and that all agents work the same, so that unequal wages only result from innate abilities. I conjecture that the span of control distributions would remain the same in a larger class of extended models with education and elastic labor supply, as differences in education abilities are in many ways ex-post isomorphic to ex-ante differences in skills, and labor supply decisions would only amplify the forces at hand in the paper. Yet for other issues, this simplification is clearly a problem. Consider optimal taxation. If abilities were all innate, a Rawlasian planner would be able to redistribute all unequal labor income in equal proportions at no efficiency cost. Yet, skills are not entirely innate, and doing so would discourage both skill investment through education as well as more labor supply from agents at the top of the distribution; because the model is purely competitive, prices provide exactly the right incentives to educate themselves more, or to work more.[13] Therefore, in its current form, the model remains essentially silent on the key efficiency-equity tradoffs. Not completely, though: since optimal taxation models are very often calibrated using an underlying Pareto distribution in agents' productivity (see, for example, Saez (2001)), the model holds promises for the structural estimation of optimal income tax rates. The model even suggests a way through which luck could (in principle) be told apart from innate skills: the Pareto in wages here only comes from luck. It is probable that education and labor supply would lead wages to behave in the same way in the cross-section, but more work remains to be done.

The model presented here attributes all heterogeneity in firm sizes to the heterogeneity in skills of the workers and managers they employ, which is consistent with some empirical evidence attributing most wage differentials to person effects instead of firm effects (for example Abowd et al. (1999)). Even though the model really matches the data very well, attributing all heterogeneity to agents and none to firms is certainly a too extreme assumption. An interesting extension of the model would therefore be to allow for firm specific heterogeneity in excess of managers' respective talents. Along these lines, an important factor that the model has neglected is firm's capital, including management capital. In the context of the model, one would model firm heterogeneity as different values for communication costs h, which can proxy for how much a firm has invested in better management tools, or extend the Garicano (2000) model to one of moral hazard between workers and entrepreneurs along the lines of Bloom et al. (2012).

The fact that the skills distribution is exogenously fixed is not only an issue for normative economics, but also for a more thorough positive investigation. This is arguably a good assumption in the short run, for example after the arrival of a new technology; or to capture the fact that some skills cannot be taught, or that some are in any case always better at learning than others. If one however allows for homogenous learning ability, then the model could help answer a number of fascinating questions. In particular, it is likely that the heterogeneity in skills generating a very unequal income distribution will not forever remain as the incentives to invest in skills are then very high. A dynamic version of the model would therefore allow to investigate the "race between technological development and education" which may have been the cause of the recent rise in inequality (Tinbergen (1974), Katz and Murphy (1992)). On a same line of reasoning, a second limitation of the model is that the good being produced is itself exogenous. If one allows for innovation, and in particular for the fact that at the beginning of the product cycle only few entrepreneurs are able to design it, then the question would be to study the incentives to spread around this knowledge, to increase span of control (which increases when the skills of lower ranking agents increases); and the countervailing forces to retain knowledge, as the wage distribution is all the more rewarding to entrepreneurs that there are relatively few of them.

On a more general note, the model confirms that competitive forces, to generate the maximum level of complementarities, can yield very high level of heterogeneity in outcomes (here, firm sizes,

[13]This would likely be reinforced in the presence of learning externalities as in human capital accumulation models.

or employment) from an arbitrarily small level of heterogeneity on primitives. In Geerolf (2013), I have shown that this was true also for a model of collateralized lending with heterogenous beliefs and endogenous margins. In that case, very optimistic investors were able to manage a lot more capital than less optimistic ones, again according to a Pareto distribution, because this arrangement was constrained optimal and maximized ex-ante welfare. A natural conjecture is that Pareto distributions for allocations in fact obtain for a more general class of assignment models with complementarities. I leave this to future research.

References

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999), 'High Wage Workers and High Wage Firms', *Econometrica: Journal of the Econometric Society* **67**(2), 251–333.

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

Antràs, P. and Helpman, E. (2004), 'Global Sourcing', *Journal of Political Economy* **112**(3), 552–580.

Arkolakis, C. (2010), 'Market Penetration Costs and the New Consumers Margin in International Trade', *Journal of Political Economy* **118**(6), 1151–1199.

Arkolakis, C., Costinot, A. and Rodríguez-Clare, A. (2012), 'New Trade Models, Same Old Gains?', *American Economic Review* **102**(1), 94–130.

Atkeson, A. and Burstein, A. (2010), 'Innovation, Firm Dynamics, and International Trade', *Journal of Political Economy* **118**(3), 433–484.

Axtell, R. (2001), 'Zipf Distribution of US Firm Sizes', *Science* **293**, 1818–1821.

Bernard, A. B., Jensen, J. B., Redding, S. J. and Schott, P. K. (2007), 'Firms in International Trade', *Journal of Economic Perspectives* **21**(3), 105–130.

Bernard, A., Redding, S. J. and Schott, P. K. (2007), 'Comparative Advantage and Heterogeneous Firms', *Review of Economic Studies* **74**(1), 31–66.

Bloom, N., Sadun, R. and Reenen, J. V. (2012), 'The Organization of Firms Across Countries', *Quarterly Journal of Economics* **127**(4), 1663–1705.

Cabral, L. and Mata, J. (2003), 'On the Evolution of the Firm Size Distribution: Facts and Theory', *American Economic Review* **93**(4), 1075–1090.

Caliendo, L., Monte, F. and Rossi-Hansberg, E. (2014), 'The Anatomy of French Production Hierarchies', *Journal of Political Economy* (forthcoming).

Caliendo, L. and Rossi-Hansberg, E. (2012), 'The Impact of Trade on Organization and Productivity', *Quarterly Journal of Economics* (3), 1393–1467.

Champernowne, D. (1953), 'A Model of Income Distribution', *The Economic Journal* **63**(250), 318–351.

Chaney, T. (2008), 'Distorted Gravity: The Intensive and Extensive Margins of International Trade', *American Economic Review* **98**(4), 1707–1721.

Cooley, T. and Quadrini, V. (2001), 'Financial Markets and Firm Dynamics', *American Economic Review* **151**(3712), 867–8.

Eaton, J., Kortum, S. and Kramarz, F. (2011), 'An Anatomy of International Trade: Evidence from French firms', *Econometrica: Journal of the Econometric Society* **79**(5), 1453–1498.

Eeckhout, J. (2004), 'Gibrat's Law for (All) Cities', *American Economic Review* **94**(5), 1429–1451.

Eeckhout, J. (2009), 'Gibrat's Law for (All) Cities : Reply', *American Economic Review* **99**(4), 1676–1683.

Eeckhout, J. and Kircher, P. (2012), 'Assortative Matching With Large Firms: Span of Control over More versus Better Workers', *Working Paper* (February 2012).

Gabaix, X. (1999), 'Zipf's Law for Cities: an Explanation', *Quarterly Journal of Economics* (August), 739–767.

Gabaix, X. (2009), 'Power Laws in Economics and Finance', *Annual Review of Economics* **1**(1), 255–293.

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

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

Garicano, L. and Rossi-Hansberg, E. (2004), 'Inequality and the Organization of Knowledge', *American Economic Review, Papers and Proceedings* **94**(2), 197–202.

Garicano, L. and Rossi-Hansberg, E. (2006), 'Organization and Inequality in a Knowledge Economy', *Quarterly Journal of Economics* **121**(4), 1383–1435.

Geerolf, F. (2013), 'A Theory of Power Law Distributions for the Returns to Capital and of the Credit Spread Puzzle', *Toulouse School of Economics* (Working Paper).

Ghironi, F. and Melitz, M. (2005), 'International Trade and Macroeconomic Dynamics with Heterogeneous Firms', *Quarterly Journal of Economics* **120**(3), 865–915.

Gibrat, R. (1931), *Les Inégalités Economiques; Applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel*, Librairie du Recueil Sirey, Paris.

Head, K., Mayer, T. and Thoenig, M. (2014), 'Welfare and Trade Without Pareto', *American Economic Review, Papers and Proceedings* **104**(5), 310–316.

Helpman, E., Itskhoki, O. and Redding, S. (2010), 'Inequality and Unemployment in a Global Economy', *Econometrica, Journal of the Econometric Society* **78**(4), 1239–1283.

Helpman, E., Melitz, M. and Yeaple, S. R. (2004), 'Export Versus FDI with Heterogeneous Firms', *American Economic Review* **94**(1), 300–316.

Hopenhayn, H. (1992), 'Entry, Exit, and Firm Dynamics in Long Run Equilibrium', *Econometrica: Journal of the Econometric Society* **60**(5), 1127–1150.

Hopenhayn, H. and Rogerson, R. (1993), 'Job Turnover and Policy Evaluation: A General Equilibrium Analysis', *Journal of Political Economy* **101**(5), 915–938.

Hsieh, C. and Klenow, P. (2007), 'Misallocation and Manufacturing TFP in China and India', *Quarterly Journal of Economics* **124**(4), 1403–1448.

Jovanovic, B. (1982), 'Selection and the Evolution of Industry', *Econometrica: Journal of the Econometric Society* **50**(3), 649–670.

Katz, L. and Murphy, K. (1992), 'Changes in Relative Wages, 1963-1987 : Supply and Demand Factors', *Quarterly Journal of Economics* **107**(1), 35–78.

Kesten, H. (1973), 'Random Difference Equations and Renewal Theory for Products of Random Matrices', *Acta Mathematica* **131**(1), 207–248.

Klette, T. and Kortum, S. (2004), 'Innovating Firms and Aggregate Innovation', *Journal of Political Economy* **112**(5), 986–1018.

Kremer, M. and Maskin, E. (1996), 'Wage Inequality and Segregation by Skill', *NBER Working Paper 5718* .

Levy, M. (2009), 'Gibrat's Law for (All) Cities: Comment', *American Economic Review* **99**(4), 1672–1675.

Liegey, M. (2014), 'Search Frictions, the Use of Knowledge and the Labor Market', *Toulouse School of Economics, mimeo* .

Lucas, R. E. (1978), 'On the Size Distribution of Business Firms', *The Bell Journal of Economics* **9**, 508–523.

Luttmer, E. (2007), 'Selection, Growth, and the Size Distribution of Firms', *Quarterly Journal of Economics* (August), 1103–1144.

Mansfield, E. (1962), 'Entry, Gibrat's Law, Innovation, and the Growth of Firms', *American Economic Review* **151**(3712), 867–8.

Melitz, M. (2003), 'The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity', *Econometrica: Journal of the Econometric Society* **71**(6), 1695–1725.

Melitz, M. and Ottaviano, G. (2008), 'Market Size, Trade, and Productivity', *Review of Economic Studies* **75**(1), 295–316.

Melitz, M. and Trefler, D. (2012), 'Gains from Trade when Firms Matter', *Journal of Economic Perspectives* **26**(2), 91–118.

Radner, R. (1992), 'Hierarchy: The Economics of Managing', *Journal of Economic Literature* **30**(3), 1382–1415.

Restuccia, D. and Rogerson, R. (2008), 'Policy Distortions and Aggregate Productivity with Heterogeneous Establishments', *Review of Economic Dynamics* **11**(4), 707–720.

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

Rosen, S. (1981), 'The Economics of Superstars', *American Economic Review* **71**(11), 845–858.

Rosen, S. (1982), 'Authority, Control, and the Distribution of Earnings', *The Bell Journal of Economics* **13**, 311–323.

Rossi-Hansberg, E. and Wright, M. (2007), 'Establishment Size Dynamics in the Aggregate Economy', *American Economic Review* **97**(5), 1640–1666.

Roy, A. (1950), 'The Distribution of Earnings and of Individual Output', *The Economic Journal* **60**(239), 489–505.

Saez, E. (2001), 'Using Elasticities to Derive Optimal Income Tax Rates', *Review of Economic Studies* **68**(1), 205–229.

Sattinger, M. (1975), 'Comparative Advantage and the Distributions of Earnings and Abilities', *Econometrica: Journal of the Econometric Society* **43**(3), 455–468.

Simon, H. (1955), 'On a Class of Skew Distribution Functions', *Biometrika* **42**(3), 425–440.

Simon, H. and Bonini, C. (1958), 'The Size Distribution of Business Firms', *American Economic Review* **48**(4), 607–617.

Sutton, J. (1997), 'Gibrat's Legacy', *Journal of Economic Literature* **35**(March), 40–59.

Terviö, M. (2008), 'The difference that CEOs make: An assignment model approach', *American Economic Review* (1999), 642–668.

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

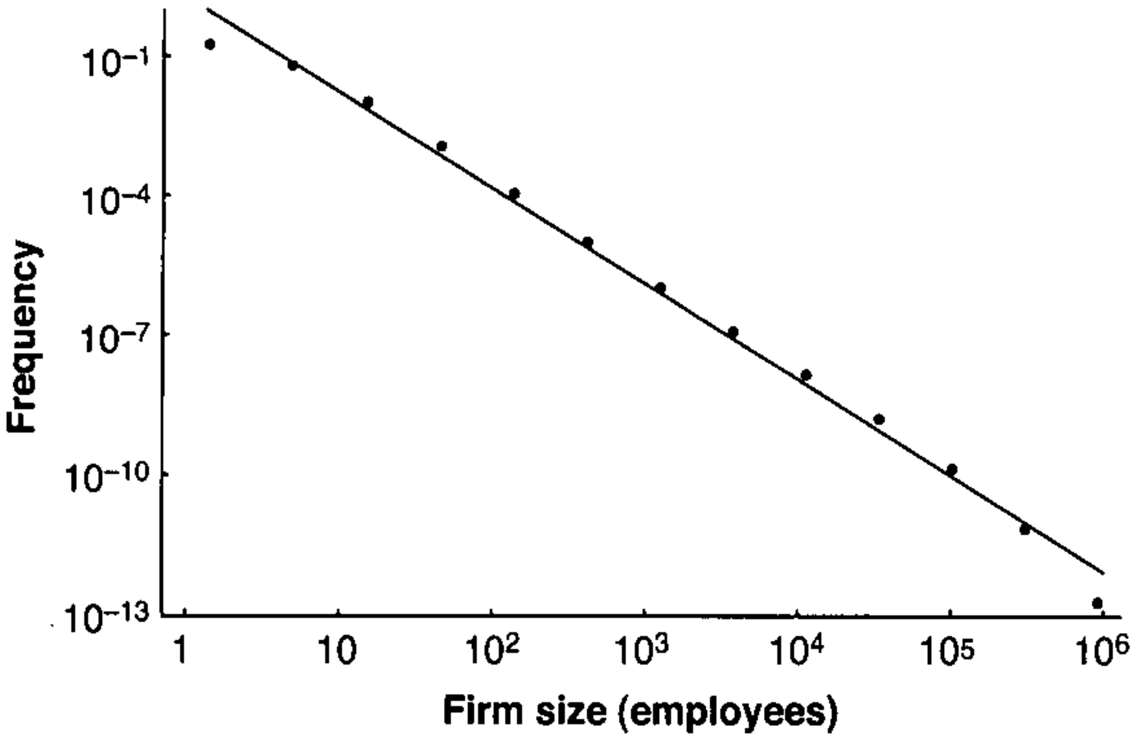
Tinbergen, J. (1974), 'Substitution of Graduate by Other Labour', *Kyklos* .

Zipf, G. K. (1949), *Human Behavior and the Principle of Least Effort*, Cambridge, UK.

A Evidence

A.1 Evidence from the Literature

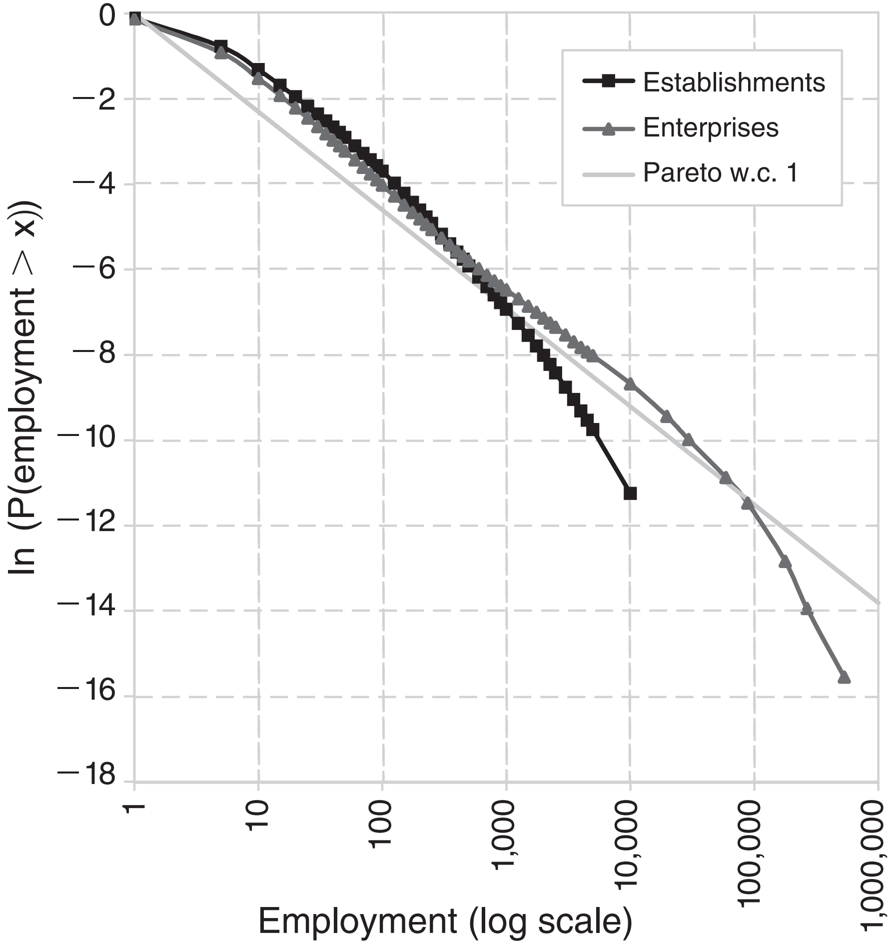
Figure 13: EMPIRICAL DISTRIBUTION OF US FIRM SIZES, LOG-LOG SCALE, AXTELL (2001)



Note: In this figure, the density of firm sizes is represented. The empirically estimated slope is 2.059 in the frequency domain, which corresponds to a Pareto tail coefficient of 1.059.

Figure 14: Distribution of US Firm Sizes. Source: Rossi-Hansberg and Wright (2007)

(a) Establishments / Enterprises



(b) Different Sectors

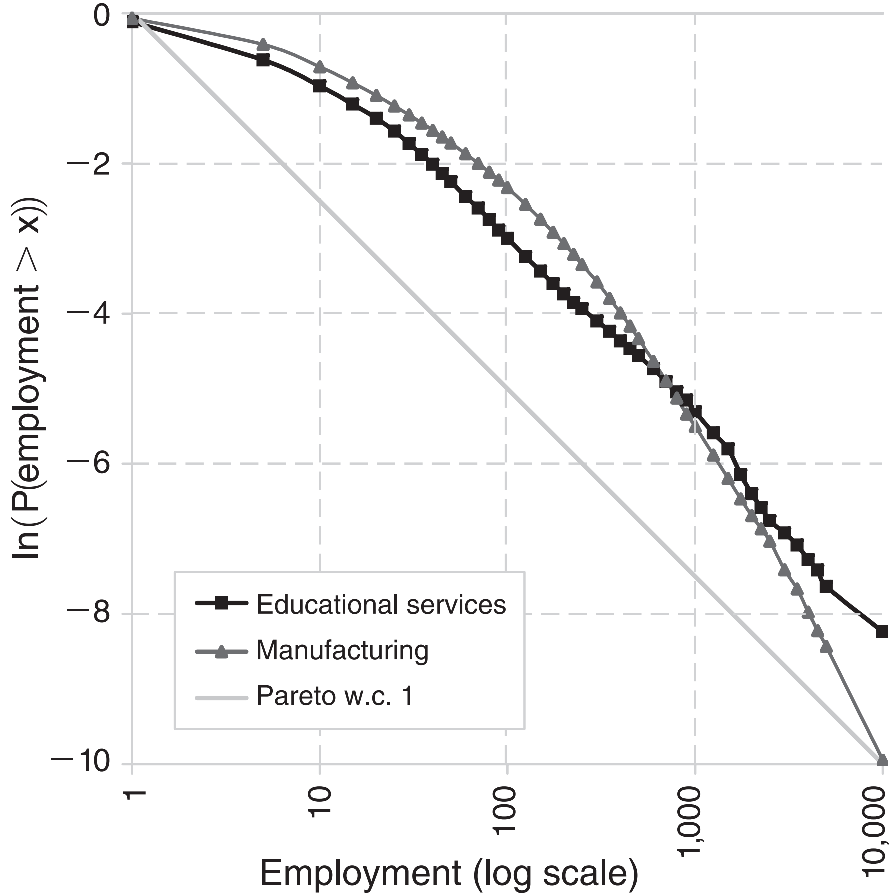
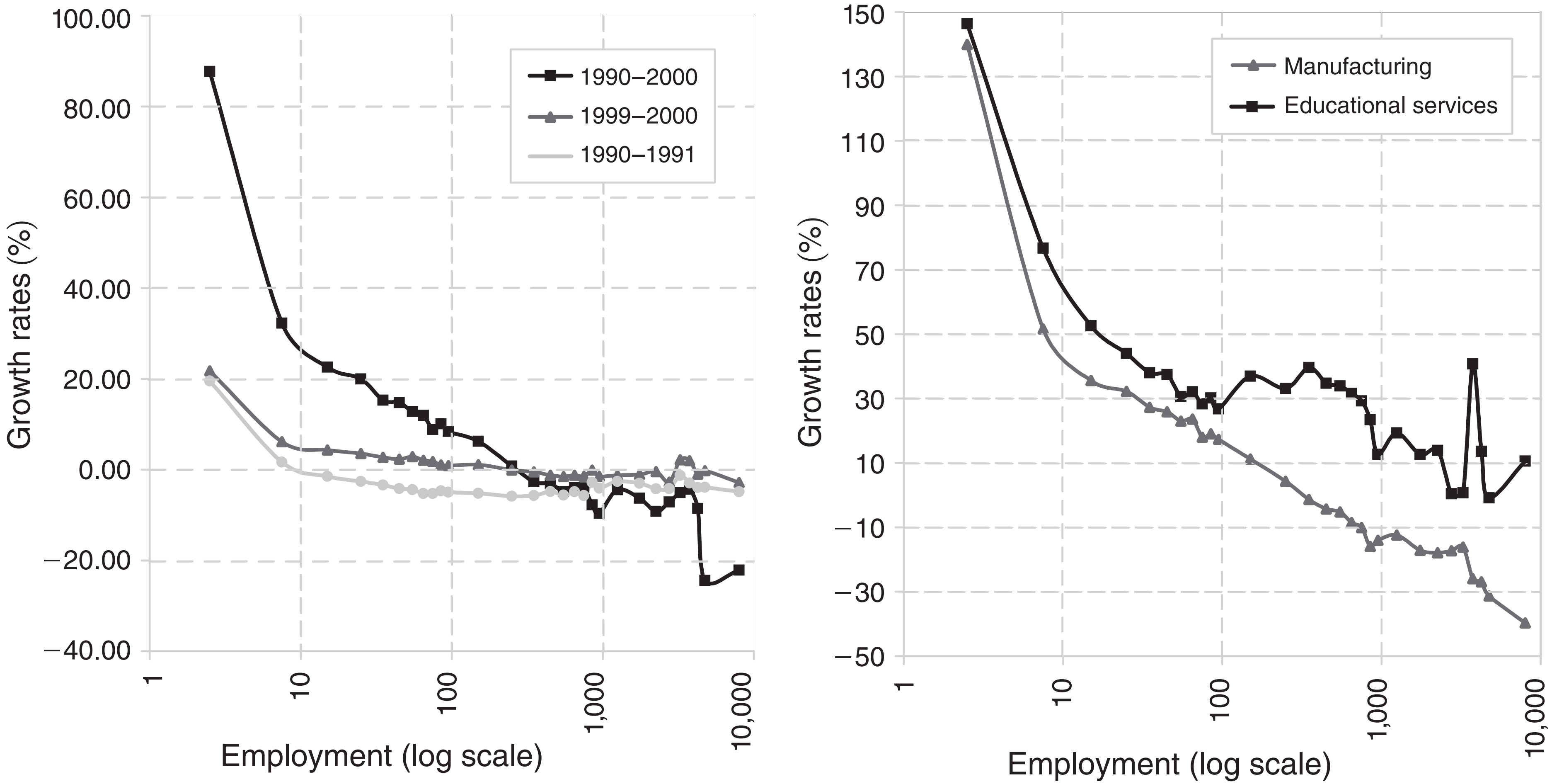


Figure 15: Distribution of US Establishment Growth Rates. Source: Rossi-Hansberg and Wright (2007)

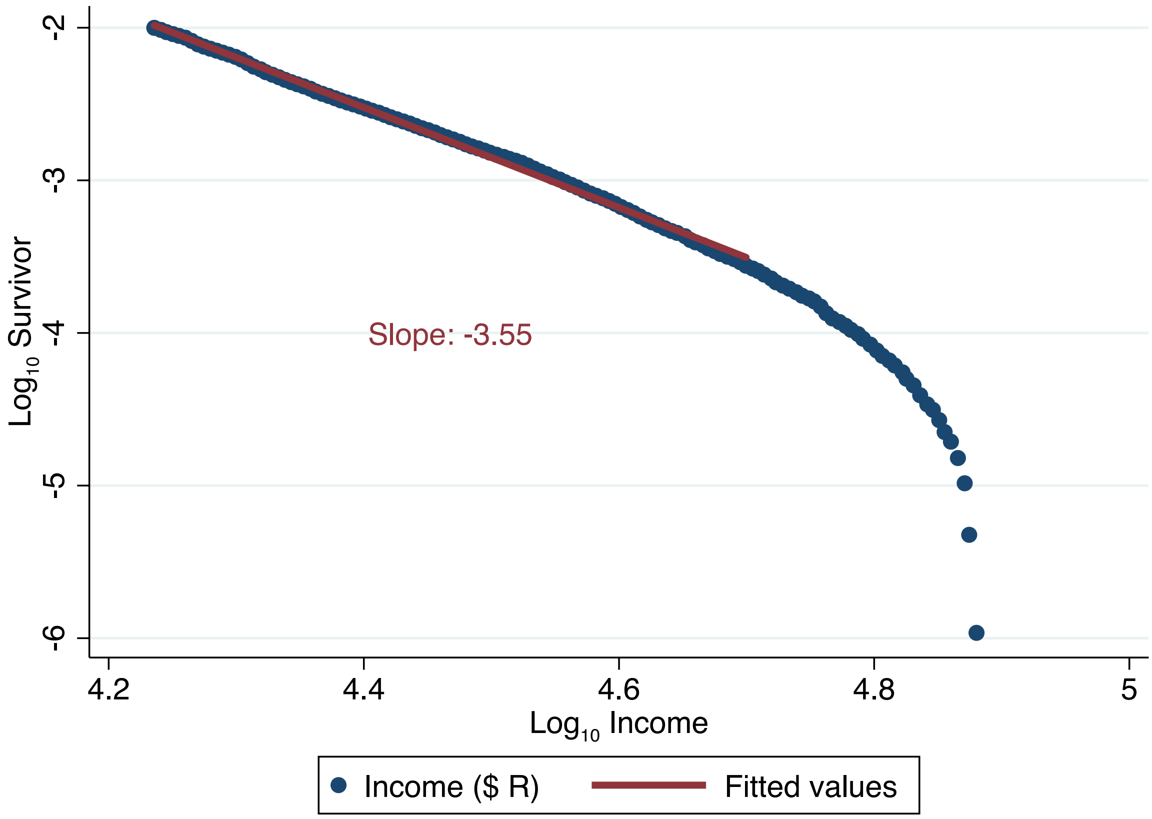
(a) Growth Rates

(b) Growth Rates by Sector



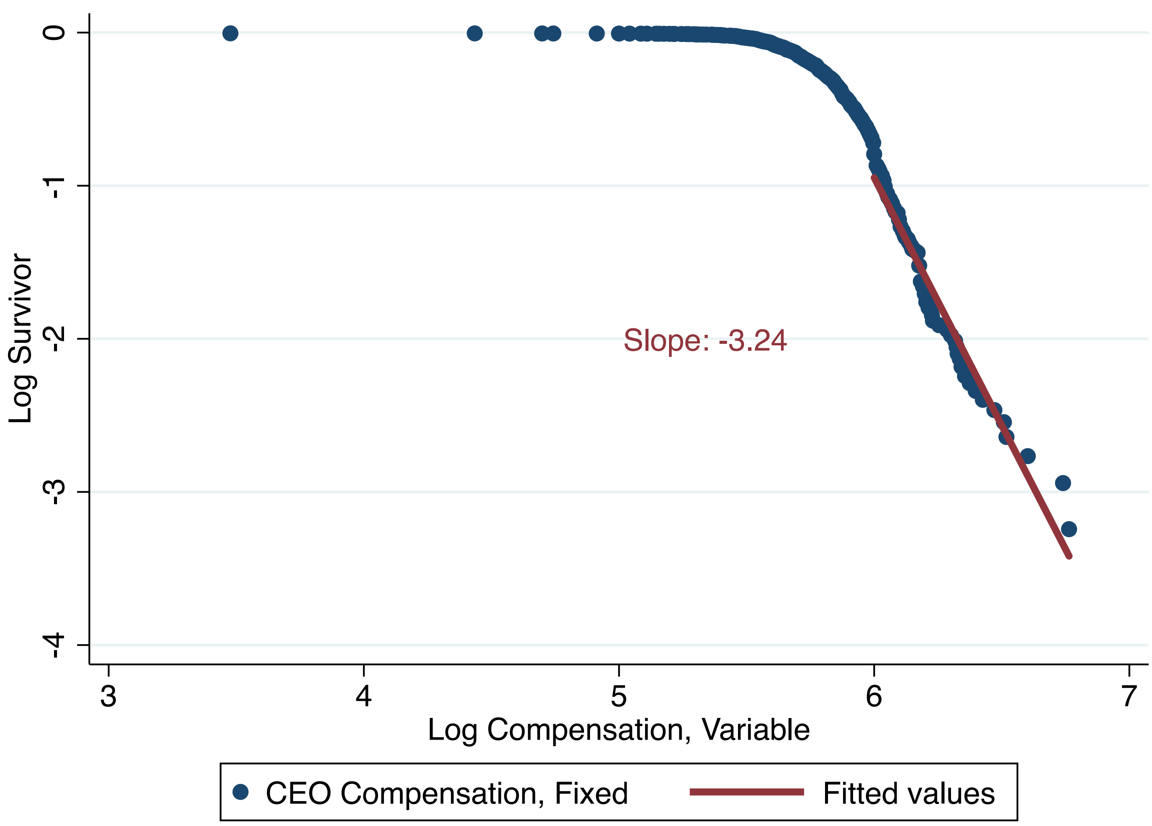
A.2 New Evidence

Figure 16: Top 1% Distribution of Brazilian Labor Incomes, Log-Log Scale. Source: Relaçao Anual de Informaçoes Sociais 2011



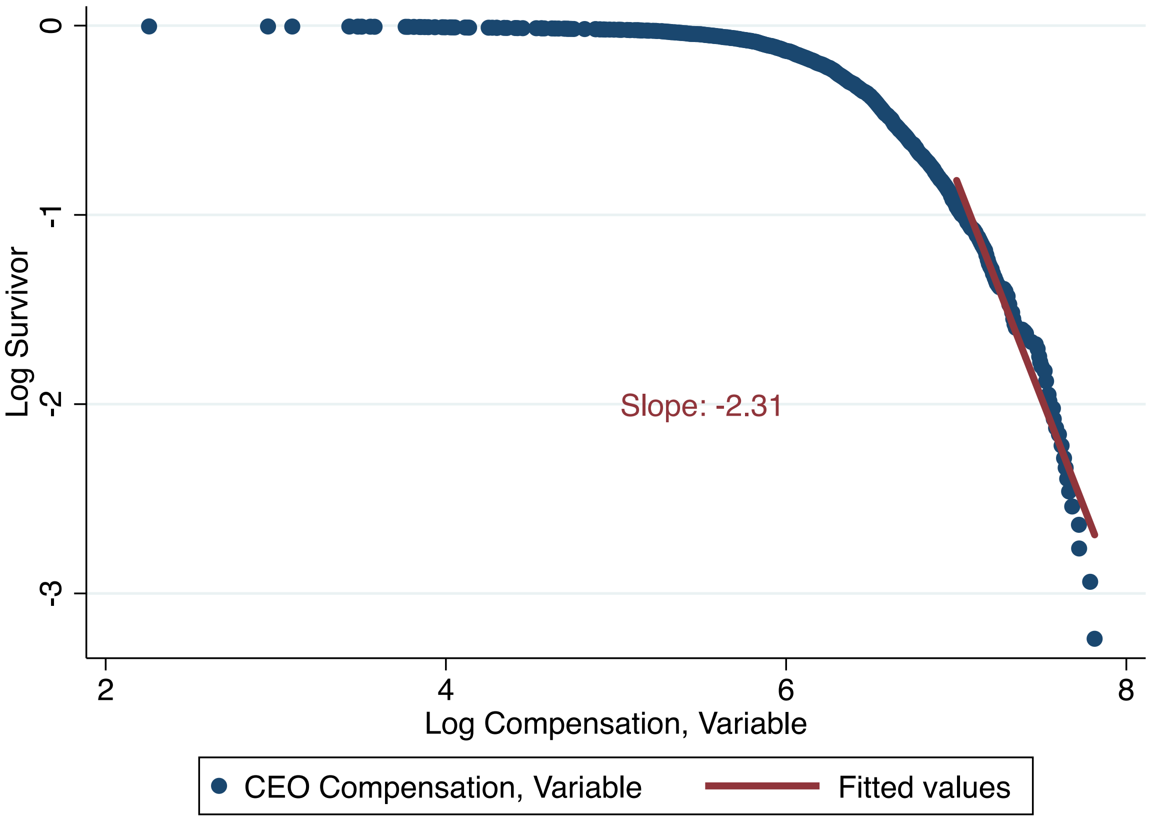
Note: In this figure, the empirical countercumulative distribution function of Brazilian wages is represented. It should be compared to one example of the theoretical wage distribution function on Figure 12.

Figure 17: US CEOs Labor Incomes, Fixed Component. Source: ExecuComp, 2005



Note: In this figure, the countercumulative distribution function of wages is represented.

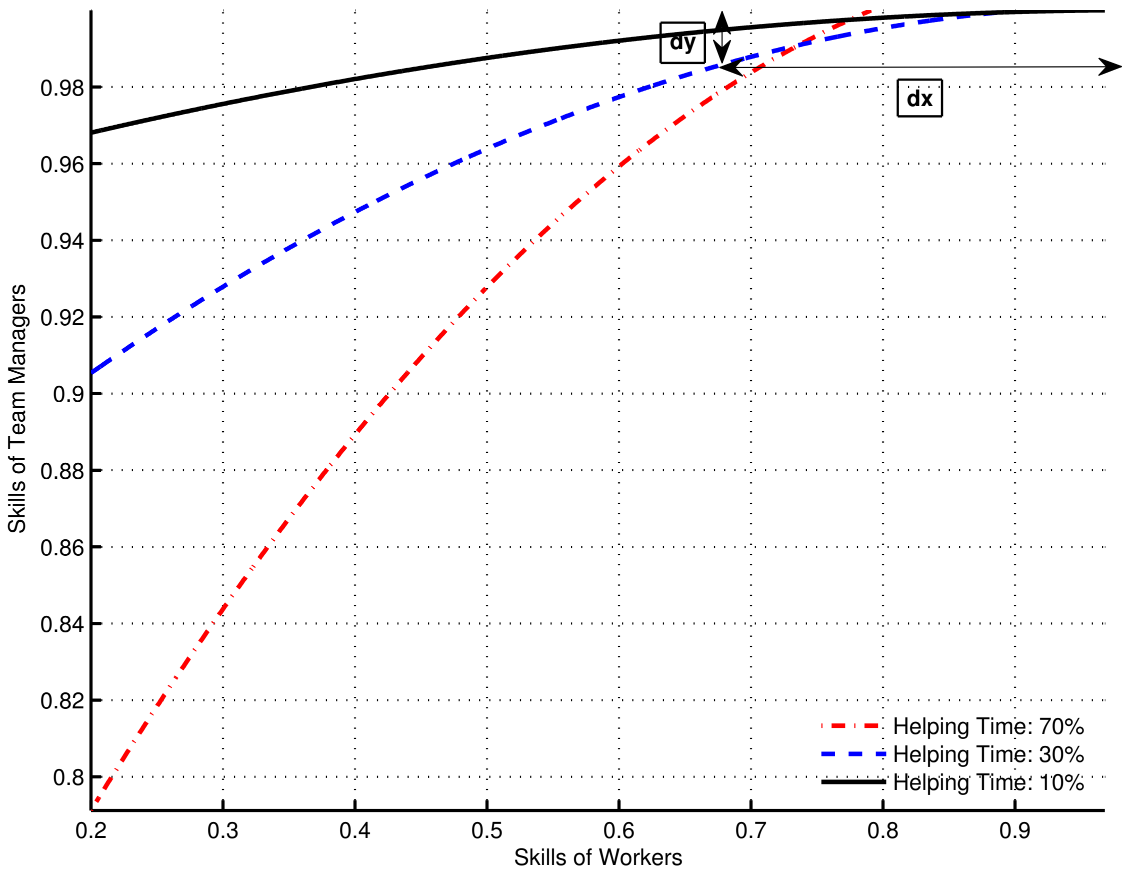
Figure 18: US CEOs Labor Incomes, Variable Component. Source: ExecuComp, 2005



Note: In this figure, the countercumulative distribution function of wages is represented.

B Simulations

Figure 19: Matching Function from Workers to Team Managers, $\Delta = 80\%$ - Comparative Statics on Helping Time



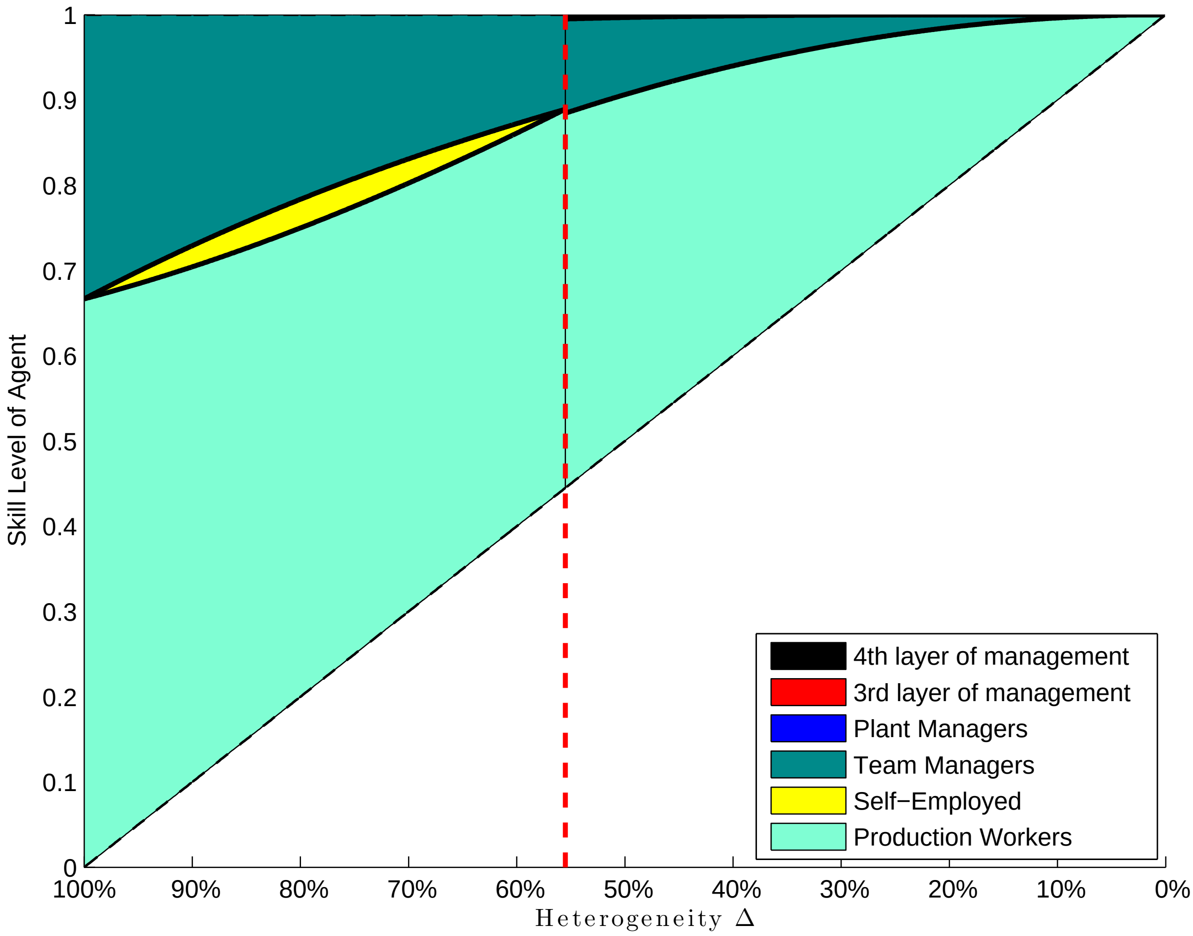
Note: This figure gives the matching function from workers to team managers derived in Section 2, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta} \mathbb{1}(1-\Delta, 1)(z) + \mathbb{1}(1, +\infty)(z)$.

Table 2: Uniform Skill Distributions: Cutoffs for Different values of Δ, h and L

	$\boldsymbol{\Delta} = \mathbf{100}\%$			$\boldsymbol{\Delta} = \mathbf{70}\%$			$\boldsymbol{\Delta} = \mathbf{30}\%$		
	z_1^L	z_2^L	z_3^L	z_1^L	z_2^L	z_3^L	z_1^L	z_2^L	z_3^L
$\mathbf{h} = \mathbf{70}\%$									
$L = 1$	0.684778			0.837717			0.968840		
$L = 2$	0.650620	0.957897		0.828537	0.989747		0.968500	0.999653	
$L = 3$	0.650000	0.957125	0.999357	0.828500	0.989706	0.999963	0.968500	0.999653	1.000000
$\mathbf{h} = \mathbf{20}\%$									
$L = 1$	0.900980			0.951238			0.991008		
$L = 2$	0.900000	0.999000		0.951000	0.999760		0.991000	0.999992	
$L = 3$	0.900000	0.999000	1.000000	0.951000	0.999760	1.000000	0.991000	0.999992	1.000000

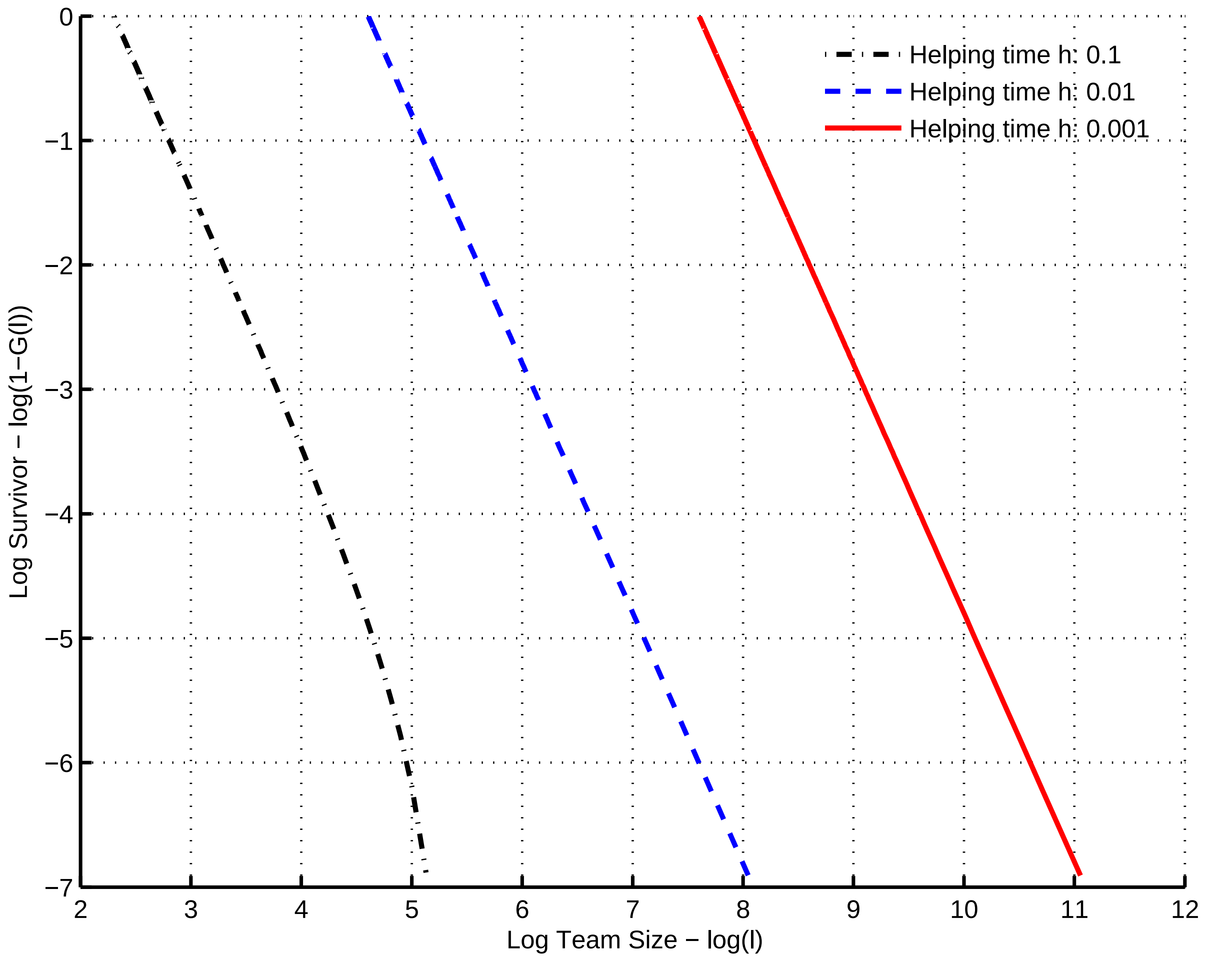
Note: In this table, the cutoffs are calculated with distribution of skills in the population given by $G(z) = \frac{z-(1-\Delta)}{\Delta}\mathbb{1}\,(1-\Delta, 1)\,(z) + \mathbb{1}\,(1, +\infty)\,(z)$. Precision up to 6 digits is displayed. Successively, the maximum number of levels is restricted to being $L = 1$, $L = 2$ or $L = 3$. The values for helping time are $h \in \{0.2, 0.7\}$, and the value for skill heterogeneity is $\Delta \in \{0.01, 0.1, 1\}$.

Figure 20: OCCUPATIONAL CHOICES FOR A UNIFORM DISTRIBUTION OF SKILLS, AND $h = 75\%$



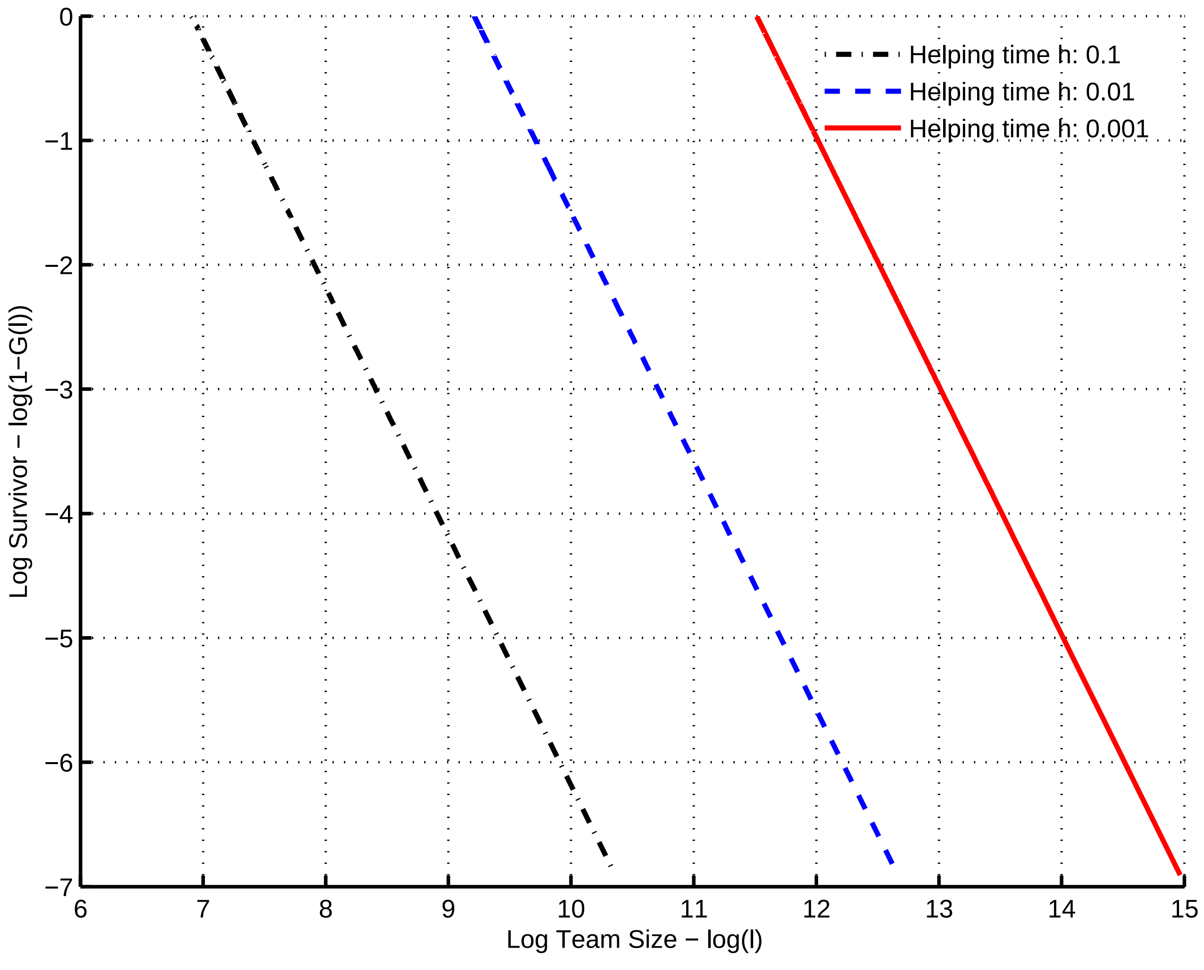
Note: This figure gives the occupational choices of agents when $h = 75\%$, and the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta}\mathbb{1}\,(1-\Delta, 1)\,(z) + \mathbb{1}\,(1, +\infty)\,(z)$. Note that the right-hand side of the graph is very similar to what is found on Figure 7. However, as shown on Figure 3, $h = 75\%$ belongs to the self-employment region for some values of Δ, numerically for $\Delta < 5/9$. In this region, there is only one level of management. The two regimes are separated by the dotted line. A third case, where the self-employment region arises for intermediary levels of heterogeneity Δ , is shown in the Online Appendix.

Figure 21: Distribution of Span of Control Down One Level of Hierarchical Organization ("Teams"), Log-Log Scale, Comparative Statics on Helping Time, High Heterogeneity



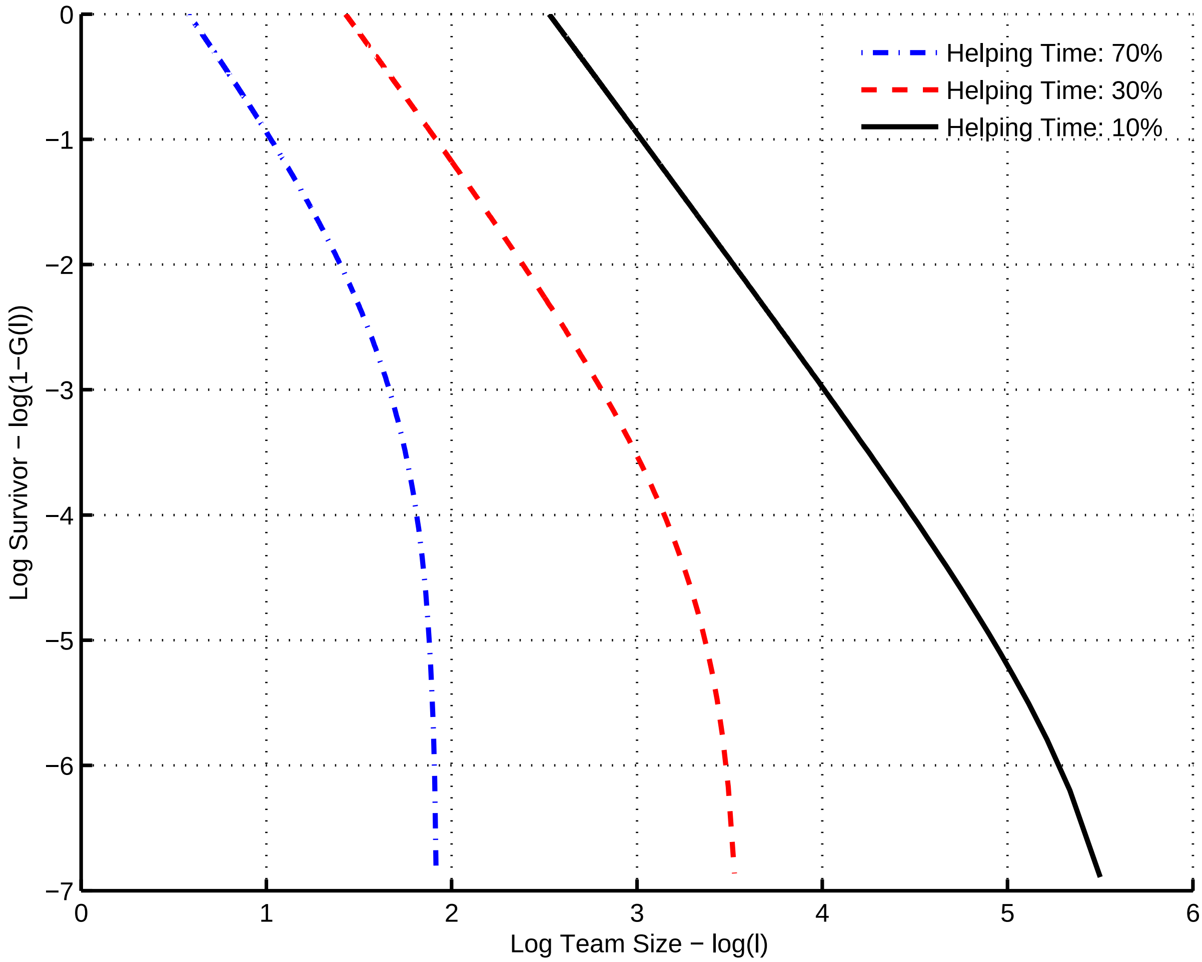
Note: This figure gives the theoretical distribution of team sizes as derived in Section 2, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta}\mathbb{1}(1-\Delta,1)(z) + \mathbb{1}(1,+\infty)(z)$, with a maximum level of skill heterogeneity $\Delta = 1$.

Figure 22: Distribution of Span of Control Down One Level of Hierarchical Organization ("Teams"), Log-Log Scale, Comparative Statics on Helping Time, Low Heterogeneity



Note: This figure gives the theoretical distribution of team sizes as derived in Section 2, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta}\mathbb{1}(1-\Delta,1)(z) + \mathbb{1}(1,+\infty)(z)$, with a low level of skill heterogeneity $\Delta = 0.01$.

Figure 23: Distribution of Team Sizes, Log-Log Scale, $\Delta = 80\%$ - Comparative Statics on Helping Time



Note: This figure gives the theoretical distribution of team sizes as derived in Section 2, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z-(1-\Delta)}{\Delta}\mathbb{1}\,(1-\Delta, 1)\,(z) + \mathbb{1}\,(1, +\infty)\,(z)$, with a low level of skill heterogeneity $\Delta = 0.01$.

C Proofs

C.1 Proof of Lemma 1

[TO BE ADDED]

C.2 Proof of Proposition 1

Proof. There is complementarity between the skills of workers and team managers, since the joint production of managers of type x_{l+1} supervising workers of type x_l is given by:

$$x_{l+1}N^l_{1+1\to l}(x_l) = \frac{x_{1+1}}{h(1-x_l)}.$$

An optimality argument allows to state that in the competitive equilibrium, there is positive sorting of managers and lower ranking managers, and managers of type one and workers. The function mapping a worker with skill x_l with a team manager of skill x_{l+1} is denoted by $m_l(.)$, defined on $[z_l^L, z_{l+1}^L]$, and such that $m_l(x_l) = x_{l+1}$.

$$\begin{aligned} \int_i t^i_W(x_l)di = 0 \quad &\Rightarrow \quad g(m_0(x_0))dm_0(x_0) = h\,(1-x_0)\,g(x_0)dx_0. \\ &\Rightarrow \quad m_0'(x_0)g(m_0(x_0)) = h\,(1-x_0)\,g(x_0). \end{aligned} \tag{M_1}$$

The market clearing equation (MC_l) for skill x_l gives a similar equation:

$$\forall x_l \in [z_l^L, z_{l+1}^L], \quad m_l'(x_l)g(m_l(x_l)) = h(1-x_l)g(x_l).$$

The positive sorting limiting condition writes, matching the less skilled and the more skilled:

$$\begin{aligned} m_l(z_l^L) &= z_{l+1}^L \\ m_l(z_{l+1}^L) &= z_{l+2}^L \end{aligned}$$

Integrating the above differential equation for $m_l(.)$ between $[z_l^L, z_{l+1}^L]$, and using these two limiting conditions give the result. □

Note that the stratification result obtains irrespective of the skill distribution. This is in contrast to Kremer and Maskin (1996) - see Garicano and Rossi-Hansberg (2006) for a discussion.

C.3 Proof of Lemma 2

[TO BE ADDED]

C.4 Closed form for Assumption 2 - Uniform Case

In the case where h or Δ are high enough, some agents remain self-employed. Self-employed agents have intermediary skills in equilibrium, because the gains from having a worker of skill

x_0 and a manager of skill x_1 work together are given by what the two produce together minus what they would have produced by themselves, that is:

$$\frac{x_1}{h(1-x_0)} - x_1 - \frac{x_0}{h\,(1-x_0)} = \frac{1}{h}\left(1 - \frac{1-x_1}{1-x_0}\right) - x_1.$$

This is clearly decreasing when the skills of workers increase, so that it is better to match the managers with the relatively less productive workers.

Figure 24: With Self-Employment in Equilibrium, One Layer.

The notations for cutoffs are introduced on Figure 24. Now the matching function $m_0(.)$ is defined on $[1-\Delta, z_1^{**}]$. An important difference also is that in that case, will not be solved independently of agents' choices. The next Appendix C.5 proves that the planner's problem would lead to the same equations as this decentralized problem.

In the decentralized problem, the two differential equations for $m_0(.)$ and $w_0(.)$ do not change compared to the case of no-self employment. However we now have four equations, not three, determining two initial conditions as well as two cutoffs. They are given by matching of the less and more skilled of workers and team managers, as previously:

$$m_0(1-\Delta) = z_1^* \qquad m_0(z_1^{**}) = 1.$$

In addition, we now have two indifference equations between being a worker and self-employed with skills z_1^{**}, and being self-employed and a team manager with skills z_1^*:

$$w_0(z_1^{**}) = z_1^{**} \qquad z_1^* = R_1(z_1^*).$$

In the case of a uniform distribution of skills, the market clearing equation for skills valid on $(1-\Delta, z_1^{**})$, together with the terminal equation $m_0(z_1^{**}) = 1$, then gives:

$$m_0'(x_0) g\left(m_0(x_0)\right) = h\left(1-x_0\right) g(x_0) \quad \Rightarrow \quad m_0'(x_0) = h(1-x_0)$$

$$\Rightarrow \quad m_0(x_0) = \frac{1}{2}\left(-hx_0^2 + 2hx_0 + h\left(z_1^{**}\right)^2 - 2hz_1^{**} + 2\right).$$

Inverting this expression, the inverse assignment function is therefore given by:

$$m_0^{-1}(x_1) = \frac{h - \sqrt{2h + h^2 - 2hx_1 - 2h^2 z_1^{**} + h^2 (z_1^{**})^2}}{h}.$$

In the case where self-employment arises in equilibrium, one must solve for the equilibrium wage function even to determine the spans of control of each team manager. One also uses $w_0(z_1^{**}) = z_1^{**} = z_1^{**}$ to integrate:

$$\left(1-x_0\right) w_0'(x_0) + x_0 w_0(x_0) = x_0 m_0(x_0)$$

$$\Rightarrow \quad w_0(x_0) = \frac{1}{2}\left(2x_0 + hx_0^2 - 2hx_0 z_1^{**} + h\left(z_1^{**}\right)^2\right)$$

Then using the two remaining $m_0(1-\Delta) = z_1^*$ and $z_1^* = R_1(z_1^*)$, and simple but lengthy algebra, one can express the cutoffs for occupational choice as a function of the heterogeneity parameter Δ and the helping time h:

$$z_1^* = -\frac{-2h + h^2\Delta + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2} \qquad z_1^{**} = \frac{-h + h^2 + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2}.$$

Replacing the cutoffs, one gets the assignment function as a function as these parameters as well:

$$m_0(x_0) = \frac{4h - 2h^2\Delta + h^3\left(-1+\Delta^2\right) - 2\sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)} + 2h^3x_0 - h^3x_0^2}{2h^2}.$$

The condition for there to be self-employment in equilibrium is that:

$$\begin{aligned}
z_1^{**} < z_1^* \quad &\Leftrightarrow \quad \frac{-h + h^2 + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2} < -\frac{-2h + h^2\Delta + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2} \\
&\Leftrightarrow \quad 3 - h\Delta - h > 2\sqrt{3 + h^2\Delta^2 - 2h(1+\Delta)} \\
&\Leftrightarrow \quad (1 + 2\Delta - 3\Delta^2)h^2 + 2(1+\Delta)h - 3 > 0 \\
&\Leftrightarrow \quad h > \frac{1 + \Delta - 2\sqrt{1 + 2\Delta - 2\Delta^2}}{-1 - 2\Delta + 3\Delta^2} \quad \text{since } h > 0.
\end{aligned}$$

When the primitives of the model are such that this is verified, we are in the case where a non trivial measure of agents are self-employed. When it is not the case, then all agents either become managers or workers. The condition is illustrated graphically on Figure 7.

C.5 Self-employment - Equivalence of Planner's and the Decentralized Problem

Even though the usual argument applies in the model (as in Garicano (2000), for that matter, and the competitive equilibrium is efficient, it is useful to compare the planner's problem with the decentralized one to gain some intuition on the model. Unlike in Garicano (2000), I do not need to use Hamiltonian-type methods to show the result.

Planner's Problem. As explained in the main text, an optimum necessarily has positive sorting between the more skilled workers and managers. Assuming that there is some self-employment in equilibrium, and using the same notations as in the main text, the planner wants to maximize total output given by:

$$\begin{aligned}
\max_{z_1^{**}} \int_{1-\Delta}^{z_1^{**}} & m_0(x_0)g(x_0)dx_0 + \int_{z_1^{**}}^{m_0(1-\Delta)} x_0 g(x_0)dx_0 \\
\text{s.t.} \quad & m_0(z_1^{**}) = 1 \\
\text{s.t.} \quad & m_0'(x_0)g\left(m_0(x_0)\right) = h(1-x_0)g(x_0) \\
\text{s.t.} \quad & z_1^{**} \le m_0(1-\Delta).
\end{aligned}$$

where z_1^* has everywhere been replaced by its expression as a function of m_0, given by $m_0(1-\Delta)$. Integrating the differential equation for the matching function between x_0 and z_1^{**} gives:

$$m_0(x_0) = G^{-1}\left(1 - \int_{x_0}^{z_1^{**}} h(1-u)g(u)du\right).$$

From there one calculates the partial derivative of the matching function with respect to the control variable z_1^{**}:

$$\frac{\partial m_0}{\partial z_1^{**}}(x_0) = -\frac{h(1-z_1^{**})g(z_1^{**})}{g(m_0(x_0))}.$$

Because $m_0(.)$ is explicit, we can eliminate the first two constraints from the optimization program. Denoting by λ the Lagrange multiplier on the last constraint, the first order condition for the above program can thus be written as follows:

$$\begin{aligned} m_0(z_1^{**})g(z_1^{**}) + \int_{1-\Delta}^{z_1^{**}} g(x_0)\frac{\partial m_0}{\partial z_1^{**}}(x_0)dx_0 - z_1^{**}g(z_1^{**}) + ... \\ ...\frac{\partial m_0}{\partial z_1^{**}}(1-\Delta)m_0(1-\Delta)g(m_0(1-\Delta)) + \lambda\left(\frac{\partial m_0}{\partial z_1^{**}}(1-\Delta) - 1\right) = 0. \end{aligned}$$

After simple algebra and using the expression above for $\frac{\partial m_0}{\partial z_1^{**}}(x_0)$, this leads to:

$$m_0(1-\Delta) + \int_{1-\Delta}^{z_1^{**}} \frac{g(x_0)}{g(m_0(x_0))}dx_0 = \frac{1}{h} + \lambda\left(\frac{\partial m_0}{\partial z_1^{**}}(1-\Delta) - 1\right).$$

There are two cases depending on whether the contraint is slack or not:

- Either the contraint is binding and the above equation only serves to determine the Lagrange multiplier. The only cutoff $z_1^{**} = z_1^*$ delimiting production and managing is determined through $z_1^* = m_0(1-\Delta)$ and $m_0(z_1^*) = 1$. This is the case with no self-employment in equilibrium.

- Or the constraint is slack $\lambda = 0$ and the above equation together with $m_0(z_1^{**}) = 1$ determines z_1^{**}, together with $m_0(.)$, and therefore z_1^* through $z_1^* = m_0(1-\Delta)$. This is the case with self-employment in equibrium.

Decentralized Problem. In the case when there is self-employment in equilibrium, the decentralized problem consists in a price system $w(.)$ for workers and managers maximizing over the skills of workers they work with, such that they choose $x_0 = m_0^{-1}(x_1)$ when of type x_1:

$$\max_{x_0} \frac{x_1 - w_0(x_0)}{h[1-x_0]} \equiv R_1(x_1).$$

This leads to a differential equation for wages:

$$\begin{aligned} & w_0'(x_0)(1-x_0) + x_0 w_0(x_0) = x_0 x_0 \\ \Rightarrow \quad & \frac{w_0(x_0)}{1-x_0} - \frac{w_0(z_1^{**})}{1-z_1^{**}} = \int_{z_1^{**}}^{x_0} \frac{u}{(1-u)^2} m_0(u)du \\ & = \left(\frac{m_0(u)}{1-u}\right)_{z_1^{**}}^{x_0} - \int_{z_1^{**}}^{x_0} \frac{m_0'(x_0)}{1-x_0}dx_0. \end{aligned}$$

Replacing the matching function in the integral using the market clearing equation for skills, and using the indifference condition $z_1^{**} = w_0(z_1^{**})$:

$$\frac{m_0(x_0) - w_0(x_0)}{1 - x_0} = 1 + \int_{z_1^{**}}^{x_0} h \frac{g(u)}{g(m_0(u))} du.$$

Then, using the indifference condition $R_1(z_1^*) = z_1^*$ yields directly:

$$m_0(1-\Delta) + \int_{1-\Delta}^{z_1^{**}} \frac{g(x_0)}{g(m_0(x_0))} dx_0 = \frac{1}{h},$$

which is the equation obtained in the planner's problem in the case where there is self-employment in equilibrium.

If in contrast there is no self-employment in equilibrium then the indifference condition between being a manager and a worker with skill z_1^{**} simply is $R_1(z_1^*) = z_1^*$, and must be regarded as an initial condition for the differential equation on $w_0(.)$.

C.6 Span of Control Distribution - Uniform Distribution - One Level

The market clearing equation for skills valid on $(1-\Delta, z_1^*)$, together with the initial equation $m_0(z_1^*) = 1$, then gives:

$$m_0'(x_0) g\left(m_0(x_0)\right) = h\left(1 - x_0\right) g(x_0) \quad \Rightarrow \quad m_0'(x_0) = h(1 - x_0)$$
$$\Rightarrow \quad m_0(x_0) = \frac{1}{2}\left(2 + 2hx_0 - hx_0^2 - 2hz_1^1 + h\left(z_1^1\right)^2\right).$$

Using then $m_0(1-\Delta) = z_1^*$, this gives a quadratic expression in z_1^* with h and Δ as exogenous parameters and therefore an expression for the threshold z_1^*, which has a simple Taylor expansion for small heterogeneity Δ and/or small helping cost h:

$$z_1^1 = \frac{1 + h - \sqrt{1 + h^2\Delta^2}}{h} \approx 1 - \frac{1}{2}h\Delta^2 + o(\Delta^2).$$

Replacing, this gives closed form expressions for the assignment equation, and its inverse:

$$m_0(x_0) = \frac{2 + 2h + h^2\left(-1 + 2x_0 - x_0^2 + \Delta^2\right) - 2\sqrt{1 + h^2\Delta^2}}{2h}$$
$$\Rightarrow \quad m_0^{-1}(x_1) = 1 - \frac{\sqrt{2 - 2h(-1 + x_1) + h^2\Delta^2 - 2\sqrt{1 + h^2\Delta^2}}}{h}$$

Span of control for a manager x_1 is then given by:

$$N_{1\to 0}^1(x_1) = \frac{1}{h\left(1 - m_0^{-1}(x_1)\right)}$$
$$N_{1\to 0}^1(x_1) = \frac{1}{\sqrt{2h(1 - x_1) + 2 + h^2\Delta^2 - 2\sqrt{1 + h^2\Delta^2}}}.$$

In the limit when heterogeneity Δ or the helping time h become small, this is a Pareto distribution of coefficient 2. Note that the convergence is very fast since:

$$2 + h^2\Delta^2 - 2\sqrt{1 + h^2\Delta^2} \approx \frac{1}{4}h^4\Delta^4 + O(h^6\Delta^6).$$