Firm Structure and Occupational Sorting

Fernando Lopes, Gabriel Toledo

March 2026

Preliminary Version
[Latest Version]

The allocation of workers across occupations is a key aspect of the labor market, and it is shaped by firms' internal organization. We document a novel empirical pattern of sorting between workers and multi-worker firms using administrative data from Germany. Looking at job switchers, we find that workers that switch to higher paying firms, on average, find themselves with more subordinates on the firm wage distribution, but at a lower relative rank within the firm. These patterns imply negative assortative matching across layers but positive assortative matching across ranks. We build a tractable many-to-one assignment model with endogenous firm structure choice that allow us to tackle to question of how structure choice shapes sorting. Firms choose how many layers they want to organize their production in, so each job is indexed by a pair of firm productivity and its position in the firm hierarchy. The model rationalizes the evidence by generating negative assortative matching across layers but positive assortative matching across ranks and delivers a parsimonious way to compare jobs across firms. We provide a sufficient condition for positive sorting, namely that task importance at higher layers cannot rise too steeply relative to lower layers. The framework nests extensions with peer effects and search frictions while matching additional moments on wage dispersion.

Lopes: University of Wisconsin-Madison fllopes@wisc.edu. Toledo: Purdue University gstoledo@purdue.edu.

1. Introduction

The allocation of workers across firms is central to labor-market performance, but occupations themselves are shaped by firms' internal organization. Workers do not only sort across firms; they also sort into different positions within the firm hierarchy. Firms, in turn, decide how to organize production: how many hierarchical layers to create, which tasks workers in each layer should execute, and how much to specialize the workforce. These organizational choices have profound implications for the sorting patterns we observe in the labor market. This paper investigates how endogenous firm structure shapes the allocation of workers across occupations and firms, and how we should compare jobs across firms when hierarchies differ.

Using German administrative matched employer–employee data (LIAB, 2010–2017), we document a novel empirical pattern. We focus on job switchers and measure two aspects of workers' positions within firms: their *layer*—defined as the count of workers in the same firm earning less, a proxy for subordinates—and their *rank*—the layer normalized by firm size, representing relative position in the hierarchy. Measuring firm quality by leave-one-out average wages, we find that when workers move to higher-quality firms, they tend to have *more subordinates* (higher layer) but a *lower relative rank* in the destination hierarchy. This pattern is robust to controls for age, tenure, occupation, industry, location, firm size, and time fixed effects, and is confirmed through binned scatter plots and kernel regressions.

These findings reveal an interesting sorting pattern: negative assortative matching (NAM) across layers but positive assortative matching (PAM) across ranks. Because many models abstract from within-firm structure, a unidimensional approach would miss or misclassify these patterns. Our results suggest that a complete understanding of labor-market sorting requires distinguishing between a worker's absolute position (layer) and relative position (rank) within heterogeneous, multi-worker firms.

To rationalize these empirical patterns, we build a tractable many-to-one assignment model with endogenous firm structure. Firms (type z) choose how many layers m to organize and, conditional on m, assign workers of skill q_{nm} to each layer $n \in [0, m]$. Production aggregates the outputs of layers through an importance function $a(n, m)$: higher layers are more consequential ($\partial a/\partial n > 0$), but any given layer's importance falls as the hierarchy becomes taller ($\partial a/\partial m < 0$). This captures the idea that as firms stretch production across more layers, each layer's marginal contribution dilutes.

Workers choose jobs $(z, n, m(z))$ to maximize wages, which equal marginal product at the optimum. The model delivers two key results. First, firm hierarchy size $m(z)$ is increasing and convex in productivity z. Second, the equilibrium assignment is linear in productivity and layer share: $\mu(z, n, m) = \mu_0(n/m(z))z$. This implies *within a firm* (fix z), higher layers hire higher-skill workers (PAM across layers), but *across firms* (fix layer n), the skill assigned to that layer *declines* with firm productivity (NAM across firms for a given layer). The intuition is that as z rises, firms optimally expand hierarchy $m(z)$ rapidly; the dilution of layer importance (via $\partial a/\partial m < 0$) can outweigh the direct productivity effect, making a given layer less valuable at higher-z firms.

The model reconciles the empirical pattern through a simple comparative static. Holding skill fixed, moving from firm z to $z' > z$ requires $n'/n = (z'/z)^{\sigma}$ (more subordinates), but in terms of rank $\phi \equiv n/m$, we have $\phi/\phi' = z'/z$ (lower relative rank). Thus, workers who move to higher-quality firms should show more subordinates but lower ranks—exactly what we observe in the data.

This framework allows us to compare workers and occupations across firms even when hierarchies differ. The classic Becker insight—PAM with supermodular match value—must be modified because firm structure $m(z)$ endogenously shifts layer importance. We derive a sufficient condition for PAM across firms: the elasticity of importance to hierarchy size cannot be too large. In a simple parameterization $a(n, m) = a_0(n/m)^{\alpha}$, PAM obtains if $\alpha < 1/(1 + \sigma)$. Intuitively, higher layers cannot become "too much" more important than lower layers as hierarchies grow, or else firms optimally substitute skill upward inside the hierarchy, inducing NAM at any fixed layer.

Beyond sorting, the model ties structure to wage dispersion in a natural way. CEO-type pay scales with $z^{1+\sigma}$, and the full wage schedule permits decomposition of within- and between-firm log-wage dispersion. Under Pareto distributions for z and q, we can match observed facts such as convex right tails in log-wage distributions and target the ratio of within to between components. A further aggregation result shows that under Pareto-distributed z, aggregate output can be represented as $Y = AL$—a representative-firm formulation—so the assignment model conciliates with canonical macro aggregates while preserving rich micro sorting.

The model also guides the interpretation of job switches. A job mover to a higher-paying firm who ends up not only with more subordinates but also a higher relative rank has, through the lens of the model, most likely experienced a real skill upgrade. When only the layer rises (but rank falls), the change reflects compositional effects from the destination firm building taller pyramids rather than genuine skill advancement.

Related Literature. This paper contributes to several strands of literature. First, it relates to the extensive literature on sorting, inaugurated by Becker (1973)'s seminal contribution and surveyed in Chade, Eeckhout, and Smith (2017). The closest resemblance is to work on many-to-one assignment models (Kelso and Crawford 1982; Kremer 1993; Kremer and Maskin 1996; Eeckhout and Kircher 2018). We contribute by allowing the "many" part of many-to-one to be endogenous while incorporating heterogeneity inside the firm in a tractable manner.

This paper also relates to recent work investigating negative sorting in many-to-one assignment models (Ahlin 2017; Chade and Eeckhout 2018; Eeckhout 2018; Boerma, Tsyvinski, and Zimin 2021). Here, negative sorting takes a simpler form enabled by multidimensional job characteristics: fixing a specific layer, the skill of the worker employed at that layer decreases as firm productivity increases. This shows that negative sorting in a multimarginal environment can be simplified by expanding the type space.

The paper connects to the literature on knowledge hierarchies (Garicano 2000; Garicano and Rossi-Hansberg 2004, 2006), where workers with distinct knowledge levels band together to solve problems. Our production technology is a reduced form of a variation of this economy, but generates a novel sorting pattern: PAM across layers and NAM across firms. This is relevant because a unidimensional sorting model might not capture complex patterns in data that does not display monotone sorting in firm-layer pairs.

We also relate to recent work on endogenous firm structure (Deming 2017; Adenbaum 2022; Freund 2022). While these papers tackle how to assign tasks to workers with multidimensional skills, we assume skill is unidimensional and workers are fully specialized within layers. This simplification provides tractability when characterizing multidimensional sorting patterns that may differ in each dimension.

Finally, the problem of linking firm size and productivity with workforce composition has a long history (Manne 1965; Lucas 1978; Jovanovic 1982; Rosen 1982). Eeckhout and Kircher (2018) unite firm size determination and sorting assuming size and worker skill are essentially substitutes, with firms hiring a single type of worker. We depart by allowing heterogeneous workers inside the firm while assuming substitutability in efficiency units between top and bottom layers, in line with Rosen (1982).

The rest of this paper is organized as follows: Section 2 presents the empirical patterns; Section 3 lays out the environment and characterizes equilibrium; Section 4 derives the supermodularity condition and compares hierarchies; Section 5 presents aggregate predictions; Section 6 discusses extensions; Section 7 concludes.

2. Empirical Patterns

This section documents novel empirical patterns of worker sorting within and across firms using German administrative data. We show that when workers switch to higher-quality firms, they systematically end up with more subordinates (higher layer) but occupy a lower relative position in the destination hierarchy (lower rank). These findings motivate our theoretical framework by demonstrating that sorting patterns depend critically on how we measure workers' positions within firms.

2.1. Data and Sample Construction

We use the LIAB (Linked Employer-Employee Data from the IAB) for the period 2010–2017, which combines administrative employment records with establishment survey data (Dauth and Eppelsheimer 2020). The data provide complete biographies of the entire workforce for a representative sample of German establishments, including information on wages, occupation (KldB 2010 classification), establishment identifiers, industry, location, age, tenure, and other worker characteristics.

Our sample focuses on West Germany, private sector establishments, and full-time workers aged 20–65. We observe the complete composition of establishments in each year, which allows us to construct measures of workers' positions within the firm hierarchy. The panel structure enables us to track job switchers—workers who move from one establishment to another between consecutive years—and compare their positions before and after the switch.

2.2. Measuring Hierarchical Position

We construct two complementary measures of a worker's position within the firm hierarchy, both based on the wage distribution within each establishment-year:

Layer. For each worker i in firm j at time t, we define layer as the count of workers in the same establishment-year who earn strictly less than worker i. This provides a measure of how many "subordinates" a worker has in the firm wage distribution. Formally:

$$\text{Layer}_{ijt} = \#\{k \in \text{firm } j \text{ at } t : w_{kjt} < w_{ijt}\}$$

where w_{ijt} is the wage of worker i in firm j at time t and the count includes all workers in the same firm-year. A higher layer indicates a position with more subordinates in the firm wage hierarchy. This measure captures the absolute position of a worker in the firm's organizational structure.

We also construct a binned version that groups workers into categories of intervals of layer (e.g., bin from 0 to 5 subordinates, bin from 6 to 10 subordinates, etc.) to reduce sensitivity to small wage differences and measurement error.

Rank. We define rank as the layer normalized by firm size, providing a measure of relative position within the hierarchy. This captures how close to the top of the organizational pyramid a worker is located:

$$\text{Rank}_{ijt} = \frac{\text{Layer}_{ijt}}{\text{Size}_{jt}}$$

where Size_{jt} is the total number of workers in firm j at time t. Rank ranges from 0 (bottom of hierarchy) to approximately 1 (top of hierarchy).

The distinction between layers and ranks is crucial. A worker can have more subordinates (higher layer) when moving to a larger, better-paying firm, yet occupy a lower relative position (lower rank) if the destination firm has a substantially taller hierarchy. This dual measurement allows us to separately identify absolute and relative hierarchical positions.

2.3. Firm Quality Measure

Following the literature on job mobility (Jarosch, Oberfield, and Rossi-Hansberg 2021; Gregory 2020), we measure firm quality using the leave-one-out mean log wage:

$$z_{-ijt} = \frac{1}{N_{jt} - 1} \sum_{k \neq i, k \in j} \log(w_{kjt})$$

This measure captures average compensation at the establishment while excluding worker i's own wage, thus avoiding mechanical correlation. Higher z_{-ijt} indicates a higher-quality firm that, on average, pays better wages.

2.4. Empirical Specification

We focus on workers who switch establishments between consecutive years and estimate how changes in firm quality affect changes in hierarchical position. For worker i moving from firm j to firm j' between periods $t-1$ and t, we estimate:

$$\log(\phi_{ij't}) - \log(\phi_{ijt-1}) = \beta_0 + \beta_1 \left[\log(z_{-ij't}) - \log(z_{-ijt-1})\right] + \mathbf{X}'_{it}\boldsymbol{\gamma} + \varepsilon_{it} \quad (1)$$

where $\phi_{ijt} \in \{\text{Layer}_{ijt}, \text{Rank}_{ijt}, \text{LayerBin}_{ijt}\}$ represents one of our hierarchical position measures, and $\mathbf{X}_{it}$ includes controls for firm size, worker age, tenure, occupation fixed effects, industry fixed effects, location fixed effects, and year fixed effects.

The coefficient β_1 captures how a percentage change in firm quality translates into a percentage change in hierarchical position. Under positive assortative matching (PAM), we would expect $\beta_1 < 0$: an increase in firm quality should lead to a lower position in the hierarchy (lower layer or rank) keeping fixed the worker's skill. In other words, a better firm requires a more skilled worker to occupy the same hierarchical position, so a given worker would need to move down the hierarchy to match with the higher-quality firm. On the other hand, under negative assortative matching (NAM), we would expect $\beta_1 > 0$: an increase in firm quality would lead to a higher position in the hierarchy (higher layer or rank) for the same worker quality.

2.5. Main Results

Table 1 presents our main regression results. The findings reveal a striking dual pattern:

Layers (NAM): The coefficient on layers is positive but statistically insignificant (0.143, $p = 0.471$). If anything, moving to higher-quality firms is associated with

TABLE 1. Effect of Firm Quality Changes on Hierarchical Position (Job Switchers)

	Layer	**Layer Bin**	**Rank**
$\Delta \log(z)$	0.143	0.065	-0.915^{***}
	(0.207)	(0.075)	(0.092)
p-value	0.471	0.473	0.000
Controls	Yes	Yes	Yes
Observations	7,177	7,177	7,177

Notes Dependent variable is the log change in hierarchical position measure. All specifications include controls for firm size, worker age, tenure, occupation fixed effects, industry fixed effects, location fixed effects, and year fixed effects. Standard errors in parentheses. *** $p < 0.01$.

weakly higher layers (more subordinates). This *rejects* positive assortative matching across layers and provides weak evidence of negative assortative matching.

Ranks (PAM): The coefficient on ranks is negative and highly significant (−0.915, $p < 0.001$). A 10% increase in firm quality is associated with a 9.15% *decrease* in relative rank. This provides strong evidence of positive assortative matching when position is measured by rank: better workers (moving to better firms) occupy relatively lower positions within larger, more complex hierarchies.

These results suggest that the answer to whether we observe PAM or NAM depends critically on whether we measure position in absolute terms (layers) or relative terms (ranks).

Figures 1 and 2 present residualized binned scatter plots that visualize the relationship between firm quality changes and hierarchical position changes, after partialling out all controls. Figure 1 shows that the layer-firm quality relationship is relatively flat or weakly positive, consistent with the regression results. In contrast, Figure 2 displays a clear negative relationship: workers who move to higher-quality firms systematically experience declines in their relative rank.

Additionally, Figures 3 and 4 present nonparametric kernel regression estimates that confirm these patterns without imposing linearity. The kernel regressions corroborate the parametric results. Figure 3 shows that the layer-firm quality gradient is essentially flat or weakly positive, providing little evidence of PAM across layers. Figure 4 demonstrates a robust negative gradient, indicating strong PAM when position is measured by rank.

These empirical patterns reveal a fundamental tension in how we conceptual-

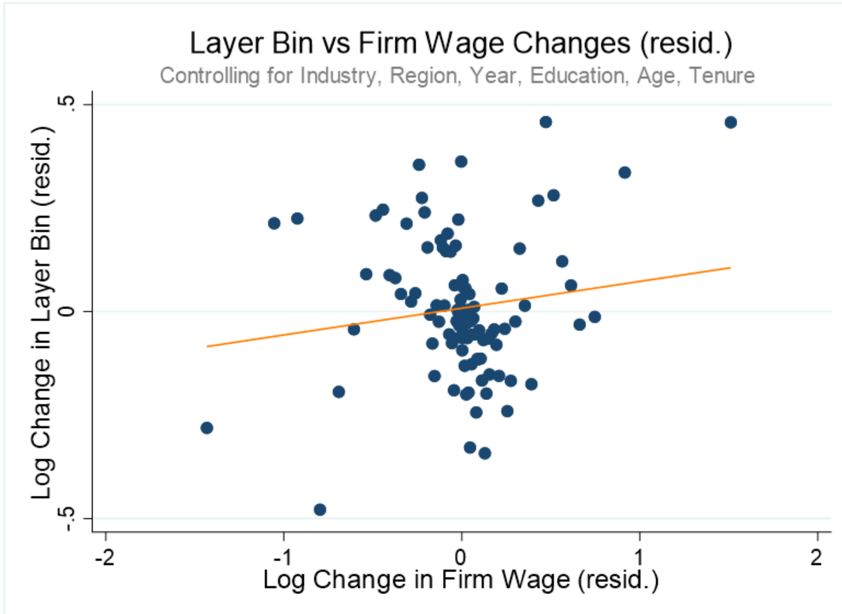

FIGURE 1. Change in Layer Bin vs. Change in Firm Quality (Residualized)

Notes: Binned scatter plot showing the relationship between changes in firm quality and changes in layer bin position after residualizing on controls. Each point represents a bin of observations. The relationship is weakly positive, indicating that moves to higher-quality firms are associated with slightly higher layers (more subordinates).

ize sorting in labor markets with heterogeneous firm structures. The conventional wisdom that better workers sort into better firms remains valid, but also crucially depend on how we measure workers' positions within rich organizational hierarchies.

When we measure position by the number of subordinates, moving to higher-quality firms does not necessarily mean moving to higher positions. A worker moving from a small, low-quality firm to a large, high-quality firm may gain subordinates in absolute terms without necessarily improving their relative position in the hierarchy. On the other hand, when we measure position by relative location in the hierarchy, we observe clear positive assortative matching. Workers who move to better firms occupy positions that are *further* from the top of the destination hierarchy. This suggests that firm quality and worker quality are complements when we account for the endogenous variation in firm size and structure.

These findings motivate our theoretical framework in Section 3, which explicitly models how endogenous firm structure choice generates these dual sorting patterns. The model will show that this empirical pattern arises naturally when: (i) more productive firms choose taller hierarchies, and (ii) the importance of

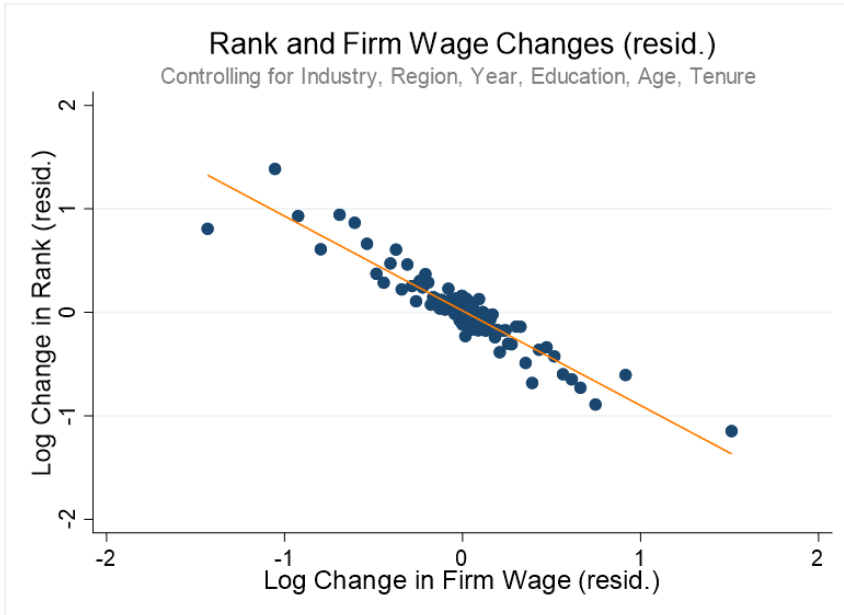

FIGURE 2. Change in Rank vs. Change in Firm Quality (Residualized)

Notes: Binned scatter plot showing the relationship between changes in firm quality and changes in rank after residualizing on controls. Each point represents a bin of observations. The relationship is strongly negative, indicating that moves to higher-quality firms are associated with substantially lower relative ranks within the destination hierarchy.

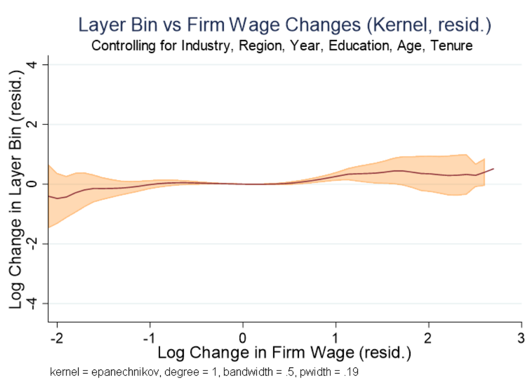

FIGURE 3. Layer Bin

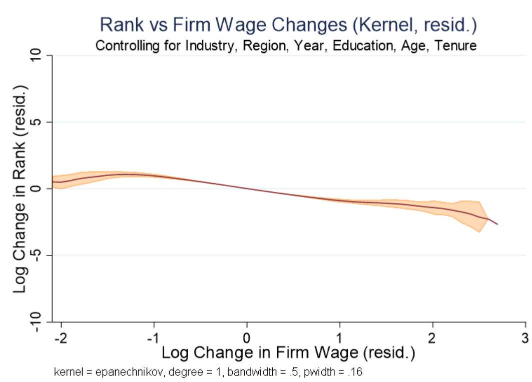

FIGURE 4. Rank

Notes: Kernel regression estimates of the relationship between changes in firm quality and changes in hierarchical position. Left panel shows layer bin relationship is nearly flat with slight positive slopes, rejecting strong PAM across layers. Right panel shows rank relationship has a consistently negative slope, with workers moving to higher-quality firms experiencing substantial declines in relative rank.

any given layer dilutes as the hierarchy expands. These mechanisms generate NAM across layers but PAM across ranks, reconciling the empirical evidence with economic intuition about complementarities between firm and worker quality.

3. Model

This is a static many-to-one assignment model of labor market sorting with structure choice. Firms choose how to organize production and which workers to assign to each layer, while workers choose where to work. This setup results in a multimarginal matching environment where the number of workers per firm is endogenous

3.1. Population

There are two types of risk-neutral agents in this economy, firms and workers. There is double-sided heterogeneity, from which the assignment problem arises. Firms have their mass normalized to 1 while workers have mass M. Firms differ in their productivity, which I denote by $z \in \mathcal{Z} := [\underline{z}, \overline{z}]$. Higher z means higher productivity. We can think of z as reflecting the amount of capital in each firm (in an environment where capital and labor are complements) or the quality of the project that a certain firm is engaged in. I assume that z is distributed according to cdf $G(z)$.

Workers differ in their skill, denoted by $q \in \mathcal{Q} := [\underline{q}, \overline{q}]$. This is a unidimensional measure of worker productivity and encompasses all the characteristics that are relevant for production: education, experience, innate ability. High skill workers have larger q, which is distributed according to cdf $H(q)$. Workers consume all their earnings, so I assume their utility is equal to the wages they earn.

3.2. Production Technology

Production is organized into different layers, where different individuals are assigned to each layer. Firms produce a homogeneous good and production requires the combination of a firm and at least one worker. The number of workers per firm is an endogenous object.

Layers are denoted by n and represent a worker's position in a hierarchy. This should not be interpreted as a job title, but simply as the number of subordinates that a worker has in a given hierarchy. The total number of layers in a firm is denoted by m, which is chosen by the firm before it enters the labor market. So the layers inside a firm form an interval $[0, m]$. Denote by q_{nm} the skill of a worker assigned to layer n in a firm with m total layers. In the baseline model, the product of each layer is independent from the others. So we can denote the output from layer n in a firm with productivity z by

$$f(z, q_{nm}) = zq_{nm}^{\sigma} \tag{2}$$

where $\sigma < 1$ is the return to worker skill.

An important assumption about the organization of firms is that they can only hire one worker for each layer. This imposes the restriction that workers cannot be side by side in a hierarchy, they must be ordered one way or the other. This is an unnatural restriction, as firms choose not only how big they want to be vertically, but horizontally as well (meaning how many workers to hire for each layer). As it is, this is a necessary assumption for tractability which is not without loss. The consequence of this assumption is that firm size and hierarchy both have the same meaning in the model, namely the variable m. As such, we cannot separate the effects of size and hierarchy on sorting. This is a drawback of this setup, but one that provides important gains in tractability that allow us to compare layers and workers across firms without having to know the size or composition of each firm.

Nevertheless, the model is not wholly inconsistent with firms choosing the size of each layer as well. From the firm's perspective, we can think of q_{nm}^{σ} as being the total effective labor units hired for layer n out of m. Then, when the firm chooses q_{nm}, it chooses the total amount of labor for that layer. The way that labor is going to be provided inside the firm is a product of the skill of all the workers in that layer. The actual number of workers in that layer then (not the effective labor supplied for that layer) will depend on whether workers are complements or substitutes when producing q_{nm}. I go into more detail on how it is possible to extend this model to allow for multiple workers in each layer in Section 7 and argue that the qualitative results I show below are consistent with a model of multi-worker layers if workers are perfect substitutes within the layer.

The production technology has the unique feature that a worker's participation in the firm's final output depends on their relative position inside the firm. Workers in different layers face different tasks and, thus, have different productivities. This is captured by function $a(n, m)$, which I call the importance function. A worker with skill q working in firm z at layer n out of m total layers will have their output scaled by $a(n, m)$ so that the output generated by layer n is given by

$a(n,m)f(z,q_{nm})$. The final output of the firm is obtained by combining linearly the output of each layer. That is, total output is given by

$$F(z,\{q_{nm}\}_{n=0}^{m}) = \int_0^m a(n,m)f(z,q_{nm})dn \tag{3}$$

The key assumption about this technology is that $a(n,m)$ is differentiable with $\frac{\partial a}{\partial n} > 0$ and $\frac{\partial a}{\partial m} < 0$. That is, as a worker moves up the hierarchy of a firm, their participation in total output grows, or their skill level becomes more important for the firm. However, as the total number of layers grows, the importance of a particular layer suffers.

This can be thought of as a reduced-form expression for a task assignment problem in a knowledge economy not too dissimilar from that of Garicano and Rossi-Hansberg (2006). Consider an environment in which problems arrive to a firm at a constant rate. These problems need to be solved in order for production to occur. Problems differ in their payoff, with some problems generating larger output if solved. Every worker has a probability of solving a problem proportional to their skill and independent of the problem's payoff. Then, due to supermodularity between worker skill and the payoff of solving a problem, firms are going to assign more important problems to higher layers. So the importance grows with n. However, as the firm adds more layers on top, existing layers only have access to lower-level problems, hence the importance decreases in m.

This knowledge economy differs from Garicano and Rossi-Hansberg (2006) in two relevant ways: there is no communication between layers and the value of problems varies. I made these assumptions (which shape the way $a(n,m)$ behaves) for two reasons. First, my focus is not on knowledge transmission inside the firm. Second, this creates a distinction between a worker's layer and their rank inside the firm, which disciplines the way we look at job transitions in Section 5.

3.3. Assignment

The key problem in the labor market is to solve for the assignment function, which dictates what worker type will be assigned to each particular layer in each firm. Then, the assignment function is a function $\mu : \mathcal{Z} \times \mathbb{R} \times \mathbb{R} \to \mathcal{Q}$ that takes as arguments the type of the firm, the layer and the total number of layers in that firm and gives us a worker type. This function is equivalent to a probability measure on the space $\mathcal{Z} \times \mathbb{R} \times \mathbb{R}$, that is, it tells us what is the density of workers of type $\mu(z, n, m)$ working at a job described by the triple (z, n, m), with the condition that the marginal distribution with respect to $\mathcal{Z}$ is G.

3.4. Firm's Problem

The problem of the firm consists of two stages. First, the firm chooses into how many layers it wishes to subdivide its production process. Second, given the number of layers and the wage that each type of worker is capable of earning in the labor market, the firm chooses the skill of each worker to hire for each layer.

We start by solving the firm's problem from the labor market, taking m as given. Then, in this stage, the problem of the firm can be written as

$$\pi(z, m) = \max_{\{q_{nm}\}_{n=0}^{m}} \; z \int_0^m a(n, m) q_{nm}^{\sigma} dn - \int_0^m w(q_{nm}) dn \tag{4}$$

The firm is choosing workers from the set Ω to assign to each layer. The firm is free to choose any subset of Ω with cardinality equal to m. In practice, each firm will hire an interval of Ω with size m. In turn, the firm has to pay each worker a wage that depends on their skill level. What determines this wage is the interplay between the worker's marginal product in a specific firm-layer pair (z, n) and also the best wage that this worker could get if they worked in another firm-layer. I will go into more detail on this when introducing the worker's problem.

Then, the FOC of the problem above is given by

$$w'(q_{nm}) = \sigma z a(n, m) q_{nm}^{\sigma-1} \tag{5}$$

The above condition relates the cost and the benefit of hiring a worker that's marginally more skilled for layer n while leaving the skill of all other workers fixed. The cost is the marginal increase in the total wage bill from hiring a better worker for that position. The benefit is the marginal product of layer n our of m.

The condition above plus the worker's optimality allows us to characterize $\mu(z, n, m)$ and $w(q_{nm})$. Armed with these objects, we can now move back to the problem of structure choice. The firm chooses m to solve

$$\max_{m} \; \pi(z, m) - c(m) \tag{6}$$

where $c(m)$ is the cost of setting up a larger or more vertical firm. The assumption on $c(m)$ is that it is differentiable and convex, so the maximization problem above is well-defined and has a unique solution. The idea behind there being an extra cost associated with m comes from the larger fixed costs of operating a large firm. More employees necessitate more office space or larger factories in order to work properly. Additionally, one can think of more verticalized organizations as being more prone to communication or coordination errors between layers, which would also be convex. I assume that these costs are independent of the skill of workers in each layer and depend only on the total number of layers. The solution to the problem above will induce a function $m(z)$, which connects a firm's productivity to its organizational structure.

3.5. Worker's Problem

Workers simply choose which job, indexed by (z, n, m), they wish to work in, provided that firm z is willing to hire that worker for layer n. Then, since workers are risk-neutral, worker q will accept a job (z, n, m) provided that

$$\begin{aligned} w(\mu(z,n,m)) \geq w(\mu(z',n',m')), \quad \forall (z',n',m') \text{ s.t. } n' \leq m' = m(z') \\ \text{and } \mu(z,n,m) = \mu(z',n',m') = q \end{aligned} \tag{7}$$

The condition above states the worker will take the job that offers the highest wage among all the jobs that are available to them. For a job to be available to a worker, we need three conditions to hold: first, the layer that the job is in must within the firm, hence $n' \leq m'$; second, m' must be a solution for firm z', hence $m' = m(z')$; finally, the worker's skill level must solve the firm's FOC, hence $\mu(z', n', m') = q$.

4. Equilibrium

With these definitions in hand, we are now ready to define a competitive equilibrium for this economy.

DEFINITION 1. *A competitive equilibrium is a set of functions $\mu : \mathbb{Z} \times \mathbb{R} \times \mathbb{R} \to \mathbb{Q}$, $w : \mathbb{Q} \to \mathbb{R}$, and $m : \mathbb{Z} \to \mathbb{R}$ such that (i) μ solves (4); (ii) m solves (6); (iii) w satisfies (7); and (iv) the allocation is feasible:*

$$\int_{\underline{z}}^{z} dG(z') = M \int_{\underline{q}}^{\mu(z,n,m(z))} dH(q'), \quad \forall (z,n) \in \mathbb{Z} \times [0, m(z)] \tag{8}$$

The feasibility constraint states that labor demand and labor supply have to be equal in equilibrium for every job denoted by (z, n). Note that m is subsumed from the definition of a job in equilibrium, as $m = m(z)$ for all jobs that are offered in equilibrium.

4.1. Allocation

Now we are ready to characterize an allocation. For the remainder of the text, I will assume functional forms for $a(n, m)$ and $c(m)$. However, this is not necessary to solve the model. In the Appendix, I provide a numerical solution method for a broader class of primitives. The method consists of performing a change of variables to write the model in terms of a worker's rank inside the firm, that is, $\frac{n}{m}$. Once this is done, firm optimality and feasibility can be used to give us a system of differential equations in μ and w. Solving this system can be straightforward depending on the distributional assumptions over G and H.

To get analytical results, I will now assume that $a(n, m) = \frac{a_0 n}{m}$. This is not a necessary assumption, the model can be characterized in closed form for a wider range of importance functions, this is purely for tractability. The assumption that will be necessary for the results that are to come is that the dependence between n and m comes purely from their ratio $\frac{n}{m}$. Solving the model without this assumption is still possible, but the comparison between workers in different firms that I present in the next Section would be compromised.

To characterize the allocation, we will focus on equilibria that are linear in z, that is, $\mu(z, n, m) = \mu(n, m)z$. Plugging this guess into the firm's first order condition and integrating over q_{nm} gives

$$w(q_{nm}) = \frac{\sigma}{1+\sigma}\frac{a(n, m)}{\mu(n, m)}q_{nm}^{1+\sigma}$$

Then, the profit $\pi(z, m)$ can be rewritten as

$$\pi(z, n) = \frac{z^{1+\sigma}}{1+\sigma}\int_0^m a(n, m)\mu(n, m)^{\sigma}dn$$

Now we make the assumption that $c(m) = \frac{\kappa m^2}{2}$. Again, this is purely for tractability, and the only key assumption on $c(\cdot)$ is that it is convex enough for the problem below to be strictly quasi-concave in m:

$$\max_{m} \frac{z^{1+\sigma}}{1+\sigma}\int_0^m a(n,m)\mu(n,m)^{\sigma}dn - \frac{\kappa m^2}{2}$$

From this, we are able to prove the following Lemma:

LEMMA 1. *Suppose* $a(n,m) = \frac{a_0 n}{m}$ *and* $c(m) = \frac{\kappa m^2}{2}$*. Then* $\mu(n,m) = \frac{\mu_0 n}{m}$*, for some constant* $\mu_0 > 0$*, is an equilibrium. Additionally,* $m(z) = \beta z^{1+\sigma}$*, for some constant* $\beta > 0$.

PROOF. See Appendix for proofs. □

Knowing the aforementioned equilibrium objects, we can go back to the labor market equilibrium to find the assignment function over (z, n). In an abuse of notation, I will call $\mu(z, n) = \mu(z, n, m(z))$ the quality of worker employed at job (z, n), since there is a one-to-one mapping between m and z. Then, we find that

$$\mu(z,n) = \frac{\mu_0 n}{\beta z^{\sigma}}$$

The most noteworthy aspect of the above assignment function is that it contains both PAM and NAM characteristics. Fixing a firm type z, we find that the allocation across occupations is PAM. That is, inside a firm, because higher layers have larger importance, the firm will hire more skilled workers.

However, if we fix the layer n, we see that the allocation across firms is NAM. That means that the skill of a worker with n subordinates in a firm with productivity z is higher than the skill of a worker with the same number of subordinates in a firm with productivity $z' > z$. This result might seem initially puzzling, but remember that the hierarchy size grows exponentially with z, according to Lemma 3.2. Consequently, as we increase z, there are two effects. First, higher productivity is a force that attracts more skilled workers. Second, higher hierarchy size reduces the importance of that layer, which attracts less skilled workers. Due to the convexity in $m(z)$, the importance of layer n is falling faster than the gains in productivity due to z, which causes job (z, n) to be ranked higher than job (z', n). On the firm side, because layer n has lower importance for firm z', that firm is willing to substitute the skill of a worker in that layer for the skill of workers in higher layers. So the firm is willing to hire a worse worker for layer n if it means that it can get hire better workers for more important layers. Hence, $\mu(z, n) > \mu(z', n)$.

This result is a direct product of the way we modeled the importance function $a(n, m)$. If instead we had assumed $\frac{\partial a}{\partial m} > 0$, then we would have PAM in both dimensions. Although this result may seem to contradict our intuition, I will discuss further in Section 5 how to use this richer sorting among pairs (z, n) to discipline the way we look at job transitions in the data.

4.2. Supermodularity

Sorting patterns are usually a result of the complementarities between the two sides of a market, as dictated by the production technology. From Becker (1973)'s seminal paper to more recent investigations into sorting with search frictions, the discussion is usually framed around a cutoff for the level of supermodularity that is sufficient to induce PAM. In the neoclassical, frictionless model (Becker 1973; Rosen 1974), if the match value is supermodular, then we have positive sorting. When search frictions are present, because there may be imperfect information about prices and types (Shimer and Smith 2000) or because there is a probability that no match is formed (Eeckhout and Kircher 2010), the supermodularity requirement is stronger.

In this particular case, the technology that relates z to q, the function f, is supermodular and there are no trading frictions in the environment. So shouldn't we expect the allocation to be PAM? The reason our initial intuition fails is because now m is an endogenous object that varies with firm productivity. Hence, the choice of firm structure affects directly labor market sorting.

Since m depends on z, we need to derive the necessary condition for the production function to be supermodular in this particular context. Using our definition for the production function, in equilibrium, for a particular layer with worker q:

$$\frac{\partial^2 F(z, \mathbf{q})}{\partial z \partial q} = \frac{\partial^2 \left(a(n, m) f(z, q)\right)}{\partial z \partial q} = \frac{\partial a(n, m)}{\partial m} \frac{\partial f(z, q)}{\partial z} m'(z) + a(n, m) \frac{\partial^2 f(z, q)}{\partial z \partial q}$$

where $\mathbf{q}$ is the set of all workers employed by this firm in equilibrium. For F to be supermodular, we need $\frac{\partial^2 F}{\partial z \partial q} > 0$. Then, rearranging the expression above, this inequality becomes

$$\frac{f_{zq}(z,q)f(z,q)}{f_z(z,q)f_q(z,q)} > \frac{|\varepsilon_{a,m}||\varepsilon_{m,z}|}{\varepsilon_{f,z}} \tag{9}$$

where f_x represents the derivative of f with respect to variable x and $\varepsilon_{r,s}$ is the elasticity of function r with respect to variable s as defined in the standard way:

$$\varepsilon_{r,s} = \frac{r'(s)s}{r(s)}$$

Equation (9) shows us exactly how strong the supermodularity between z and q have to be in order for the productivity gains to dominate the loss in importance from increased firm size. On the left-hand side, we have the cross-derivative of $\log f$. On the right-hand side, we have a measure of how fast the importance function changes as we increase z. First, m changes at a rate $|\varepsilon_{m,z}|$ as z increases. Second, this feeds into the importance function, which falls at a rate $|\varepsilon_{a,m}|$ in response. This is corrected by the larger output generated per unit of worker skill, which grows at rate $\varepsilon_{f,z}$.

Another aspect of this equation that is noteworthy is that it is possible that the right-hand side exceeds one, which would require f to satisfy a condition that is potentially stronger than log-supermodularity. As discussed above, log-supermodularity is usually a strong condition that is sufficient for PAM in applications with search frictions. This suggests that substitutability between layers induces a submodularity effect that is potentially stronger than the substitutability of the matching function in a search context. In the latter case, in a directed search context, a high skill worker can search for jobs in low productivity firms and obtain a match with higher probability. If the two sides of the market are sufficiently substitutes, then the increased probability of matching will be high enough to counteract the lower match value. The logic in the former case is entirely through the match value: a high skill worker accepts a job at a firm with lower productivity because the effective productivity of that job, $a(n, m(z))z$, is higher.

Below, I show a condition on model primitives for a restricted class of importance functions that highlights the main mechanisms behind how supermodularity works in this particular model.

PROPOSITION 1. *Suppose* $a(n,m) = a_0\left(\frac{n}{m}\right)^{\alpha}$*. Then, the allocation exhibits PAM if*

$$\alpha < \frac{1}{1+\sigma}.$$

When a worker considers accepting a job at a firm z' instead of $z < z'$, three forces act over the match value. First, it increases linearly with z. Second, m increases at a rate $1 + \sigma$. Hence, the expression on the right-hand side of the condition above reflects the adjusted growth rate in match value that comes from the increase in firm productivity. Finally, the productivity of the job that the worker is going to execute falls at a rate α, equal to the elasticity of $a(n,m)$. The condition above shows how parameters α and σ must interact in order to have the productivity growth be higher than the importance decrease. In particular, $a(n,m)$ needs to be "concave enough" in the sense that the importance of high level layers is not too high compared to the importance of low level layers.

5. Comparing Hierarchies

One problem that arises in empirical work is how to compare different occupations in firms with different productivity. Since the organizational structure is firm-specific, the same occupation label may involve the execution of a different set of tasks and may include a larger or smaller managerial content (influencing how much that worker interacts with peers). Having a disciplined way of comparing workers across firms is relevant to the understanding of job transitions as well as measuring the extent of sorting in the labor market.
Now I tackle the problem of comparing workers across firms in a way that allows us to make predictions about job transitions and how that relates to sorting. With the allocation in hand, we can now use the assignment function as a basis for comparison between firms of different productivity. The way to do this is to look for the set of firms that a worker can be assigned to in equilibrium. Since the assignment function takes both z and n as an argument, this would give us the relationship between firm productivity and layer that is necessary to keep worker quality fixed.

Suppose we wish to compare layers across firms z and z'. Take a worker at some layer n in firm z. If this worker were to change jobs to work in firm z', the layer n' to which this worker would be assigned must satisfy

$$\mu(z, n) = \mu(z', n') \Leftrightarrow \frac{n'}{n} = \left(\frac{z'}{z}\right)^{\sigma} \tag{10}$$

It has to be optimal for firm z' to hire this worker at layer n', hence we are using the assignment function (which, remember, satisfies the firm FOC) evaluated at (z', n'). Since the skill of the worker doesn't change when moving from one firm to the other, we need $\mu(z', n') = \mu(z, n)$.

The equality above is another way to express NAM across firms: as a worker moves to a more productive firm, we expect that worker to have *more*, not *less* subordinates in the higher firm. Perhaps more interestingly, this is true even though the average skill of workers inside a firm, given by

$$\overline{q}(z) := \int_0^m \mu(z, n) dn \propto z^{2+\sigma},$$

is increasing in productivity. Then, in this thought experiment, this worker moves to a firm in which the average worker is more productive than they most likely are, but they still manage to have more subordinates than before. The reason for this, as explained before, is the relative importance of layer n' (and of all layers below n') in the new firm is low enough that the firm is willing to hire a worker that's possibly worse than its average employee.

Now we repeat this exercise, but using as a basis for comparison the rank of the worker in each firm. Define $\varphi := \frac{n}{m}$. Then, the same worker is hired in two different firms as long as

$$\frac{\varphi}{\varphi'} = \frac{z'}{z} \tag{11}$$

Then, we see that in ranks, the allocation is PAM across firm productivity. This means that, if a worker were to move from firm z to z', even though they would work in a higher layer and have more subordinates at z', their relative rank would decrease in the new firm.

This result is useful for two reasons. The first one is that it allows us to think of the correct notion of what positive sorting means in an environment with differentiated multi-worker firms. As I have shown above, looking at the way workers sort across n or φ delivers different predictions on the type of sorting we expect to see. If we were to use a model of knowledge hierarchies in the vein of Garicano and Rossi-Hansberg (2006), conflating n and φ would come without loss and we would expect to see PAM in both cases. Therefore, in a dataset that exhibits NAM across the n dimension, we could inadvertently reject the model or draw wrong conclusions about the skill ordering of workers in different firms. The main takeaway here is that the context in which that worker is inserted matters for the notion of sorting. Then, if we were to use a model like Garicano and Rossi-Hansberg (2006) to calibrate parameters in the data, a measure of φ would more properly align with the predictions of the knowledge hierarchy model. This measure could be the worker's occupation label if, for example, we have a way of credibly ordering occupations by their managerial content. Then, we could use the occupation label as a proxy for someone's rank in the firm, i.e., a lead project manager has a higher φ than a project manager.

The second reason is that it recontextualizes job transitions in a way that gives a fuller description of the way the job ladder operates. Suppose we can observe both firm productivity and the number of subordinates for each particular position at the firm level (that is, the pair (z, n) is observable). Then, when we see a worker move from a (z, n) to (z', n'), with $z' > z$ and $n' > n$, cursory intuition would tell us that this worker is moving up the job ladder. However, in light of this model, this transition is expected for this worker and doesn't represent any move up. In fact, if we don't observe worker skill directly, concluding that a worker employed in (z', n') is better than a worker in (z, n) would in fact not be merited in the context of this model. The only way we would be able to say without a doubt that the worker moved up from one firm to the other (or if one worker is more skilled than the other) is if their *rank* also improved, that is, if $z' > z$ and $\varphi' > \varphi$. Therefore, this model gives us a way of comparing workers across firms and understanding whether a move represents an improvement for the worker in a way that integrates the way sorting operates in both the productivity and occupation dimensions.

Since the context in which workers are inserted (meaning, their skill relative to that of their co-workers) is relevant for understanding sorting, a natural question that arises is what role peer effects play in this comparison. One could think that if a worker moves from a firm with less skilled workers on average to a firm with highly skilled workers, then the peer effects of those workers would increase their marginal product enough to merit a higher rank in the more productive firm. However, as I show below, this is not true.

Consider a version of this model with peer effects obeying the following production function:

$$F\left(z, \{q_{nm}\}_{n=0}^{m}\right) = z\left(\int_0^m q_{nm} dn\right)^{\gamma} \int_0^m a(n, m) q_{nm}^{\sigma} dn \tag{12}$$

with $\gamma > 0$. This functional form, when taken in logs, results in a similar model to that of Manski (1993)'s reflection problem, where the firm represents the worker's reference group and γ is the intensity of the peer effects. Then, the proposition below follows.

PROPOSITION 2. *Suppose the production function exhibits peer effects in the form of (12). Then, in equilibrium, for* $\mu(z, n, m) = \mu(z', n', m')$*,*

$$\frac{\varphi}{\varphi'} = \left(\frac{z'}{z}\right)^{b(\gamma)}$$

with $b(\gamma)$ *increasing in* γ*.*

The proposition above states that higher peer effects exacerbate the "PAM-ness" we expect to see in ranks. The reason is that, despite the benefit of working with more skilled peers, the boost that a worker gets from those peers is proportional to their own skill level. Then, when moving to a firm with a higher average skill worker, the decrease in their own relative skill level is more intense, since their new co-workers enjoy a higher boost from peer effects. This is easy to see if we write

$$f(z, q_{nm}) = z\overline{q}(z) q_{nm}^{\sigma}$$

Then, we can define $\tilde{z} := z\overline{q}(z)$ and redefine jobs in $(\tilde{z}, \varphi)$ space. With peer effects, the move from a firm z to $z' > z$ represents a more-than-proportional move in $\tilde{z}$ jobs, which means that the worker must move further down the ranks to be hired by firm $\tilde{z}'$. Bonhomme (2021) cites the incorporation of complementarities into models of multi-worker firms as an interesting avenue for future research and this could be a simple, yet effective way of doing this in a model of structure choice.

6. Aggregate Implications

In this Section, we are interested in aggregate cross-sectional properties of this model. For this, we are required to make some distributional assumptions on z and q.

6.1. Power Laws

I start by showing what distributions we expect to see in the data for some equilibrium objects assuming z and q follow power law distributions. For the purposes of this application, a power law is a probability distribution for some random variable X satisfying

$$\Pr(X > x) = S(x)x^{1-\eta}, \quad \eta > 1$$

where $S : \mathbb{R}_+ \to \mathbb{R}_+$ is a slowly varying function. A slowly varying function is any measurable function satisfying

$$\lim_{x\to\infty} \frac{S(ax)}{S(x)} = 1$$

for all $a > 0$. The Pareto distribution is, of course, a particular case of a power law distribution. With this definition, it is possbile to show the proposition below.

PROPOSITION 3. *If z follows a power law, then $m(z)$ also follows a power law. In particular, if z is Pareto with scale parameter z_{min} and shape parameter $\eta > 1 + \sigma$, then $m(z)$ is Pareto with scale $\beta z_{min}^{1+\sigma}$ and shape $\frac{\eta}{1+\sigma}$.*

The proposition above states that if we can observe the distribution of m in the data, under the assumption that it follows a Pareto distribution, we would be able to recover the distribution of z, $G(\cdot)$, provided we have a calibrated value for σ.

Now, we move on to q. One important and robust fact that shows up in wage data is that the right tail of the distribution of log wages is convex. See Figure 5, for example, the distribution of log wages for São Paulo, the most populous state in Brazil, for the year 2017. From around the 20th percentile onward, the distribution is clearly convex.

This model is able to match this empirical fact. Again, I assume q follows a power law and show the following proposition.

PROPOSITION 4. *If q follows a power law, then $w(q)$ also follows a power law. In particular, if q is Pareto with scale parameter q_{min} and shape parameter $\eta_q > 1 + \sigma$, then $w(q)$ is Pareto with scale $\frac{\sigma a_0}{\mu_0} q_{min}^{1+\sigma}$ and shape $\frac{\eta_q}{1+\sigma}$. Furthermore, the distribution of log*

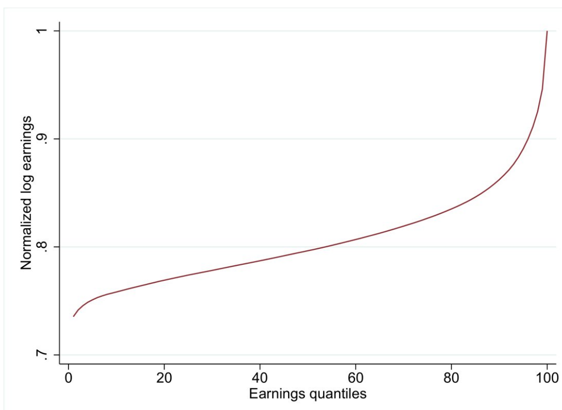

FIGURE 5. Distributions of log wages (normalized to be between 0 and 1) for the state of São Paulo, Brazil, in the year 2017. Source: RAIS.

wages is convex if, and only if, the percentile function $Q(p)$, as defined implicitly by

$$\Pr_W\left(W \leq w\left(Q(p)\right)\right) = p,$$

is log-convex, where $\Pr_W$ *is the induced probability measure of the wage distribution.*

The proposition above has two parts. The first one is similar to the previous proposition about z, that is, we can recover $H(\cdot)$ as long as we assume that wages follow a Pareto distribution and we have a credible calibration for σ. The second part relates to the convexity of log wages. It imposes a restriction on the percentile function $Q(p)$ that the skill distribution must satisfy in order for model results to be compatible with the data. Perhaps more importantly, it disciplines the types of distributions that are reasonable for q if we are working with wage data that exhibits a convex distribution for log wages.

6.2. Wage Dispersion

This model can also be used to target wage dispersion moments in the data. Using the functional form we found for $\mu(z, n, m)$, we can plug it into the wage equation to write wages as a function of jobs (z, n):

$$w(z, n) = \frac{\sigma}{1+\sigma}\frac{a_0 \mu_0^{\sigma} n^{1+\sigma}}{m(z)^{1+\sigma}} z^{1+\sigma}$$

With the expression above, we are able to compute firm-specific measures of wage dispersion. For example, since every firm has $n = 0$ as its lowest layer, a growing CEO wage is indicative of increasing wage dispersion among firms, in a max-min sense. The CEO wage is given by evaluating the equation above at $n = m$:

$$w(z, m) = \frac{\sigma}{1+\sigma} a_0 \mu_0^{\sigma} z^{1+\sigma}$$

which is obviously increasing in firm productivity.

However, a more sophisticated measure of wage dispersion is in order. With $w(z, n)$, we can compute within- and between-firm log wage dispersion. Within-firm log wage dispersion is given by

$$\Delta^{within} = \mathbb{E}\left[Var\left(\log w|z\right)\right]$$

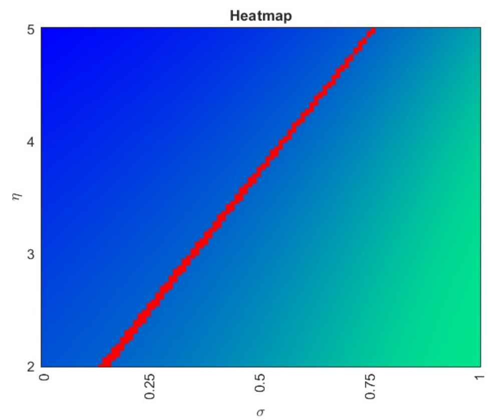

FIGURE 6. Heatmap of all the values of $\frac{\Delta^{within}}{\Delta^{between}}$ for a range of combinations of (σ, η). Green cells are higher values and blue cells are lower values. Red cells indicate that $\frac{\Delta^{within}}{\Delta^{between}} \approx 1.5$ within a tolerance level of 10^{-3}.

That is, the mean of firm-specific variance of loq wages across all firms. Between-firm log wage dispersion is given by

$$\Delta^{between} = Var\left(\mathbb{E}\left[\log w|z\right]\right)$$

That is, the total variance in firm-specific average log wages across all firms.

As an illustration, I am going to target the ratio between within- and between-firm log wage dispersion. Song et al. (2019) find a ratio of $\frac{\Delta^{within}}{\Delta^{between}}$ close to 1.5 for the US economy for the year 2013. Maintaining the assumption that z follows a Pareto distribution and using their estimate as a target, I show all the consistent values for parameters in (σ, η)-space. Figure 6 plots the ratio $\frac{\Delta^{within}}{\Delta^{between}}$ for a wide range of combinations of σ and η, given a parameterization of a_0, κ and z_{min}. As expected, higher dispersion in firm productivity (higher η) raises $\Delta^{between}$ relative to Δ^{within}. On the other hand, a higher return to worker skill (higher σ) makes wages more convex in q. Since there is PAM inside the firm, this raises Δ^{within} relative to $\Delta^{between}$.

The red line in the figure is of particular interest. It marks all the combinations of (σ, η) that generate a ratio $\frac{\Delta^{within}}{\Delta^{between}}$ close enough to 1.5, within a tolerance level. This exercise shows that if we are able to credibly calibrate η, it is also possible to obtain a calibration for the return to worker skill σ by targeting the relative sizes of within- and between-firm log wage dispersion.

6.3. Aggregation

Finally, I present a result that reconciles the baseline model with canonical production functions widely used in the firm dynamics literature. In particular, we are interested in a representation of aggregate output that is consistent with neoclassical models with representative firms and labor. Define

$$Y = \int_{\mathcal{Z}} F\left(z, \{\mu(z,n)\}_{n=0}^{m(z)}\right) dG(z)$$

as aggregate output. Note that Y is the equilibrium output, taken at the equilibrium allocation of workers to firms. Then, it is possible to show that

PROPOSITION 5. *Suppose z follows a Pareto distribution with minimum value z_{min}. Then $Y = AL$, where*

$$L = \int_{z_{min}} \int_0^{m(z)} \mu(z,n,m) dndG(z)$$

is the total effective units of labor in the economy and A is a constant.

The proposition above states that it is possible the represent the output resulting from this particular assignment with multi-worker firms as the output resulting from the profit maximization problem of a representative firm and a representative worker. Note that, since we have positive profits in the model, we can further match the results of the assignment model by assuming convex adjustment costs for the representative firm. Then, the problem of the representative firm could be written as

$$\tilde{\pi} = \max_L \; AL - \theta \frac{L^2}{2}$$

The profit resulting from this problem is $\tilde{\pi} = \frac{A^2}{2\theta}$. The adjustment cost parameter that is consistent with the level of profits in the assignment model is

$$\theta = \frac{A^2}{2}\left[\int_{z_{min}} \left(\pi(z, m(z)) - \frac{\kappa m(z)^2}{2}\right) dG(z)\right]^{-1}$$

Therefore, with the results of the assignment model in hand, we can fully parameterize a model with a representative firm that results in the same aggregate outcomes.

7. Extensions

In this Section, I present two possible extensions of this model and briefly discuss how they could be implemented. None of these extensions are fully characterized and this Section mainly serves as robustness analysis for some of the key assumptions of the model, as well as possible topics for future research.

7.1. Multi-Worker Layers

One of the most salient assumptions in this model is that firms can only hire one worker per layer. As discussed, this imposes an unnatural restriction on firms and confers double meaning to variable m, which serves both as firm size and hierarchy size. Below, I present one possible way of extending the model to allow for multi-worker layers and argue that some qualitative results are not lost with the assumption of single-worker layers.

Consider a version of the model where q_{nm} represents the effective labor units employed in layer n out of m. Then, q_{nm} depends on the skill and the number of workers in that layers. Suppose that q_{nm} is produced through a production function

$$q_{nm} = \chi\left(\{q_{tnm}\}_{t=0}^{T_{nm}}\right)$$

where q_{tnm} is the skill of the t-th worker in layer n out of m, and T_{nm} is the total number of workers in that layer. The way T_{nm} is determined and how it interacts with q_{nm} is equilibrium would depend on the level of complementarity between workers in the same, as determined by function χ.

As a robustness test, take as an example

$$\chi\left(\{q_{tnm}\}_{t=0}^{T_{nm}}\right) = \int_0^{T_{nm}} q_{tnm}^{\nu} dt, \quad \nu > 1$$

That is, workers inside a layer are perfect substitutes. Then, the firm's problem in the inner loop becomes

$$\max_{\{\{q_{tnm}\}_{t=0}^{T_{nm}}, T_{nm}\}_{n=0}^{m}} \quad z \int_0^m a(n,m) \left(\int_0^{T_{nm}} q_{tnm}^{\nu} dt \right)^{\sigma} dn - \int_0^m \int_0^{T_{nm}} w(q_{tnm}) dt dn$$

Although this may seem like a much more complicated problem on the surface, note that the FOC for all q_{tnm} in the same layer n satisfies

$$w'(q_{tnm}) = \sigma\nu z a(n,m) q_{nm}^{\sigma-1} q_{tnm}^{\nu-1}$$

We can focus on equilibria where $q_{tnm} = \mu(n,m)z$, that is, the assignment doesn't vary with t. Since all worker skills in the same layer must satisfy the same FOC, then $q_{tnm} = q_{t'nm} = \tilde{q}_{nm}$ for all (t,t') in all layers. Additionally, $q_{nm} = T_{nm}\tilde{q}_{nm}^{\nu}$. Then,

$$w'(\tilde{q}_{nm}) = \sigma \frac{\nu a(n,m)}{\mu(n,m)} \frac{q_{nm}^{\sigma}}{T_{nm}}$$

which confirms that the same $\tilde{q}_{nm}$ satisfies the FOC of the firm for all t in n.

Knowing this, the problem of the firm becomes

$$\max_{\{\tilde{q}_{nm}, T_{nm}\}_{n=0}^{m}} \quad z \int_0^m a(n,m) T_{nm}^{\sigma} \tilde{q}_{nm}^{\sigma\nu} dn - \int_0^m T_{nm} w(\tilde{q}_{nm}) dt dn$$

The FOC for T_{nm} yields

$$w(\tilde{q}_{nm}) = \sigma z a(n,m) T_{nm}^{\sigma-1} \tilde{q}_{nm}^{\sigma\nu}$$

Combining with the FOC for worker skill, we obtain

$$T_{nm} = \frac{1+\sigma}{\nu \tilde{q}_{nm}^{\nu-1}}$$

Since $\tilde{q}_{nm}$ is increasing in n, we get that this model results in a pyramidal hierarchy, as in Garicano and Rossi-Hansberg (2006). Moreover, the relationship $q_{nm} = T_{nm}\tilde{q}_{nm}^{\nu} = \frac{1+\sigma}{\nu}\tilde{q}_{nm}$ allows us to exactly compare the skill of worker in each layer between the baseline model and this richer version. This exercise shows that the assumption that we may only have one worker per layer is troublesome insofar we are not willing to assume that workers are perfect substitutes within a layer. Otherwise, the results in the baseline version can be understood as the reduced form of a more complicated version of the model where firms also choose the number of workers in each layer and the hierarchy is pyramid-shaped.

7.2. Search Frictions

One possible criticism that would render this model unfit for empirical applications is that it ignores a crucial aspect of the labor market, that is, it is frictional.

However, it is straightforward to extend this model to accommodate search frictions as long as we keep the assumption of separability between layers. Once layers are no longer separable, the probability that the firm will match with other workers affects the decision of workers to search in a particular layer and the problem becomes intractable. Therefore, for the next part, I am going to maintain the assumption that there are no peer effects between workers.

Suppose workers direct their search over jobs, indexed by (z, n, m). Call $\lambda(z, n, m)$ the tightness ratio in each submarket, where it is defined as the number of workers per job. Firms post a wage share $\omega(z, n, m)$, which determines the share of output that workers take home as a wage. Suppose there is an aggregate CRS matching function that induces a probability $p(\lambda(z, n, m))$ of a firm matching with a worker. Then, the firm's problem in the inner loop is

$$\max_{\{\omega(z,n,m)\}_{n=0}^{m}} \; z \int_0^m p(\lambda(z, n, m)) (1 - \omega(z, n, m)) a(n, m) q_{nm}^{\sigma} dn$$

While workers choose in which submarket to search for jobs to maximize expected utility (assuming workers are risk-neutral):

$$U(q) = \max_{(z,n,m)} \frac{p(\lambda(z, n, m))}{\lambda(z, n, m)} \omega(z, n, m) a(n, m) q^{\sigma}$$

Plugging $U(q)$ into the firm's problem, we can recast the problem of the firm as choosing a tightness ratio and the skill of the worker for each layer:

$$\max_{\{\lambda(z,n,m), q_{nm}\}_{n=0}^{m}} \; z \int_0^m p(\lambda(z, n, m)) a(n, m) q_{nm}^{\sigma} dn - \int_0^m \lambda(z, n, m) U(q_{nm}) dn$$

The FOC for the problem above are

$$\begin{aligned} & z p'(\lambda(z, n, m)) a(n, m) \mu(z, n, m)^{\sigma} = U(\mu(z, n, m)) \\ & \sigma z p(\lambda(z, n, m)) a(n, m) \mu(z, n, m)^{\sigma-1} = \lambda(z, n, m) U'(\mu(z, n, m)) \end{aligned}$$

By performing a change of variable to φ and a similar procedure to that in Eeckhout and Kircher (2010), we obtain a system of differential equations again, this time in μ and λ. This system can be solved numerically for each φ to obtain the allocation in this market under some distributional assumption about G and H.

Moreover, we can find a condition on the production function that is necessary for supermodularity of F. Taking the cross-derivative with respect to z and q and rearranging, we obtain

$$\frac{f_{zq}(z,q)f(z,q)}{f_z(z,q)f_q(z,q)} > \frac{|\varepsilon_{a,m}||\varepsilon_{m,z}| + |\varepsilon_{p,\lambda}||\varepsilon_{\lambda,z}|}{\varepsilon_{f,z}}$$

Now, the supermodularity requirement on $f(z,q)$ is even stronger: not only does the complementarity between z and q have to grow faster than the reduction in importance, but it also needs to overcome the elasticity of the matching function. The reason for this is that workers can trade off a higher value match for a higher probability of matching. So now the increase in output from moving to a more productive firm has to be large enough to cover the lower relative importance of that worker in the new firm and the longer queue for that job.

8. Conclusion

This paper documents a novel empirical pattern using German administrative data: when workers switch to higher-quality firms, they systematically have more subordinates (higher layer) but occupy lower relative positions (lower rank) within the hierarchy. This reveals negative assortative matching across layers but positive assortative matching across ranks. To rationalize these findings, we construct a tractable model of multi-worker sorting across firms and occupations where firms endogenously choose their hierarchical structure. Due to the key assumption that the importance function is decreasing in the total size of the hierarchy, the allocation exhibits PAM across layers within firms and NAM across firms for fixed layers. We use this feature to discipline the comparison of workers in different jobs across firms and highlight the necessary degree of supermodularity to obtain PAM across firms in this environment.

As it is, this model carries a number of simplifying assumptions made to obtain closed-form solutions. However, this model could also be solved in a computationally cheap way by solving a simple system of first-order differential equations. This feature of the model allows it to be used as a module that can easily be incorporated into larger macro models that wish to make use of the distinct sorting patterns across firms and layers.

Some avenues for future research include the use of this model for policy analysis and the extension to multidimensional skills. For the first application, this environment would have to be integrated into a richer structural model. However, the endogenous firm structure could be used to analyze how size-based policies, for example, could distort labor market sorting through the distortion in m (which, in turn, would distort the importance function and sorting inside the firm).

For the second application, it would be interesting to expand the notion of skill to something that captures workers' ability in different tasks (routine and non-routine, for example). The characterization of jobs along two dimensions delivered some insights about how workers may move along these two dimensions independently, so a model that marries multidimensional skills with multidimensional jobs could potentially provide ways to rethink what sorting actually means in this context. Will workers sort with jobs according to their comparative advantages, or how well they transition from one task to another? That will depend on whether production is Ricardian, CES or something else entirely. Much of the recent literature on endogenous firm structure focuses on the level of specialization of workers inside the firm and this model could potentially deliver some rich sorting patterns on that front as well.

References

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

Adenbaum, Jacob. 2022. "Endogenous firm structure and worker specialization." PhD dissertation, University of Minnesota.

Ahlin, Christian. 2017. "Matching patterns when group size exceeds two." *American Economic Journal: Microeconomics* 9 (1): 352–384.

Becker, Gary S. 1973. "A theory of marriage: Part I." *Journal of Political Economy* 81 (4): 813–846.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin. 2021. "Sorting with team formation."Technical report, National Bureau of Economic Research.

Bonhomme, Stéphane. 2021. "Teams: Heterogeneity, sorting, and complementarity." *arXiv preprint arXiv:2102.01802*.

Bonhomme, Stéphane, Thibaut Lamadon, and Elena Manresa. 2019. "A distributional framework for matched employer employee data." *Econometrica* 87 (3): 699–739.

Chade, Hector, and Jan Eeckhout. 2018. "Matching information." *Theoretical Economics* 13 (1): 377–414.

Chade, Hector, Jan Eeckhout, and Lones Smith. 2017. "Sorting through search and matching models in economics." *Journal of Economic Literature* 55 (2): 493–544.

Dauth, Wolfgang, and Johann Eppelsheimer. 2020. "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: a guide." *Journal for Labour Market Research* 54 (1): 1–14.

Deming, David J. 2017. "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics* 132 (4): 1593–1640.

Eeckhout, Jan. 2018. "Sorting in the labor market." *Annual Review of Economics* 10: 1–29.

Eeckhout, Jan, and Philipp Kircher. 2010. "Sorting and decentralized price competition." *Econometrica* 78 (2): 539–574.

Eeckhout, Jan, and Philipp Kircher. 2018. "Assortative matching with large firms." *Econometrica* 86 (1): 85–132.

Freund, Lukas. 2022. "Superstar Teams: The Micro Origins and Macro Implications of Coworker Complementarities." *Available at SSRN 4312245*.

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

Garicano, Luis, and Esteban Rossi-Hansberg. 2004. "Inequality and the Organization of Knowledge." *American Economic Review* 94 (2): 197–202.

Garicano, Luis, and Esteban Rossi-Hansberg. 2006. "Organization and inequality in a knowledge economy." *The Quarterly Journal of Economics* 121 (4): 1383–1435.

Gregory, Victoria. 2020. "Firms as learning environments: Implications for earnings dynamics and job search." *FRB St. Louis Working Paper* (2020-036).

Hagedorn, Marcus, Tzuo Hann Law, and Iourii Manovskii. 2017. "Identifying equilibrium models of labor market sorting." *Econometrica* 85 (1): 29–65.

Jarosch, Gregor, Ezra Oberfield, and Esteban Rossi-Hansberg. 2021. "Learning from coworkers." *Econometrica* 89 (2): 647–676.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica*: 649–670.

Kelso, Alexander S, and Vincent P Crawford. 1982. "Job matching, coalition formation, and gross substitutes." *Econometrica*: 1483–1504.

Kremer, Michael. 1993. "The O-ring theory of economic development." *The Quarterly Journal of Economics* 108 (3): 551–575.

Kremer, Michael, and Eric Maskin. 1996. "Wage inequality and segregation by skill."

Lucas, Robert E. 1978. "On the size distribution of business firms." *The Bell Journal of Economics*: 508–523.

Manne, Henry G. 1965. "Mergers and the Market for Corporate Control." *Journal of Political Economy* 110 (10.2307): 1829527110.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies* 60 (3): 531–542.

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: product differentiation in pure competition." *Journal of Political Economy* 82 (1): 34–55.

Rosen, Sherwin. 1982. "Authority, control, and the distribution of earnings." *The Bell Journal of Economics*: 311–323.

Shimer, Robert, and Lones Smith. 2000. "Assortative matching and search." *Econometrica* 68 (2): 343–369.

Song, Jae, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2019. "Firming up inequality." *The Quarterly Journal of Economics* 134 (1): 1–50.

Appendix

Solution Method

The following solution method for the multi-worker assignment problem assumes an importance function $a(n,m)$ that depends only on the ratio between n and m. As an abuse of notation for the next part, I am going to write the importance function as $a\left(\frac{n}{m}\right)$.

First, define $\varphi = \frac{n}{m}$ and rewrite the firm's problem to be independent from m:

$$\max_{\{q_\varphi\}_{\varphi=0}^{1}} \int_0^1 a(\varphi) f(z, q_\varphi) d\varphi - \int_0^1 w(q_\varphi) d\varphi$$

We also redefine the assignment function to depend only on φ, that is, $\mu(z,n,m) = \mu(z,\varphi)$, by slightly abusing notation again. Taking the FOC of the problem above and evaluating at the equilibrium allocation:

$$w'(\mu(z,\varphi)) = a(\varphi) f_q(z, \mu(z,\varphi))$$

This is a differential equation in w and it will be the first block of the equilibrium system that we need to solve.

Again using the definition of φ and assuming G and H admit density functions, we can rewrite the feasibility condition as

$$\int_{\underline{z}}^{z} \int_0^{\varphi} g(z') d\varphi' dz' = M \int_{\underline{q}}^{\mu(z,\varphi)} h(q') dq', \quad \forall z$$

By applying the Leibniz integral rule, we get the second differential equation, this time in μ:

$$\frac{\partial \mu(z,\varphi)}{\partial z} = \varphi \frac{g(z)}{Mh(\mu(z,\varphi))}$$

Then, labor market equilibrium is given by a family of functions $\{w_\varphi, \mu_\varphi\}_{\varphi=0}^{1}$ that satisfy the following system of differential equations:

$$\frac{dw_\varphi(z)}{dz} = a(\varphi) f_q(z, \mu_\varphi(z)) \frac{d\mu_\varphi}{dz}(z)$$
$$\frac{d\mu_\varphi(z)}{dz} = \varphi \frac{g(z)}{h(\mu_\varphi(z))}$$

By solving the system above for each $\varphi \in [0, 1]$, we get the allocation and the wage for workers in any combination (z, n). Using this allocation and transforming the functions back into (z, n, m) space, we can solve the structure choice problem.

Proofs

Proof of Lemma 1

Use the guess $\mu(n, m) = \frac{\mu_0 n}{m}$ to plug wages and worker skills back into $\pi(z, n)$ to obtain

$$\pi(z, n) = \frac{z^{1+\sigma} a_0 \mu_0^{\sigma} m}{(1+\sigma)(2+\sigma)}$$

The structure choice problem is then given by

$$\max_m \frac{z^{1+\sigma} a_0 \mu_0^{\sigma} m}{(1+\sigma)(2+\sigma)} - \frac{\kappa m^2}{2}$$

Taking the FOC wrt m and rearranging, we get

$$m(z) = \frac{a_0 \mu_0^{\alpha} z^{1+\alpha}}{\kappa(1+\alpha)(2+\alpha)} = \beta z^{1+\alpha}$$

Now that we know $m(z)$, we can plug it into $\mu(n, m(z))$ and use feasibility to find the μ_0 that clears the market. On the left-hand side, we have total labor demand, while on the right-hand side, we have total labor supply:

$$\int_Z \int_0^{m(z)} \frac{\mu_0 n}{m(z)} z dn dG(z) = M \Rightarrow \mu_0 > 0$$

Then, the wage equation can be written as $w(q) = \omega q^{1+\sigma}$, for some constant $\omega > 0$. Since it only depends on q, then worker optimality is satisfied. Finally, since $\mu_0 > 0$, that implies $\beta > 0$ as well, which concludes the proof. □

Proof of Proposition 1

First, we prove a Lemma in the same vein as 3.2, but for our new importance function.

Lemma A.1. Suppose $a(n, m) = a_0 \left(\frac{n}{m}\right)^{\alpha}$ and $c(m) = \frac{\kappa m^2}{2}$. Then $\mu(n, m) = \mu_0 \left(\frac{n}{m}\right)^{\alpha}$, for some constant $\mu_0 > 0$ is an equilibrium. Additionally, $m(z) = \beta z^{1+\sigma}$, for some constant $\beta > 0$.

PROOF. Using the FOC plus our guess and plugging into the $\pi(z,m)$, we find

$$\pi(z,m) = \frac{z^{1+\sigma}a_0\mu_0}{1+\sigma}\int_0^m \left(\frac{n}{m}\right)^{\alpha(1+\sigma)} dn$$

Solving the above, we find

$$\pi(z,m) = \frac{z^{1+\sigma}a_0\mu_0 m}{(1+\sigma)(1+\alpha(1+\sigma))}$$

Then, the structure choice problem gives

$$m(z) = \frac{z^{1+\sigma}a_0\mu_0}{\kappa(1+\sigma)(1+\alpha(1+\sigma))} = \beta z^{1+\sigma}$$

Verifying that this solution is indeed an equilibrium:

$$\int_Z \int_0^{m(z)} \mu_0 \left(\frac{n}{m(z)}\right)^{\alpha} z dn dG(z) = M \Rightarrow \mu_0 > 0$$

which implies $\beta > 0$ □

Now, notice that the condition for PAM can be written as

$$\frac{\partial\left(\mu(n,m)z\right)}{\partial z} = \frac{\partial\mu(n,m)}{\partial m}\frac{\partial m(z)}{\partial z}z + \mu(n,m) > 0$$

Rearranging this expression, we find the condition over elasticities

$$|\varepsilon_{\mu,m}||\varepsilon_{m,z}| < 1$$

Note that, from Lemma A.1, $|\varepsilon_{\mu,m}| = |\varepsilon_{a,m}| = \alpha$ and $|\varepsilon_{m,z}| = 1+\sigma$. Then, the condition for PAM is equivalent to

$$\alpha(1+\sigma) < 1$$

which proves the proposition. □

Proof of Proposition 2

For this version of the model we need to assume a cost function that is more convex in m in order to have a well-defined structure choice problem. I assume $c(m) = \frac{m^{\kappa+1}}{\kappa+1}$, with $\kappa > \sigma + 2\gamma$. Taking the FOC of the labor choice problem, we get

$$w'(q_{nm}) = z\left[\gamma\left(\int_0^m q_{n'm}dn'\right)^{\gamma-1}\int_0^m a(n',m)q_{n'm}^{\sigma}dn' + \left(\int_0^m q_{n'm}dn'\right)^{\gamma}\sigma a(n,m)q_{nm}^{\sigma-1}\right]$$

Conjecturing $q_{nm} = \mu(n,m)z$ and integrating over q_{nm}, we get

$$w(q_{nm}) = \frac{q_{nm}^2}{2\mu(n,m)}\gamma\left(\int_0^m q_{n'm}dn'\right)^{\gamma-1}\int_0^m a(n',m)q_{n'm}^{\sigma}dn' + \left(\int_0^m q_{n'm}dn'\right)^{\gamma}\sigma\frac{a(n,m)}{\mu(n,m)}\frac{q_{nm}^{1+\sigma}}{1+\sigma}$$

Then, revenue can be written as

$$z^{1+\sigma+\gamma}\left(\int_0^m \mu(n,m)dn\right)^{\gamma}\int_0^m a(n,m)\mu(n,m)^{\sigma}dn$$

While the total wage bill is

$$\int_0^m w(q_{nm})dn = z^{1+\sigma+\gamma}\int_0^m \frac{\gamma}{2}\mu(n,m)\left(\int_0^m \mu(n,m)dn\right)^{\gamma-1}\left(\int_0^m a(n,m)\mu(n,m)^{\sigma}dn\right)dn + \int_0^m \frac{\sigma}{1+\sigma}a(n,m)\mu(n,m)^{\sigma}\left(\int_0^m \mu(n,m)dn\right)^{\gamma}dn$$

which, after manipulating the integrals, yields

$$\int_0^m w(q_{nm})dn = \delta z^{1+\sigma+\gamma}\left(\int_0^m \mu(n,m)dn\right)^{\gamma}\int_0^m a(n,m)\mu(n,m)^{\sigma}dn$$

where $\delta = \frac{\gamma}{2} + \frac{\sigma}{1+\sigma}$. Then,

$$\pi(z,n) = (1-\delta)z^{1+\sigma+\gamma}\left(\int_0^m \mu(n,m)dn\right)^{\gamma}\int_0^m a(n,m)\mu(n,m)^{\sigma}dn$$

For this version of the problem, we need to use a slightly different version of the assignment function, that is, $\mu(n,m) = \mu_0 n$. Then, we get

$$\pi(z,m) = \frac{(1-\delta)a_0\mu_0^{\gamma+\sigma}z^{1+\gamma+\sigma}}{2^{\gamma}(2+\gamma)}m^{1+\sigma+2\gamma}$$

Taking the FOC of the structure choice problem, we get

$$m(z) = \beta z^{\frac{1+\sigma+\gamma}{\kappa-\sigma-2\gamma}}$$

where β is some positive constant. Then,

$$\mu(z,n) = \mu_0 nz = \mu_0 \varphi m(z) z = \mu_0 \varphi z^{b(\gamma)}$$

where $b(\gamma) = \frac{\kappa-\gamma+1}{\kappa-\sigma-2\gamma}$, with

$$b'(\gamma) \propto -(\kappa-\sigma-2\gamma) + 2(\kappa+1-\gamma) = \kappa+\sigma+2 > 0.$$

Then, two different firms will hire the same worker if and only if

$$\mu(z,\varphi) = \mu(z',\varphi') \Leftrightarrow \frac{\varphi}{\varphi'} = \left(\frac{z'}{z}\right)^{b(\gamma)}$$

which concludes the proof. □

Proof of Proposition 3

First, suppose z is distributed according to a power law. Then,

$$\begin{aligned}
\Pr(m(z) > m) &= Pr(\beta z^{1+\sigma} > m) \\
&= Pr\left(z > \left(\frac{m}{\beta}\right)^{\frac{1}{1+\sigma}}\right) \\
&= S\left(\left(\frac{m}{\beta}\right)^{\frac{1}{1+\sigma}}\right)\left(\frac{m}{\beta}\right)^{-\frac{\eta-1}{1+\sigma}}
\end{aligned}$$

We can rewrite this as

$$\Pr(m(z) > x) = \tilde{S}(x) x^{\eta_m - 1}$$

where $\eta_m = \frac{\eta+\alpha}{1+\alpha}$ and $\tilde{S}(x) = \beta^{1-\eta_m} S\left(\left(\frac{x}{\beta}\right)^{\frac{1}{1+\sigma}}\right)$. Now, we only need to show that

$$\lim_{x\to\infty} \frac{\tilde{S}(rx)}{\tilde{S}(x)} = \frac{\beta^{1-\eta_m} S\left(\left(\frac{rx}{\beta}\right)^{\frac{1}{1+\sigma}}\right)}{\beta^{1-\eta_m} S\left(\left(\frac{x}{\beta}\right)^{\frac{1}{1+\sigma}}\right)} = 1, \quad r > 0$$

To see this, we can just perform a change of variable to $y = \left(\frac{x}{\beta}\right)^{\frac{1}{1+\sigma}}$ and note that, since S is a slowly varying function, then

$$\lim_{x\to\infty}\frac{\tilde{S}(rx)}{\tilde{S}(x)} = \lim_{y\to\infty}\frac{S(ry)}{S(y)} = 1, \quad r > 0$$

Now, suppose z follows a Pareto distribution with scale parameter z_{min} and shape parameter $\eta > 1 + \sigma$. Then,

$$\begin{aligned}
\Pr(m(z) > m) &= Pr\left(z > \left(\frac{m}{\beta}\right)^{\frac{1}{1+\sigma}}\right)\\
&= \left(\frac{z_{min}}{(m/\beta)^{\frac{1}{1+\sigma}}}\right)^{\eta}\\
&= \left(\frac{\beta z_{min}^{1+\sigma}}{m}\right)^{\frac{\eta}{1+\sigma}}
\end{aligned}$$

which is the survival function of a Pareto distribution with parameter $\beta z_{min}^{1+\sigma}$ and shape parameter $\frac{\eta}{1+\sigma}$. □

Proof of Proposition 4

The first part of the proof is analogous to the proof for z and m. So suppose q is Pareto with scale q_{min} and shape $\eta_q > 1 + \sigma$. Then,

$$\begin{aligned}
\Pr(w(q) > w) &= Pr\left(q > \left(\frac{\mu_0 w}{\sigma a_0}\right)^{\frac{1}{1+\sigma}}\right)\\
&= \left(\frac{q_{min}}{(\mu_0 w/\sigma a_0)^{\frac{1}{1+\sigma}}}\right)^{\eta_q}\\
&= \left(\frac{\frac{\sigma a_0}{\mu_0} q_{min}^{1+\sigma}}{w}\right)^{\frac{\eta_q}{1+\sigma}}
\end{aligned}$$

which is the survival function of a Pareto distribution with parameter $\frac{\sigma a_0}{\mu_0} q_{min}^{1+\sigma}$ and shape parameter $\frac{\eta_q}{1+\sigma}$.

Finally, we need to show when the distribution of $\log w(q)$ is convex. To be precise, convexity in this case means

$$\frac{d^2 \log w}{dp^2} > 0$$

where p is the percentile of the distribution of wages. Since in equilibrium w depends solely on q, the condition above depends on how worker skill q changes with the percentile of the distribution of wages. This is given by the function $Q(p)$ implicitly defined by

$$\Pr_W\left(W \leq w\left(Q(p)\right)\right) = p,$$

Then, we have

$$\log w\left(Q(p)\right) = c + (1+\sigma)\log Q(p)$$

where c is a constant. Then, $\log w$ is convex in p whenever

$$\frac{d\log w}{dp} = (1+\sigma)\frac{Q'(p)}{Q(p)}$$

is increasing in p. This is true if, and only if, $Q(p)$ is log-convex in p, that is,

$$\frac{d^2\log Q(p)}{dp^2} = \frac{d\left(\frac{Q'(p)}{Q(p)}\right)}{dp} > 0$$

which gives the condition we were looking for. □

Proof of Proposition 5

First, we compute the total aggregate output in the economy. Plugging $m(z)$ back into the production function, we get the output of a firm with productivity z:

$$y(z) = \frac{a_0\mu_0^\sigma m(z)z^{1+\sigma}}{2+\sigma} = Bz^{2(1+\sigma)}$$

Then, aggregate output is

$$\begin{aligned}
Y &= \int_{z_{min}} y(z)dG(z) \\
&= B\eta\int_{z_{min}} z^{2(1+\sigma)}\frac{z_{min}^\eta}{z^{\eta+1}}dz \\
&= \frac{B\eta}{\eta-2(1+\sigma)}z_{min}^{2(\eta-\sigma-1)} \\
&= Cz_{min}^{2(\eta-\sigma-1)}
\end{aligned}$$

Now, we compute the total effective labor units in the economy:

$$
\begin{aligned}
L &= \int_{z_{min}} \int_0^{m(z)} \mu(z,n,m)\,dndG(z) \\
&= \frac{\mu_0 \eta}{\beta} \int_{z_{min}} \int_0^{m(z)} ndn \frac{z_{min}^{\eta}}{z^{\sigma} z^{\eta+1}} dz \\
&= \frac{\mu_0 \eta \beta}{2} \int_{z_{min}} z^{2(1+\sigma)} \frac{z_{min}^{\eta}}{z^{\eta+\sigma+1}} dz \\
&= \frac{\mu_0 \eta \beta}{2(\eta - \sigma - 2)} z_{min}^{2\eta-\sigma-2} \\
&= D z_{min}^{2\eta-\sigma-2}
\end{aligned}
$$

Then, with a little algebra, we can see

$$
Y = AL
$$

with $A = \frac{C}{D z_{min}^{\sigma}}$. □