A Static and Microfounded Theory of Zipf’s Law for Firms
and of the Top Labor Income Distribution

François Geerolf
University of California, Los Angeles

VERY PRELIMINARY AND INCOMPLETE
This version: September 2014 (First version: April 2014) [Last Version]

Abstract

I propose a theory of Zipf’s Law for firm sizes and Pareto distributions for top labor incomes based on arbitrarily small differences in skills between workers, and no functional form assumption. A version of Garicano (2000)’s knowledge-based production hierarchies model is shown to generate Pareto distributions for high span of control of intermediary managers with coefficients equal to \(2^L/(2^L - 1)\) between \(L\) levels of management, under very limited assumptions on the underlying distribution of agents’ skills. This breakdown of the aggregate firm size distribution receives considerable empirical support in the French matched employer-employee administrative data. The model provides a method to structurally estimate the relative importance of different factors in the recent rise in labor income inequality: the tail index of the Pareto top labor income distribution depends on a notion of race between education and technology, while the scale parameter depends on communication costs and heterogeneity. It also has striking implications for the literature in trade and macroeconomics on firm heterogeneity: in the model, heterogeneity in firm sizes grows when underlying productivity heterogeneity decreases.

Keywords: Pareto distributions, Inequality, Hierarchies
JEL classification: D2, L2

*I thank David Atkin, Simon Board, Ariel Burstein, Harold Cole, Jacques Crémer, Jan Eeckhout, Xavier Gabaix, Christian Hellwig, Adriana Lleras-Muney, Martí Mestieri, François de Soyres, Robert Ulbricht, Pierre-Olivier Weill and numerous seminar participants at UCLA, UPenn and TSE for comments. I thank Toulouse School of Economics, where this project was started, for their hospitality; and I am particularly indebted to Maxime Liegey for presenting a graph on French team and plant sizes during a student workshop there. 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 very preliminary, all comments are welcomed.
Introduction

Economists usually agree that firm size’s distribution approximately follows Zipf’s law in the upper tail. To some, it is even one of the few quantitative laws in economics, holding across time and countries. Models explaining this Pareto distribution all at some level rely on a stochastic random growth process with statistical frictions, following Gibrat (1931)’s law, or on some other Pareto distribution underlying the economy’s primitives, whether it be entrepreneurial skill or firm’s productivity. In this paper, I propose a static mechanism, able to microfound the existence of Zipf’s law, based on a production hierarchies model a la Garicano (2000), and which makes no functional form on the underlying distribution of primitives, in this model agents’ skills. The model makes ancillary predictions about the distribution of span of controls between intermediary levels of management, which are borne out by the French data, as Figure 1 below shows. In theory, the span of control of managers over employees down one level of hierarchical organization should follow a Pareto distribution in the upper tail with a coefficient equal to two, $\frac{4}{3} \approx 1.33$ down two levels of management, $\frac{8}{7} \approx 1.14$ down three levels, getting closer to a Pareto with coefficient one, or Zipf’s law, as the number of levels of management increases. In the French data, the point estimates for the corresponding numbers are 1.96, 1.30 and 1.14 respectively. The model also is able to account for the fact that Pareto actually holds only for the upper tail, that very big firms are smaller than Zipf’s law would predict, and that establishment size is also Pareto distributed with a coefficient equal to about 1.33, not just in France, but also in the United States, as Figure 2 shows on publicly-available data.

![Figure 1: French Production Hierarchies](image)

<table>
<thead>
<tr>
<th>Pareto coeff., DADS Data:</th>
<th>1 level: 1.96</th>
<th>2 levels: 1.30</th>
<th>3 levels: 1.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pareto coeff., Theory:</td>
<td>1 level: 2</td>
<td>2 levels: $\frac{4}{3}$</td>
<td>3 levels: $\frac{8}{7}$</td>
</tr>
</tbody>
</table>

**Note:** Source: French matched employer-employee Data (DADS) reproduced in Liegey (2014). The construction of hierarchical levels follows Caliendo et al. (2014)’s methodology.
Figure 2: US Firm and Establishment Sizes

Pareto coeff., Census Data: Establishments: 1.33 Firms: 1.01
Pareto coeff., Theory: 2 levels: 4/3


Figure 3: Distribution of Labor Incomes in France

Note: Source: French matched employer-employee data, year 2007. (Déclarations Annuelles de Données Sociales) The bins with less than 5 different employees are not shown for confidentiality reasons (in the upper tail). The distribution of Pareto incomes is well approximated by a Pareto distribution with coefficient 2.59 for annual salaries in the range [23,000 €; 355,000 €], and coefficient 2.00 for annual salaries greater than 355,000 €. (see Figure 20 in the Appendix for a plot of the fitted slopes)
That it is possible to microfound Zipf’s law in this way is not just interesting for its own sake but because many issues crucially rely on understanding why firm sizes are so heterogeneous in the economy. The most important of them is perhaps that high levels of income inequality are sometimes justified on the basis that top managers have high span of control, so that slight improvements in upper level decisions have an enormous influence as a whole by affecting the productivity of lesser ranking workers (Rosen (1982)). As a matter of fact, the upper tail of the wage distribution is also well approximated by a Pareto distribution as shown on French data, in Figure 3. In these theories, the reason for why managers earn very high and very heterogeneous salaries, from ten to more than a thousand times as large as that of the average worker, in the end remains all about the underlying heterogeneity in the firms they run. In other words, one can explain why labor incomes might be heterogeneous, but only given a certain span of control distribution arises in the first place. In the random growth tradition, big firms are those which have been exceptionally lucky, and which have experienced a long and continued sequence of good idiosynchratic shocks. The very rich workers are then those who are lucky enough to be running these firms. Another tradition following Champernowne (1953) has consisted in assuming random shocks to human capital at the individual level, and to view the workers with high earnings as the exceptionally lucky ones. However, quoting Mincer (1958): "From the economist’s point of view, perhaps the most unsatisfactory feature of the stochastic models... is that they shed no light on the economics of the distribution process. It is difficult to see how the factor of individual choice can be disregarded in analyzing personal income distribution." (p. 283) Finally, explanations relying on some primitive being distributed Pareto are a bit tautological, whether they rely on the productivity of firms or the talent of managers being distributed Pareto. IQ or other measures of ability are certainly not distributed Pareto, but rather according to distributions with finite support.

The theory I develop in this paper suggests that the very high heterogeneity in firm sizes and labor incomes does not need to come from the heterogeneity in underlying primitives, or from any notion of "luck". Rather, the competitive equilibrium can generate on its own very large levels of inequalities in outcomes, from only an epsilon amount of ex-ante heterogeneity. The mechanism actually develops on an idea which was put forward a long time ago by Tinbergen (1956) and then developed by Sattinger (1975): that a matching process with complementarities can lead to wage schedules which are convex functions of skills. In this paper, I take this insight to its furthest extent, in order to explain how such a mechanism can amplify ex-ante, even infinitesimal differences, and lead to Pareto distributions in allocations as well as in wage schedules, in line with the evidence presented on the previous three figures.

The analysis also sheds a new light on the growing body of work which emphasizes heterogeneity in firm productivities as a potential source of misallocation across firms, or of reallocation of resources from least to most productive firms consecutive to openness to trade, as in Melitz (2003) type models. The model suggests on the contrary that underlying heterogeneity can actually be bounded, and even infinitesimal, and nonetheless generate Zipf’s law for firms in the upper tail; which could suggest a much less important role for the effects just mentioned. Worse,
the comparative statics one can draw from my analysis actually go in a rather counterintuitive
direction: the less heterogeneity in firm primitives, the more heterogeneity in firm sizes. The
reason is that when managers’ and workers’ skills become closer, the managers’ span of control
increases because workers need them less. This can suggest a new reason for why developing
economies have fewer big firms than developed ones, or why the returns to education are lower
there: the underlying skill heterogeneity is higher in developing countries.

The static model which forms the basis of this paper and generates a Zipf’s law for firm
sizes is a well-known Garicano (2000) production hierarchies model, where time can be used for
production as well as for communicating one’s knowledge. In such a model, production workers
encounter problems when producing. When they do not know the solution to this problem,
you can potentially ask someone more knowledgeable than them for a solution. If the time of
communicating knowledge, called "helping time", is lower than the time of producing by oneself,
higher production can sometimes be achieved by having more knowledgeable agents specialize in
communicating solutions to problems. Because workers ask these agents only when they do not
know the solution themselves, these agents are called managers. The theoretical contribution
of this paper is to show that in general, the distribution of span of control for managers over
agents down $L$ levels is a Pareto distribution of coefficient $2^L/(2^L - 1)$ in the upper tail, under
minimal assumptions for the distribution of problems as well as the distribution of workers’ skills.
Hence, Zipf’s law obtains in the upper tail with a sufficiently large number of hierarchical levels.

While Garicano (2000) and subsequent papers like Garicano and Rossi-Hansberg (2006) have
looked mostly at the impact of organization and decreases in costs of communication on wage
polarization, in models either with learning or with an exogenously low number of hierarchical
organization, I show that taking the distribution of skills as exogenous as well as allowing for
multiple layers of management naturally gives rise to the Pareto properties described above.

Compared to their work, my main methodological contribution is to show that some very precise
things can be said about the upper tail, both of the span of control distribution, and of the wage
distribution, irrespective of the underlying distribution of skills. Moreover, I will show that the
kind of skill distribution Garicano and Rossi-Hansberg (2006) could not lead to very high labor
income inequality, as they were bounded away from zero for very high skills.

The intuition for why a Pareto distribution for span of control obtains in the upper tail is that
the most knowledgeable managers work with the most knowledgeable workers in the competitive
equilibrium of this model. This is because of a complementarity between workers’ and managers’
skills: at an optimal allocation, the most skilled managers must be shielded from answering easy
questions, as they can make a more productive use of their time. To the limit, when heterogeneity
goes to zero, or when the number of layers of hierarchical organization increases, the most
productive workers are able to solve almost all the problems by themselves, so that they almost
never ask managers. Those very productive managers can therefore handle many workers, who
could almost produce by themselves, and only ask for their help only when the question is very
difficult. However, not all managers can work with those very productive, almost self-sufficient
workers, for there is only a limited supply of very skilled workers. Less knowledgeable managers therefore have to "tap" progressively into the knowledge of less knowledgeable workers. The way in which they have to "tap" into this knowledge is given by a Pareto distribution of tail coefficient equal to two, for the same mathematical reason as in Geerolf (2013). Just as in Geerolf (2013), the distribution of the largest firms’ sizes corresponding to the most productive workers and the most productive managers does not depend on the functional forms underlying the primitives - distribution of arriving problems, and of agents’ skills - because relatively few managers and few workers underly this distribution, so that to the limit, the distribution can be approximated by a uniform distribution. When there are not just one, but many levels of hierarchical organization, the formula for the Pareto’s endogenous tail coefficient is \( \frac{2^L}{(2^L - 1)} \) in general. The intuition for that is that managers at the top of the firm manage other intermediary managers, who themselves manage other agents. The tail for span of control is thus thicker with multiple levels of intermediary management than with just one of them. For example, in the case of two layers of management, I show that the number of workers that the intermediary managers supervise is given by a Pareto distribution of coefficient four in the space of top managers: there are fewer managers than intermediary managers. Since their own span of control over intermediary management is given by a Pareto with coefficient two, total span of control therefore has a tail Pareto exponent given by \( \frac{1}{1/2 + 1/4} = \frac{4}{3} \), since it is just a multiplication of the two above distributions. This reasoning generalizes straightforwardly to a case with \( L \) layers, with an exponent of \( \frac{2^L}{(2^L - 1)} \) in general.

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

An important take from the model is that heterogeneity in the primitives, here agents’ skills, can very well have finite support, even a support with infinitesimally small measure, and yet heterogeneity in outcomes appear unbounded. To the best of my knowledge, this contrasts with all existing literature which attributes outcomes with infinite support to very heterogeneous causes, ones with infinite support as well. Among many examples, managers’ skills in Lucas (1978) have unbounded support - they are even distributed Pareto; or productivities are Pareto distributed in some models of heterogeneous firms with decreasing returns to scale and fixed costs. In contrast, the model will allow to map an empirically very unequal labor income distribution

---

1There is a clear analogy between lenders and borrowers’ leverage ratios in Geerolf (2013) and workers and managers’ span of control in this paper.
to an underlying potentially very equal distribution in skills. Even on a purely methodological ground, whether the "true model" of the economy has an unbounded dispersion of primitives (productivity, demand, skills), or a bounded one, as this model would suggest, is likely to matter not just quantitatively, but also qualitatively. I will come back to this point in the literature review below when referring to the relevant literature on models with heterogenous firms.

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

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

The paper most chiefly belongs to the span of control literature, developed since Lucas (1978), who assumes Pareto distribution in skills of managers together with homogeneity of workers' skills to explain Pareto distributions in span of control; and to the literature on the "economics of superstars", initiated by Rosen (1981). It also speaks to the literature on organizational structure, a survey of which is given in Radner (1992) for the older literature. The paper builds on Garicano (2000)'s production hierarchies model to investigate the distribution of firm sizes in the economy. More precisely, it draws heavily from the applied production hierarchies models developed in Garicano and Rossi-Hansberg (2004), Garicano and Rossi-Hansberg (2006) and Antràs et al. (2006), and Caliendo and Rossi-Hansberg (2012). It uses the same intuition, and a similar methodology, as Geerolf (2013) to generate a Pareto distribution of tail coefficient equal

---

Gibrat (1931) wanted to explain why the distribution of firm sizes was approximately log-normal, so he did not need statistical frictions. We now know that Pareto fits the distribution much better in the upper tail, although log-normal gives a better fit for the rest of the distribution.
to two. However, the mechanism to generate more skewed Pareto distributions (with lower tail coefficients) is different, and here relies on a static mechanism, and not from a dynamic one as in Geerolf (2013). The use of French data for production hierarchies, and in particular the occupation code as an indication of workers’ position in a firm’s hierarchical structure follows Caliendo et al. (2014) and Liegey (2014).

The paper is an alternative to random growth theory to microfound with a very limited number of assumptions on primitives why firms’ sizes tend to be distributed Pareto. It is therefore potentially relevant to far more distant pieces of the economics literature, since research on firm heterogeneity takes place at the intersection of macroeconomics, labor economics, trade and industrial organization. For example a very large literature in trade with heterogenous firms following Melitz (2003) uses evidence of Pareto firm sizes to justify Pareto distributions in productivity. There is also a large literature in macroeconomics on misallocation, speaking to the welfare consequences of microeconomic and macroeconomic distortions, using abundantly the firm size distribution to back out productivity differences, among which Hopenhayn and Rogerson (1993), Restuccia and Rogerson (2008), Hsieh and Klenow (2009). In the light of this paper, it seems on the contrary that the shape of firm size’s distribution says very little on the underlying distribution of productivity across them, even on the relative importance of their dispersion.

Finally, the model is a microfoundation for the Pareto nature of the top labor income distribution in the upper tail. The model is part of the competitive assignment literature, comprising Roy (1950), Tinbergen (1956), Rosen (1974), Sattinger (1975), Rosen (1981), Teulings (1995), Terviö (2008) and Gabaix and Landier (2008). In particular, a crucial assumption underlying Garicano (2000) as well as in similar models of sorting is that quality and quality of workers are imperfect substitutes, and that there is no such thing as a concept of "efficiency units of labor" (Eckhout and Kircher (2012)). The model can potentially speak to the causes of the recent rise in labor income inequality, documented in particular by Piketty and Saez (2003) for the United States, and then for other countries (see Piketty (2014)). It can potentially try to address the different explanations which have been put forward for this rise: skill-biased technological change (Acemoglu (2011)), race between education and technology (Katz and Murphy (1992)), decrease in communication costs (Garicano and Rossi-Hansberg (2006)).

1 Evidence on the Firm Size Distribution and Firm Dynamics

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

Stylized facts 1-4 concern the overall firm size distribution and are illustrated on Figure 24 taken from Axtell (2001). Stylized fact 1 states the Zipf’s law character of the firm size distribution, that random growth theory seeks to explain.

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

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

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

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

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

1.2 Firm Dynamics

Ever since the random growth literature, Zipf (1949)’s law has been seen as indirect evidence for scale independence in the growth of firms. However, it has been known for a long time, at least since Mansfield (1962), that Gibrat’s law for firm growth does not hold so well in the data. For example, Figures 25a and 25b are taken from Rossi-Hansberg and Wright (2007).

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

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

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

The perhaps most important stylized facts that this paper is after concern the disaggregated size of production units. Stylized fact 7 concerns the distribution of plant sizes in the US data, highlighted in particular in Rossi-Hansberg and Wright (2007) (pp 1948), and illustrated on Figure 2. To be precise, Rossi-Hansberg and Wright (2007) do not make explicit the exact Pareto coefficient on the distribution of establishments. Only do they note that the size distribution of establishments in general has a thinner tail than the Pareto distribution of firms.3

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

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

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

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

2 Model

In this section and the following, I develop and study the upper tail properties of a Garicano (2000) model of production hierarchies. In order to get quickly at the main results of the model, I will assume a fixed, exogenous distribution of skills - although endogenous skill acquisition is potentially a fruitful extension. Unlike Antràs et al. (2006) however, I will not here "force" agents to join teams. They will work in cooperation with others only when they have an interest in doing so, that is when they gain more in team than by being self-employed: I will show this is the case when heterogeneity in skills, or the cost of communicating solutions to problems, is sufficiently low. Since occupational choice is a crucial element for the formation of Pareto distributions for span of control, I don’t want to restrict the choice set exogenously in this dimension. Again, unlike in Antràs et al. (2006), I do not here restrict the number of layers of hierarchical organization to be one in the maximum, but the number of layers $L$ will be determined endogenously. This higher number of layers is crucial to getting the Pareto results.

3Rossi-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."
I consider a static economy, with a continuum of agents indexed by $i$. All agents value consumption in the same way according to a linear utility function. Agents are endowed with one unit of time, which they supply for production or for helping others solve problems. When producing, they encounter problems, whose set is indexed on $[0, 1]$. The arrival of these problems is uniformly distributed on this interval, and the problems’ indexes are higher when these problems are harder to solve - that is, less agents know how to solve them. This is without loss of generality, as from any initial draw of problems and their probability distributions, one can just relabel them such that this is verified. As in this literature, I shall denote the cumulative distribution function for the arrival of problems by $F(x) = x$ on $[0, 1]$.

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

An agent $i$ has a skill which is indexed by the hardest problem he can solve: an agent $i$ with skill $z_i$ can solve all problems contained in set $[0, z_i]$. The distribution of agents’ skills over problems is given by a cumulative distribution function $G(.)$, with density $g(.)$ over $[0, 1]$. It is assumed that the complexity of producing the good in the economy is such that only the most skilled of them would know how to produce the good by themselves. In other words that the density function has support $[1 - \Delta, 1]$, where $\Delta$ indexes, without summarizing in general, the level of skill heterogeneity in the economy.\(^4\)

**Production.** The organization of production closely follows Garicano (2000): apart from producing with their time as a factor of production, agents can also transmit solutions to other agents. When production workers know the solution to these problems, they can produce one unit of output per unit of time. But if they do not, they can ask their managers. Regardless of whether the manager knows the solution or not, it takes $h < 1$ units of time for a manager to communicate a solution to the worker. $h < 1$ is assumed for agents to find it (sometimes) optimal to form teams: if the time of communicating knowledge is greater than the time of producing, then it is always better to engage in self-production. $h$ is called the helping time. Potentially, managers can also ask other agents to solve the problems if they do not know the solutions to them, so that hierarchies can form.

A crucial assumption is that workers cannot diagnose or "label" problems when they cannot solve them and therefore do not know who may know the solution to problems they cannot solve,\(^4\)One could generalize this to having a non-trivial measure of agents able to solve all problems, but this would only add sub-cases, without adding much intuition to the model. On the other hand, the results rely on some agents being able to solve all problems arising in production, at least to the limit. It is doubtful that agents would in equilibrium agree to produce a good which would pose unsolvable problems in production even to the most skilled of them. Another option would be to endogenize the choice of the good being produced.
otherwise the organization would never have a pyramidal structure, and agents would directly go to the agents who know. I call manager of type \( l \) a manager who answers questions which have been transmitted \( l \) times: managers of type 1 directly supervise production workers, managers of type 2 answer questions of managers of type 2, and so on.

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

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

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

\[
\begin{align*}
\max_{\{t^i_S, t^i_W(.), t^i_M_l(.)\}, t^i_M_L} & \quad z^i t^i_S + \int_z w_0(z) t^i_W(z) dz + \sum_{l=1}^{L-1} \int_z w_l(z) t^i_M_l(z) dz + z^i t^i_M_L & \quad (I) \\
\text{s.t.} & \quad (TC) \quad t^i_S + \int_z \max\{t^i_W(z), 0\} dz + \sum_{l=1}^{L-1} \int_z \max\{t^i_M_l(z), 0\} dz + t^i_M_L \leq 1 \\
\text{s.t.} & \quad (CTC_0) \quad \int_{z \neq z^i} h(1 - z) t^i_W(z) dz \leq t^i_M_l \\
\text{s.t.} & \quad (CTC_l) \quad \forall l \in \{1, \ldots, L - 1\}, \quad \int_{z \neq z^i} h(1 - z) t^i_M_l(z) dz \leq t^i_M_{l+1} \\
\text{s.t.} & \quad \forall z \neq z^i, \quad t^i_W(z) \leq 0 \\
\text{s.t.} & \quad \forall l \in \{1, \ldots, L - 1\}, \forall z \neq z^i, \quad t^i_M_l(z) \leq 0
\end{align*}
\]

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

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

\[
\forall z, \quad \int_i t^i_W(z) \, di = 0. \quad (MC_0)
\]
\[
\forall l \in \{1, \ldots, L-1\}, \quad \forall z, \quad \int_i t^i_{Ml}(z) \, di = 0. \quad (MC_l)
\]

Because the program is linear in \( \left( t^i_W(.), \{t^i_{Ml}(.)\}_{l=1}^{L-1}, t^i_S \right) \), it will be optimal in equilibrium for agents to do one of three things: use their whole time producing in a team, or managing and communicating solutions at a given level of management, or engage in self-employment. In other words, the model I am working with is very similar to that in the seminal Garicano (2000), or Garicano and Rossi-Hansberg (2006). I will now solve for the equilibrium of this model and state the main result of the paper about the distribution for spans of control of managers. Following Caliendo and Rossi-Hansberg (2012), I will assume that a top manager can only work in one firm, and not manage multiple firms. 5

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

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

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

---

5 I could as well assume that firms can be run by two or three agents, or that agents can run at most an integer number of firms, as long as it is a fixed number. In practice, conflict of interest or management issue would certainly arise in that case, so that the constraint can be thought of as a reduced form of an upper level model with an explicit decision-making process.
As can be inferred from Definition 1, a Competitive Equilibrium of the model defined above is potentially a high dimensional object. In this section, I will study independently allocations in this model with no mention of the equilibrium prices sustaining these allocations. As in Garicano (2000), all of these allocations are Pareto-optimal and can be decentralized using labor markets inside the firm something I turn to in Section 4.

### 3.1 Ruling out Self-Employment

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

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

**Proof.** See Appendix C.1.

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

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

To get an idea of how restrictive this assumption might be, one can look at the shape of Assumption 2 in the case of a uniform distribution of skills with disagreement $\Delta$ (see Definition 2). The set $\mathcal{A}_2$ of $(\Delta, h)$ values such that the assumption is verified can then be expressed in
closed form:

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

Figure 4: Validity of Assumption 2 in the Uniform Case

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

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

3.2 Allocations

Proposition 1 describes, given a maximum number of layers of management, the allocations in this model: the occupational choice of agents, and who works for whom (defining the boundaries of the firm). Proposition 2 gives the equilibrium number of layers, for a finite number of workers - and when the skill distribution with probability distribution function given by $G(.)$ is the continuous limit of the underlying discrete skill distribution. Together, Propositions 1 and 2 uniquely define the allocations of a Competitive Equilibrium defined in Definition 1.
Before stating Proposition 1, it is convenient to denote the limiting bounds for the support of beliefs for any number of layers \( L \) as \([z_0^L, z_{L+1}^L] = [1 - \Delta, 1]\).

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

\[
z_0^L = 1 - \Delta \quad z_{L+1}^L = 1.
\]

**Proposition 1** (Allocations, given \( L \)). Assume a maximum number of layers \( L \) of hierarchical organization. There exists \( L \) endogenous cutoffs \( \{z_l^L\}_{l=0}^{L} \) uniquely given by a set of \( L \) equations:

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

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

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

*Proof.* See Appendix C.2.

---

**Figure 5: Equilibrium Allocations: Notations**

- \(z_0^L, x_0\) to \(z_1^L, x_1\) to \(z_2^L, x_2\) to \(\cdots\) to \(z_L^L, x_L\) to \(z_{L+1}^L, x_{L+1}\)
- \(1 - \Delta\) to \(1\)
- Workers managers of type 1 managers of type \( L \)
- \(x_1 = m_0(x_0)\)

*Note:* Only "workers" use their time producing, and draw problems in production. Managers of type \( l \) for \( l \in \{1, \ldots, L\} \) do not draw problems, but only spend their time communicating solutions to problems. Workers do not know how hard the problem is, so they cannot go directly to top managers (of type \( L \)) if the problem is very hard. Instead they must waste every intermediary manager’s time before the question reaches the top.

Note that the matching functions depend on the assumed maximum number of layers \( L \), just as the cutoffs, and so should normally be denoted by \( \{m_i^L(\cdot)\}_{i=0}^{L-1} \), but this is omitted for conciseness (see Figure 5 above for the detail of the notations).

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

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

*Proof.* This results straightforwardly from Proposition 1. See Appendix C.3. \( \square \)

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

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

\[
L = \max_{L \in \mathbb{N}} \left\{ L \ s.t. \ 1 - z_L^L \geq \frac{1}{N_w} \right\} \quad (L)
\]

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

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

### 3.3 Span of Control Distributions

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

\[
h (1 - x_l) N_{l+1 \to l}^{l+1}(x_{l+1}) = 1 \quad \Rightarrow \quad N_{l+1 \to l}^{l+1}(x_{l+1}) = \frac{1}{h (1 - m_l^{-1}(x_{l+1}))}.
\]
Table 1: Uniform Skill Distributions: Cutoffs for Different values of $L$, with $\Delta = 100\%$ and $h = 75\%$

<table>
<thead>
<tr>
<th>$L$</th>
<th>$z_1^L$</th>
<th>$z_2^L$</th>
<th>$z_3^L$</th>
<th>$z_4^L$</th>
<th>$z_5^L$</th>
<th>$z_6^L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.666667</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.625993</td>
<td>0.948538</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.625000</td>
<td>0.947266</td>
<td>0.998958</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.625000</td>
<td>0.947266</td>
<td>0.998957</td>
<td>1.000000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.625000</td>
<td>0.947266</td>
<td>0.998957</td>
<td>1.000000</td>
<td>1.000000</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.625000</td>
<td>0.947266</td>
<td>0.998957</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
</tr>
</tbody>
</table>

Note: In this table, the cutoffs are calculated with distribution of skills in the population given by $G(z) = \frac{z - (1 - \Delta)}{\Delta} \mathbb{I} (1 - \Delta, 1) (z) + \mathbb{I} (1, +\infty) (z)$. Precision up to 6 digits is displayed. Successively, the maximum number of levels is restricted to being $L = 1$, $L = 2$, $L = 3$, $L = 4$, $L = 5$, $L = 6$. Heterogeneity in skills is taken to be maximal $\Delta = 100\%$ and so is helping time $h = 75\%$. For comparative statics on $h$ and $\Delta$, and the resulting cutoffs, refer to Table 2 in Appendix E for numerical values and Figures 8 and 29 for a graphical illustration.

Note in passing that $N_{l+1 \rightarrow l}^l = N_{l+1 \rightarrow l}^{l+1} \circ m_l$. All the other span of control functions $N_{l_1 \rightarrow l_2}^{l_2} (x_{l_3})$ result from these fundamental span of control functions between intermediary levels of management, up sometimes to a small change of variable through matching functions $m_l(.)$ for some $l \in \{1, ..., L - 1\}$. The notations for different span of control distributions are illustrated in the case of a firm with $L = 2$ layers on Figure 5, inspired by Figure 1 in Rosen (1982)’s seminal contribution.

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

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

---

6This is because in Geerolf (2013), the leverage factor depended on the price of bond contracts sold, which were a equilibrium outcome of the assignment process.
Note: This chart is inspired from Rosen (1982)’s Figure 1. It represents a firm with $L = 2$ layers, whose top manager (or manager of type-2) has skill $x_2$, supervises intermediary managers (managers of type-2) of skill $x_1 = m_1^{-1}(x_2)$, who themselves supervise workers of skill $x_0 = m_0^{-1}(x_1)$. The span of control of the top managers over intermediary managers is denoted by $N^1_{2 \rightarrow 1}(x_1)$ as a function of intermediary managers’ skills, and $N^2_{2 \rightarrow 1}(x_2)$ as a function of their own skill (so that $N^2_{2 \rightarrow 1} \circ m_1 = N^1_{2 \rightarrow 1}$). Total Span of Control of a top manager over both intermediary managers and workers is given by $N^2_{2 \rightarrow 1}(x_2) \ast (N^1_{1 \rightarrow 0}(x_2) + 1)$ which is behaves like $N^2_{2 \rightarrow 1}(x_2) \ast N^1_{1 \rightarrow 0}(x_2)$ when span of control $N^1_{1 \rightarrow 0}(x_2)$ is large.

### 3.3.1 Special Case: Uniform Distribution of Skills

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

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

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

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

#### Lemma 3 (One Level, Uniform Case)

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

given by a shifted Pareto distribution of coefficient equal to two:

$$N_{l+1}^{l+1}(x_{l+1}) = \frac{1}{h \sqrt{(1 - z_{l+1}^L)^2 + \frac{2}{\pi}(z_{l+2}^L - x_{l+1})}} \sim \text{Pareto}(2).$$

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

\[
N^l_{l+1}(x_{l+1}) \sim \frac{1}{\sqrt{2h}} \frac{1}{\sqrt{z^L_{l+2} - x_{l+1}}}
\]

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

Figure 8: Occupational Choices for a Uniform Distribution of Skills, and \( h = 70\% \)

Note: This figure gives the occupational choices of agents when \( h = 70\% \), and the distribution of skills in the population is given by \( G(z) = \frac{z^{-1}(1-\Delta)}{\Delta} (1 - \Delta, 1) (z) + \mathbb{I} (1, +\infty) (z) \). Note that the 3rd layer and the 4th layer of management can barely be seen on the graph, because their measures are very small, even for high heterogeneity. This gives an intuition for why the Pareto distribution for span of control obtains for any value of overall skill heterogeneity: managers in the upper levels of management endogenously are little different in terms of skills. Note that the decrease in the measure of team managers, plant managers, is not an artefact due to the decrease in heterogeneity \( \Delta \) (see the Graph 30): the measure of managers decreases faster than heterogeneity. In other work, the fraction of managers relative to that of workers goes to zero: this is because production workers are now able to solve almost all problems by themselves.

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

The intuition for why the span of control of top managers becomes very high, even more so when there is an arbitrarily small heterogeneity in skills in the economy, comes from the fact that all workers ask relatively few questions when heterogeneity goes to zero: they are then able to solve almost all problems by themselves. The fraction of managers required to answer these questions also goes to zero, in the limit, as can be seen on Figure 8. The marginal worker, who is just indifferent between being a worker and a manager, then is really able to solve almost all problems by himself, and the corresponding manager has a very high span of control. This reasoning actually also holds for relatively high values of heterogeneity: again as in Figure 8, adding new layers of management has the same effect as reducing heterogeneity in skills between the upper layer.

The fact that span of control is the distributed like a shifted Pareto distribution in the upper tail comes from the mathematical reasoning above: the span of control is an inversely proportional function of the skill of the corresponding worker. This Pareto distribution has an endogenous tail coefficient equal to two, which can be seen on Figure 9, even for relatively high values of heterogeneity: on this Figure the slope of the linear part is $-2$ on a log-log scale when the survivor function is plotted against the value for span of control. Intuitively, this coefficient two comes from the shape of the matching function for the most skilled managers and workers, which becomes flat for those agents when heterogeneity goes to zero, as can be seen on Figure 7 above. The intuition for why this function is flat is that managers with high skills, close to one, are matched with workers who also have relatively high skills, and can in effect solve almost all problems by themselves. It is all the more true that heterogeneity is relatively low. Again, for non zero helping time or heterogeneity, the distribution of span of control we obtain could reasonably be called a "shifted Pareto": in particular, it has an upper bound (though it is very high for most values of the parameters) and does not quite have the Pareto property in the very upper tail. This fact can also be seen on Figure 7, the matching function is not exactly flat for non-zero heterogeneity. Note that this behavior is exactly what is observed in the data, where real economic variables are bounded, and the distribution of firms indeed has fewer firms in the upper tail than the Pareto benchmark would suggest (Stylized Fact 3).\footnote{There is an ongoing fierce debate in the literature about whether the distribution of firm sizes is Pareto or log-normal. The same debate rages about the distribution of city sizes (see Eeckhout (2004), Levy (2009)), as well as about that of trade flows (Head et al. (2014)). The model presented here generates a shifted Pareto only in the upper tail: in particular, the distribution of small firms depends on the distribution of skills; and the way in which the Pareto is shifted in the upper tail depends on the level of skill heterogeneity as well as on helping time.} The way in which the distribution of span of control differs from the full Pareto distribution is illustrated through
Figure 9: Negative Relationship between Heterogeneity in Agents’ Productivities and Heterogeneity in Firm Sizes

Note: This figure gives the theoretical distribution of span of control down one level of hierarchical organization as derived in Section 2, on a log-log-scale, with \( h = 70\% \), for different values of \( \Delta \). The distribution of skills is taken as to be uniform. As heterogeneity \( \Delta \) goes to zero, the distribution approaches the full Pareto distribution in the upper tail. When the number of layers of hierarchical organization is endogenous, the number of agents in the last and before last layers endogenously go to zero, so that the full Pareto is obtained.

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

\[
N_{L \rightarrow L-1}^L(x_L) = \frac{1}{h \sqrt{(1-x_L)^2 + \frac{2}{h}(1-x_L)}} \sim \frac{1}{\sqrt{2h}} \frac{1}{\sqrt{1-x_L}}.
\]
Of course, when \( L > 1 \), the agents below the top managers themselves manage other workers. Their span of control in turn is given by:

\[
N_{L-1 \rightarrow L-2}^{L-1}(x_{L-1}) = \frac{1}{h \sqrt{(1 - z_{L-1}^{L-1})^2 + \frac{2}{h} (z_{L-1}^{L-1} - x_{L-1})}}
\]

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

\[
N_{L-1 \rightarrow L-2}^L(x_L) = N_{L-1 \rightarrow L-2}^{L-1}(m_1^{-1}(x_L))
\]

\[
N_{L-1 \rightarrow L-2}^L(x_L) = \frac{1}{h \sqrt{(1 - z_L^{-2})^2 - \frac{2}{h} (1 - z_L^{-2}) + \frac{2}{h} (1 - x_L)}}
\]

Similarly, this becomes arbitrarily close to a Pareto distribution of tail coefficient equal to four, when \( \Delta \) or \( h \) become small, or when the number of layers increases, as for \( x_L \in [z_L^{-2}, 1] \):

\[
N_{L-1 \rightarrow L-2}^L(x_L) \sim \frac{1}{\sqrt{8h}} \frac{1}{\sqrt{1 - x_L}}.
\]

**Figure 10: Span of Control: Pareto Coefficients, \( L = 2 \) Layers**

![Diagram](image_url)

**Note:** This chart complements Figure 6 in giving an intuition for the formula \( 2^L/(2^L - 1) \) in the case \( L = 2 \). The Pareto distribution for total span of control has a coefficient equal to \( 2/3 = 2^2/(2^2 - 1) \), because of the multiplication of a Pareto with coefficient 2, that coming from the span of control of intermediary managers \( N_{2 \rightarrow 1} \), and one with coefficient 4, that originating from the span of control of intermediary managers on workers, in the space of top managers’ skills \( N_{1 \rightarrow 0} \).

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

\[
N_{L \rightarrow L-2}(x_L) = N_{L \rightarrow L-1}^L(x_L)N_{L-1 \rightarrow L-2}^L(x_L)
\]

\[
N_{L \rightarrow L-2}^L(x_L) = \frac{1}{h^2 \sqrt{(1 - z^L_{L-1})^2} + \frac{2}{h}(1 - x_L) \sqrt{(1 - z^L_{L-1})^2} - \frac{2}{h}(1 - z^L_{L}) + \frac{2}{h}\sqrt{(1 - z^L_{L})^2} + \frac{2}{h}(1 - x_L)}.
\]

Again, one can approximate this as \(\Delta\) or \(h\) become small, or when the number of layers increases by:

\[
N_{L \rightarrow L-2}(x_L) \sim \frac{1}{\sqrt{2h} \sqrt[4]{1 - x_L}} \frac{1}{\sqrt{8h} \sqrt[4]{1 - x_L}} \sim \frac{1}{\sqrt{2h} \sqrt[4]{8h}} \frac{1}{\sqrt[4]{1 - x_L}}.
\]

This is a Pareto distribution with coefficient equal to \(4/3\). The steps of the reasoning for \(L = 2\) levels of Hierarchical Organization are illustrated on Figure 10, where the firm has only two levels of hierarchical organization, in total. The total span of control of the top manager in that case is given by the multiplication of his span of control of managers of type 1 and of each one of these managers of type 1 on workers. The first distribution of span of control is given by a Pareto of coefficient two, as was explained before, and the second one if given by a Pareto of coefficient four.

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

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

\[
N_{n+1 \rightarrow n}^L(x_L) \sim \text{Pareto}(2^{L-n}).
\]

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

\[
N_{L-1 \rightarrow L-2}^L(x_L) = \frac{1}{h \sqrt{(1 - z^L_{L-2})^2} - \frac{2}{h}(1 - z^L_{L}) + \frac{2}{h}\sqrt{(1 - z^L_{L})^2} + \frac{2}{h}(1 - x_L)}.
\]

The shifted Paretos are given explicitly by the formula:

\[
N_{n+1 \rightarrow n}^L(x_L) = \frac{1}{h (1 - m^{-n}_L \circ \ldots \circ m^{-1}_L(x_L))} = \frac{1}{h \left[ 1 - \left( \bigotimes_{l=n}^{L-1} m^{-1}_l \right) (x_L) \right]},
\]

25
Figure 11: Decomposing Zipf’s Law into its Components

Note: This Figure represents the theoretical distribution of span of control for top managers over employees down one to four levels of hierarchical organization, with the assumption that the distribution of skills in the population is given by

\[ G(z) = \frac{z - (1 - \Delta)}{\Delta} (1 - \Delta, 1) (z) + \Pi (1, +\infty) (z). \]

The graph is drawn on a log-log scale, with \( \Delta = 7/10 \), \( h = 8/10 \) and the maximum \( L = 4 \). The slopes on a log-log scale are \(-2\), \(-4/3\), \(-8/7\) and \(-16/15\) respectively. It can be compared with Figure 1 corresponding to the data counterpart.

where each one of the assignment functions are given as a function of the endogenous cutoffs \( \{z^L_i\}_{i=1}^L \) in Proposition 1, namely:

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

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

\[ N^L_{n+1,n}(x_L) = \frac{1}{h (1 - m^{-1}_n \circ \cdots \circ m^{-1}_{L-1}(x_L))} = \frac{1}{h \left[ 1 - \left( \bigotimes_{l=n+1}^{L-1} m^{-1}_l \right)(x_L) \right]} \]

\[ \sim \frac{1}{h \sqrt{1 - \bigotimes_{l=n+1}^{L-1} (m^{-1}_l)(x_L)}} \quad \text{from Lemma 3.} \]
By induction hypothesis:

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

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

\[
N_{n+1\to n}^L(x_L) \sim \text{Pareto}(2^{L-n}),
\]

which proves the proposition for \(L\) and concludes the proof by induction.

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

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

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

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

\[
N_{L\to 0}^L(x_L) = N_{L\to L-1}^L(x_L) * N_{L-1\to L-2}^L(x_L) * ... * N_{1\to 0}^L(x_L) = \prod_{l=0}^{L-1} N_{L-l\to L-l-1}^L(x_L).
\]

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

\[
\frac{1}{\alpha_L} = \sum_{l=0}^{L-1} \frac{1}{2^{l+1}} = \frac{1}{2} \sum_{l=0}^{L-1} \left( \frac{1}{2} \right)^l = \frac{2L - 1}{2^L} \Rightarrow \alpha_L = \frac{2L}{2^L-1}.
\]

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

<table>
<thead>
<tr>
<th>(L)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>...</th>
<th>(+\infty)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\alpha_L) (exact)</td>
<td>2</td>
<td>4/3</td>
<td>8/7</td>
<td>16/15</td>
<td>32/31</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>(\alpha_L) (approx.)</td>
<td>2.00</td>
<td>1.33</td>
<td>1.17</td>
<td>1.07</td>
<td>1.03</td>
<td>...</td>
<td>1.00</td>
</tr>
</tbody>
</table>
These different coefficients are illustrated on Figure 11. In particular, the theoretical coefficients are close to those observed in the data (see Figure 1), and converge to 1 as the number of layers increases. Of course, the uniform distribution of skills is a very particular one. However I show in the next section that what was shown here is more generally true in the upper tail of the span of control distributions for any smooth distribution function, provided its density is continuous and bounded away from zero for higher skills. This is an arguably rather limited set of assumptions.

3.3.2 General Case, Boundedness from Zero

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

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

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

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

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

\[
\forall \epsilon > 0, \quad \exists \eta > 0, \quad \forall x_{l+1} \in [z_{l+2}^L - \eta, z_{l+2}^L],
\]

\[
\frac{1}{h\sqrt{(1 - x_{l+1})^2 + \frac{2}{Ah(1-\epsilon)}(x_{l+1})}} \leq N_{l+1}^{l+1}(x_{l+1}) \leq \frac{1}{h\sqrt{(1 - x_{l+1})^2 + \frac{2}{Ah(1+\epsilon)}(x_{l+1})}}.
\]

This will be denoted as:

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

28
The shifted Pareto on the left-hand side is a lower bound in the upper tail:

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

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

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

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

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

As an illustration, Figure 12 plots the equivalent of Figure 9 in the case where the distribution of skills is given by an increasing function. One can see how the behavior in the lower tail of the distribution differs a bit from the uniform case, but that the upper tail behavior, and in particular the slope \(-2\) for the survivor function on a log-log scale, remains unchanged. (see Appendix D, for more Figures in the increasing case)

Proof. From the market clearing equation for skills:

\[m_i'(x_i) = h(1 - x_i) \frac{g(x_i)}{g(m_i(x_i))}\]

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

\[Ah(1 - x_i)(1 - \epsilon) \leq m_i'(x_i) \leq Ah(1 - x_i)(1 + \epsilon)\]

Integrating between \([x_i, z_{L+1}^l]\), this gives:

\[Ah(1 - \epsilon) \left(z_{L+1}^L - \frac{(z_{L+1}^L)^2}{2} - x_i + \frac{x_i^2}{2}\right) \leq z_{L+2}^L - m_i(x_i) \leq Ah(1 + \epsilon) \left(z_{L+1}^L - \frac{(z_{L+1}^L)^2}{2} - x_i + \frac{x_i^2}{2}\right)\]

\[\Rightarrow 2 \frac{z_{L+2}^L - m_i(x_i)}{Ah(1 + \epsilon)} - 2z_{L+1}^L + (z_{L+1}^L)^2 \leq x_i^2 - 2x_i \leq 2 \frac{z_{L+2}^L - m_i(x_i)}{Ah(1 - \epsilon)} - 2z_{L+1}^L + (z_{L+1}^L)^2\]

\[\Rightarrow 2 \frac{z_{L+2}^L - m_i(x_i)}{Ah(1 + \epsilon)} + (1 - z_{L+1}^L)^2 \leq (1 - x_i)^2 \leq 2 \frac{z_{L+2}^L - m_i(x_i)}{Ah(1 - \epsilon)} + (1 - z_{L+1}^L)^2\]

\[\Rightarrow \sqrt{(1 - z_{L+1}^L)^2 + \frac{2}{Ah(1 + \epsilon)}(z_{L+2}^L - x_{l+1})} \leq 1 - m_i^{-1}(x_{l+1}) \leq \sqrt{(1 - z_{L+1}^L)^2 + \frac{2}{Ah(1 - \epsilon)}(z_{L+2}^L - x_{l+1})}\]

\[\Rightarrow \frac{1}{h\sqrt{(1 - z_{L+1}^L)^2 + \frac{2}{Ah(1-\epsilon)}(z_{L+2}^L - x_{l+1})}} \leq N_{l+1}^{l+1}(x_{l+1}) \leq \frac{1}{h\sqrt{(1 - z_{L+1}^L)^2 + \frac{2}{Ah(1+\epsilon)}(z_{L+2}^L - x_{l+1})}}.\]

Similarly, Proposition 4 generalizes Lemma 4 and Proposition 5 generalizes Lemma 5, to a case where Assumption 3 holds.
Figure 12: INDEPENDENCE OF THE UPPER TAIL BEHAVIOR FROM THE UNDERLYING SKILL DISTRIBUTION

Note: This figure gives the theoretical distribution of team sizes, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{(z-(1-\Delta))^2}{\Delta^2} 1_{[1-\Delta,1]}(z) + 1_{[1,\infty]}(z)$, instead of the uniform case. One can compare to Figure 9 for an equivalent in the uniform case: still the graph is on a log-log scale, down One Level of Hierarchical Organization ("Teams") with always $h = 70\%$. Note that the lower tail behavior differs a bit with this increasing distribution. (as in the data, see Figure 24) However, similarly as in the uniform case, as heterogeneity $\Delta$ goes to zero, the distribution approaches the full Pareto distribution with coefficient two in the upper tail.

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

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

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

Proof. These two Propositions result directly from Proposition 3 in the same way as Lemmas 4 and 5 result from Lemma 3. The proof for the tail coefficient of the approximating shifted Pareto proceeds by induction, using the case for $L = 1$ as a starting statement, assuming that the statement is true for some $L - 1$ with $L \geq 2$ and showing it is true for $L$. Similarly, the proof of Proposition 5 follows from multiplying individual span of control distributions together, which gives the same formula for the total span of control distributions’ tail coefficient $\alpha_L = 2^L/(2^L - 1)$. Refer to the proofs of Lemmas 4 and 5 for details, replacing $\sim$ by $\simeq$ equivalence relations. 

3.3.3 General Case, Non-Boundedness from zero

Let us drop now Assumption 3, and look at what happens when $g(1) = 0$ instead. In that case, it will be useful to approximate the density function in the neighborhood of 1 as follows:

$$g(x) = A(1-x)^\rho + O \left( (1-x)^{\rho+1} \right).$$

Note: The polynomial density functions for skills, for which closed form solutions are obtained, are drawn for $\rho = 0$ (uniform density function), $\rho = 1/2$, $\rho = 1$ and $\rho = 2$ respectively, with $\Delta = 10\%$. $\rho$ governs the behavior of the density function for high skills. Only $\rho = 0$ corresponds to Assumption 3.

As before, the only relevant thing for the shape of the span of control distribution in the upper tail is coefficient $\rho$. Symmetrically to the benchmark of a uniform distribution the case of
boundedness from zero, it suffices to look at the following simple polynomial functions:

$$g(x) = \frac{\rho + 1}{\Delta^{\rho + 1}} (1 - x)^\rho.$$  

They integrate up to one since the corresponding cumulative distribution function is given by:

$$G(x) = 1 - \frac{(1 - x)^{\rho + 1}}{\Delta^{\rho + 1}}.$$  

In that case, the matching function is given through the differential equation:

$$\frac{\rho + 1}{\Delta^{\rho + 1}} (1 - m(x))^\rho m'(x) = h(1 - x) \frac{\rho + 1}{\Delta^{\rho + 1}} (1 - x)^\rho$$

$$\Rightarrow \left[ \frac{(1 - m(u))^\rho + 1}{\rho + 1} \right] z_{l+1} = h \left[ - \frac{(1 - u)^{\rho + 2}}{\rho + 2} \right] z_{l+1}.$$  

The span of control distribution down 1 level of hierarchical organization is therefore given by:

$$N_{t+1-s}(x_{l+1}) = \frac{1}{h \left[ (1 - z_{l+1})^{\rho + 2} - \frac{1}{h^{1+\rho}} [(1 - x_{l+1})^{\rho + 1} - (1 - z_{l+1})^{\rho + 1}] \right]} \sim \text{Pareto} \left( \frac{2 + \rho}{1 + \rho} \right).$$  

Note that for $\rho = 0$, one obtains the formula derived earlier for a uniform distribution of skills. Symmetrically as before, one can generalize this to the case of $L$ levels of hierarchical organizations, in which case the Pareto coefficient on $N_L(x_L)$, such that $N_L(x_L) \sim (1 - x_L)^{1/\alpha_L}$ is given by:

$$\frac{1}{\alpha_L} = \sum_{l=0}^{L-1} \left( \frac{1 + \rho}{2 + \rho} \right)^{l+1} = \frac{1 + \rho}{2 + \rho} \left( \frac{1 + \rho}{2 + \rho} \right)^L \Rightarrow \alpha_L = \frac{1}{1 + \rho} \left( \frac{2 + \rho}{2 + \rho} \right)^L.$$  

Note that the Pareto coefficient for the span of control distribution as measured in the data, given by $N_{L-0}(x) \sim (1 - x)^{1/\beta_L}$ in turn is given by:

$$\beta_L = \frac{(2 + \rho)^L}{(2 + \rho)^L - (1 + \rho)^L}.$$  

This is because:

$$N_{L-0}(x) = N_L \left( G^{-1}(x) \right) \sim (1 - G^{-1}(x))^{1/\alpha_L} \sim (1 - x)^{1/(\alpha_L^{\rho + 1})}.$$  

The last equivalence comes from:

$$1 - G(x) \sim \frac{(1 - x)^{\rho + 1}}{\Delta^{\rho + 1}} \Rightarrow 1 - G^{-1}(x) \sim \Delta (1 - x)^{\frac{1}{\rho + 1}}.$$  

This allows to state Proposition 6 in the general case where Assumption 3 does not hold.
Proposition 6 \((L \text{ Levels, General Case, } g(1) = 0)\). Assume that in the neighborhood of 1, \(g(x) \sim_{x\to 1} (1 - x)^\rho\) for some \(\rho\). Then total span of control of managers of type \(n\) on levels of hierarchical organization down \(L\) levels, in the space of top managers’ skills, denoted by \(N_{n\to n-L}^n(x_n)\) is arbitrarily close to the shifted Pareto distribution of coefficient equal to that obtained in the polynomial of degree \(\rho\) case:

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

The span of control distribution of top managers as measured in the data then follows:

\[
N_{n+L\to n}^{top}(x) \simeq\text{Pareto}\left(\frac{(2 + \rho)^L}{(2 + \rho)^L - (1 + \rho)^L}\right).
\]

4 Top Labor Income Distribution [Very Preliminary and Incomplete]

The equilibrium spans of control for managers was derived without any reference to equilibrium prices. This is because the solution is more easily found through the planner’s problem. However one can calculate how this competitive equilibrium arises in practice, which is decentralized through internal labor markets. Interestingly, when the race between technology and education parameter \(\rho\) parameter is higher than a certain number, Pareto distributions for top incomes arise.

4.1 Theory

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

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

\[
\forall x_L \in [z^L, z^L_{L+1}], \quad w_L(x_L) = x_L.
\]

The return of a manager of level \(l\) is then given by the number of problems that the hierarchy solves on the aggregate minus what is given back, for each unit of problem, to lower level workers:

\[
R_{l+1}(x_{l+1}) \equiv (w_{l+1}(x_{l+1}) - w_l(x_l)) N_{l+1\to 0}^l(x_l).
\]

Note that the total wage of the workers in layer \(l\) are actually given by how many of those problems are solved by the hierarchy on top of which they are:

\[
W_l(x_l) \equiv N_{l\to 0}^l(x_l)w_l(x_l).
\]
By convention, let us similarly define the return function for a production worker as:

\[ \forall x_0 \in [z^L_0, z^L_1], \quad R_0(x_0) = w_0(x_0). \]

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

**Proposition 7** (Internal Labor Markets). The implicit wages received by each layer of the hierarchy are given through inverse recursion, starting from \( w_L(x_L) = x_L \) by:

\[ \forall l \in \{0, ..., L - 1\}, \quad w_l(x_l) = \frac{1}{N^l_{l+1 \to 0}(x_l)} \int_{z^L_l}^{x_l} w_{l+1}(m_l(u)) \frac{dN^l_{l+1 \to 0}(u)}{dx_l} du + \frac{N^l_{l+1 \to 0}(z^L_l)}{N^l_{l+1 \to 0}(x_l)} w_l(z^L_l). \]

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

\[ \forall l \in \{0, ..., L - 1\}, \quad R_{l+1}(z^L_{l+1}) = R_l(z^L_l). \]

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

\[ m^{-1}_l(x_{l+1}) = \arg \max_{x_l} (w_{l+1}(x_{l+1}) - w_l(x_l)) N^l_{l+1 \to 0}(x_l) \]

\[ \Rightarrow \frac{d}{dx_l} \left( w_l(x_l) N^l_{l+1 \to 0}(x_l) \right) = w_{l+1}(m_l(x_l)) \frac{dN^l_{l+1 \to 0}(u)}{dx_l} du. \]

\[ \Rightarrow w_l(x_l) N^l_{l+1 \to 0}(x_l) - w_l(z^L_l) N^l_{l+1 \to 0}(z^L_l) = \int_{z^L_l}^{x_l} w_{l+1}(m_l(u)) \frac{dN^l_{l+1 \to 0}(u)}{dx_l} du. \]

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

\[ \forall l \in \{0, ..., L - 1\}, \quad R_{l+1}(z^L_{l+1}) = R_l(z^L_l). \]

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

Of particular interest is the theoretical distribution of top managers’ wages, both because it does not depend on the underlying distribution of skills as for the distribution of span of control, and because inequality issues have gained some prominence in the public debate recently. We are therefore interested in the distribution of \( R_L(x_L) \), which obtains rather straightforwardly from Proposition 7.

**Proposition 8** (Top Labor Income Distribution). The incomes at the top are given as a function of the skill of top managers \( x_L \) by:

\[ R_L(x_L) = R_L(z^L_L) + \int_{z^L_L}^{x_L} N^L_{L \to 0}(u) du. \]
In fact, more generally, total returns of managers are given as:

\[ R_{l+1}(x_{l+1}) = R_{l+1}(z_{l+1}^L) + \int_{z_{l+1}^L}^{x_{l+1}} w_{l+1}(u) N_{l+1-0}^{l+1}(u) du. \]

One can get a justification for this formula through a very straightforward envelope condition. Since, again, managers of type \( l + 1 \) optimize over their expected revenue \( R_{l+1}(x_{l+1}) \):

\[ \max_{x_l} R_{l+1}(x_{l+1}) = (w_{l+1}(x_{l+1}) - w_l(x_l)) N_{l+1-0}^{l+1}(x_l). \]

We have, at the optimum:

\[ \frac{\partial R_{l+1}(x_{l+1})}{x_{l+1}} = w_{l+1}'(x_{l+1}) N_{l+1-0}^{l+1}(x_l) = w_{l+1}'(x_{l+1}) N_{l+1-0}^{l+1}(x_{l+1}). \]

Integrating straightforwardly gives the result. And Proposition 8 is just an application of this result for \( l = L - 1 \), in which case \( w_L'(x_L) = F'(x_L) = 1 \) by definition. One may however feel uncomfortable with the successive changes of variables, as well as wonder whether an envelope theorem can so directly be applied in this sequential maximization problem, so I provide a lengthier, though perhaps less problematic, proof in Appendix C.5.

**Polynomial Density Functions.** Assume that the distribution of skills is taken in the family of polynomial functions defined above \( g(x) = (\rho + 1)(1 - x)^\rho/(1 - \Delta)^\rho \). From the formula in Proposition 8:

\[ R_L(x_L) = R_L(z_L^L) + \int_{z_L^L}^{x_L} N_{L-0}^{L}(u) du. \]

From the fact that:

\[ N_{L-0}^{L}(x_L) \approx \text{Pareto}(\alpha_L) \quad \text{with} \quad \alpha_L = \left( \frac{1}{1 + \rho} \left( \frac{2 + \rho}{2 + \rho} \right)^L \right) \]

From straightforward integration, and in the case where \( \alpha_L \), we have that:

\[ R_L(x_L) \approx \text{Pareto} \left( \frac{1}{\alpha_L - 1} \right). \]

Finally, the top income distribution as measured in the data is arbitrarily close to a Pareto in the upper tail with:

\[ R_{\text{top}}(x_L) \approx \text{Pareto} \left[ \frac{(1 + \rho)(2 + \rho)^L}{\rho(2 + \rho)^L - (1 + \rho)^{L+1}} \right]. \]

The Pareto Coefficients for Top Incomes therefore vary as a function of \( \rho \) and \( L \), in a manner illustrated on Figure 13.

**General Distribution of Skills.** Again, these cases can be generalized to the case of sufficiently regular functions, which can be approximated around one by one of the above polynomials as a Taylor expansion.
Figure 13: Theoretical Top Labor Incomes Pareto Coefficients, as a function of ρ and L

Note: This chart represents the Pareto Coefficients for Top Incomes as a function of the race between technology and education, embodied in the density of highest skills at the top, and the equilibrium number of layers of hierarchical organization. Darker colors correspond to lower Pareto coefficients for Top Incomes (converging to 1), so a relatively fatter tail and more inequality. The red dot represents the maximum level of inequality attainable for ρ ∈ [0, 20] and L ∈ [1, 20], with a Pareto coefficient equal to 1.26.

4.2 Comparative Statics Exercises

Three comparative statics are of interest and can all lead to an increase in top income inequality, or rise of the top 1% income share. Only the last one is illustrated by a Figure in the main text (Figure 14), as the first two are neither completely new nor surprising.

Decrease in communication costs. The comparative statics exercises involving a decrease in communication costs are rather straightforward. As in Garicano and Rossi-Hansberg (2006),
decreases in communication costs lead to larger and more heterogenous span of control for managers, and therefore rising income inequality. (in fact, it leads also to rising wage polarization)

This corresponds to Figure 34 in Appendix E.

**Increase in** $\rho$. Comparative statics with respect to $\rho$ are a little bit less straightforward. In particular, interesting and not completely intuitive is that when $\rho = 0$, that is when the distribution is bounded away from zero for higher skills, income inequality is at a low. One can see that straightforwardly on the top incomes distribution formula in Proposition 8. Integrating the Pareto distribution with a coefficient higher than 1 leads to a distribution that is no longer Pareto, as the integral of $N_{L_{>0}}(\cdot)$ converges. The intuition is that managers then compete to attract the best workers, and they end up giving almost all the surplus from the match to them. In contrast, for higher values of $\rho$, that is when talented individual are in relatively scarce supply,
everything happens "as if" they had some market power: the distribution of the surplus shifts towards the managers. This force is actually stronger than a countervailing force: because there are fewer good managers, the workers also are relatively less competent, and so the span of control of managers is lower. These comparative statics are illustrated on Figure 35 in Appendix E.

**Decrease in heterogeneity.** Just as the span of control distribution is more heterogenous when agents’ skills are more similar (Δ decreases), the comparative statics of the labor income distribution with respect to heterogeneity are, again, quite counterintuitive. The intuition again is that when heterogeneity decreases, workers that managers hire are relatively more competent and can almost answer all questions by themselves, so that they allow managers to achieve a higher span of control.

Figure 15: **Empirical Top Incomes Pareto Coefficients in a Sample of Advanced Economies (1920-2010).**

---

*Note:* Source: World Top Income Database. This plot shows the evolution of Pareto-Lorenz coefficients α, as taken from the World Top Income Database (Atkinson et al. (2009), see also Alvaredo et al. (2013)). The sample is restricted to the countries having the longest series on record.
4.3 Evidence [Preliminary]

One would ideally need micro-level data in order to assess the different factors having contributed to the recent rise in labor income inequality. Preliminary evidence seems to suggest that Pareto coefficients for top incomes have lowered, pointing to a race between education and technology (see the US series on Figure 15). The Survey of Consumer Finances data on Figures 23(a) and 23(b) seems to point to the same tendency. This assessment however remains preliminary at this stage. The top 1% income share, on which data is far more reliable, is not a sufficient statistic to investigate this issue. In the model, it is consistent with a decrease in communication costs, a larger race between education and technology than before, or (a new insight) a decrease in average worker skill.

5 Conclusion

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

As Garicano and Rossi-Hansberg (2006) have already noted, the model outlined above can perhaps help better understand the increase in top income inequality which has occurred in recent years, in particular due to the decrease in communication costs made possible by information technology. Through comparative statics on parameter $h$, the model can therefore give some flesh to the skill-biased technical change hypothesis, presenting an alternative to the macroeconomic model with skilled and unskilled labor appearing as imperfect substitutes in a neoclassical production function (Katz and Murphy (1992)). It can also explain why a macroeconomic analysis leads to attribute a large share of skill-biased technological change to a rise in capital equipment, if the latter is complementary with skilled labor as in Krusell et al. (2000). Indeed, decreases in communication technology both appear as more capital equipment in firms that now use information technology more intensively, and lead skilled individuals to manage more workers through larger firms.

Unfortunately, the model is very much a first pass in many respects. Its most important
limitation is perhaps that skills are assumed to be a given in the model, and that all agents work the same, so that unequal wages only result from innate abilities. I conjecture that the span of control distributions would remain the same in a larger class of extended models with education and elastic labor supply, as differences in education abilities are in many ways ex-post isomorphic to ex-ante differences in skills, and labor supply decisions would only amplify the forces at hand in the paper. Yet for other issues, this simplification is clearly a problem. Consider optimal taxation. If abilities were all innate, a Rawlasian planner would be able to redistribute all unequal labor income in equal proportions at no efficiency cost. Yet, skills are not entirely innate, and doing so would discourage both skill investment through education as well as more labor supply from agents at the top of the distribution; because the model is purely competitive, prices provide exactly the right incentives to educate themselves more, or to work more. Therefore, in its current form, the model remains essentially silent on the key efficiency-equity tradeoffs. Not completely, though: since optimal taxation models are very often calibrated using an underlying Pareto distribution in agents’ productivity (see, for example, Diamond (1998), Saez (2001), Tuomala (1990)), the model holds promises for the structural estimation of optimal income tax rates.

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

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

---

8 This would likely be reinforced in the presence of learning externalities as in human capital accumulation models.
one allows for innovation, and in particular for the fact that at the beginning of the product cycle only few entrepreneurs are able to design it, then the question would be to study the incentives to spread around this knowledge, to increase span of control (which increases when the skills of lower ranking agents increases); and the countervailing forces to retain knowledge, as the wage distribution is all the more rewarding to entrepreneurs that there are relatively few of them.

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


43


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


A Evidence

A.1 New Evidence

This data is assembled from French Déclarations Annuelles de Données Sociales (DADS), from a sample in year 2007. The sample originally contains 55,979,881 observations of employee-employer matched pairs. In order to classify workers into different layers of hierarchical organization, I follow Caliendo et al. (2014) and use the first digit of PCS (Profession, Catégorie Socioprofessionnelle), which allows me to divide workers into subgroups of high ranking managers, middle managers, workers, manual workers. More precisely, the first digit of PCS is one of these six alternatives:

1. Farmers.
2. Self-employed / Owners. Example: Plumbers, Film directors, Chief Executive Officers.
3. Senior staff or Top management positions. Example: Chief Financial Officers, heads of Human Resources, purchasing managers.
4. Employees at the supervisor level. Example: Quality control technicians, Sales supervisors.
5. Clerical, White collar employees. Example: Secretaries, Human resource or Accounting, Sales Employees.
PCS 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 a first digit PCS equal to 3, 4, 5, or 6.

Figure 17: **Wholesale Trade, Accommodation, Food (Administrative Data)**

![Graph 1](image1)

Figure 18: **Team Size Distribution in France (Administrative Data)**

![Graph 2](image2)

Source: *Déclarations Annuelles de Données Sociales, Year 2007*
Figure 19: **Plants Per Firms Distribution (Administrative Data)**

![Plants Per Firms Distribution](image)

Slope: -1.32

Figure 20: **Pareto Coefficients for Top Labor Incomes in France (Administrative Data)**

![Pareto Coefficients](image)

Slope: -2.59

Slope: -2

Fitted: 23000€ ≤ w ≤ 355000€

Fitted: w ≥ 355000€
Figure 21: Roberts (1956)’s law - ExecuComp Database, United States

![Graph showing the relationship between the logarithm of the number of employees and the logarithm of fixed plus variable salary. The graph includes data points and a fitted line.

Source: ExecuComp Database, Year 2005]

Figure 22: Labor Incomes in the United States (Survey Data)

![Graph showing the relationship between the logarithm of wage income and the logarithm of survivor. The graph includes data points and a fitted line.

Source: ExecuComp Database, Year 2005]

Note: Source: Survey of Consumer Finances, 2007. This evidence must be interpreted with a lot of caution, since surveys capture the top incomes very imperfectly.
Figure 23: Comparing the Labor Income Distribution in 1989 and 2010 in SCF

(a) 1989

(b) 2010

Note: This evidence must be interpreted with extreme caution, since surveys capture the top incomes very imperfectly.
A.2 Evidence from the Literature

Figure 24: Zipf’s Law for US Firm Sizes, Log-Log Scale, Axtell (2001)

Note: In this figure, the density of firm sizes is represented. The estimated slope is 2.059 in the frequency domain, corresponding to a Pareto tail index of 1.059. The distribution of firms between 1 and 10 employees is more concave than Pareto, which has sparked a vigorous debate about whether the distribution was closer to a log-normal or to a Pareto (see Eeckhout (2004)). The static model I develop predicts a Zipf’s law only in the upper tail (see Figure 12). Similarly, big firms are smaller than Zipf’s law would predict, again a prediction of the model for non-zero heterogeneity.
Figure 25: Puzzle for Random Growth Theory: Distribution of US Establishment Growth Rates. Source: Rossi-Hansberg and Wright (2007)

(a) Growth Rates

(b) Growth Rates by Sector

Notes:

1652 THE AMERICAN ECONOMIC REVIEW
B Computational Appendix

B.1 Calculating the Cutoffs \( \{z_{L}^{l}\}_{l=1}^{L-1} \) and the Matching Functions \( \{m_{l}(.)\}_{l=0}^{L-1} \)

From market clearing for time, the differential equations for the matching functions are:

\[
\forall l \in \{0, ..., L - 1\}, \quad \forall x \in [z_{L}^{l}, z_{L+1}^{l}], \quad m_{l}^{'}(x)g(m_{l}(x)) = h(1 - x)g(x) \quad \text{and} \quad m_{l}(z_{L}^{l}) = z_{L}^{l+1}.
\]

Integrating, this is equivalent to:

\[
\forall l \in \{0, ..., L - 1\}, \quad \forall x \in [z_{L}^{l}, z_{L+1}^{l}], \quad m_{l}(x) = G^{-1}\left[G(z_{L+1}^{l}) + \int_{z_{L}^{l}}^{x} h(1 - u)g(u)du\right].
\]

The computational procedure follows the uniqueness proof. Taking \( z_{1}^{L} \) as given, all the remaining \( \{z_{L}^{l}\}_{l=2}^{L} \) can be expressed recursively as a function of \( z_{1}^{L} \) using the above equations as well as:

\[
\forall l \in \{0, ..., L - 1\}, \quad z_{L+1}^{l} = m_{l}(z_{L+1}^{l}).
\]

In particular, this gives a final value for \( z_{L+1}^{L}(z_{1}^{L}) \) as a function of \( z_{1}^{L} \), which defines \( z_{1}^{L} \) implicitly through:

\[
z_{L+1}^{L}(z_{1}^{L}) = 1.
\]

The matching functions and cutoffs are then expressed as a function of \( z_{1}^{L} \), and all the primitives in the model.

**Polynomial Functions.** In the case of the family of polynomial functions, these are even expressed in closed form as a function of \( z_{1}^{L} \) because:

\[
\int_{1-\Delta}^{x} (1 - u)g(u)du = \frac{\rho + 1}{\Delta^{\rho+1}} \int_{1-\Delta}^{x} (1 - u)^{\rho+1}du
\]

\[
= \frac{\rho + 1}{\Delta^{\rho+1}} \left[ \frac{(1 - u)^{\rho+2}}{\rho + 2} \right]_{1-\Delta}^{x}
\]

\[
H(x) \equiv \int_{1-\Delta}^{x} (1 - u)g(u)du = \frac{\rho + 1}{\rho + 2} \Delta \left[ 1 - \frac{(1 - x)^{\rho+2}}{\Delta^{\rho+2}} \right].
\]

The inverse of this function is also expressed in closed form:

\[
H^{-1}(x) = 1 - \Delta \left( 1 - \frac{\rho + 2}{\Delta(\rho + 1)} \right)^{1/\rho+2}.
\]

Finally, \( g \) and \( G \) are also closed form expressions as well as \( G^{-1}(.) \):

\[
G^{-1}(x) = 1 - \Delta(1 - x)^{1/\rho+1}.
\]

Span of control distributions are similarly expressed in closed form in that case, as the inverse of the matching functions are given by:

\[
G(x) - G(z_{1}^{L}) = h \left( m_{L-1}^{-1}(x) - H(z_{L-1}^{L}) \right)
\]

\[
\Rightarrow \quad m_{L-1}^{-1}(x) = H^{-1}\left[H(z_{L-1}^{L}) + \frac{G(x) - G(z_{1}^{L})}{h}\right].
\]
Computationally speaking, solving the competitive equilibrium completely thus only relies on finding one scalar \(\{z^l\}\), which determine the allocations completely. Allocations, matching functions and span of control distributions are even given in closed form in the case of polynomial functions.

**B.2 Calculating Initial Conditions of the Wage Functions \(\{a_l\}^{L-1}_{l=0}\), and the Return Functions \(\{R_l\}^L_{l=0}\)**

Denote the initial conditions for the wage functions as \(\{a_l\}^{L-1}_{l=0}\), that is:

\[
\forall l \in \{0, ..., L-1\}, \quad a_l = w_l(z^l),
\]

The wage functions are entirely defined recursively in integral form as a function of \(\{a_l\}^{L-1}_{l=0}\) as well as \(\{z^l\}^L_{l=1}\) through:

\[
\forall x \in [z^l, z^{l+1}], \quad w_l(x) = \frac{1}{N^{l+1}_{l+1-0}(x) \int_{z^l}^{x} w_{l+1}(m_l(u)) \frac{dN^{l}_{l+1-0}}{dx_l}(u)du + \frac{N^{l}_{l+1-0}(z^l)}{N^{l+1}_{l+1-0}(x)}w_l(z^l)}
\]

s.t. \(w_l(z^l) = a_l\).

For backward recursion the wage function of manager of type \(L\) is here defined as \(w_L(x_l) = F(x_l) = x_l\). By backward recursion, starting from given by the identity function, one finds the whole sequence of wage functions \(\{w_l(.)\}^{L-1}_{l=0}\) as a function of the \(\{a_l\}^{L-1}_{l=0}\). The \(\{a_l\}^{L-1}_{l=0}\) are in turn given by the \(1+L-1 = L\) indifference equations:

\[
R_l(z^l) = w_0(z^l),\quad \forall l \in \{1, ..., L-1\}, \quad R_{l+1}(z^{l+1}) = R_l(z^{l+1}).
\]

Denote by \(\{b_l\}^{L-1}_{l=0}\) the following quantities:

\[
\forall l \in \{0, ..., L-1\}, \quad b_l = w_l(z^{l+1}).
\]

Then the \(L\) previous equations simplify as follows:

\[
R_l(z^l) = w_0(z^l) \quad \Rightarrow \quad b_0 = (a_1 - a_0)N^0_{l-0}(z^l).
\]

\[
\forall l \in \{1, ..., L-1\}, \quad R_{l+1}(z^{l+1}) = R_l(z^{l+1}) \quad \Rightarrow \quad (w_{l+1}(z^{l+1}) - w_l(z^{l})) N^l_{l+1-0}(z^l) = (w_l(z^{l+1}) - w_{l-1}(z^{l})) N^{l-1}_{l-0}(z^l)
\]

\[
\Rightarrow \quad (a_{l+1} - a_l) N^l_{l+1-0}(z^l) = (b_l - b_{l-1}) N^{l-1}_{l-0}(z^l).
\]

Note that for symmetry, I have defined \(a_L\) as the initial condition of the hypothetical wage function of managers of type \(L\), equal to \(F(.)\) so that \(a_L = w_L(z^L) = z^L\).

The numbers \(\{b_l\}^{L-1}_{l=0}\) are given as a function of \(\{a_l\}^{L-1}_{l=0}\) by using the above differential equations between \([z^L, z^{L+1}]\) so that:

\[
\forall l \in \{0, ..., L-1\}, \quad N^l_{l+1-0}(z^{l+1})b_l = \int_{z^l}^{z^{l+1}} w_{l+1}(m_l(u)) \frac{dN^l_{l+1-0}}{dx_l}(u)du + N^l_{l+1-0}(z^l)a_l.
\]
By recursion:

\[ N_{L-1}^{L-1}(z_{L-1}^L) b_{L-1} = \int_{z_{L-1}^L}^{Z_{L-1}^L} m_{L-1}(t) \frac{dN_{L-1}^{L-1}}{dx_{L-1}}(t) dt + N_{L-1}^{L-1}(z_{L-1}^L)a_{L-1} \]

\[ N_{L-1}^{L-2}(z_{L-1}^L) b_{L-2} = \int_{z_{L-1}^L}^{Z_{L-1}^L} w_{L-1}(m_{L-2}(v)) \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv + N_{L-1}^{L-2}(z_{L-1}^L)a_{L-2} \]

With:

\[ w_{L-1}(m_{L-2}(v)) = \frac{1}{N_{L-1}^{L-1}(m_{L-2}(v))} \int_{z_{L-1}^L}^{Z_{L-1}^L} m_{L-1}(t) \frac{dN_{L-1}^{L-1}}{dx_{L-1}}(t) dt + \frac{1}{N_{L-1}^{L-1}(m_{L-2}(v))} N_{L-1}^{L-1}(z_{L-1}^L)a_{L-1} \]

Therefore:

\[ N_{L-1}^{L-2}(z_{L-1}^L) b_{L-2} = \int_{z_{L-2}^L}^{Z_{L-2}^L} \frac{1}{N_{L-1}^{L-1}(m_{L-2}(v))} \int_{z_{L-1}^L}^{Z_{L-1}^L} m_{L-1}(t) \frac{dN_{L-1}^{L-1}}{dx_{L-1}}(t) dt \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv + \ldots \]

\[ \ldots \left[ \int_{z_{L-2}^L}^{Z_{L-2}^L} \frac{1}{N_{L-1}^{L-1}(m_{L-2}(v))} \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv \right] N_{L-1}^{L-1}(z_{L-1}^L)a_{L-1} + N_{L-1}^{L-2}(z_{L-2}^L)a_{L-2} \]

Finally:

\[ N_{L-1}^{L-3}(z_{L-2}^L) b_{L-3} = \int_{z_{L-3}^L}^{Z_{L-3}^L} w_{L-2}(m_{L-3}(y)) \frac{dN_{L-1}^{L-2}}{dx_{L-3}}(y) dy + N_{L-1}^{L-3}(z_{L-2}^L)a_{L-3} \]

With:

\[ w_{L-2}(m_{L-3}(y)) = \frac{1}{N_{L-1}^{L-2}(m_{L-3}(y))} \int_{z_{L-2}^L}^{Z_{L-2}^L} w_{L-1}(m_{L-2}(v)) \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv + \ldots \]

\[ \ldots \frac{1}{N_{L-1}^{L-2}(m_{L-3}(y))} N_{L-1}^{L-2}(z_{L-2}^L)a_{L-2} \]

Therefore:

\[ N_{L-1}^{L-3}(z_{L-2}^L) b_{L-3} = \int_{z_{L-3}^L}^{Z_{L-3}^L} \frac{1}{N_{L-1}^{L-2}(m_{L-3}(y))} \int_{z_{L-2}^L}^{Z_{L-2}^L} w_{L-1}(m_{L-2}(v)) \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv ... \]

\[ \ldots \frac{dN_{L-1}^{L-2}}{dx_{L-3}}(y) dy + \left[ \int_{z_{L-2}^L}^{Z_{L-2}^L} \frac{1}{N_{L-1}^{L-2}(m_{L-3}(y))} \frac{dN_{L-1}^{L-3}}{dx_{L-3}}(y) dy \right] N_{L-1}^{L-2}(z_{L-2}^L)a_{L-2} + N_{L-1}^{L-3}(z_{L-3}^L)a_{L-3} \]

Finally:

\[ N_{L-1}^{L-3}(z_{L-2}^L) b_{L-3} = \int_{z_{L-3}^L}^{Z_{L-3}^L} \frac{1}{N_{L-1}^{L-2}(m_{L-3}(y))} \int_{z_{L-2}^L}^{Z_{L-2}^L} \frac{1}{N_{L-1}^{L-3}(m_{L-3}(y))} \int_{z_{L-1}^L}^{Z_{L-1}^L} m_{L-1}(t) \frac{dN_{L-1}^{L-1}}{dx_{L-1}}(t) dt \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv \frac{dN_{L-1}^{L-3}}{dx_{L-3}}(y) dy + \ldots \]

\[ \ldots \left[ \int_{z_{L-1}^L}^{Z_{L-1}^L} m_{L-1}(t) \frac{dN_{L-1}^{L-1}}{dx_{L-1}}(t) dt \frac{dN_{L-1}^{L-2}}{dx_{L-2}}(v) dv \frac{dN_{L-1}^{L-3}}{dx_{L-3}}(y) dy \right] N_{L-1}^{L-1}(Z_{L-1}^L)a_{L-1} \]

\[ \ldots \left[ \int_{z_{L-2}^L}^{Z_{L-2}^L} \frac{1}{N_{L-1}^{L-2}(m_{L-3}(y))} \frac{dN_{L-1}^{L-3}}{dx_{L-3}}(y) dy \right] N_{L-1}^{L-2}(z_{L-2}^L)a_{L-2} + N_{L-1}^{L-3}(z_{L-3}^L)a_{L-3} \]
The first closest terms to the diagonal of \( \{\lambda_i\}_{i=0}^{L-1} \) and the first terms of \( \{\mu_i\}_{i=0}^{L-1} \) can therefore be written as follows:

\[
\mu_{L-1} - \int_{z_{L-1}}^{z_L} m_{L-1}(t) \frac{dN_{L-0}^{L-1}(t)}{dx_{L-1}} dt \\
\mu_{L-2} = \int_{z_{L-2}}^{z_{L-1}} \frac{1}{N_{L-1}^{L-0}(m_{L-2}(v))} \int_{z_{L-2}}^{m_{L-2}(v)} \frac{dN_{L-1}^{L-1}(v)}{dx_{L-2}} dv \int_{z_{L-1}}^{m_{L-2}(v)} m_{L-1}(t) \frac{dN_{L-0}^{L-1}(t)}{dx_{L-1}} dt \\
\mu_{L-3} = \int_{z_{L-3}}^{z_{L-2}} \frac{1}{N_{L-1}^{L-0}(m_{L-3}(y))} \int_{z_{L-2}}^{m_{L-3}(y)} \frac{dN_{L-1}^{L-1}(y)}{dx_{L-2}} dv \int_{z_{L-1}}^{m_{L-2}(v)} m_{L-1}(t) \frac{dN_{L-0}^{L-1}(t)}{dx_{L-1}} dt \\
\lambda_{L-2} - \int_{z_{L-2}}^{z_L} \frac{1}{N_{L-0}^{L-0}(m_{L-2}(v))} \int_{z_{L-2}}^{m_{L-2}(v)} \frac{dN_{L-1}^{L-1}(v)}{dx_{L-2}} dv \\
\lambda_{L-3} - \int_{z_{L-3}}^{z_{L-2}} \frac{1}{N_{L-0}^{L-0}(m_{L-3}(y))} \int_{z_{L-2}}^{m_{L-3}(y)} \frac{dN_{L-1}^{L-1}(y)}{dx_{L-2}} dv \\
\lambda_{L-4} - \int_{z_{L-3}}^{z_{L-2}} \frac{1}{N_{L-0}^{L-0}(m_{L-4}(z))} \int_{z_{L-2}}^{m_{L-4}(z)} \frac{dN_{L-1}^{L-1}(z)}{dx_{L-2}} dv \\
\lambda_{L-5} - \int_{z_{L-3}}^{z_{L-2}} \frac{1}{N_{L-0}^{L-0}(m_{L-5}(w))} \int_{z_{L-2}}^{m_{L-5}(w)} \frac{dN_{L-1}^{L-1}(w)}{dx_{L-2}} dv \\
\lambda_{L-6} - \int_{z_{L-3}}^{z_{L-2}} \frac{1}{N_{L-0}^{L-0}(m_{L-6}(v))} \int_{z_{L-2}}^{m_{L-6}(v)} \frac{dN_{L-1}^{L-1}(v)}{dx_{L-2}} dv \\
\lambda_{L-7} - \int_{z_{L-3}}^{z_{L-2}} \frac{1}{N_{L-0}^{L-0}(m_{L-7}(y))} \int_{z_{L-2}}^{m_{L-7}(y)} \frac{dN_{L-1}^{L-1}(y)}{dx_{L-2}} dv
\]

Example with \( L = 4 \). The above equations consist of \( L \) linear equations in \( \{a_i\}_{i=0}^{L-1} \). Calculating the coefficients of these linear equations involves a computation of \( L(L+1)/2 \) integral quantities \( \{\lambda_i\}_{i=0}^{L-1} \) and \( \{\mu_i\}_{i=0}^{L-1} \), which give \( \{b_i\}_{i=0}^{L-1} \) as a function of \( \{a_i\}_{i=0}^{L-1} \) through:

\[
\begin{bmatrix}
N_1^{0\rightarrow0}(z_L^0) b_0 \\
N_2^{0\rightarrow1}(z_L^0) b_1 \\
N_3^{0\rightarrow2}(z_L^0) b_2 \\
N_4^{0\rightarrow3}(z_L^0) b_3
end{bmatrix} = 
\begin{bmatrix}
1 & \lambda_{01} & \lambda_{02} & \lambda_{03} \\
0 & 1 & \lambda_{12} & \lambda_{13} \\
0 & 0 & 1 & \lambda_{23} \\
0 & 0 & 0 & 1
end{bmatrix}
\begin{bmatrix}
N_1^{0\rightarrow0}(z_L^0) a_0 \\
N_2^{1\rightarrow0}(z_L^0) a_1 \\
N_3^{2\rightarrow0}(z_L^0) a_2 \\
N_4^{3\rightarrow0}(z_L^0) a_3
end{bmatrix} + 
\begin{bmatrix}
\mu_0 \\
\mu_1 \\
\mu_2 \\
\mu_3
end{bmatrix}.
\]

With the \( \{\mu_i\}_{i=0}^{L-1} \) as follows:

\[
\mu_0 = \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \frac{1}{N_1^{0\rightarrow0}(m_1(u))} \frac{1}{N_2^{1\rightarrow0}(m_1(u))} \frac{1}{N_3^{2\rightarrow0}(m_1(u))} \frac{1}{N_4^{3\rightarrow0}(m_1(u))} \frac{1}{N_5^{4\rightarrow0}(m_1(u))} dt dv dx du dy dz
\]

\[
\mu_1 = \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \frac{1}{N_1^{0\rightarrow0}(m_1(u))} \frac{1}{N_2^{1\rightarrow0}(m_1(u))} \frac{1}{N_3^{2\rightarrow0}(m_1(u))} \frac{1}{N_4^{3\rightarrow0}(m_1(u))} \frac{1}{N_5^{4\rightarrow0}(m_1(u))} dt dv dx du dy dz
\]

\[
\mu_2 = \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \frac{1}{N_1^{0\rightarrow0}(m_1(u))} \frac{1}{N_2^{1\rightarrow0}(m_1(u))} \frac{1}{N_3^{2\rightarrow0}(m_1(u))} \frac{1}{N_4^{3\rightarrow0}(m_1(u))} \frac{1}{N_5^{4\rightarrow0}(m_1(u))} dt dv dx du dy dz
\]

\[
\mu_3 = \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \int_{z_L^0}^{z_L} \frac{1}{N_1^{0\rightarrow0}(m_1(u))} \frac{1}{N_2^{1\rightarrow0}(m_1(u))} \frac{1}{N_3^{2\rightarrow0}(m_1(u))} \frac{1}{N_4^{3\rightarrow0}(m_1(u))} \frac{1}{N_5^{4\rightarrow0}(m_1(u))} dt dv dx du dy dz
\]
And the closest terms to the diagonal \((j - i = 1)\) in \(\{\lambda_i\}_{i<j=0}^{L-1}\) are given by:

\[
\begin{align*}
\lambda_{01} &= \int_{z_0}^{z_1} \frac{1}{N_{2 \rightarrow 0}^1 (m_0(u))} \frac{dN_{1 \rightarrow 0}^0 (u)}{dx_0} du \\
\lambda_{12} &= \int_{z_1}^{z_2} \frac{1}{N_{3 \rightarrow 0}^2 (m_1(y))} \frac{dN_{2 \rightarrow 1}^1 (y)}{dx_1} dy \\
\lambda_{23} &= \int_{z_2}^{z_3} \frac{1}{N_{4 \rightarrow 0}^3 (m_2(v))} \frac{dN_{3 \rightarrow 2}^2 (v)}{dx_2} dv
\end{align*}
\]

The next closest terms \((j - i = 2)\) are:

\[
\begin{align*}
\lambda_{02} &= \int \int_{\substack{y \in [z_1, m_0(u)] \cap u \in [z_0, z_1]}} \frac{1}{N_{2 \rightarrow 0}^1 (m_0(u))} \frac{dN_{1 \rightarrow 0}^0 (u)}{dx_0} du \frac{1}{N_{3 \rightarrow 0}^2 (m_1(y))} \frac{dN_{2 \rightarrow 1}^1 (y)}{dx_1} dy du \\
\lambda_{13} &= \int \int_{\substack{y \in [z_1, m_1(y)] \cap y \in [z_{1, -} z_2]}} \frac{1}{N_{3 \rightarrow 0}^2 (m_1(y))} \frac{dN_{2 \rightarrow 1}^1 (y)}{dx_1} dy \frac{1}{N_{4 \rightarrow 0}^3 (m_2(v))} \frac{dN_{3 \rightarrow 2}^2 (v)}{dx_2} dv dy
\end{align*}
\]

Finally, the term further away from the diagonal \((j - i = 2)\) is:

\[
\lambda_{03} = \int \int \int_{\substack{y \in [z_2, m_1(y)] \cap y \in [z_1, m_0(u)] \cap u \in [z_{0, -} z_1]}} \frac{1}{N_{2 \rightarrow 0}^1 (m_0(u))} \frac{dN_{1 \rightarrow 0}^0 (u)}{dx_0} du \frac{1}{N_{3 \rightarrow 0}^2 (m_1(y))} \frac{dN_{2 \rightarrow 1}^1 (y)}{dx_1} dy \frac{1}{N_{4 \rightarrow 0}^3 (m_2(v))} \frac{dN_{3 \rightarrow 2}^2 (v)}{dx_2} dv dy du.
\]

Remember that for uniform distributions, the matching functions have closed form solutions. \(\{a_i\}_{i=0}^{L-1}\) obtain then very easily from the preceding, also in closed form, apart from \(z_i^1\). And wage functions can be calculated by inverse recursion beginning by \(w_L(.) = F(.)\) through the previous expressions.

# C Proofs

## C.1 Proof of Lemma 1

![Diagram showing skill levels and self-employment](image)

To prove that there is no self-employment in equilibrium requires to use the notations which are later used to solve the model in the paper (reminded above). The fact that self-employment does not arise when communication costs are sufficiently low is quite intuitive, as \(h\) governs the strength of the complementarities in this model: when \(h\) is low, it becomes valuable for agents with even very close skills to form teams and work together. However, the behavior with respect to \(\Delta\) is perhaps less intuitive.

Of course, there are some gains to forming teams of managers and workers: when a worker with low skill does not know a solution to a problem, he asks a manager with a higher skill which increases the probability that the problem is going to be solved, and therefore output in the
economy. However, there is also a cost: the managers spend some time away from production. For it to be profitable, the formation of a team of worker with skill \( x_0 \) together with managers of type \( l \) with type \( x_l \) it has to be that the first term is higher than the second. The gains are decomposed as follows:

\[
x_L g(x_0) dx_0 - \sum_{i=0}^{L} x_i g(x_i) = (x_L - x_0) g(x_0) dx_0 - \sum_{i=1}^{L} x_i g(x_i).
\]

Output together "Increased Productivity" of workers Output alone "Lost production time" of managers

Summing over the interval of workers, the total gains are therefore given by:

\[
\int_{z_l^0}^{z_{l+1}^l} (m_{L-1} \circ m_{L-2} \circ \ldots \circ m_0(x_0) - x_0) g(x_0) dx_0 - \sum_{i=1}^{L} \int_{z_l^i}^{z_{l+1}^l} x_i g(x_i) dx_i.
\]

It will be shown later that when \( \Delta \to 0 \), \( m_{L-1} \circ m_{L-2} \circ \ldots \circ m_0(.) \) converges to 1 in a \( \Delta^2 \) term. The first term therefore is greater than a linear term in \( \Delta \), while the second is smaller than a quadratic term:

\[
\sum_{i=1}^{L} \int_{z_l^i}^{z_{l+1}^l} x_i g(x_i) dx_i \leq \sum_{i=1}^{L} \int_{z_l^i}^{z_{l+1}^l} g(x_i) dx_i = 1 - G(z_{l+1}^l).
\]

Intuitively, when heterogeneity goes to 0, there really does not need a lot of managers to increase the productivity of all workers. The lost production time is becoming infinitesimally small relative to the increased gains from trade.

### C.2 Proof of Proposition 1

Managers with skill \( x_{l+1} \) maximize their revenue \( R_{l+1}(x_{l+1}) \):

\[
R_{l+1}(x_{l+1}) = \max_{x_l} \left\{ N^l_{1+l \to 0}(x_l) \left( w_{l+1}(x_{l+1}) - w_l(x_l) \right) \right\}
\]

with \( N^l_{1+l \to 0}(.) = N^l_{1+l \to l}(.) * N^l_{1 \to l}(.) * \ldots * N^l_{1 \to 0}(.) \).

There is complementarity between the skills of workers and team managers, since the joint surplus of managers of type \( x_{l+1} \) supervising workers of type \( x_l \) is supermodular \( x_{l+1} \) and \( x_l \) and given by:

\[
\frac{\partial^2 (N^l_{1+l \to 0}(x_l) w_{l+1}(x_{l+1}))}{\partial x_l \partial x_{l+1}} = w_{l+1}'(x_{l+1}) \frac{\partial N^l_{1+l \to 0}(x_l)}{\partial x_l} > 0
\]

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

\[
\int_{z_l^l}^{z_{l+1}^l} t_W(x_l) d_i = 0 \quad \Rightarrow \quad g(m_l(x_l)) dm_l(x_l) = h(1 - x_l) g(x_l) dx_l.
\]

\[
\Rightarrow \quad m_l'(x_l) g(m_l(x_l)) = h(1 - x_l) g(x_l).
\]
The positive sorting limiting condition writes, matching the less skilled and the more skilled:

\[ m_l(z^L) = z^L_{l+1} \]
\[ m_l(z^L_{l+1}) = z^L_{l+2} \]

Integrating the above differential equation for \( m_l(.) \) between \([z^L_l, z^L_{l+1}]\), and using these two limiting conditions gives the result.

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

C.3 Proof of Lemma 2

[TO BE ADDED]

C.4 Self-Employment or No Self-Employment? Closed Form Expression for Assumption 2 (Uniform Density)

In the case where \( h \) or \( \Delta \) are high enough, some agents remain self-employed. Self-employed agents have intermediary skills in equilibrium, because the gains from having a worker of skill \( x_0 \) and a manager of skill \( x_1 \) work together are given by what the two produce together minus what they would have produced by themselves, that is:

\[
\frac{x_1}{h(1-x_0)} - x_1 - \frac{x_0}{h(1-x_0)} = \frac{1}{h} \left( 1 - \frac{1-x_1}{1-x_0} \right) - x_1. 
\]

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

Figure 26: WITH SELF-EMPLOYMENT IN EQUILIBRIUM, ONE LAYER.

The notations for cutoffs are introduced on Figure 26. Now the matching function \( m_0(.) \) is defined on \([1 - \Delta, z_1^{**}]\). An important difference also is that in that case, will not be solved independently of agents’ choices.

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

\[ m_0(1 - \Delta) = z_1^* \quad m_0(z_1^{**}) = 1. \]
In addition, we now have two indifference equations between being a worker and self-employed with skills $z_1^*$, and being self-employed and a team manager with skills $z_1^*$:

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

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

$$m'_0(x_0)g(m_0(x_0)) = h(1 - x_0)g(x_0) \quad \Rightarrow \quad m'_0(x_0) = h(1 - x_0)$$

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

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

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

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

$$(1 - x_0)w'_0(x_0) + x_0w_0(x_0) = x_0m_0(x_0)$$

$$\Rightarrow \quad w_0(x_0) = \frac{1}{2} \left(2x_0 + hx_0^2 - 2hx_0z_1^{**} + h(z_1^{**})^2\right).$$

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

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

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

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

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

$$z_1^{**} < z_1^* \quad \Leftrightarrow \quad \frac{-h + h^2 + \sqrt{h^2 (3 + h^2 \Delta^2 - 2h(1 + \Delta))}}{h^2} < \frac{-2h + h^2 \Delta + \sqrt{h^2 (3 + h^2 \Delta^2 - 2h(1 + \Delta))}}{h^2}$$

$$\Leftrightarrow \quad 3 - h\Delta - h > 2\sqrt{3 + h^2 \Delta^2 - 2h(1 + \Delta)}$$

$$\Leftrightarrow \quad (1 + 2\Delta - 3\Delta^2)h^2 + 2(1 + \Delta)h - 3 > 0$$

$$\Leftrightarrow \quad h > \frac{1 + \Delta - 2\sqrt{1 + 2\Delta - 2\Delta^2}}{-1 - 2\Delta + 3\Delta^2} \quad \text{since} \ h > 0.$$ 

When the primitives of the model are such that this is verified, we are in the case where a non trivial measure of agents are self-employed. When it is not the case, then all agents either become managers or workers. The condition is illustrated graphically on Figure 8.
C.5 Alternative Proof of Proposition 8

In the main text, I provide a proof of Proposition 8 based on an envelope condition. Here is another proof, where the role of changes of variables perhaps appears more clearly.

**Proof.** The first result follows straightforwardly from the second, which is more general: set $l = L - 1$ and use $w_L(x_L) = x_L$ so that $w'_L(x_L) = 1$ for all $x_L \in [z_L^L, 1]$. To prove the second result, we need to integrate result in Proposition 7 by parts:

$$\int_{z_l^l}^{x_l} w_{l+1}(m_l(u)) \frac{dN_{l+1-0}^{l}(u)}{dx_l} du = \left[ w_{l+1}(m_l(u)) N_{l+1-0}^{l}(u) \right]_{z_l^l}^{x_l} - \int_{z_l^l}^{x_l} m_l'(u) w_{l+1}'(m_l(u)) N_{l+1-0}^{l}(u) du. $$

Using the previous integration by parts, this simplifies into:

$$\int_{z_l^l}^{x_l} m_l'(u) w_{l+1}'(m_l(u)) N_{l+1-0}^{l}(u) du = \int_{z_l^l}^{x_l} w_{l+1}'(v) N_{l+1-0}^{l+1}(v) dv. $$

Now, replacing $w_l(x_l)$ in $R_{l+1}(x_{l+1})$:

$$R_{l+1}(x_{l+1}) = (w_{l+1}(x_{l+1}) - w_l(x_l)) N_{l+1-0}^{l}(x_l) = w_{l+1}(x_{l+1}) N_{l+1-0}^{l}(x_l) - w_l(x_l) N_{l+1-0}^{l}(x_l)$$

$$R_{l+1}(x_{l+1}) = w_{l+1}(x_{l+1}) N_{l+1-0}^{l}(x_l) - \int_{z_l^l}^{x_l} w_{l+1}(m_l(u)) \frac{dN_{l+1-0}^{l}(u)}{dx_l} du - w_l(x_l) N_{l+1-0}^{l}(x_l)$$

$$= w_{l+1}(x_{l+1}) N_{l+1-0}^{l}(x_l) - \left[ w_{l+1}(m_l(u)) N_{l+1-0}^{l}(u) \right]_{z_l^l}^{x_l} + ...$$

$$... \int_{z_l^l}^{x_l} w_{l+1}'(v) N_{l+1-0}^{l+1}(v) dv - w_l(x_l) N_{l+1-0}^{l}(x_l).$$

Using the previous integration by parts, this simplifies into:

$$R_{l+1}(x_{l+1}) = R_{l+1}(z_{l+1}^l) + \int_{z_l^l}^{x_l} w_{l+1}'(u) N_{l+1-0}^{l+1}(u) du.$$
D  Linearly Increasing Distribution for Skills

Because the results in the main paper hold for all sufficiently smooth distribution of skills, and that in particular the Pareto coefficients do not depend on them, it is useful to illustrate the equilibrium of the model with another example than the uniform distribution used throughout the main paper.

I illustrate this result using a linearly increasing distribution for the frequency of skills. In other words, relatively high skilled agents are relatively more numerous than relatively less skilled agents, so that:

\[
G(z) = \frac{(z-(1-\Delta))^2}{\Delta^2} \mathbb{1}_{[1-\Delta, 1]}(z) + \mathbb{1}_{[1, +\infty]}(z).
\]

For example, one can see on Figure 27 some comparative statics with respect to heterogeneity.

Figure 27: Matching Function from Workers to Team Managers, \(h = 70\%\) - Comparative Statics on Heterogeneity

**Note:** This figure gives the matching function from workers to team managers, with the assumption that the distribution of skills in the population is given by \(G(z) = \frac{(z-(1-\Delta))^2}{\Delta^2} \mathbb{1}_{[1-\Delta, 1]}(z) + \mathbb{1}_{[1, +\infty]}(z)\).
E Comparative Statics

E.1 Matching Functions

Figure 28: Matching Function from Workers to Team Managers, $\Delta = 80\%$ - Comparative Statics on Helping Time

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

#### Table 2: Uniform Skill Distributions: Cutoffs for Different values of $\Delta$, $h$ and $L$

<table>
<thead>
<tr>
<th>$\Delta = 100%$</th>
<th>$\Delta = 70%$</th>
<th>$\Delta = 30%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L = 1$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z_1^L$</td>
<td>0.684778</td>
<td>0.837717</td>
</tr>
<tr>
<td>$z_2^L$</td>
<td>0.650620</td>
<td>0.828537</td>
</tr>
<tr>
<td>$z_3^L$</td>
<td>0.650000</td>
<td>0.828500</td>
</tr>
</tbody>
</table>

| $L = 2$          |                  |                  |
| $z_1^L$          | 0.900980         | 0.951238         | 0.991008        |
| $z_2^L$          | 0.900000         | 0.951000         | 0.991000        |
| $z_3^L$          | 0.900000         | 0.951000         | 0.991000        |

| $L = 3$          |                  |                  |
| $z_1^L$          | 0.900000         | 0.951000         | 0.991000        |
| $z_2^L$          | 0.900000         | 0.951000         | 0.991000        |
| $z_3^L$          | 0.900000         | 0.951000         | 0.991000        |

**Note:** In this table, the cutoffs are calculated with distribution of skills in the population given by $G(z) = \frac{z - (1-\Delta)}{\Delta}$ $(1 - \Delta, 1) (z) + \mathbb{I} (1, +\infty) (z)$. Precision up to 6 digits is displayed. Successively, the maximum number of levels is restricted to being $L = 1$, $L = 2$ or $L = 3$. The values for helping time are $h \in \{0.2, 0.7\}$, and the value for skill heterogeneity is $\Delta \in \{0.01, 0.1, 1\}$. 
Figure 29: Occupational choices for a Uniform Distribution of Skills, and $h = 75$

Note: This figure gives the occupational choices of agents when $h = 75\%$, and the distribution of skills in the population is given by $G(z) = \frac{2(z-\Delta)}{\Delta} \mathbb{1}(1-\Delta,1) (z) + \mathbb{1}(1, +\infty) (z)$. Note that the right-hand side of the graph is very similar to what is found on Figure 8. However, as shown on Figure 4, $h = 75\%$ belongs to the self-employment region for some values of $\Delta$, numerically for $\Delta < 5/9$. In this region, there is only one level of management. The two regimes are separated by the dotted line. A third case, where the self-employment region arises for intermediary levels of heterogeneity $\Delta$, is illustrated in Figure 31.
Figure 30: **Occupational choices for a Uniform Distribution of Skills, and \( h = 70\% \) in Percentile Terms**

Note: This figure gives the occupational choices of agents when \( h = 70\% \), and the distribution of skills in the population is given by \( G(z) = \frac{(1-\Delta)z}{\Delta} \mathbb{1}_{[1-\Delta, 1]}(z) + \mathbb{1}_{[1, +\infty]}(z) \), as a function of the skill level in terms of percentile. The occupational choice of agents is shown in the case of a uniform distribution of beliefs, as a function of the skill percentile instead of the skill level. One notes that the decreasing measure of managers of higher rank is not an artefact of the decrease in heterogeneity.
Figure 31: **Occupational choices for a Uniform Distribution of Skills, and \( h = 73.5\% \)**

Note: This figure gives the occupational choices of agents when \( h = 73.5\% \), and the distribution of skills in the population is given by \( G(z) = \frac{z(1-\Delta)}{\Delta} \mathbb{1}_{[1-\Delta,1]}(z) + \mathbb{1}_{[1,\infty]}(z) \). Note that self-employment arises in this limit case for intermediary levels of heterogeneity \( \Delta \in \left[ \frac{2}{161} (347 - 2\sqrt{409}) , \frac{2}{161} (347 + 2\sqrt{409}) \right] \approx [69.5\%, 87.8\%] \).
E.3 Span of Control Distributions

Figure 32: Distribution of Team Sizes, Log-Log Scale, $\Delta = 80\%$ - Comparative Statics on Helping Time

Note: This figure gives the theoretical distribution of team sizes as derived in Section 2, with the assumption that the distribution of skills in the population is given by $G(z) = \frac{z^{(1-\Delta)}}{\Delta} \mathbb{1}(1-\Delta, 1)(z) + \mathbb{1}(1, +\infty)(z)$, with a low level of skill heterogeneity $\Delta = 0.01$. 
Figure 33: DISTRIBUTION OF SPAN OF CONTROL DOWN ONE LEVEL OF HIERARCHICAL ORGANIZATION ("Teams"), LOG-LOG SCALE, COMPARATIVE STATICS ON HELPING TIME, HIGH HETEROGENEITY

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

$\Delta = 1$
E.4 Top Labor Incomes Distributions

Figure 34: Negative Relationship between Communication Costs and Top Labor Incomes Inequality

Note: The distribution of skills is given by the polynomial density and $\rho = 1$. Heterogeneity is assumed to be 40%. The baseline salary is normalized to be equal to $10000$. 

70
Figure 35: Comparative Statics of the Top Labor Income Distribution with Respect to $\rho$

Note: The distribution of skills is given by the polynomial density and $\rho = 1$. Heterogeneity is assumed to be 40%. The baseline salary is normalized to be equal to $10000$. 

71