A Theory of Pareto Distributions*

François Geerolf †
UCLA

September 2017

Abstract

A strong empirical regularity is that the distributions of firm size and labor income are Pareto in the upper tail. This paper shows that Pareto tails may arise from some production functions in static assignment models with complementarities. Under limited assumptions on primitives' distribution, these models then generate Pareto tails for the span of control of CEOs and intermediary managers, and Zipf's law for firm size. This novel justification for Pareto tails sheds new light on why firm size and labor income are so heterogeneous despite small observable differences. The model receives substantial support in French matched employer-employee data.

Keywords: Pareto, Zipf's Law, Complementarities.
JEL classification: D31, D33, D2, J31, L2

*First version: April 2014. I thank Daron Acemoğlu, Pol Antràs, Andy Atkeson, Saki Bigio, Pablo Fajgelbaum, Xavier Gabaix, Sam Kortum, Thierry Mayer, Ezra Oberfield, Aleh Tsyvinski, Pierre-Olivier Weill for valuable suggestions, and especially Erzo Luttmer for detailed comments on a previous draft of this paper. I also thank numerous seminar participants at the 2016 NBER EF&G Meeting (Fall), the 2015 SED Meetings in Warsaw, the Bank of France Conference on "Granularity", and at the Chicago Fed, MIT, Polytechnique, UC Berkeley, UCLA, UC Riverside, UPenn, TSE, and Yale for comments. This project owes to the research environment at Toulouse School of Economics, and particularly to Maxime Liégey, François de Soyres, and Martí Mestieri. This work is supported by a public grant overseen by the French Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD).

†E-mail: fgeerolf@econ.ucla.edu. This is a revised version of an earlier draft entitled "A Static and Microfounded Theory of Zipf's Law for Firms and of the Top Labor Income Distribution."

Introduction

In the 1890s, Vilfredo Pareto studied income tax data from several countries. He plotted the number of people earning an income above a certain threshold against the respective threshold on double logarithmic paper and revealed a linear relationship. The corresponding income distribution was more skewed and heavy-tailed than bell-shaped curves: Pareto felt that he had discovered a new type of "universal law" that was the result of underlying economic mechanisms. Since then, Pareto's discovery has been confirmed and generalized to the distribution of firm size (Axtell (2001)) and wealth, which also follow Pareto distributions in the upper tail. According to Gabaix (2016), this is one of the few quantitative "laws" in economics that hold across time and countries. A striking qualitative feature of Pareto tails is large inequality: for example, the top 1% gets about 20% of pre-tax income in the United States. Therefore, understanding the source of Pareto distributions is a first order question for the economics of inequality.

In this paper, I offer a new explanation for Pareto distributions. I show that Pareto distributions can be generated endogenously by some production functions, in static assignment models with complementarities. Assuming these production functions, endogenous Pareto distributions in firm size and income occur under very limited assumptions on the distribution of underlying primitives. Unlike in previous theories, large firms or incomes can appear instantaneously and result from an arbitrarily small level of ex ante heterogeneity.

In contrast, economists' current understanding of why Pareto distributions emerge falls into two categories. The first theory works through a "transfer of power law". One Pareto distribution can be explained by assuming that some other variable is distributed according to a Pareto distribution; for example, entrepreneurial skills (Lucas (1978)), firm productivities (Helpman et al. (2004), Lucas and Moll (2014), Perla and Tonetti (2014)), or firm size (Gabaix and Landier (2008)). A functional form for the production function also needs to be assumed, which preserves the Pareto functional form, such as a power function. The second theory holds that Pareto distributions result from a dynamic, proportional, "random growth" process, following Gibrat's (1931) law. In this theory, many firms or incomes are large because they have been hit by a long and unlikely continued sequence of good idiosyncratic shocks (Champernowne (1953), Simon and Bonini (1958), or Luttmer (2007)). To the best of my knowledge, this paper is the first which generates Pareto distributions in firm size and labor income that arise solely from production functions.[1]

[1] In Geerolf (2015), I also generate a Pareto distribution for leverage ratios and the returns to capital in a model of frictional asset markets. To the best of my knowledge, Geerolf (2015) is the first example in the economics literature of Pareto distributions that arise solely from production functions, and not the distribution of primitives. The understanding of Pareto distributions in firm size and labor income is probably more important, as it has implications for research on heterogeneous firms (misallocation, international trade) and labor income inequality. I discuss this point further in the literature review.

Even though the argument only hinges on a special form for the production function, I choose to develop the argument around a particular version of Garicano's (2000) problem-solving model, which provides a readily available microfoundation for a Pareto generating production function. In Garicano (2000), production consists in solving problems, with varying levels of difficulty. Managers help workers deal with the problems that they are unable to solve. In this model, managers' time constraint is a source of limited managerial attention, and limits the maximum size of firms. This paper shows that the organization of firms can then be decomposed into elementary Pareto distributions for span of control. For example, if an economy has two-layer firms with managers and workers, then the span of managers' control over workers is a Pareto distribution with a tail coefficient equal to two in the upper tail. When the number of layers of management increases, and CEOs can also supervise other intermediary managers, the Pareto coefficient of the firm size distribution converges to one. Thus, the model provides a new justification for why firm sizes follow Zipf's (1949) law. This decomposition of the firm size distribution into intermediary hierarchies' size distribution is supported empirically by French matched employer-employee data. It also explains well-known evidence – for example, on the distribution of establishment sizes in the US.

I show that the model also allows us to take a fully microfounded approach to the labor income distribution of managers derived by Terviö (2008) and applied to the Pareto distribution of firms by Gabaix and Landier (2008), as skill prices also follow an endogenous Pareto distribution in the competitive equilibrium. Unlike in those two papers, Zipf's law for firm size is not assumed but obtained endogenously. An important advantage, relative to the latter reduced form approach, is that one can then relate last decades' increase in firm size to deep parameters of the model, such as the costs of communication, or the change in the underlying skill distribution, and understand why span of control has increased by so much.

The argument is most easily developed with the help of mathematics, and results from a "change of variable near the origin." Mathematically, the reason why only the form of the production function matters for the upper tail of the span of control and top income distributions is that through this change of variable, the upper tail corresponds to a very small portion of the underlying density of skills. But the mathematics also convey a powerful new economic intuition. The key is to recognize that in a hierarchical organization, managers end up doing what workers don't do, either by lack of skill or because of specialization. In terms of abilities, it does not matter much whether a worker will not do 0.01% of the work or 0.001% of the work – if they were working independently on an island, their productivities would barely be distinguishable. However, the reduction in time cost for a manager is given by a factor of ten. Thus, a manager working with the latter type of workers rather than the former can supervise ten times as many workers, and lever up his higher productivity accordingly. This simple reasoning can help explain why a hierarchical organization of production can lead to unbounded differences in measured wages, even when underlying differences are actually quite small.

At the same time, these developments around Garicano's (2000) model of a knowledge economy should make clear that the argument is more general, and only relies on the special form of the production function. This kind of production function has been encountered in previous work by Geerolf (2015), a sorting model of financial markets with complementarities between borrowers and lenders entertaining heterogeneous beliefs. The simplicity of these two models, and the fact that Garicano's (2000) model was not purposefully developed to generate Pareto distributions, in fact suggest that there might be something more general about joint production with complementarities that leads to Pareto generating production functions. I conjecture that the mechanism at play is, for example, more general than production based on knowledge.

More generally, my analysis has striking implications for the literature on firm heterogeneity: Pareto distributions are in fact the benchmark distributions that arise in the case of perfect homogeneity, while heterogeneity in primitives should be backed out in the deviations from Pareto distributions. Empirically, it is well known that firm productivities across large and small firms appear much smaller than what is implied by an assumption of Pareto distributed productivities, which would imply unboundedly increasing productivity (for example, Combes et al. (2012)). Moreover, the theory provides an interpretation for why the distributions of firm size and income never quite follow Pareto distributions exactly. First, empirical Pareto distributions are truncated, and second, Pareto is a good approximation only for the upper tail. With power law production functions, Pareto distributions are everywhere, and they are in fact a signature of the homogeneous benchmark. On the contrary, deviations from Pareto distributions indicate that heterogeneity in skills is strictly positive. So far, researchers working on firm heterogeneity have ignored these deviations from the Pareto benchmark because they were considered to be second order, compared to the importance of heterogeneity implied by a Pareto distribution. If Pareto distributions ultimately come from production functions rather than the distribution of primitives, as this paper suggests, then only deviations from Pareto they may in fact provide information on heterogeneity.[2]

Regarding top income inequality, it has been known for a long time that assignment models of the labor market à la Tinbergen (1956) or Sattinger (1975) could amplify

[2]This result is counterintuitive. In fact, that very heterogeneous outcomes, such as Zipf's (1949) law, must eventually come from at least somewhat heterogeneous primitives is a qualitative insight that could perhaps be seen as a matter of "common sense." This paper on the contary argues that for managers to be able to supervise many workers, it must be that workers are not much less skilled than they are themselves. Large firms would not be manageable if at least parts of the firm could not function largely independently. Thus, the decreasing relationship between heterogeneity in primitives and heterogeneity in outcomes is counterintuitive only superficially.

inequality, as wages are generally a convex function of abilities. I take this insight to its furthest extent by demonstrating that with a power law production function, sorting models lead to the amplification of even very small ex ante differences and lead endogenously to Pareto distributions. This fact may explain why residual wage inequality is so high in the data. At the top of the distribution, observables are very similar, yet small differences in abilities lead to large differences in pay. More generally, it also offers a new intuition for the so-called Pareto principle: that in the social sciences, roughly 80% of the effects come from 20% of the causes.[3]

The rest of the paper proceeds as follows. Section 1 reviews the literature. Section 2 develops a particular version of the Garicano (2000) model to illustrate the main point of the paper: power law production functions can generate Pareto distributions for firm size and labor income. It also provides a new explanation for Zipf's (1949) law for firms. Section 3 reconciles this theory with the leading theory generating endogenous Pareto distributions through a dynamic random growth process, and shows that the proposed theory leads endogenously to stationarity and Gibrat's law. Section 4 provides empirical support for the disaggregation of Zipf's law into intermediary span of control distributions in the data. Section 5 derives the distribution of skill prices endogenously generated by Zipf's law. Section 6 concludes.

1 Literature

I employ a means to generate Pareto (1895) distributions that is already known to some physicists, although it is somewhat marginal in the field. Sornette (2006) gives an overview of these "Power Laws Change of Variable Close to the Origin" (section 14.2.1), and Sornette (2002) and Newman (2005) offer surveys of this approach.

According to Pareto (1895), social institutions could not be the underlying reason for these regularities, as they were observed in very different societies. He also dismissed random chance, as chance does not produce such thick tails. He concluded that Pareto distributions must arise from "human nature." The modern economics literature has used either random growth models or the distribution of primitives to explain the emergence of Pareto distributions. In random growth models, the stochastic process is assumed to be scale independent (Gibrat's (1931) law), and one looks for stationary distributions created by that process. Gibrat's law also intuitively leads to scale independence in the stationary distribution created by the process – thus a power law distribution – and Zipf's (1949) law when frictions become small. Champernowne (1953) is perhaps the first of such random growth models for incomes, and Simon and

[3]This is only true when the tail coefficient is equal to $\log(5)/\log(4) \approx 1.16$. This is a good approximation for firm size, cities, and ownership of land in nineteenth century Italy, according to Pareto.

Bonini's (1958) for firms. Kesten (1973), Gabaix (1999), and Luttmer (2007) are other examples of this approach. There is also a literature that links hierarchies to power law distributions, in the case of both firms and cities – for example Lydall (1959) for firms and Beckmann (1958) for cities. Hsu (2012) is a microfoundation of Beckmann (1958) using central place theory, and in which a multiplicative process occurs at the spatial level rather than in a time dimension. Another way the literature has generated Pareto distributions is by using Pareto as primitives' distribution, such as Lucas (1978), Helpman et al. (2004), Chaney (2008), Terviö (2008) and Gabaix and Landier (2008).

The boundaries of the firm are defined as in the span of control model of Lucas (1978), who first formalizes that the limits to the boundaries of the firm can arise from limited managerial attention. Garicano (2000) is more explicit about what management is; his model leads to the kind of production functions that I emphasize in this paper.[4] Garicano and Rossi-Hansberg (2006) is certainly the most closely related paper: They investigate the implications of the Garicano (2000) model on the distribution of incomes, but make no mention of Pareto distributions. Caliendo et al. (2015) develop a measure of hierarchies in the French matched employer-employee data, which I use in this paper, and which I connect to Pareto distributions.

This paper is very closely related to the literature on the "economics of superstars" pioneered by Rosen (1981) and applied to the CEO market by Terviö (2008). Rosen argues in favor of imperfect substitution among sellers and consumption technologies with scale economies "with great magnification if the earnings-talent gradient increases sharply near the top of the scale," which gives rise to a winner-takes-all phenomenon. Gabaix and Landier (2008) observe that since the distribution of firm size is given by Zipf's law, CEOs face a Pareto distribution for the size of stakes. Under limited assumptions on the distribution of abilities, this leads to a Pareto distribution for their labor income. In this paper, Zipf's law for firms obtains endogenously without assuming any functional form on distributions, but instead results from the production function. In a nutshell, my contribution is to show that some production functions lead to a Pareto-like distribution of the size of stakes regardless of primitives' distribution, which Rosen (1981), Terviö (2008) and Gabaix and Landier (2008) took as given. Section 5 compares the fully microfounded and the earlier reduced form approach in more depth.

Finally, this new microfoundation of Pareto distributions provides a framework which can perhaps help shed a new light on the recent changes in inequality at the very top of the income distribution, particularly in the US, documented in Piketty and Saez (2003). It also connects to recent empirical work on the determinants of wage inequality across and within firms, such as Abowd et al. (1999) and Song et al. (2015).

[4] Lucas (1978) anticipated this: "The description of management is a shallow one ... it does not say anything about the nature of the tasks performed by managers, other than that whatever managers do, some do it better than others."

2 A static theory of Zipf's (1949) law for firms

In Section 2.1, I set up the simplest possible Garicano (2000) economy, restricting attention to two-layer firms. This allows to present most of the new insights about Pareto distributions. In section 2.2, I spend some time on the uniform distribution, a very special case but which allows to understand the intuition behind most of the results in this paper. In Section 2.3 and Section 2.4, I show that Pareto tails occur much more generally under very general regularity condition on the density function, and way outside of the uniform benchmark. In Section 2.5, I consider L-layer firms, which is necessary to understand why the tail coefficient on the firm size distribution is one (Zipf's (1949) law).

2.1 A Garicano (2000) economy with two-layer firms

I consider a static economy populated by a continuum of agents, each endowed with one unit of time. Production consists of solving problems, which are drawn randomly from a unit interval $[0,1]$. An agent with skill z is able to solve problems in interval $[0,z] \subset [0,1]$, and therefore fails to solve a measure $1-z$ of problems. The distribution of agents' skills (or abilities) is F, with density f and support $[1-\Delta,1]$. Δ is thus a measure of skill heterogeneity.[5] A key assumption here, and one that distinguishes this paper from previous work on hierarchy models, is that skills are exogenous and that some agents can solve all problems by themselves. The second assumption is crucial for all the results, while the first assumption is needed to obtain Pareto wages.

In this section, I constrain firms to have only two layers exogenously.[6] Firms are then composed of one manager who employs a set of workers. Agents can choose to become workers or managers. As workers, agents use their time to draw a unit measure of problems. This unit measure ensures that an agent is able to solve a deterministic fraction of problems, and allows to leave aside issues of stochastic arrival of problems, which could lead to queuing. If they are managers, they listen to problems that their workers fail to solve in time $h<1$ per measure of problem. h is a key parameter, called "helping time"; the lower the helping time, the greater potential complementarities in production.[7]

[5] It is not a sufficient statistic however, as the whole shape of the distribution f matters. But it will prove useful in comparative static exercises to change Δ while keeping the shape of the distribution fixed.

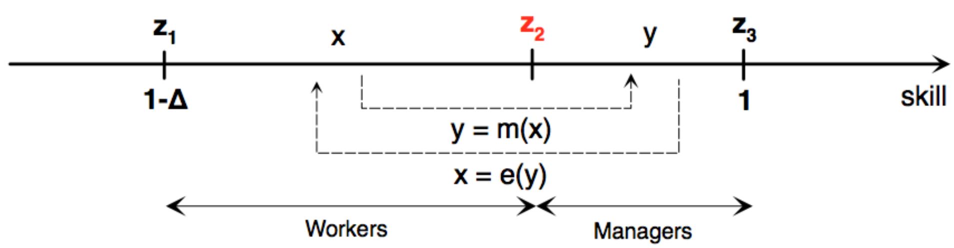
[6] Talking about "firms" is perhaps a slight abuse of terminology, given that all objects in this model are continuums. More precisely, I use the following terms to interpret this continuous model in a discrete way. I refer to one manager and the measure of workers who work for him as "one firm".

[7] In the original Garicano (2000) model, agents can also choose to remain self-employed. In Appendix B.2, I consider the case of self-employment, when agents are allowed to draw and solve problems without the help of managers. As long as helping time is sufficiently low, the help of a manager is sufficiently cheap, and in equilibrium there are only managers and workers. To streamline exposition, I focus directly on this case.

This model has a block recursive property. That is, optimal allocations – occupational choice, and the composition of firms – can be solved without reference to skill prices, which determine the division of surplus inside the firm. Supporting skill prices will not be derived before Section 5. The competitive equilibrium is more easily solved by looking at the planner's problem.

Managers have a chance to try to solve more problems, as it takes $h < 1$ units of time to listen, and 1 unit of time to draw. It is thus optimal that agents with higher skill become managers. Managers and workers are seggregated by skill (Kremer and Maskin (1996)), as shown formally in Garicano and Rossi-Hansberg (2006) and Antràs et al. (2006). The endogenous cutoff z_2 splits $[1-\Delta, 1] \equiv [z_1, z_3]$ into workers and managers: Managers spend time on problems whose difficulty is at least z_2.

Figure 1: Two-layer firms: Notations



It is also straightforward to see that the planner's problem has positive sorting between managers and workers: There is complementarity between the skills of managers and that of workers. The problems a more skilled worker cannot solve are harder statistically, so it is optimal for him to report these problems to a more skilled manager, while easier problems are best left in the hands of less skilled managers. A more skilled manager supervises more workers, because more skilled workers need little listening. This positive sorting result can also be obtained in a more mathematical way. When she hires workers with skill x, a manager with skill y needs to answer a number $1-x$ of problems. Since she takes h units of time to listen to one of these problems, she can oversee $1/[h(1-x)]$ workers. The production of this manager together with her workers will thus be given by $y/[h(1-x)]$. Since:

$$\frac{\partial^2}{\partial y \partial x}\left(\frac{y}{h(1-x)}\right) = \frac{1}{h(1-x)^2} > 0,$$

there is complementarity between the skills of managers and that or workers: More skilled managers benefit relatively more from higher span of control, and thus from working with more competent workers. The marginal gain of a more skilled worker is greater for marginally better managers, which translates into a positive cross-partial derivative of output. The planner thus wants more skilled managers and more skilled workers to work together.

Denote by $m(.)$ the matching function from workers with skill x to managers with skill $y = m(x)$. Because of positive sorting, $m(.)$ is a strictly increasing function from $[1 - \Delta, z_2]$ onto $[z_2, 1]$. Moreover, the less skilled workers are hired by the less skilled managers, so that:

$$m(1 - \Delta) = z_2. \tag{1}$$

The matching function is such that the time of managers with skills in $[y, y + dy]$, given by $f(y)dy = f(m(x))m'(x)dx$, is used to answer the problems of workers with skills in $[x, x + dx]$, who draw $f(x)dx$ problems, a fraction $1 - x$ of which they cannot solve, and who require $h(1 - x)f(x)dx$ of listening time, so that:

$$f\,(m(x))m'(x) = h(1 - x)f(x). \tag{2}$$

Equation (2) is a market clearing equation for managers' time: the matching function is more steep ($m'(.)$ high) if there are relatively fewer managers at the given skill level ($f\,(m(x))$ low), if there are many workers available to be matched ($f\,(x)$ high), if the communication cost is high (h high), or these workers are not very skilled ($1 - x$ high). Given z_2, $m(.)$ is determined on $[1 - \Delta, z_2]$ by equations (1) and (2), constituting an initial value problem. Finally, because the most skilled workers are hired by the most skilled managers, z_2 is a solution to:

$$m(z_2) = 1. \tag{3}$$

Managers' and workers' behavior, who is matched to whom, and the span of control of each manager are uniquely characterized by the matching function $m(.)$ on $[z_1, z_2]$, as well as by the endogenous cutoff z_2. Note that the span of control $n(y)$ of a manager with skill y hiring workers with skills $x = m^{-1}(y)$ is given as a function of parameters and the endogenous matching function by:

$$n(y) = \frac{1}{h(1 - m^{-1}(y))}.$$

2.2 Example: uniform distribution for skills

The uniform distribution for skills may appear very special at first, but in fact it allows us to get at the main results of the paper, because it is a local approximation to any smooth density function. With a uniform distribution for skills f on $[1 - \Delta, 1]$, and for $x \in [z_1, z_2]$, span of control $n(y)$ has a closed form expression because, with f constant, equation (2) is a simple ordinary differential equation: $m'(x) = h(1 - x)$.[8]

[8] I try to keep as close as possible to the earlier literature in terms of notation. However, both analytically and computationally (when one works with arbitrary density functions), it may prove

Proposition 1. *(Two-layer firms, uniform distribution)*

(a) With two-layer firms and a uniform distribution for skills on $[1-\Delta, 1]$*, the distribution of the span of managers' control over workers is a* ***truncated Pareto distribution with a coefficient equal to 2****. That is, the probability that span of control is higher than* n *for* $n \in [\underline{n}, \bar{n}]$ *is given by:*

$$\mathbb{P}[N \geq n] = \frac{\underline{n}^2}{1 - (\underline{n}/\bar{n})^2} \left(\frac{1}{n^2} - \frac{1}{\bar{n}^2} \right),$$

with the minimum and maximum span of control being given respectively by:

$$\underline{n} = \frac{1}{h\Delta} \quad \text{and} \quad \bar{n} = \frac{1}{\sqrt{1 + h^2\Delta^2} - 1}.$$

(b) When $\Delta \to 0$*,* $\bar{n} \to \infty$ *and* $\frac{\bar{n}}{\underline{n}} \to \infty$ *and the distribution of the span of managers' control over workers becomes a* ***full Pareto distribution with coefficient 2****:*

$$\mathbb{P}[N \geq n] \sim_{\Delta \to 0} \frac{\underline{n}^2}{n^2}.$$

More precisely, let $U = N/\underline{n}$ *be the "scaled" span of control; then* U *converges in distribution to a Pareto distribution with a coefficient equal to 2 when heterogeneity* Δ *goes to 0, as:*

$$\forall u \geq 1, \qquad \mathbb{P}[U \geq u] \to_{\Delta \to 0} \frac{1}{u^2}.$$

Proof. See Appendix A.1. □

Proposition 1 contains the main result of the paper, which is generalized later. Part (a) of the proposition says that the production function in the Garicano (2000) model produces truncated Pareto distributions for span of control with location parameters $\underline{n}$ (scale) and $\bar{n}$ (truncation) from a uniform distribution of skills. Note that firm size is just span of control plus one, in this case of two-layer firms. In the following, I therefore use "span of control" and "firm size" interchangeably.

The expression in Proposition (1a) corresponds to a truncated Pareto distribution with tail index 2 and location parameters $\underline{n}$ and $\bar{n}$, with the same density as the Pareto distribution, except above the maximum size $\bar{n}$, where the density is identically equal to zero, and the density is appropriately renormalized (see Appendix B.1).

The expressions for the scale parameter $\underline{n}$ and the truncation parameter $\bar{n}$ are also interesting. First, the helping time h and heterogeneity parameter Δ enter symmetrically. When the helping time h decreases, complementarities increase, and both the

more convenient to work directly with the "employee" function, which is the inverse of the "matching" (or manager) function: $e^{-1} = m$. I show this in Appendix B.3.

scale parameter and the truncation parameter ($\underline{n}$ and $\bar{n}$) increase: The whole Pareto distribution of firm size shifts out, and one moves closer to a full Pareto distribution as $\bar{n}/\underline{n}$ increases. What is perhaps more surprising is that the same happens when heterogeneity Δ decreases: The less heterogeneity in skills, the more heterogeneity in firm size.

This detour also allows to understand that Part (b) of Proposition 1 is a straightforward implication of Part (a). The first statement is expressed in terms of the span of control distribution, but it is heuristic in that the support of this distribution is moving as heterogeneity goes to zero (both the lowest and the highest sizes go to $+\infty$). The next statement is more rigorous, in that if one defines a new random variable, given by the relative size of a firm compared to the smallest firm $U \equiv N/\underline{n}$, then from proposition (1a):

$$\mathbb{P}[U \geq u] = \frac{1}{1-(1/\bar{u})^2}\left(\frac{1}{u^2} - \frac{1}{\bar{u}^2}\right).$$

The relative span of control follows a truncated Pareto distribution with a coefficient equal to two, and shape parameters given by $\underline{u}$ and $\bar{u}$ such that:

$$\underline{u} = 1 \quad \text{and} \quad \bar{u} = \frac{h\Delta}{\sqrt{1+h^2\Delta^2}-1}.$$

Then when $\Delta \to 0$, we have that $\bar{u} \to \infty$. For all u, we also have that:

$$\mathbb{P}[U \geq u] \to_{\Delta \to 0} \frac{1}{u^2}.$$

Therefore, the scaled distribution of firm size converges in distribution to a Pareto distribution with a coefficient equal to 2. The result is quite counterintuitive when stated as follows:

Claim 1. (Truncated Pareto distributions) With a power law production function, the concavity of the distribution of span of control in the upper tail is positively related to skill heterogeneity. In the limit, untruncated Pareto distributions correspond to no heterogeneity in underlying primitives.

Claim 1 is only an implication of Proposition 1. With non-zero heterogeneity, Pareto distributions are truncated, which results in a concave part on a log-log scale. This truncation of the Pareto distribution is pervasive in the empirical literature. Among many examples, the concave part is visible in Axtell's (2001) famous evidence about Zipf's law for firm size.

The results below will only confirm this insight to the case of very general density functions and multiple layers. Again, this result is potentially important because a substantial body of work in trade and firm heterogeneity has more generally used Pareto distributed primitives as an input, arguing that these were necessary to understand Pareto distributed firm size. If Pareto distributions instead come from power law production functions – for example, of the Garicano (2000) type – then how important is the truncation for the upper tail may be the only information to emerge.

[INSERT FIGURE 5 ABOUT HERE]

[INSERT FIGURE 6 ABOUT HERE]

Similarity to Physics. To use Sornette's (2006) taxonomy, this generation of endogenous Pareto distributions works through "Power Laws Change of Variable Close to the Origin." According to Newman (2005), "one might argue that this mechanism merely generates a power law by assuming another one: the power-law relationship between x and y generates a power-law distribution for x. This is true, but the point is that the mechanism takes some physical power-law relationship between x and y – not a stochastic probability distribution – and from that generates a power-law probability distribution. This is a non-trivial result." In this model, span of control is a change of variable close to the origin of the number of problems that a worker with skill x is unable to solve, as:

$$n(y) = \frac{1}{h(1-x)}.$$

To paraphrase Newman (2005), one can argue that this mechanism takes some economic power-law relationship between span of control and skill, and from that generates a power-law probability distribution. Of course, this is not exactly the mechanism that occurs in physics, because here the denominator never quite reaches zero, which leads to a truncated Pareto distribution at maximum size $\bar{n}$. Moreover, it is important to recognize that the generation of truncated Pareto distributions ultimately results from agents' optimizing choices (for example, that they sort through prices), and not solely from a mechanical production function. To the best of my knowledge, this model is the first in the economics literature, to generate Pareto distributions in firm size out of production functions.

Another proof using size-biased distributions. We can understand why two-layer firms follow Pareto distributions even more directly by using the concept of size-biased distributions. This distribution expresses the distribution of firm size from the workers' point of view: If a firm has 50 workers, then a size-biased distribution will count 50 firms of size 50. Workers of type x are in a firm with the following span of control:

$$n(m(x)) = \frac{1}{h(1-x)}.$$

With a uniform distribution for skill, and in the limit where Δ becomes very small, $1-z_2 = \dfrac{\sqrt{1+\Delta^2 h^2}-1}{h} \approx \dfrac{1}{2}\Delta^2 h \ll 1-z_1 = \Delta$. To a first approximation, therefore, one can see x as a uniform in $[z_1, 1]$. Thus the size-biased distribution of span of control if a Pareto distribution with a coefficient equal to 1. Since a size-biased Pareto distribution of coefficient α corresponds to a Pareto distribution of coefficient $\alpha+1$, this shows that the distribution of span of control is a Pareto distribution of coefficient two.

2.3 Bounded away from 0 density functions

At this point, one could argue that the uniform distribution is very special. After all, if the Pareto result ultimately relies on the distribution of skills' being uniform, then one would be left to explain where the uniform distribution ultimately comes from. But this is not the case. In fact, it turns out that under a mild assumption on the distribution of skills, the distribution is always a Pareto distribution with a coefficient equal to 2, in the upper tail. Using Mandelbrot's (1960) terminology, the distribution of span of control then follows the weak law of Pareto.

Proposition 2. *(Two-layer firms, density bounded away from 0 near 1)*

(a) With two-layers firms, and if the density function is bounded away from 0 near 1 (for example, the density is continuous and $f(1) \neq 0$), then the distribution of the span of managers' control over workers is a ***truncated Pareto distribution with a coefficient equal to 2 in the upper tail****. That is, the probability that span of control is higher than n for $n \in [\underline{n}, \bar{n}]$ is such that:*

$$\frac{\mathbb{P}[N \geq n]}{\frac{\bar{n}}{2}\frac{f(z_2)(1-z_2)}{1-F(z_2)}\left(\frac{1}{n^2}-\frac{1}{\bar{n}^2}\right)} \rightarrow_{n\rightarrow\bar{n}} 1.$$

(b) When $\Delta \rightarrow 0$, $\bar{n} \rightarrow \infty$ and the distribution of span of control becomes a ***full Pareto distribution with coefficient 2 in the upper tail (weak form of the law of Pareto)****. That is, for small enough Δ, and large enough n:*

$$\mathbb{P}[N \geq n] \sim_{\Delta\rightarrow 0} \frac{f(1)}{2\int_0^1 (1-y) f\left(1-\Delta+\Delta y\right) dy} \frac{\underline{n}^2}{n^2}.$$

More precisely, let $U = N/\underline{n}$ be the "scaled" span of control; then U converges in distribution to a Pareto distribution with a coefficient equal to 2 in the upper tail when heterogeneity Δ goes to 0. That is, for small enough Δ and for large enough *u:*

$$\mathbb{P}[U \geq u] \sim_{\Delta \to 0} \frac{f(1)}{2\int_0^1 (1-y) f(1-\Delta+\Delta y)\, dy} \frac{1}{u^2}.$$

Proof. See Appendix A.2. □

The intuition for this result is straightforward. Locally, one can always approximate any density function by a uniform distribution. Thus, locally, the distribution of span of control cannot be too far from the distribution obtained with a uniform density. The key is then to recognize that when heterogeneity in skills goes to zero, as in Proposition (1b), the production function transforms a bounded distribution of skills into an unbounded distribution for span of control. Thus, locally near 1 in the space of skills x corresponds to locally in the neighborhood of $+\infty$ in the space of span of control $n(y)$. Therefore, a Pareto distribution in the upper tail arises regardless of the underlying distribution of skills. Of course, how far in the upper tail Pareto turns out to be a good approximation is a quantitative question. Figures 7 and 8 show that quantitatively, for the case of an increasing distribution, given by:

$$F(z) = \frac{(z-(1-\Delta))^2}{\Delta^2} \mathbb{1}\left[1-\Delta, 1\right](z) + \mathbb{1}\left[1, +\infty\right](z),$$

Pareto is a very good approximation for the bulk of the distribution (at least when the density is rather slowly varying).

[INSERT FIGURE 7 ABOUT HERE]

[INSERT FIGURE 8 ABOUT HERE]

Proposition 2 allows us to state a second result.

Claim 2. (Deviations from Pareto distributions in the lower tail) With a power law production function, the upper tail of the span of control distribution is a truncated Pareto distribution. The upper tail of the span of control distribution corresponds to a very small part of the underlying skill distribution, which can be approximated by a uniform density to a first order. Deviations from Pareto distributions in the lower tail correspond to deviations from a uniform density of skills.

Claim 2 is an implication of Proposition 2. If Pareto distributions come from a power law production function, then Pareto distributions are not really informative on the distribution of skills, because they correspond to small parts of the underlying distribution. Again, this may explain the ubiquity of the Pareto distribution. Together, result 1 and result 3 show that the only informative parts in the span of control distribution are those that do not follow Pareto distributions: the lower tail and the very upper tail. In Section 2.5, this result will be confirmed with more layers.

2.4 General density functions

What happens when the assumption that the density is bounded away from zero is violated? I show that firms are then smaller, and thus do not appear in the upper tail. A natural starting point is to look at the sister of the uniform density in that case, which is a polynomial density for skills f on $[1-\Delta, 1]$, $f(x) = \frac{\rho+1}{\Delta^{\rho+1}}(1-x)^{\rho}$ and which also leads to closed-form solutions:

$$\begin{aligned}
&f(m(x))m'(x) = h(1-x)f(x)\\
&\qquad\Rightarrow\quad \frac{\rho+1}{\Delta^{\rho+1}}(1-m(x))^{\rho}\, m'(x) = h\frac{\rho+1}{\Delta^{\rho+1}}(1-x)^{\rho+1}\\
&\qquad\Rightarrow\quad \left[\frac{(1-m(u))^{\rho+1}}{\rho+1}\right]_x^{z_2} = h\left[-\frac{(1-u)^{\rho+2}}{\rho+2}\right]_x^{z_2}\\
&\qquad\Rightarrow\quad (1-m(x))^{\rho+1} = h\frac{\rho+1}{\rho+2}\left[(1-x)^{\rho+2} - (1-z_2)^{\rho+2}\right]\\
&\qquad\Rightarrow\quad 1-m^{-1}(y) = \left[\frac{1}{h}\frac{\rho+2}{\rho+1}(1-y)^{\rho+1} + (1-z_2)^{\rho+2}\right]^{\frac{1}{\rho+2}}.
\end{aligned}$$

The span of control distribution for two-layer firms is therefore given by:

$$n(y) = \frac{1}{h}\left[\frac{1}{h}\frac{\rho+2}{\rho+1}(1-y)^{\rho+1} + (1-z_2)^{\rho+2}\right]^{-\frac{1}{\rho+2}}.$$

I show in Appendix A.4 that in this case, the distribution of span of managers' control over workers is a truncated Pareto distribution with a coefficient equal to $\rho+2$ (uniform is a special case with $\rho = 0$). However, the maximum span of managers' control over workers in this case is then much lower than in the case in which the density is bounded away from zero, so that one should not expect to encounter firms resulting from such a density in the upper tail of the firm-size distribution, at least when heterogeneity in small. This is also shown in Appendix A.4, where it is shown that the maximum size is such that:

$$\bar{n} \sim_{\Delta\to 0} \frac{1}{h\Delta^{\frac{\rho+2}{\rho+1}}}.$$

Thus for small heterogeneity, the maximum size for $\rho > 0$ is always negligible compared to the maximum size for $\rho = 0$:

$$\frac{\bar{n}(\rho > 0)}{\bar{n}(\rho = 0)} = \Delta^{\frac{\rho}{\rho+1}} \to_{\Delta\to 0} 0.$$

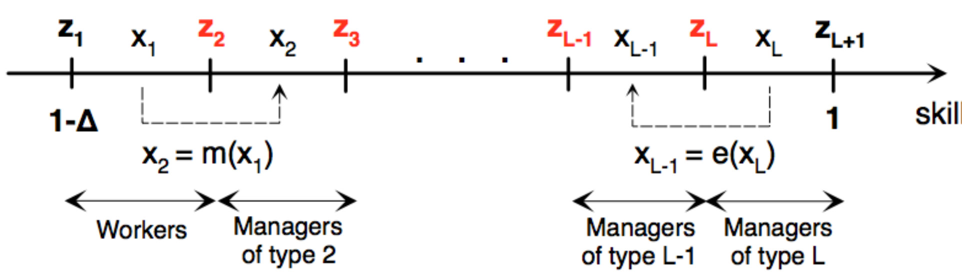
This is reinforced by the fact that larger firms are much less frequent in a Pareto with a higher tail coefficient equal to $\rho+2$, which is higher than ρ (the tail is thinner). It is also intuitive that all this reasoning results only from a local approximation, and that any density that has a Taylor expansion near 1 will lead to a Pareto for the distribution of span of control, but with much lower firm sizes. A counterexample to Pareto distributions for span of control is the density function given by $f(x) = \exp\left[-\frac{1}{(1-x)^2}\right]$, which goes so fast to zero near 1 that it does not have a Taylor expansion, and would not lead to Pareto distributions for span of control. However, there is no reason to expect the distribution of skills to be this badly behaved. Moreover, given the very small mass near the top of the distribution, the firm sizes corresponding to such a distribution would be infinitesimally small.

[INSERT FIGURE 9 ABOUT HERE]

2.5 An economy with L-layer firms: A static theory of Zipf's law

In the previous section, I have exogenously imposed that firms had only two layers, with managers and workers. Garicano's (2000) model, however, suggests that there should be an incentive for managers who supervise workers to report the hardest problems they face to a higher level manager. Production then works as follows: If the manager does not know how to solve the problem, she can also ask her own manager how to solve it – unless this manager is at the top of the hierarchy.[9] The main purpose of this section is to show that when allowing for multiple layers in the organization of firms, we obtain a new static theory of Zipf's law in the upper tail of the size distribution of firms when the number of layers of hierarchical organization becomes large.

Figure 2: L-LAYER FIRMS: NOTATIONS



For the same reason as in the two-layer case, there is positive sorting at all layers of the firm. Problems that a manager with higher skill cannot solve are harder statistically, regardless of his hierarchical position in the firm. In terms of span of control, however, there is a small twist to the two-layer case. For example, a manager of type 2 with skill x_2 receives problems of difficulty at least equal to $m^{-1}(x_2)$. The conditional probability that he is unable to solve them is given by $\frac{1-x_2}{1-x_1}$. Thus the manager's function is such that the time of managers of type 3 with skills in $[x_3, x_3 + dx_3]$ is used to answer the problems that managers of type 2 are unable to solve, according to the following time

[9] This is another improvement of the Garicano (2000) model over the Lucas (1978) model, because it provides a clear role for a pyramidal managerial structure.

constraint:

$$f(x_3)dx_3 = h\frac{1-x_2}{1-x_1}f(x_2)dx_2 \quad \Rightarrow \quad f\left(m(x_2)\right)m'(x_2) = h\frac{1-x_2}{1-m^{-1}(x_2)}f(x_2).$$

Cutoffs are similarly determined by the conditions that (a) the less skilled of each occupation type are matched with all others less skilled, and (b) the most skilled of each occupation type are matched with others more skilled. The first layer is special, in that the problems drawn by workers have not yet been sorted by anyone, or equivalently that they were sorted by someone who has chosen to pass on all problems (in other words, $m^{-1}\left([z_1, z_2]\right) = 0$. Thus these first-order differential equations for the managers' functions all collapse into:

$$\forall x \in]z_1, z_L[\setminus \{z_2, ..., z_{L-1}\}, \qquad f\left(m(x)\right)m'(x) = h\frac{1-x}{1-m^{-1}(x)}f(x)$$
$$\text{with} \qquad m^{-1}\left([z_1, z_2]\right) = 0, \quad m \text{ continuous}$$

As in the two-layer case, the span of control of the top manager with ability x_L is given by the multiplication of intermediary span of control distributions:

$$n(x_L) = \frac{1-x_{L-2}}{h(1-x_{L-1})} * \frac{1-x_{L-3}}{h(1-x_{L-2})} * ... * \frac{1-x_1}{h(1-x_2)} * \frac{1}{h(1-x_1)} = \frac{1}{h^{L-1}\left(1-m^{-1}(x_L)\right)}.$$

Proposition 3. *(L-layer firms, density bounded away from* 0 *near* 1*)*

(a) With L-layer firms, and if the density function is bounded away from 0 *near* 1 *(for example, the density is continuous and* $f(1) \neq 0$*), then the distribution of the span of managers' control over workers is a* ***truncated Pareto distribution with a coefficient equal to*** $\mathbf{1 + \frac{1}{L-1}}$ ***in the upper tail****. Denoting by* $\bar{n}$ *the maximum span of managers' control over workers, the measure of firms with a size higher than* n *is such that, for some constant* A_L*:*

$$\forall \epsilon > 0, \quad \exists A, \quad \forall n \geq A, \qquad \left| \mathbb{P}[N \geq n] - \frac{A_L}{h^L} f\left(1 - \frac{1}{\bar{n}h}\right)\left(\frac{1}{n^{1+\frac{1}{L-1}}} - \frac{1}{\bar{n}^{1+\frac{1}{L-1}}}\right)\right| < \epsilon.$$

(b) When $\Delta \to 0$*,* $\bar{n} \to \infty$ *and the distribution of span of control becomes a* ***full Pareto distribution with coefficient*** $\mathbf{1 + \frac{1}{L-1}}$ ***in the upper tail****:*

$$\forall \epsilon > 0, \quad \exists A, \quad \exists \eta, \quad \forall n \geq A, \quad \forall \Delta < \eta, \qquad \left| \mathbb{P}[N \geq n] - \frac{f(1)A_L}{h^L}\frac{1}{n^{1+\frac{1}{L-1}}}\right| < \epsilon.$$

Proof. See Appendix A.3. □

A qualitative intuition for this result is as follows. The overall span of managers' control over workers of type $L-1$ (CEOs of the firm) is given by the multiplication of intermediary span of control distributions. Intuitively, the distribution of span of control for multiple layers is fatter than that for only one layer. Proposition 3 states that the distribution of the span of managers' control over workers is a Pareto distribution with a coefficient equal to $\mathbf{1} + \frac{\mathbf{1}}{\mathbf{L-1}}$ (which is fatter-tailed than a Pareto distribution with a coefficient equal to 2, for any $L > 2$). Proposition 2 is a special case with $L = 2$, in which case the Pareto has a coefficient equal to 2. As the number of layers increases, the coefficient above approaches 1, which corresponds to Zipf's law for firm sizes. This allows us to state the following result:

Claim 3. (Zipf's Law from the multiplication of intermediary Pareto distributions) With a power law production function, the distribution of firm sizes follows a truncated Pareto distribution with a coefficient equal to $\mathbf{1} + \frac{\mathbf{1}}{\mathbf{L-1}}$ in the upper tail. When the number of layers becomes large, the coefficient of the Pareto distribution goes to 1, and therefore the distribution of firm size approaches Zipf's law.

[INSERT FIGURE 10 ABOUT HERE]

[INSERT FIGURE 11 ABOUT HERE]

Endogenous L. Of course, the number of layers is itself an endogenous object. There are many ways to endogenize the number of layers. One would be to assume a fixed cost of adding a new layer. Another is to determine this endogenous object by looking at a discrete counterpart of the continuous types model, with a given population of agents N. One then must assume that a manager needs to work full time at the top of his organization, which pins down the size of firms in the economy as well as their number of layers. L is then determined as a solution to the following:

$$L = \max_{L} \left\{ L \quad \text{s.t.} \quad 1 - z_L \geq \frac{1}{N} \right\}.$$

Table 1 shows the endogenous cutoffs in the case in which the density of skills is uniform. If there are one million workers, then the number of layers according to the above formula is given by $L = 6$.

[INSERT TABLE 1 ABOUT HERE]

3 Endogenous scale independent growth and stationarity

Section 2 offered an alternative theory of endogenous Pareto distributions, one that does not rely on a dynamic random growth process. In this section, I demonstrate that the two theories can actually be reconciled, as the static theory in Section 2 leads to endogenous scale independence in growth rates as well as to endogenous stationarity.

3.1 Endogenous Gibrat's law

In random growth theory, an exogenous "random growth" process is assumed in which the growth of firms is assumed to be independent of size – following the empirical observation, attributed to Gibrat (1931). Zipf's law arises, then, as the stationary solution to such a process. Intuitively, the stationary distribution that results from a scale independent process itself has to be scale independent. For example, Gabaix (1999) gives an intuition for Zipf's law for cities along these lines (p. 744). Symmetrically, it is easy to understand why, if a static mechanism produces Zipf's law, then the growth of firms generated by such a static mechanism will produce scale-independent growth over time, or Gibrat's (1931) law.

One example is to look at comparative statics, in which agents would improve their skills to equal those of the most skilled agents, so that the support $[1-\Delta, 1]$ would shift over time. Then it is easy to see that all firms, whether they are small or large, would grow at the same rate. This can, for example, be seen in Figure 5. Another example is to look at the comparative statics on firm-size distribution when the helping time h changes. Again, this would lead small firms to grow at the same rate as large firms.

Even though these results are rather straightforward from those in Section 2, one must not forget how challenging it is to many theories of the firm that large firms would on average grow at the same rate as small firms. Indeed, in many theories of the firm, the average cost curve determines how large a firm is. Once a firm has reached its optimal size, there is no reason why it should grow further (or as fast). In contrast, a theory of the firm-size distribution based on a power law production function does deliver scale-independent growth. The intuition is that as workers improve, they take less and less time to manage, and that this can help even large firms to grow in an unbounded fashion.

Some scholars have argued that random growth theory is not microfounded, and does not provide a sound economic model. For example, Penrose (1955) writes that random growth models "leave no room for human motivation and conscious human decision, and I think should be rejected on that ground." In contrast, the above model is microfounded, and helps explain why even large firms appear to be able to grow in an unbounded fashion.

3.2 Endogenous stationarity

Another key assumption in random growth theory is that one needs to look for a stationary distribution. However, most scale-independent random growth processes in fact lead to nonstationary distributions, and in the end this is a restriction that is imposed on the process itself. It is important to note that stationarity is an assumption in those models, and that they do not explain why, despite overwhelming transformations of the economy, the firm size distribution has remained the same over centuries. In contrast, because my model is static, the fact that the exact same distribution (Zipf's law) is always obtained is a conclusion, not an assumption. Of course, this assumes that the power law production function is sufficiently fundamental that it also did not change over the centuries.

4 Empirics

The model presented in this paper yields highly precise and testable predictions about how Zipf's law arises in the data. In particular, one distinctive implication of the model is that not only does firm-size distribution follow a Pareto distribution of coefficient 1, but that intermediary span of control distributions also follow Pareto distributions, with higher tail coefficients. In particular, I show that in theory, two-layer firms should follow Pareto distributions with a coefficient equal to 2 in the upper tail. In Section 4.1, I show how the theory can help explain some of the puzzles in the empirical literature on firm growth and size. In Section 4.2, I present new evidence from the French matched employer-employee data, which provides strong support for the mechanism that I propose.

4.1 Related literature

The explanation for power laws in this paper allows us to connect different dots in the existing empirical literature concerning firm growth and firm and establishment size.

Firm birth. Cabral and Mata (2003) show, using Portuguese micro-level data, that the distribution of firm size is already very skewed to the right at the time of birth. More importantly, some have argued that the faster growth rate of small firms compared to large firms is still consistent Gibrat's law if there is high 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 size. The model presented above can make sense of these facts, as Pareto distributions arise out of a static model.

Distribution of establishment size. It has long been known that the distribution of establishments is less fat tailed than that of firms. For example, Luttmer (2010) writes, in his review of the random growth literature: "In the United States, the right tail of the size distribution of establishments is noticeably thinner than that of firms." Similarly, 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.

The model I presented is well suited to address this issue, as it is precisely how Zipf's law is generated: through the multiplication of Pareto distributions with a higher tail coefficient. Figure 17 illustrates the fact that establishment size is also Pareto distributed in the upper tail in the US, albeit with a higher tail coefficient, through publicly available US Census data. To the best of my knowledge, no random growth mechanism to date has been able to explain simultaneously the distribution of establishment size and firm size. This provides a key empirical validation of the model.

[INSERT FIGURE 17 ABOUT HERE]

4.2 French matched employer-employee data

In addition to the empirical evidence, the French matched employer-employee administrative data also strongly supports this proposition. I follow Caliendo et al. (2015) closely in defining hierarchies in this dataset.

Data. I use a sample of French Déclarations Annuelles de Données Sociales (DADS). I report only the data from a sample in year 2007; however other time periods have very similar profiles. The sample originally contains $55,979,881$ observations of employee-employer matched pairs.

I follow Caliendo et al. (2015) and use the first digit of the PCS variable (PCS stands for "Profession, Catégorie Socioprofessionnelle," or social class based on occupation) as a proxy for the hierarchical position of workers in a firm. This allows me to divide workers into subgroups of high-ranking managers, middle managers, workers, and manual workers. More precisely, the first digit of the PCS variable is one of six alternatives, as follows: one, the farmers; two, self-employed and owners (for example, plumbers, firm directors, CEOs); three, senior staff or top management positions (for example, chief financial officers, heads of human resources, purchasing managers); four, employees at the supervisor level (for example, quality control technicians, sales supervisors); five, clerical and other white-collar employees (for example, secretaries, human resource or accounting, and sales employees); and six, blue-collar workers (for example, assemblers, machine operators, maintenance workers).

Unfortunately, the PCS variable is not available for all workers, and farmers and manual workers are in separate categories (first digit equals to 1 and 2), so I drop them.

This leaves me with employer-employee matches corresponding to the PCS variable having a first digit equal to 3, 4, 5, or 6.

Results. In Section 2, I argued that with a power law production function, the distribution of span of control down one level of hierarchical organization should be close to a Pareto distribution with coefficient 2, whatever the underlying distribution of skills (provided it is bounded away from zero for high skills). Figure 18 plots the corresponding distribution of span of control that is obtained empirically, using the dataset described above. The total number of employees in a given layer, as defined by the first digit of the PCS variable, is divided by the total number of employees occupying the layer above them (again, this is seen through the PCS variable). This measure is calculated at the establishment level. These ratios are referred to on the x-axis as "team size." Their distribution is shown on a log-log plot, with rank on the y axis (plotting the survivor function would give a similar picture). As can be seen in Figure 18, the data seems to point to a Pareto distribution for large spans of control in the upper tail, and a coefficient close to that predicted by the theory, 1.96.

[INSERT FIGURE 18 ABOUT HERE]

Figures 19 and 20 add further support to the theory developed in this paper. They are constructed using a similar methodology as Figure 18. Figure 19 shows that the distribution of firm size is indeed more fat tailed than that of teams, establishments, or the number of plants per firm. Figure 20 demonstrates that this pattern is not specific to a particular sector. Indeed, the sector "wholesale trade, accommodation, food" shows very similar patterns as the overall economy, and so do other two-digit NAICS industries in unreported graphs. Finally, Figure 21 shows that the distribution of the number of establishments per firm is strikingly close to what the theory predicts. As shown in Section 2, the distribution is expected to get closer to a Pareto distribution as one approaches the highest levels of hierarchical organization, as the uniform approximation for the distribution of skills becomes a very good approximation for the top (given that the corresponding cutoffs z_l are very close to 1). This is exactly what can be seen on Figure 21 – compared, for example, to Figure 18 – where Pareto is a good approximation for most of the distribution and not just for the upper tail. Moreover, note also how the Pareto is shifted for very high levels of span of control (that is, concave, and bounded), just as the theory predicts for non-zero heterogeneity in skills. The coefficient estimated for this distribution (1.33) further seems to suggest that there are two levels of hierarchical organization between the CEO of the firm and each one of the establishment managers.

[INSERT FIGURE 19 ABOUT HERE]

[INSERT FIGURE 20 ABOUT HERE]

[INSERT FIGURE 21 ABOUT HERE]

5 Labor income distribution

Section 2 offers a theory of the equilibrium span of control distribution, without any reference to supporting skill prices. However, it is also a theory of the division of surplus between agents with different skills, across and within firms.

In this section, I focus attention on the top labor income distribution, for two reasons. First, the distribution of labor incomes is least sensitive to distributional assumptions in the tail, where it is dominated by a functional form assumption on the production function – which is the focus of this paper. Second, the increase in top income inequality, particularly at the very top (top 1%, top 0.1%), has received much attention since the work of Piketty and Saez (2003). However, the reader should keep in mind that the model also implies a theory of the whole labor income distribution, at different levels of the hierarchy of the firm.

Section 5.1 derives the top labor income distribution, which displays Pareto-like behavior under some conditions. These results are very similar to those obtained in Terviö (2008) and Gabaix and Landier (2008), who study the one-to-one matching of CEOs to a distribution of firms whose sizes exogenously follow Zipf's law. Section 5.2 discusses some advantages of endogenizing Zipf's law for firms, as opposed to taking that distribution as given.

5.1 Top labor income distribution

The optimal allocations derived in Section 2 result from managers' optimal choices. For example, with two-layer firms, as in Section 2.1, and the notations on Figure 1, the wage of managers with skill $w(y)$ is given by the solution to managers' optimal choice of workers' ability x:

$$w(y) = \max_x \frac{y - w(x)}{h(1-x)},$$

where the wage of a worker with ability x is $w(x)$. These skill prices allow us to sustain the match between the best managers and the best workers, which was derived in Section 2. The envelope condition writes:

$$w'(y) = \frac{1}{h(1-m^{-1}(y))} = n(y).$$

This assignment equation has a straightforward interpretation. The difference in wages $dw(y)$ between two managers with skill y and $y + dy$ is given by the difference in their respective talents dy, multiplied by the size of the firm on which they exert that differential talent $n(y)$ (that is also equal to the number of problems to which this differential solving ability can be applied). This explains why $dw(y) = n(y)dy$.

More generally, with L layers, as in Section 2.5, and the notations on Figure 2, the envelope condition for a manager with skill x_L is as follows:

$$w'(x_L) = \frac{1}{h^{L-1}(1 - m^{-1}(x_L))} = n(x_L).$$

Proposition 4 simply integrates this classic assignment equation (Sattinger (1975)).

Proposition 4. *(Top labor income distribution) The incomes at the top are given by the integral of the span of control $n(y)$:*

$$w(x_L) = w(z_L) + \int_{z_L}^{x_L} n(y)dy,$$

where the distribution of span of control of CEOs $n(.)$ is given by a truncated Pareto distribution in the upper tail, as shown in Proposition 3. As in Gabaix and Landier (2008), the top income distribution thus displays Pareto-like behavior.

Matching CEOs to preexisting firms: Comparing assignment equations. The expression in Proposition 4 corresponds exactly to the formula obtained in earlier work by Terviö (2008) and Gabaix and Landier (2008). These two papers consider a one-to-one assignment problem in which there are firms of different sizes on one side of the market (given by Zipf's law), and managers with different skills on the other side of the market. Assuming some multiplicative complementarity between skill and size (potentially to some power), and denoting the rank of firms and CEOs by n, the size of firms by $S(n)$ and the talent of CEOs by $T(n)$, they obtain the following assignment equation (equation (5), p. 57):

$$w'(n) = CS(n)^{\gamma}T'(n),$$

with C a constant. In the Garicano (2000) model above, one can similarly relabel agents' talents by $y(n)$, with n the rank of the manager in terms of ability; rewriting the equation above allows us to recognize the usual assignment equation:

$$\frac{dw(y)}{dy} = n(y) \quad \Rightarrow \quad \frac{dw(y(n))}{dn} = \underbrace{n(y(n))}_{\equiv S(n)} \underbrace{\frac{dy(n)}{dn}}_{T'(n)}.$$

Note however that, compared to Gabaix and Landier (2008), the present model provides a microfoundation for why $\gamma = 1$ and why the talent of CEOs and the sizes of their firms enter multiplicatively.

Relation to the economics of superstars. A key difference between this model and earlier literature that matches CEOs to preexisting firms is that the model also explains why the size of the stakes is so spread out across managers. Endogenizing Zipf's law thus also allows us to understand how "economics of superstars" (Rosen (1981)) arise. Even without any increasing returns in production or consumption, market forces lead the most skilled to have a high span of control, simply because they hire people who can almost work independently. This, in turn, allows them to specialize in tasks they have expertise for. The distribution of the size of these stakes is given by Zipf's law when the number of layers of management becomes large for the reason explained in detail in Section 2.

Example: Uniform-polynomial density. Consider the following density function:

$$f(x) = \begin{cases} A_1 & \text{if} \quad x \in [1-\Delta_1-\Delta_2, 1-\Delta_2] \\ A_2(\rho+1)(1-x)^\rho & \text{if} \quad x \in [1-\Delta_2, 1] \end{cases}$$

I assume two-layer firms. One can easily show that the inverse of the matching function is then such that:

$$1 - m^{-1}(y) = \sqrt{(1-z_2)^2 + \frac{2}{h}\frac{A_2}{A_1}(1-y)^{\rho+1}}.$$

Thus, this gives span of control $n(y)$ in closed form, which is given by a truncated Pareto distribution with coefficient 2. The distribution of wages $w(y)$ is the integral of this span of control distribution:

$$w(y) = w(z_2) + \int_{z_2}^{y} \frac{du}{h\sqrt{(1-z_2)^2 + \frac{2}{h}\frac{A_2}{A_1}(1-u)^{\rho+1}}}.$$

In fact, the wage function can also be expressed in closed form, using hypergeometric functions. Figures 12 and 13 show that the integration of this truncated Pareto distribution leads to a distribution for CEO income that is quite close to a Pareto distribution in the upper tail, when $\rho > 0$.

[INSERT FIGURE 12 ABOUT HERE]

[INSERT FIGURE 13 ABOUT HERE]

In contrast, when $\rho = 0$, one gets a very compressed distribution for wages, as managers compete too hard for the best workers. Section 5.2 shows that these comparative statics with respect to ρ lead to a potential disconnect between the largest firms and the largest incomes, and help explain the noise in the size-pay relationship in the data.

5.2 Endogenous Zipf's law versus reduced-form approach

Section 5.1 demonstrated that the Garicano (2000) model leads to the same assignment equation as the reduced form approach used by Terviö (2008) and Gabaix and Landier (2008) . What are the differences between the full structural approach and a more reduced form for the study of top labor incomes?

Integrating truncated Pareto distributions. One slight difference with regard to Gabaix and Landier (2008) is that they do not consider truncated Zipf's laws, but rather full Zipf's laws, and thus the wage of CEOs results from integrating a truncated Pareto distribution instead of an exact one.

This in fact, turns out to lead to quite different comparative statics, as the change in average firm size (for example, through a change in h) leads to both a change in the scale of the top labor income distribution and a change in the tail index of that distribution. This is potentially interesting, because empirically, the tail index of the top labor income distribution has also evolved in the direction of more inequality (lowering of the tail index).

Largest firms and largest incomes. Terviö (2008) obtains a strong result. The rank of CEO's wage and CEO's span of control should be perfectly correlated: The largest firm should pay its CEO the most, etc. However, the data speak more ambiguously to the relationship between size and pay. This relationship is positive, but also quite noisy, as seen in the Execucomp data in Figure 16. Some firms are very large, but their CEOs' incomes are not correspondingly large; in contrast, some CEOs have very large incomes, but their firms are not particularly large.

[INSERT FIGURE 16 ABOUT HERE]

Endogenizing Zipf's law for firm size allows us to offer a possible explanation for this. Imagine that there are multiple sectors in the economy, and that workers cannot move across sectors (that is, the CEO of a manufacturing company cannot become a CEO for a tech company). Assume that different sectors have different parameters, h and Δ. Then, as shown in Figures 14 and 15, this would lead to an ambiguous relationship between size and pay.

[INSERT FIGURE 14 ABOUT HERE]

[INSERT FIGURE 15 ABOUT HERE]

Wages in other layers of the firm. Another advantage of endogenizing the size and organization of firms is that one can look at the wages of workers in all layers of the firm. With L layers, one can show that the wage function is a solution of the following system of ordinary differential equations (together with pasting conditions at the cutoffs, expressing the fact that the wage function needs to be continuous by arbitrage):

$$\forall x \in]z_1, z_L[\setminus \{z_2, ..., z_{L-1}\}, \qquad w'(x) = h \frac{w\,(m(x))}{1 - m^{-1}(x)}$$
$$\text{with} \qquad w\,([z_L, z_{L+1}]) = \frac{1}{h^{L+1}}, \quad w \text{ continuous.}$$

Managers compete for workers (especially the best, who allow the greatest increase in span of control), so that increased wages for CEOs usually occur together with increased wages for all workers. One advantage of a general equilibrium model with endogenous firm size is that it allows us to study these "trickle-down" effects.

6 Conclusion

This paper has shown using a knowledge-based hierarchies model à la Garicano (2000), that Pareto distributions can arise from production functions rather than solely from a random growth proportional process or assumed Pareto heterogeneity in primitives. This model offers a new microfounded justification for the existence of Zipf's law for firm size, and provides a new intuition for why many economic variables tend to follow Pareto distributions.

However stylized the model may be, I believe that it captures a powerful amplification mechanism that goes beyond production based on knowledge. Managers hire people to do most of the work of an organization and focus their own efforts on the most difficult tasks, for which they have expertise. This endogenously leads to Zipf's law for firm size when the number of layers of hierarchical organization becomes large.

In terms of labor income distribution, the model really concerns CEOs' ability to delegate more tasks to senior management when these employees are relatively more competent. For Robinson Crusoe living on an island, being able to solve 99.9% of 99.99% of problems would not make much of a difference. For a manager, hiring the latter rather than the former would allow him to grow his firm by a factor of 10, and to expand his particular knowledge accordingly. This intuition – that Pareto distributions may simply result from a "power law change of variable close to the origin" – is (to the best of my knowledge) new to economics. It is also simple, and may be at the heart of many Pareto distributions that are observed empirically.

References

Abowd, John M., Francis Kramarz, and David N. Margolis, "High Wage Workers and High Wage Firms," *Econometrica*, 1999, *67* (2), 251–333.

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

Axtell, Robert L., "Zipf Distribution of U.S. Firm Sizes," *Science*, 2001, *293*, 1818–1821.

Beckmann, Martin J., "City Hierarchies and the Distribution of City Size," *Economic Development and Cultural Change*, 1958, *6* (3), 243–248.

Cabral, Luis M.D. and José Mata, "On the Evolution of the Firm Size Distribution: Facts and Theory," *American Economic Review*, 2003, *93* (4), 1075–1090.

Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg, "The Anatomy of French Production Hierarchies," *Journal of Political Economy*, 2015, *123* (4), 1–75.

Champernowne, David G., "A Model of Income Distribution," *Economic Journal*, 1953, *63* (250), 318–351.

Chaney, Thomas, "Distorted Gravity: The Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux, "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection," *Econometrica*, 2012, *80* (6), 2543–2594.

Gabaix, Xavier, "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics*, 1999, *114* (3), 739–767.

_ , "Power Laws in Economics: An Introduction," *Journal of Economic Perspectives*, 2016, *30* (1), 185–206.

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

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

_ **and Esteban Rossi-Hansberg**, "Organization and Inequality in a Knowledge Economy," *Quarterly Journal of Economics*, 2006, *121* (4), 1383–1435.

Geerolf, François, "Leverage and Disagreement," *Working Paper*, 2015, pp. 1–61.

Gibrat, Robert, *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*, Paris: Librairie du Recueil Sirey, 1931.

Helpman, Elhanan, Marc J. Melitz, and Stephen R Yeaple, "Export Versus FDI with Heterogeneous Firms," *American Economic Review*, 2004, *94* (1), 300–316.

Hsu, Wen-Tai, "Central Place Theory and City Size Distribution," *Economic Journal*, 2012, *122* (563), 903–932.

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670.

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

Kremer, Michael and Eric Maskin, "Wage Inequality and Segregation by Skill," *NBER Working Paper 5718*, 1996.

Lucas, Robert E., "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 1978, *9* (2), 508–523.

_ **and Benjamin Moll**, "Knowledge Growth and the Allocation of Time," *Journal of Political Economy*, 2014, *122* (1), 1–51.

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

_ , "Models of Growth and Firm Heterogeneity," *Annual Review of Economics*, sep 2010, *2* (1), 547–576.

Lydall, Harold F., "The Distribution of Employment Incomes," *Econometrica*, 1959, *27* (1), 110–115.

Mandelbrot, Benoît, "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review*, 1960, *1* (2), 79–106.

Newman, Mark, "Power Laws, Pareto Distributions and Zipf's law," *Contemporary Physics*, 2005, *46* (1), 323–351.

Pareto, Vilfredo, "La Legge della Domanda," *Giornale Degli Economisti*, 1895, *10* (6), 59–68.

Penrose, Edith, "Limits to the Growth and Size of Firms," *American Economic Review, Papers and Proceedings*, 1955, *45* (2), 531–543.

Perla, Jesse and Christopher Tonetti, "Equilibrium Imitation and Growth," *Journal of Political Economy*, 2014, *122* (1), 1–21.

Persky, Joseph, "Retrospectives: Pareto's Law," *Journal of Economic Perspectives*, 1992, *6* (2), 181–192.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913-1998," *Quarterly Journal of Economics*, 2003, *118* (1), 1–39.

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

Rossi-Hansberg, Esteban and Mark L.J. Wright, "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*, 2007, *97* (5), 1640–1666.

Sattinger, Michael, "Comparative Advantage and the Distributions of Earnings and Abilities," *Econometrica*, 1975, *43* (3), 455–468.

Simon, Herbert A. and Charles P. Bonini, "The Size Distribution of Business Firms," *American Economic Review*, 1958, *48* (4), 607–617.

Song, Jae, David J. Price, Nicholas Bloom, Jae Song, and Till Von Wachter, "Firming Up Inequality," *NBER Working Paper No. 21119*, 2015.

Sornette, Didier, "Mechanism for Powerlaws Without Self-Organization," *International Journal of Modern Physics*, 2002, *13* (2), 3.

_ , *Critical Phenomena in Natural Sciences - Chaos, Fractals, Self-Organization and Disorder: Concepts and Tools*, Springer Verlag, 2006.

Terviö, Marko, "The Difference that CEOs Make: An Assignment Model Approach," *American Economic Review*, 2008, *2* (98), 642–668.

Tinbergen, Jan, "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 1956, *77*, 155–175.

Zipf, George K., *Human Behavior and the Principle of Least Effort*, Cambridge, MA: Addison-Wesley Press, 1949.

A Main Proofs

A.1 Proof of Proposition 1

Proof. Proposition (1a) results from the expression for span of control $n(y)$:

$$n(y) = \frac{1}{h\left(1 - m^{-1}(y)\right)} \quad \Rightarrow \quad \mathbb{P}[N \geq n] = \mathbb{P}\left[\frac{1}{h\left(1 - m^{-1}(y)\right)} \geq n\right].$$

The maximum span of control is:

$$\bar{n} = n(1) = \frac{1}{h(1 - z_2)}.$$

The minimum span of control is:

$$\underline{n} = n(z_2) = \frac{1}{h(1 - z_1)} = \frac{1}{h\Delta}.$$

With f a uniform density, we have:

$$m'(x) = h(1 - x) \quad \Rightarrow \quad m(z_2) - m(x) = 1 - m(x) = h\frac{(1-x)^2}{2} - h\frac{(1-z_2)^2}{2}.$$

Evaluated at z_1, with $m(z_1) = z_2$, this gives:

$$h^2(1 - z_2)^2 + 2h(1 - z_2) - h^2\Delta^2 = 0 \quad \Rightarrow \quad \frac{1}{\bar{n}^2} + \frac{2}{\bar{n}} - \frac{1}{\underline{n}^2} = 0 \quad \Rightarrow \quad \frac{\bar{n}}{2} = \frac{\underline{n}^2}{1 - (\underline{n}/\bar{n})^2}.$$

This also shows that:

$$1 - z_2 = \frac{\sqrt{1 + h^2\Delta^2} - 1}{h}.$$

The highest span of control $\bar{n}$ is:

$$\bar{n} = \frac{1}{h(1 - z_2)} = \frac{1}{\sqrt{1 + h^2\Delta^2} - 1}$$

Since y is distributed uniformly over $[1 - \Delta, 1]$:

$$\begin{aligned}
\mathbb{P}[N \geq n] &= \mathbb{P}\left[y \geq m\left(1 - \frac{1}{nh}\right) \,\middle|\, y \geq z_2\right] \\
&= \frac{1}{1 - z_2}\left[1 - m\left(1 - \frac{1}{nh}\right)\right] \\
&= \frac{1}{1 - z_2}\left[\frac{h}{2}\frac{1}{n^2h^2} - \frac{h}{2}(1 - z_2)^2\right] \\
&= \frac{\bar{n}}{2}\left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right) \\
\mathbb{P}[N \geq n] &= \frac{\underline{n}^2}{1 - (\underline{n}/\bar{n})^2}\left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right).
\end{aligned}$$

This proves Proposition (1a). Proposition (1b) looks at the no-heterogeneity limit $\Delta \to 0$. When Δ approaches 0, we have that:

$$\bar{n} \sim \frac{2}{h^2\Delta^2}.$$

This shows simultaneously that when Δ goes to 0, $\bar{n}$ goes to infinity and $\bar{n}/\underline{n}$ goes to infinity, and gives the result on the untruncated Pareto with a tail index equal to 2. Note that $n(y)$

also has a closed form expression as a function of y for f a uniform density, as:

$$m'(x) = h(1-x) \quad \Rightarrow \quad m(z_2) - m(x) = 1 - m(x) = h\frac{(1-x)^2}{2} - h\frac{(1-z_2)^2}{2}$$
$$\Rightarrow \quad 1 - m^{-1}(y) = \sqrt{(1-z_2)^2 + \frac{2}{h}(1-y)} \quad \Rightarrow \quad n(y) = \frac{1}{\sqrt{2h}} \frac{1}{\sqrt{(1-y) + \dfrac{h(1-z_2)^2}{2}}}.$$

Finally, a more rigorous statement of proposition (1b) follows from writing everything in terms of the scaled size distribution U, with $U = N/\underline{n}$, so that Proposition (1a) writes:

$$\mathbb{P}[U \geq u] = \frac{1}{1-(1/\bar{u})^2}\left(\frac{1}{u^2} - \frac{1}{\bar{u}^2}\right),$$

which proves convergence in distribution:

$$\forall u \geq 1, \qquad \mathbb{P}[U \geq u] \to_{\Delta \to 0} \frac{1}{u^2}.$$

□

Note that one can even prove a stronger result: There is convergence in law according to norm L^1 (for example) to a full Pareto distribution, as, denoting by $F_\Delta(.)$ the c.d.f. of U when heterogeneity is Δ and F the c.d.f. of the Pareto distribution with scale 1 and tail index 2:

$$\int |F_\Delta(u) - F(u)| du \to_{\Delta \to 0} 0.$$

Indeed, we have that:

$$\int |F_\Delta(u) - F(u)| du = \int_1^{\bar{u}} \left(F_\Delta(u) - F(u)\right) du + \int_{\bar{u}}^{+\infty} \left(1 - F(u)\right) du$$

with:

$$\int_1^{\bar{u}} \left(F_\Delta(u) - F(u)\right) du = \int_1^{\bar{u}} \left[(1-F(u)) - (1 - F_\Delta(u))\right] du$$
$$= \int_1^{\bar{u}} \frac{\dfrac{1}{\bar{u}^2}}{1 - \dfrac{1}{\bar{u}^2}} \left(1 - \frac{1}{u^2}\right) du$$
$$\int_1^{\bar{u}} \left(F_\Delta(u) - F(u)\right) du = \frac{\dfrac{1}{\bar{u}} + \dfrac{1}{\bar{u}^3} - \dfrac{2}{\bar{u}^2}}{1 - \dfrac{1}{\bar{u}^2}}.$$

and:

$$\int_{\bar{u}}^{+\infty} \left(1 - F(u)\right) du = \frac{1}{\bar{u}}.$$

After some algebra:

$$\int |F_\Delta(u) - F(u)| du = \frac{2}{1+\bar{u}} \to_{\Delta \to 0} 0.$$

A.2 Proof of Proposition 2

Proof. The first part of proposition (2a) is a local result. It also results from:

$$n(y) = \frac{1}{h\left(1 - m^{-1}(y)\right)} \quad \Rightarrow \quad \mathbb{P}[N \geq n] = \mathbb{P}\left[\frac{1}{h\left(1 - m^{-1}(y)\right)} \geq n \,\middle|\, y \geq z_2\right].$$

The maximum span of control is given by:

$$\bar{n} = n(1) = \frac{1}{h(1 - z_2)} \quad \Rightarrow \quad z_2 = 1 - \frac{1}{\bar{n}h}.$$

The probability that span of control is higher than n for $n \in [\underline{n}, \bar{n}]$ is given by:

$$\begin{aligned}
\mathbb{P}[N \geq n] &= \mathbb{P}\left[y \geq m\left(1 - \frac{1}{nh}\right) \,\middle|\, y \geq z_2\right] \\
&= \frac{1}{\mathbb{P}\left[1 - F(y) \leq 1 - F(z_2)\right]} \mathbb{P}\left[1 - F(y) \leq 1 - F\left[m\left(1 - \frac{1}{nh}\right)\right]\right] \\
&= \frac{1 - F\left[m\left(1 - \frac{1}{nh}\right)\right]}{1 - F(z_2)} \\
\mathbb{P}[N \geq n] &= \frac{1}{1 - F(z_2)} \int_{1-\frac{1}{nh}}^{1-\frac{1}{\bar{n}h}} h(1-u) f(u) du
\end{aligned}$$

As a first step, let us prove that:

$$\frac{\mathbb{P}[N \geq n]}{\frac{\bar{n}}{2} \frac{f(z_2)(1 - z_2)}{1 - F(z_2)} \left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right)} \to_{n \to \bar{n}} 1.$$

Indeed:

$$\begin{aligned}
\int_{1-\frac{1}{nh}}^{1-\frac{1}{\bar{n}h}} h(1-u) du &= \frac{h}{2} \frac{1}{n^2 h^2} - \frac{h}{2} \frac{1}{\bar{n}^2 h^2} \\
&= \frac{1}{2h}\left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right) \\
\int_{1-\frac{1}{nh}}^{1-\frac{1}{\bar{n}h}} h(1-u) du &= \frac{\bar{n}}{2}(1 - z_2)\left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right)
\end{aligned}$$

Thus:

$$\left|\mathbb{P}[N \geq n] - \frac{\bar{n}}{2} \frac{f(z_2)(1 - z_2)}{1 - F(z_2)} \left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right)\right| \leq \frac{1}{1 - F(z_2)} \int_{1-\frac{1}{nh}}^{1-\frac{1}{\bar{n}h}} h(1-u) \left|f(u) - f(z_2)\right| du$$

For n close enough to $\bar{n}$, $|f(u) - f(z_2)|$ is arbitrarily small; thus for any epsilon, there exists a neighborhood of $\bar{n}$ such that for n in this neighborhood:

$$\frac{1}{1 - F(z_2)} \int_{1-\frac{1}{nh}}^{1-\frac{1}{\bar{n}h}} h(1-u) \left|f(u) - f(z_2)\right| du \leq \epsilon \frac{\bar{n}}{2} \frac{f(z_2)(1 - z_2)}{1 - F(z_2)} \left(\frac{1}{n^2} - \frac{1}{\bar{n}^2}\right).$$

Proposition (2b) results from the first. Integrating equation (2) between z_1 and z_2:

$$1 - F(z_2) = h \int_{1-\Delta}^{z_2} (1 - x) f(x) dx \sim_{\Delta \to 0} h \int_{1-\Delta}^{1} (1 - x) f(x) dx.$$

For $\Delta \to 0$, we have:

$$1 - F(z_2) \sim_{\Delta \to 0} f(1)(1 - z_2).$$

Thus:

$$\frac{\bar{n}}{2} = \frac{1}{2h(1-z_2)} \sim_{\Delta \to 0} \frac{1}{2h^2} \frac{f(1)}{\displaystyle\int_{1-\Delta}^{1} (1-x)f(x)dx}.$$

Using that:

$$\underline{n}^2 = \frac{1}{h^2\Delta^2}, \quad \text{and} \quad \frac{f(z_2)(1-z_2)}{1-F(z_2)} \to_{\Delta \to 0} 1.$$

we have:

$$\mathbb{P}[N \geq n] \sim_{\Delta \to 0} \frac{f(1)\Delta^2}{\displaystyle\int_{1-\Delta}^{1} 2(1-x)f(x)dx} \frac{\underline{n}^2}{n^2}.$$

The distribution of scaled span of control is given by:

$$\mathbb{P}[U \geq u] \sim_{\Delta \to 0} \frac{f(1)\Delta^2}{\displaystyle\int_{1-\Delta}^{1} 2(1-x)f(x)dx} \frac{1}{u^2}.$$

Finally, using the change of variable $x = 1 - \Delta + \Delta y$, one gets:

$$\mathbb{P}[U \geq u] \sim_{\Delta \to 0} \frac{f(1)}{\displaystyle\int_{0}^{1} 2(1-y)f\left(1-\Delta+\Delta y\right) dy} \frac{1}{u^2}$$

Note that Proposition (1b) is a special case with $f(x) = 1/\Delta$ for $x \in [1-\Delta, 1]$, since then:

$$\frac{f(1)}{\displaystyle\int_{0}^{1} 2(1-y)f\left(1-\Delta+\Delta y\right) dy} = 1.$$

□

A.3 Proof of Proposition 3

Proof. The span of control distribution with L layers results from:

$$n(y) = \frac{1}{h^{L-1}\left(1-m^{-1}(y)\right)} \quad \Rightarrow \quad \mathbb{P}[N \geq n] = \mathbb{P}\left[\frac{1}{h^{L-1}\left(1-m^{-1}(y)\right)} \geq n\right].$$

$$\mathbb{P}[N \geq n] = \mathbb{P}\left[y \geq m\left(1-\frac{1}{nh^{L-1}}\right)\right] \sim f(1)\left[1-m\left(1-\frac{1}{nh^{L-1}}\right)\right].$$

One thus needs to find a first-order approximation to $1-m(.)$ for $y \in [z_L, z_{L+1}]$. Given the recursive nature of the ordinary differential equations, one needs to integrate from the bottom to the top of the hierarchy. Managers of type 1 of type x_2 are matched with workers of type x_1 according to:

$$f(x_2)dx_2 = h(1-x_1)f(x_1)dx_1 \quad \Rightarrow \quad 1-x_2 \sim \frac{h}{2}(1-x_1)^2.$$

Let us prove more generally by recursion that, for some sequence $\{A_l\}_{l=2}^{+\infty}$:

$$1-x_L \sim A_L\left(1-x_{L-1}\right)^{\frac{L}{L-1}}$$

The previous calculations show that this proposition is true for $L = 2$ with $A_2 = \frac{h}{2}$. At a second iteration, one can write:

$$f(x_3)dx_3 = h\frac{1-x_2}{1-x_1}f(x_2)dx_2 \quad \Rightarrow \quad f(x_3)dx_3 = \frac{h^{3/2}}{\sqrt{2}}\sqrt{1-x_2}f(x_2)dx_2$$
$$\Rightarrow \quad 1 - x_3 \sim \frac{\sqrt{2}}{3}h^{3/2}(1-x_2)^{3/2}$$

Thus the proposition is true for $L = 3$ and with $A_3 = \frac{\sqrt{2}}{3}h^{3/2}$. By iteration, we want to show more generally that:

$$1 - x_L \sim A_L\left(1 - x_{L-1}\right)^{\frac{L}{L-1}} \quad \text{where} \quad A_{L+1} = h\frac{L}{L+1}A_L^{\frac{L-1}{L}}$$

Assume that this is true for L; let us show that for $L+1$, it is the case that:

$$1 - x_{L+1} \sim A_{L+1}\left(1 - x_L\right)^{\frac{L+1}{L}},$$

with the above defined A_{L+1}.

We have:

$$f(x_{L+1})dx_{L+1} = h\frac{1-x_L}{1-x_{L-1}}f(x_L)dx_L.$$

The hypothesis is:

$$1 - x_L \sim A_L\left(1 - x_{L-1}\right)^{\frac{L}{L-1}} \qquad \Rightarrow \qquad 1 - x_{L-1} \sim A_L^{-\frac{L-1}{L}}\left(1 - x_L\right)^{\frac{L-1}{L}}.$$

From the previous differential equation:

$$f(x_{L+1})dx_{L+1} = h\frac{1-x_L}{A_L^{-\frac{L-1}{L}}(1-x_L)^{\frac{L-1}{L}}}f(x_L)dx_L$$

Thus:

$$1 - x_{L+1} \sim hA_L^{\frac{L-1}{L}}\frac{L}{L+1}(1-x_L)^{\frac{L+1}{L}}.$$

which proves the hypothesis for $L + 1$ layers.

This allows us to conclude, as the span of control distribution is given by:

$$\mathbb{P}[N \geq n] = \frac{1}{\Delta}\left(\frac{A_L}{h^L}\frac{1}{n^{\frac{L}{L-1}}} - \frac{A_L}{h^L}\frac{1}{\bar{n}^{\frac{L}{L-1}}}\right)$$
$$\mathbb{P}[N \geq n] = \frac{A_L}{\Delta h^L}\left(\frac{1}{n^{\frac{L}{L-1}}} - \frac{1}{\bar{n}^{\frac{L}{L-1}}}\right),$$

which shows that the distribution of span of control is a truncated Pareto distribution of coefficient $\mathbf{1} + \frac{\mathbf{1}}{\mathbf{L-1}}$. $\square$

A.4 Polynomial Functions: 2 layers

The proof goes as that of Proposition 1. Since y is distributed according to the polynomial function with $1 - F(y) = \frac{(1-y)^{\rho+1}}{\Delta^{\rho+1}}$:

$$\mathbb{P}[N \geq n] = \mathbb{P}\left[y \geq m\left(1 - \frac{1}{nh}\right)\right]$$

$$
\begin{aligned}
&= \frac{1}{\Delta^{\rho+1}}\left[1 - m\left(1 - \frac{1}{nh}\right)\right]^{\rho+1} \\
&= \frac{h(\rho+1)}{(\rho+2)\Delta^{\rho+1}}\left(\left(\frac{1}{nh}\right)^{\rho+2} - (1-z_2)^{\rho+2}\right) \\
\mathbb{P}[N \geq n] &= \frac{\rho+1}{\rho+2}\frac{1}{(h\Delta)^{\rho+1}}\left(\frac{1}{n^{\rho+2}} - h^{\rho+2}(1-z_2)^{\rho+2}\right).
\end{aligned}
$$

From $y = 1$, the highest span of control $\bar{n}$ is such that:

$$
\bar{n} = \frac{1}{h(1-z_2)} \quad \Rightarrow \quad h^{\rho+2}(1-z_2)^{\rho+2} = \frac{1}{\bar{n}^{\rho+2}}.
$$

Thus, in the end:

$$
\mathbb{P}[N \geq n] = \frac{\rho+1}{\rho+2}\frac{1}{(h\Delta)^{\rho+1}}\left(\frac{1}{n^{\rho+2}} - \frac{1}{\bar{n}^{\rho+2}}\right).
$$

The endogenous cutoff z_2 is solution to $m(1-\Delta) = z_2$ so that:

$$
(1-z_2)^{\rho+2} + \frac{1}{h}\frac{\rho+2}{\rho+1}(1-z_2)^{\rho+1} - \Delta^{\rho+2} = 0
$$

When $\Delta \to 0$, $1 - z_2 \to 0$. Therefore $(1-z_2)^{\rho+2} \ll (1-z_2)^{\rho+1}$, thus $1 - z_2 \approx \Delta^{\frac{\rho+2}{\rho+1}}$. Therefore:

$$
\bar{n} = \frac{1}{h(1-z_2)} \approx \frac{1}{h\Delta^{\frac{\rho+2}{\rho+1}}}.
$$

The maximum span of control thus is greater when ρ is minimum, which corresponds to $\rho = 0$ and a density bounded away from zero near 1.

The reasoning with 2 layers can also again be obtained through:

$$
\begin{aligned}
&f(x_2)dx_2 = h(1-x_1)f(x_1)dx_1 \quad \Rightarrow \quad \frac{(1-x_2)^{\rho+1}}{\rho+1} \sim h\frac{(1-x_1)^{\rho+2}}{\rho+2} \\
&\Rightarrow \quad 1 - x_2 \sim \left(h\frac{\rho+2}{\rho+1}\right)^{\frac{1}{\rho+1}}(1-x_1)^{\frac{\rho+2}{\rho+1}} \quad \Rightarrow \quad 1 - x_1 \sim \left(h\frac{\rho+2}{\rho+1}\right)^{-\frac{1}{\rho+1}}(1-x_2)^{\frac{\rho+1}{\rho+2}}.
\end{aligned}
$$

B Other Mathematical Derivations

B.1 Full and truncated Pareto distributions

Pareto distributions. By definition, a (untruncated) Pareto distribution for firm size is such that when the measure of firms with a number of employees higher than a certain number is plotted against that number on a log-log scale, the relationship is linear. In fact, this is how Pareto discovered the regularity that now bears his name.[10] Analytically, if $\bar{F}_N(n) = 1 - F_N(n) = \mathbb{P}(N \geq n)$ is the tail distribution, or complementary cumulative distribution function (c.c.d.f.), then a Pareto distribution with a coefficient (or tail index) α and a minimum size $\underline{n}$ has c.c.d.f.:

$$
\log\left(\bar{F}_N(n)\right) = \alpha\log(\underline{n}) - \alpha\log(n),
$$

[10] In the second volume of his Cours d'Economie Politique (p.305), Pareto plots the number of people earning more than a certain income against that income on a log-log scale, for Great Britain and Ireland in 1893-94. For a historical account, see Persky (1992): "Mitchell et al. (1921) in discussing Pareto's law, suggested that double logarithmic paper was commonly used by engineers. Pareto trained as an engineer and for several years practiced that profession. Perhaps this background influenced his choice of graph paper."

and thus $\log(\mathbb{P}(N \geq n))$ is a linear function of $\log(n)$. This corresponds to the following c.c.d.f. and p.d.f.:

$$1 - F_N(n) = \underline{n}^\alpha \frac{1}{n^\alpha} \qquad f_N(n) = \frac{\alpha \underline{n}^\alpha}{n^{\alpha+1}}.$$

Truncated Pareto distributions. In contrast, a truncated Pareto distribution with tail index α and location parameters $\underline{n}$ and $\bar{n}$ has a density function equal to that of the Pareto distribution for $n \leq \bar{n}$, and identically equal to zero for $n > \bar{n}$, appropriately renormalized:

$$f_N(n) = \begin{cases} \dfrac{\alpha \underline{n}^\alpha}{1 - (\underline{n}/\bar{n})^\alpha} \dfrac{1}{n^{\alpha+1}} & \text{if } n \in [\underline{n}, \bar{n}] \\ 0 & \text{if } n \notin [\underline{n}, \bar{n}]. \end{cases}$$

The survivor function of the truncated Pareto distribution is given as follows:

$$\mathbb{P}(N \geq n) = 1 - F_N(n) = \int_n^{\bar{n}} f_N(N) dN = \frac{\underline{n}^\alpha}{1 - (\underline{n}/\bar{n})^\alpha} \left(\frac{1}{n^\alpha} - \frac{1}{\bar{n}^\alpha} \right)$$

Thus the distribution in Proposition (1a) is a truncated Pareto distribution with a tail index equal to 2.

A Pareto plot is a representation of $\log(1 - F_N(n))$ on the y axis as a function of $\log(n)$ on the x axis. The slope of this Pareto plot is given by:

$$\frac{d \log(1 - F_N(n))}{d \log n} = -\frac{n}{dn} \frac{1}{\dfrac{1}{n^\alpha} - \dfrac{1}{\bar{n}^\alpha}} \frac{\alpha}{n^{\alpha+1}} dn = -\alpha \frac{1}{1 - \dfrac{n^\alpha}{\bar{n}^\alpha}} = -\alpha \frac{1}{1 - \dfrac{1}{\bar{n}^\alpha} e^{\alpha \log n}}.$$

B.2 Self-Employment

Claim 4. (No self-employment) If h or Δ are low enough, then there is no self-employment in equilibrium.

This result is actually quite intuitive. When h is low enough, complementarities are sufficiently strong. Managers' time is very productive, as they can communicate the answers to many problems. Self-employment thus is not an optimum of the planner's problem.

When Δ is low enough, then workers can solve all of the problems almost by themselves, so that workers need a shrinking fraction of managers to solve their problems. This fraction of managers goes to 0 as $O(\Delta^2)$. In contrast, workers' problems get solved with an increased probability, which is a $O(\Delta)$. Thus, for low enough Δ, the gains of working in firms outweigh the costs.

In the case of a uniform density function, the self-employment and no self-employment regions can actually be calculated in closed form, and are represented in Figure 3.

In the case in which 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 and a manager of skill y work together are given by what the two produce together minus what they would have produced by themselves. That is:

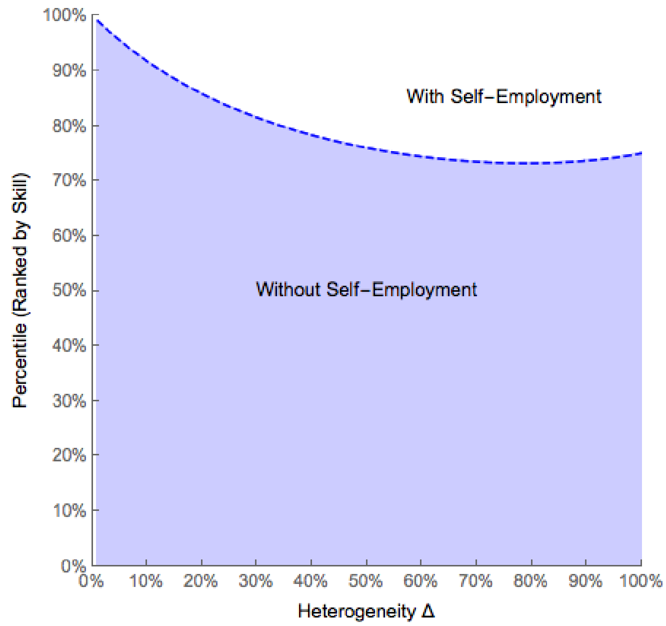
$$\frac{y}{h(1-x)} - y - \frac{x}{h(1-x)} = \frac{1}{h} \left(1 - \frac{1-y}{1-x} \right) - y.$$

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

The notations for cutoffs are introduced in Figure 4. Now the matching function $m(.)$ is defined on $[1 - \Delta, z_2^{**}]$. Another important difference is that in that case, allocations are not solved independently from agents' choices.

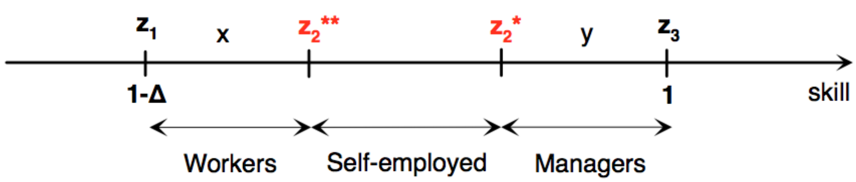
In the decentralized problem, the two differential equations for $m(.)$ and $w(.)$ do not change compared to the case of no self-employment. However, I now have four equations, not three, that determine two initial conditions as well as two cutoffs. They are given by matching the

Figure 3: Self-Employment 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 is no self-employment in equilibrium if heterogeneity Δ is sufficiently low.

Figure 4: With Self-Employment in Equilibrium, One Layer



less and more skilled workers and team managers, as previously:

$$m(1-\Delta) = z_2^* \qquad m(z_2^{**}) = 1.$$

In addition, I now have two indifference equations between being a worker and self-employed with skills z_2^{**}, and being self-employed and a team manager with skills z_2^*:

$$w(z_2^{**}) = z_2^{**} \qquad z_2^* = R(z_2^*).$$

In the case of a uniform distribution of skills, the market clearing equation for skills valid on $(1-\Delta, z_2^{**})$, together with the terminal equation $m(z_2^{**}) = 1$, then gives:

$$m'(x) f\left(m(x)\right) = h\left(1-x\right) f(x) \quad \Rightarrow \quad m'(x) = h(1-x)$$
$$\Rightarrow \quad m(x) = \frac{1}{2}\left(-hx^2 + 2hx + h\left(z_2^{**}\right)^2 - 2hz_2^{**} + 2\right).$$

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

$$m^{-1}(y) = \frac{h - \sqrt{2h + h^2 - 2hy - 2h^2 z_2^{**} + h^2 (z_2^{**})^2}}{h}.$$

In the case in which 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 $w_0(z_2^{**}) = z_2^{**}$ to integrate:

$$
\begin{aligned}
&(1-x)\,w'(x) + xw(x) = xm(x) \\
&\qquad\qquad \Rightarrow \quad w(x) = \frac{1}{2}\left(2x + hx^2 - 2hxz_2^{**} + h\left(z_2^{**}\right)^2\right)
\end{aligned}
$$

Then using the two remaining $m(1-\Delta) = z_2^*$ and $z_2^* = R(z_2^*)$, 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:

$$
\begin{aligned}
z_2^* &= -\frac{-2h + h^2\Delta + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2} \\
z_2^{**} &= \frac{-h + h^2 + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2}.
\end{aligned}
$$

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

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

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

$$
\begin{aligned}
z_2^{**} < z_2^* \quad &\Leftrightarrow \quad \frac{-h + h^2 + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2} \\
&\qquad\qquad < -\frac{-2h + h^2\Delta + \sqrt{h^2\left(3 + h^2\Delta^2 - 2h(1+\Delta)\right)}}{h^2} \\
&\Leftrightarrow \quad 3 - h\Delta - h > 2\sqrt{3 + h^2\Delta^2 - 2h(1+\Delta)} \\
&\Leftrightarrow \quad (1 + 2\Delta - 3\Delta^2)h^2 + 2(1+\Delta)h - 3 > 0 \\
&\Leftrightarrow \quad h > \frac{1 + \Delta - 2\sqrt{1 + 2\Delta - 2\Delta^2}}{-1 - 2\Delta + 3\Delta^2} \quad \text{since } h > 0.
\end{aligned}
$$

When the primitives of the model are such that this is verified, we are in the case in which a nontrivial measure of agents are self-employed. When this is not the case, then all agents either become managers or workers.

B.3 Employee Function

In keeping with the literature on Garicano (2000) type models, I have expressed all endogenous variables in terms of the manager function in the main text. However, both analytically and computationally, it is actually easier to work with the employee function, which matches managers to workers, and is related to the manager function through $m^{-1} = e$. In the case of two-layer firms as in Section 2.1, and a uniform distribution for skills, the employee function such that $e(y) = x$ is given by:

$$
\begin{aligned}
&h(1-x)f(x)dx = f(y)dy \quad \Rightarrow \quad h\left(1 - e(y)\right)e'(y) = 1 \\
&\Rightarrow \quad h\left[-\frac{\left(1-e(y)\right)^2}{2}\right]_y^1 = 1 - y \quad \Rightarrow \quad h\frac{\left(1-e(y)\right)^2}{2} - \frac{\left(1-z_2\right)^2}{2} = 1 - x
\end{aligned}
$$

The span of control $N(x_2)$ of managers with skills x_2 is given by:

$$
n(y) = \frac{1}{h\left(1-x\right)} = \frac{1}{h\left(1-e(y)\right)} = \frac{1}{h\sqrt{(1-z_2)^2 + \frac{2}{h}(1-y)}}.
$$

C Figures

Figure 5: **Span of Control Distribution**, Pareto Plot, f = Uniform, $h = 70\%$

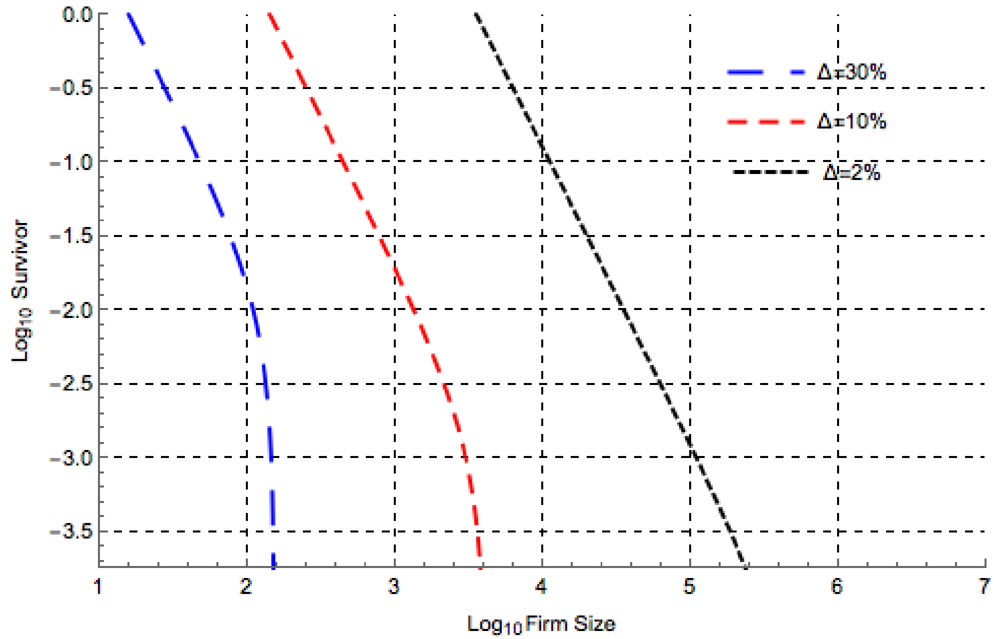


Figure 6: **Matching Function**, f = Uniform, $h = 70\%$

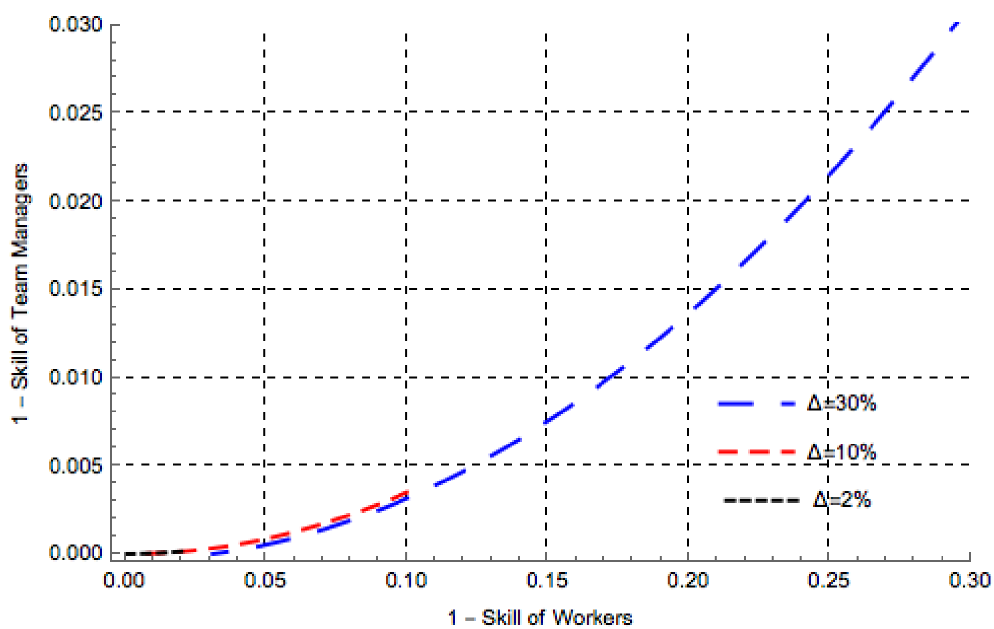


Figure 7: **Span of Control Distribution**, Pareto Plot, f = Increasing, $h = 70\%$

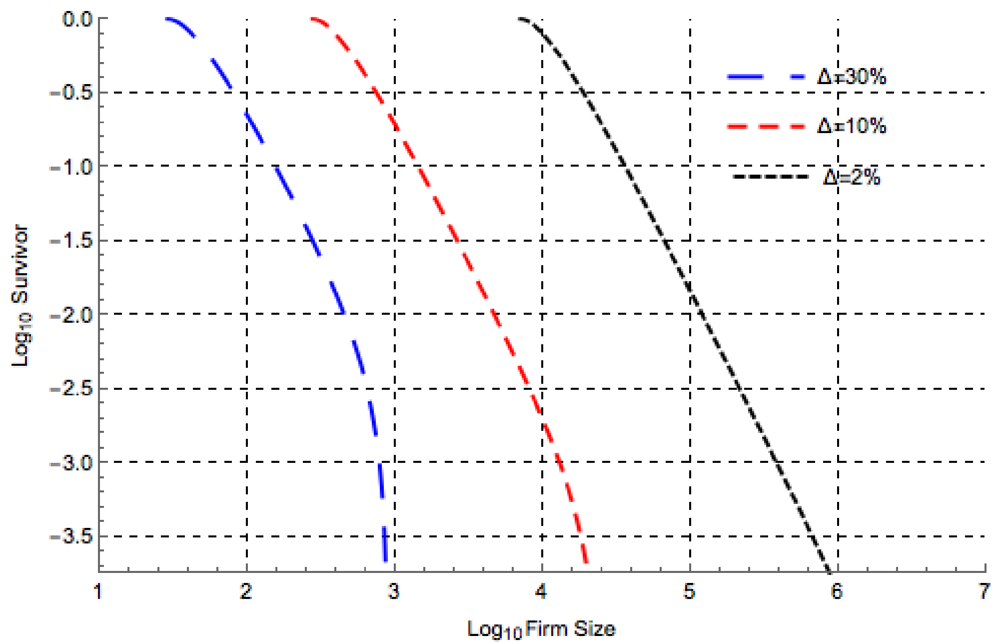


Figure 8: **Matching Function**, f = Increasing, $h = 70\%$

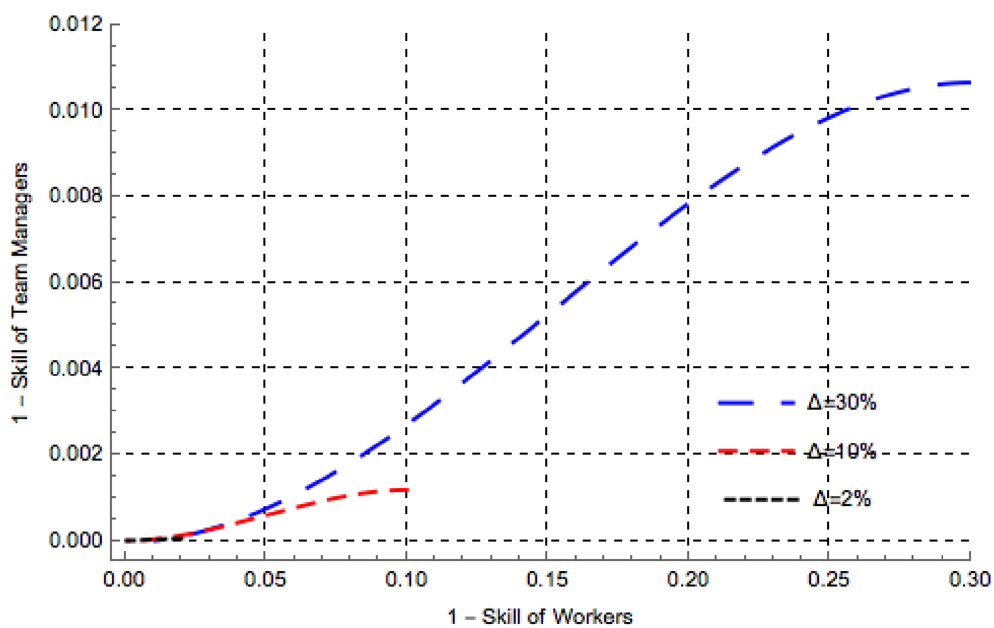


Figure 9: **Span of Control Distribution**, Pareto Plot, f = Beta(1, 1), $h = 70\%$

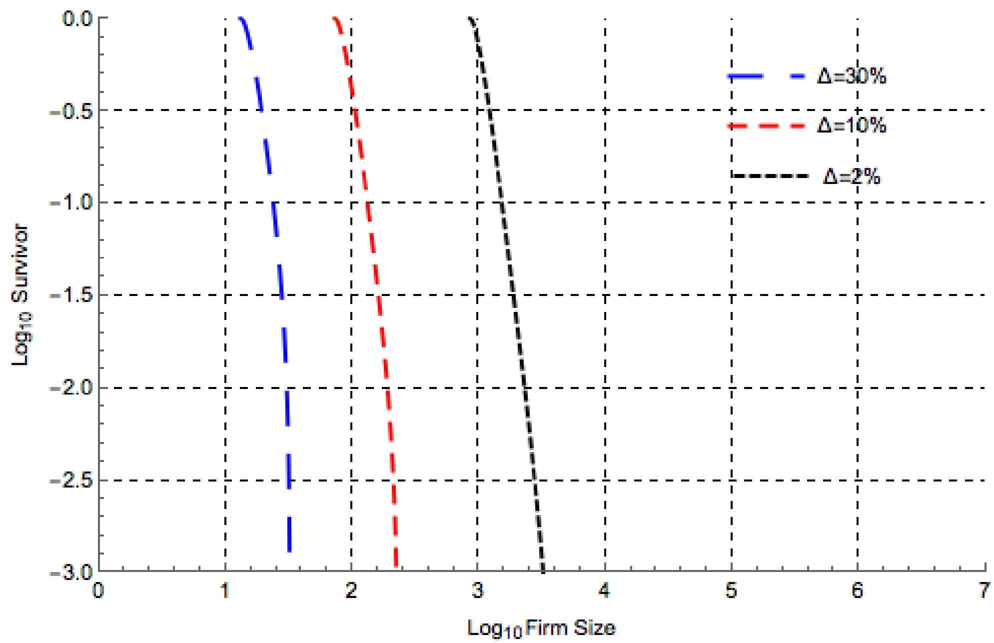


Figure 10: **Span of Control Distribution**, Pareto Plot, $h = 70\%$

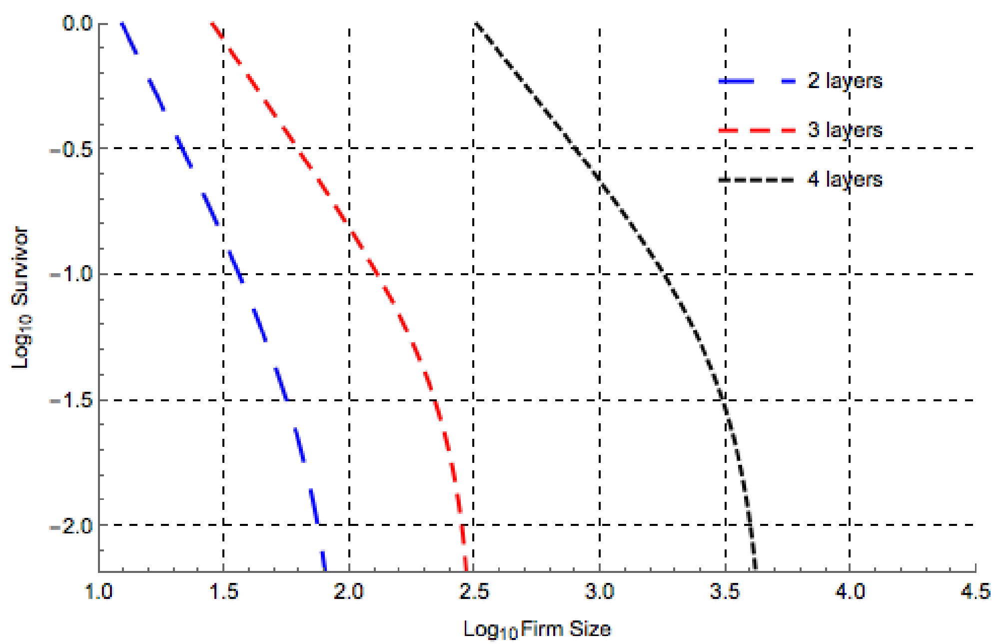


Figure 11: **Span of Control Distribution**, Pareto Plot, $h = 70\%$

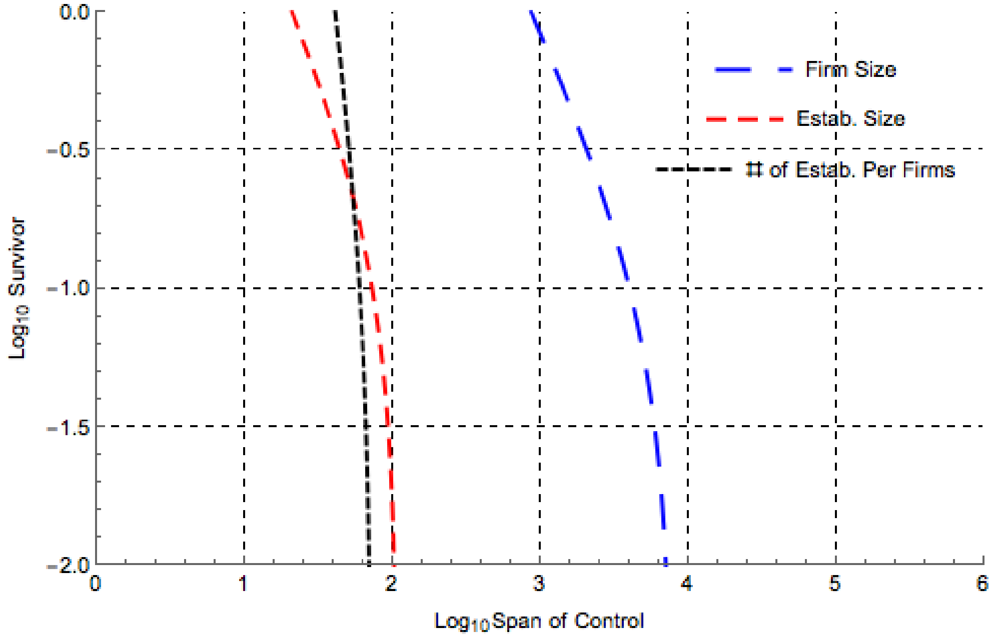


Table 1: Uniform Skill Distributions: Cutoffs for Different values of L, with $\Delta = 100\%$ and $h = 70\%$

	$\mathbf{z_2}$	$\mathbf{z_3}$	$\mathbf{z_4}$	$\mathbf{z_5}$	$\mathbf{z_6}$	$\mathbf{z_7}$
$\mathbf{L = 2}$	0.684778					
$\mathbf{L = 3}$	0.637648	0.941694				
$\mathbf{L = 4}$	0.631346	0.933779	0.993365			
$\mathbf{L = 5}$	0.630850	0.933155	0.993344	0.999630		
$\mathbf{L = 6}$	0.630828	0.933128	0.993344	0.999614	0.999987	
$\mathbf{L = 7}$	0.630828	0.933127	0.993344	0.999614	0.999986	1.000000

Figure 12: **Labor Income Distribution**, Pareto Plot, Two Part Distribution, $h = 3\%$

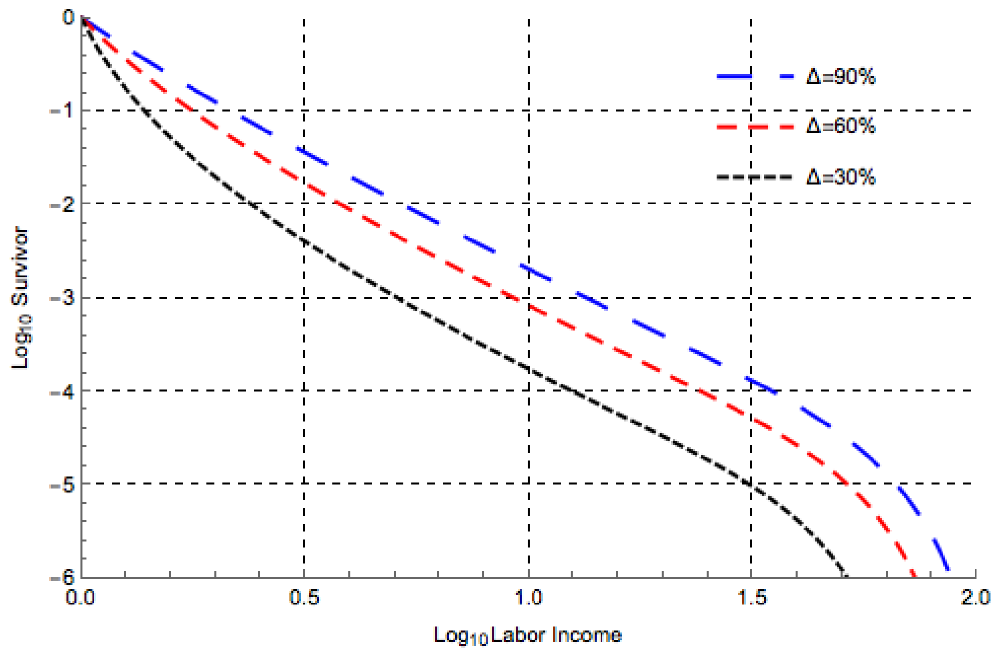


Figure 13: **Labor Income Distribution**, Pareto Plot, Two Part Distribution, $\Delta = 90\%$

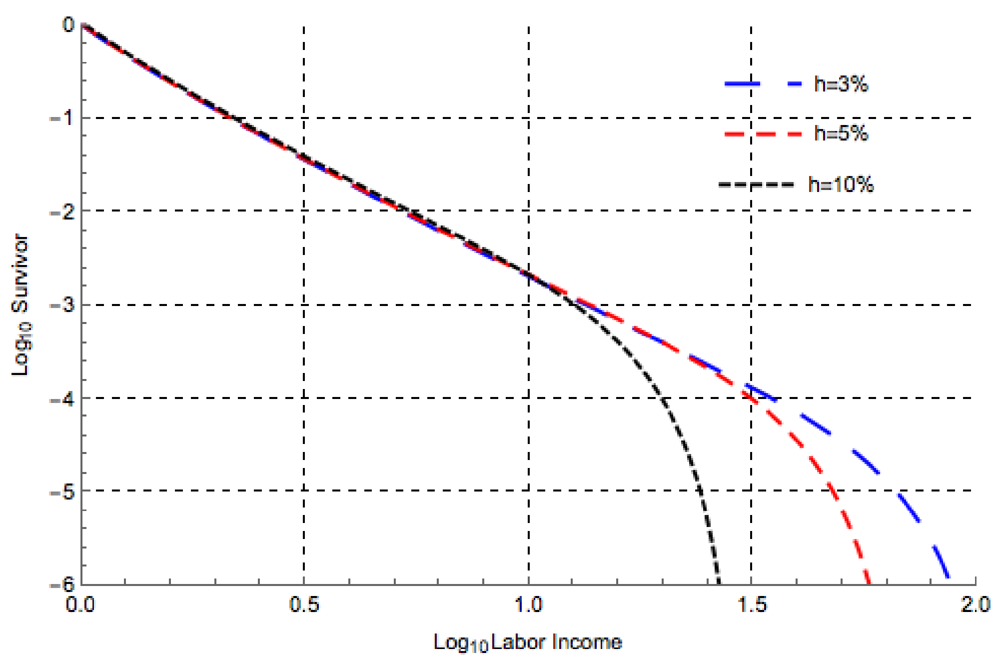


Figure 14: **Roberts' Law**, Pareto Plot, Two Part Distribution, $h = 3\%$

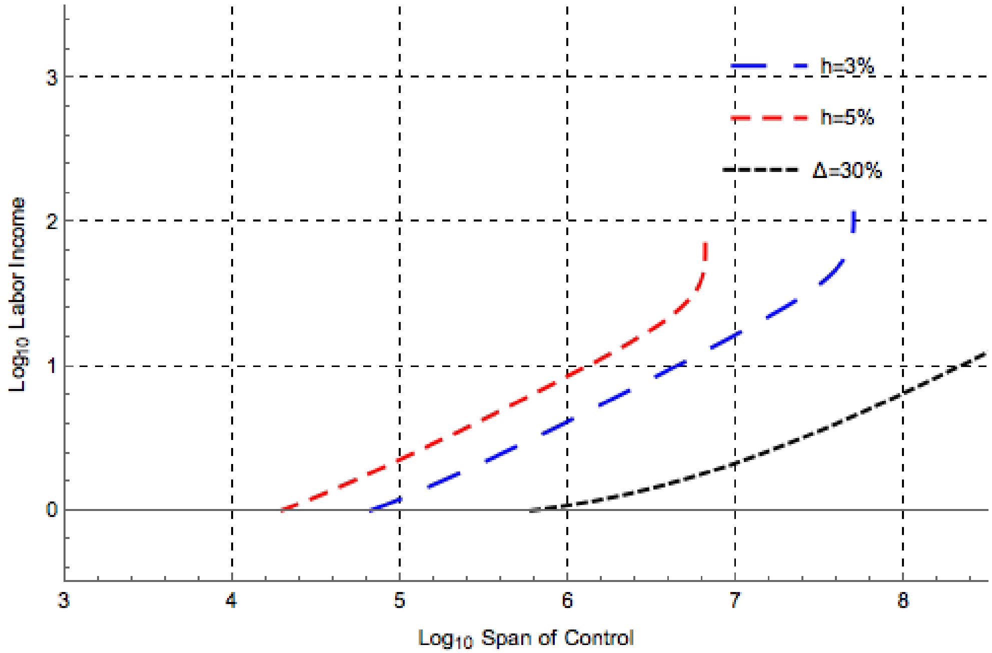
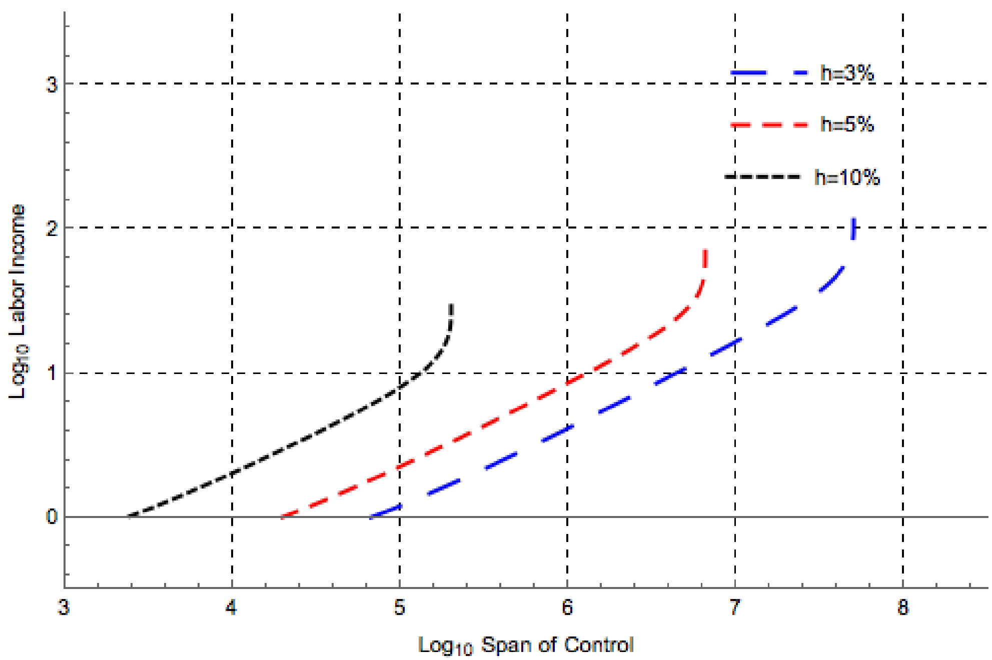


Figure 15: **Roberts' Law**, Pareto Plot, Two Part Distribution, $\Delta = 90\%$



D Evidence

Figure 16: Roberts' Law in the Data

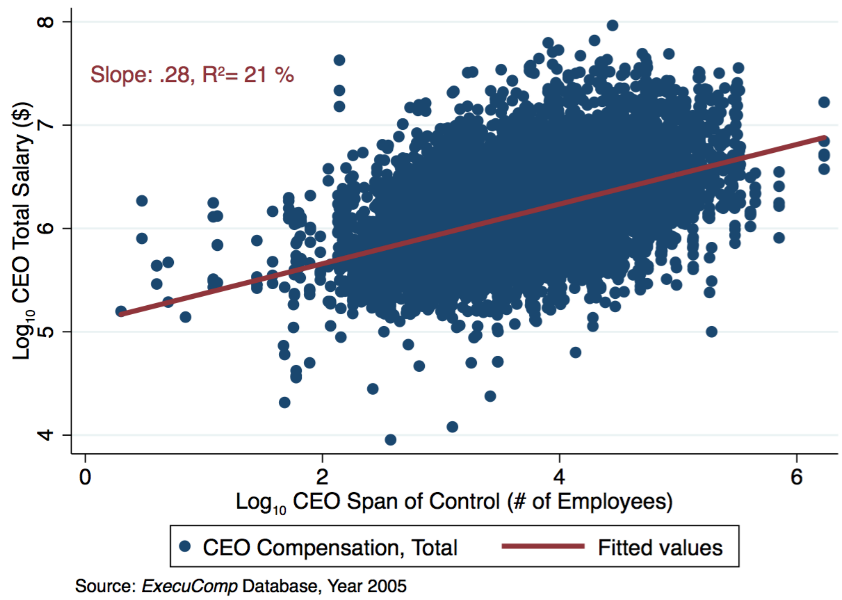
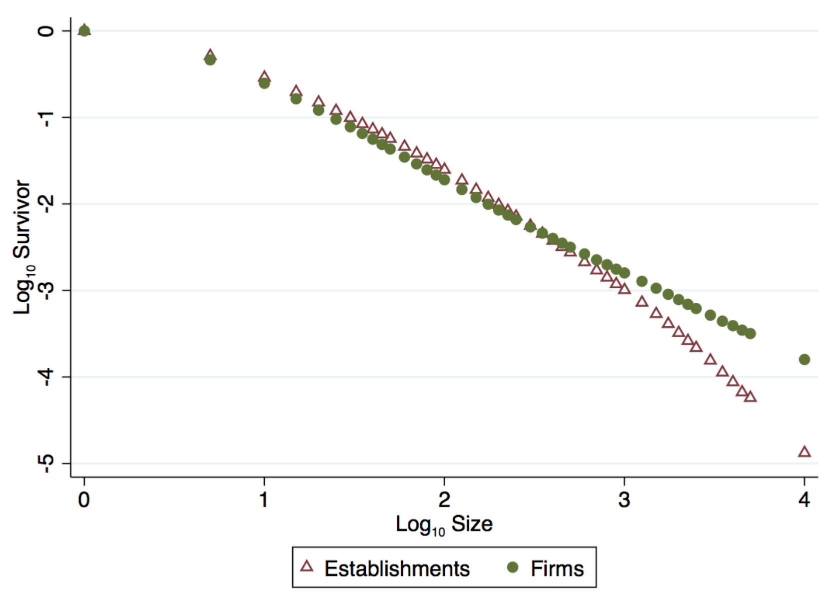
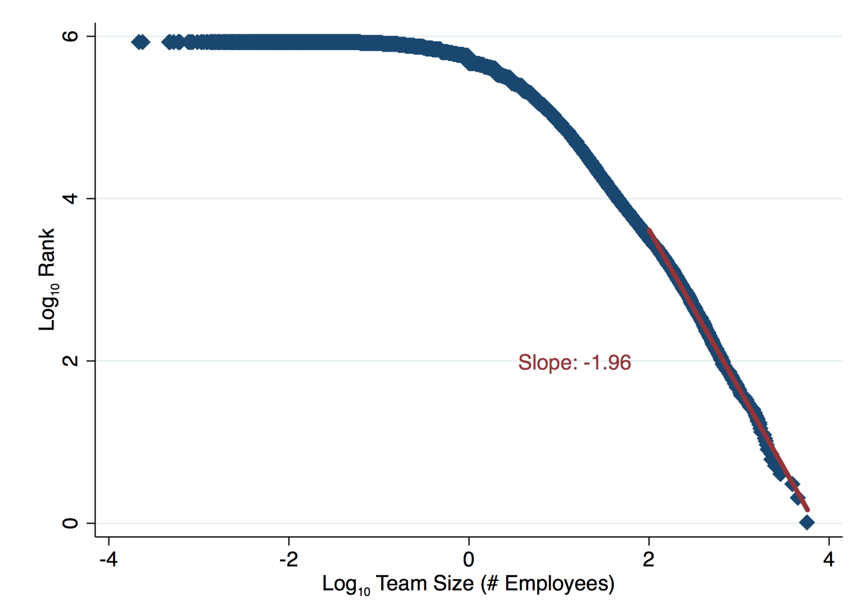


Figure 17: US Firm and Establishment Sizes



Note: Source: Census Bureau, Statistics of US Businesses, 1990.

Figure 18: Span of Control down one level of Hierarchical Organization



Note: Source: French matched employer-employee data, year 2007.

Figure 19: All sectors (Administrative Data)

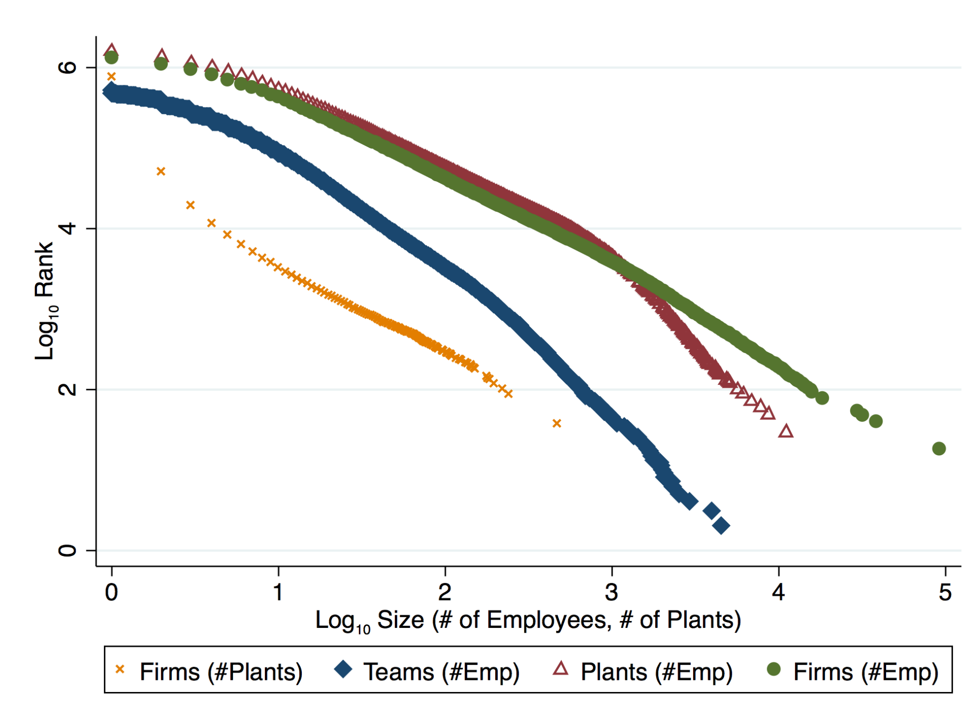
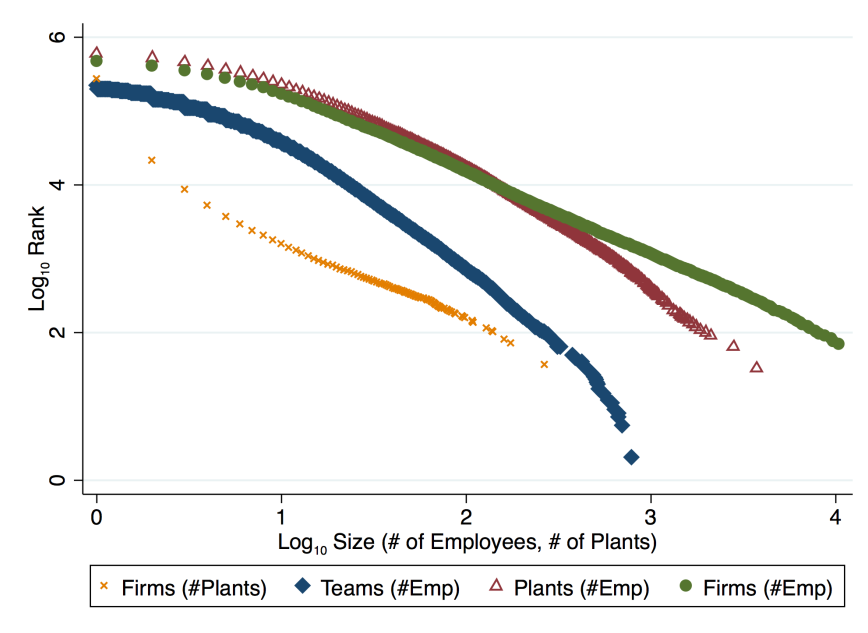
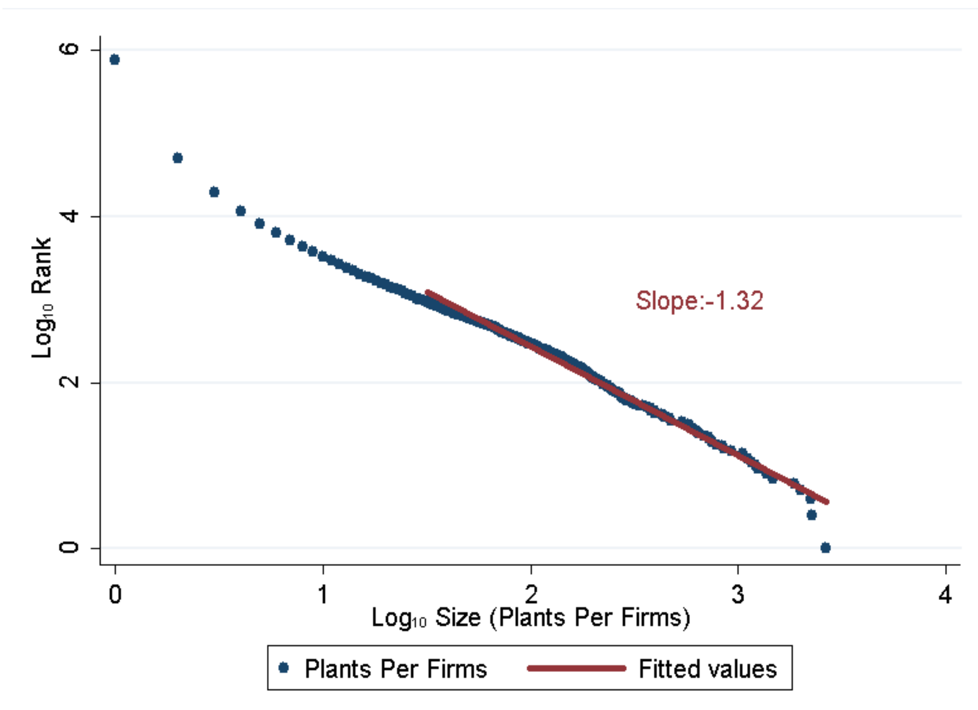


Figure 20: Wholesale Trade, Accommodation, Food (Administrative Data)



Note: Source: French matched employer-employee data, year 2007.

Figure 21: Plants Per Firm Distribution (Administrative Data)



Note: Source: French matched employer-employee data, year 2007.