Linking Firm Size and Skill Composition: Theory and Evidence from Australia

Chase Abram*

Job Market Paper

Updated frequently, please click here to read the most recent version

February 26, 2025

Abstract

I examine the relationship between firm size and skill composition, using microdata from Australian workers and firms between 2011 and 2020. While larger firms generally employ higher shares of high-skill workers, all firms tend to hire relatively more low-skill workers as they grow. To explain these patterns, I develop a model that distinguishes total factor productivity (TFP) from skill-biased productivity (SBP), in which firms choose their scale and workforce composition. I validate the model mechanisms using evidence from a payroll tax policy change in South Australia, demonstrating how firms adjust both scale and skill mix in response to cost changes. I use a quantitative model of the Australian labor market to explore the implications of shifts in aggregate skill composition on the distributions of firm size and earnings inequality. I find that an 11 percentage point increase in the educated share of the workforce decreases equilibrium skill premia within all firms. Despite this, the aggregate skill premium increases, because firms that employ higher shares of high-skill workers raise earnings levels more for all their workers. The increased share of educated workers also leads to employment gains of 3% in the largest firms and an aggregate reallocation of 1% of workers to the largest firms. The results highlight how accounting for employment composition decisions by firms is crucial for understanding observed patterns of worker skill distribution and earnings across firms.

*abram@uchicago.edu, chaseabram.com, I cannot overstate my gratitude to my advisor, Greg Kaplan, and my committee, Erik Hurst and Christina Patterson, for their support and advice. I am also grateful for the many helpful conversations with and comments from Jack Buckley, Craig Chikis, Arjun Gopinath, Jonathan Hambur, Hugo Lopez, Jordan Rosenthal-Kay, Esteban Rossi-Hansberg, Camilla Schneier, Rob Shimer, Marcos Sora, Dalton Zhang, and the seminar participants at the Capital Theory, Applied Macroeconomic Theory, and Student Applied Micro seminars at the University of Chicago. All errors are my own.

1 Introduction

Worker skill is not uniformly distributed across firms. Some firms employ a small number of high-skill workers alongside many low-skill workers (e.g., retail), while others employ mostly high-skill workers (e.g., consulting). Firms also differ widely in scale: most employ fewer than 50 workers, but a small subset employ over 5000 workers. In choosing a workforce, firms must determine both their *scale* (how many total workers?) and *composition* (what types of workers?) of employment. These decisions depend not only on their own productive capacity, but also the supplies of each type of skill, which depend on the abundance of each skill type and the labor demands of all other firms. For example, a firm that finds high-skill workers highly productive will tend to hire more of them in a labor market where high-skill workers are more abundant, or where most other firms do not find high-skill workers as productive, and hence the competition for such workers is weaker.

As decisions of composition and scale are made jointly, an increase in the abundance of skill in the aggregate workforce may lead to changes in employment concentration, and the aggregate skill premium will be affected by both a wage effect and a distribution effect. Employment concentration may increase because firms that find high-skill labor relatively more productive will increase their quantity of high-skill labor the most, and may even increase their demand for low-skill workers, given that they are imperfect substitutes for high-skill workers. The wage effect for the aggregate skill premium is that the rise in high-skill supply will lead to absolute gains in earnings for all workers, but the size of these gains will not be equally distributed across firms with varying skill compositions. When the largest absolute gains occur in firms which employ relatively more high-skill labor, this provides a force for increasing the aggregate skill premium. The distribution effect is that labor will be reallocated across firms with varying skill premia. When firms with the largest skill premia see the largest gains in total employment, this provides another force for increasing the aggregate skill premium. Skill premia within each firm will decline simply because the rise in high-skill supply will lower the relative wages of high-skill workers, but if the wage and distribution effects are large enough to dominate the within-firm decrease, then an increase in the abundance of high-skill workers can lead to an *increase* in the aggregate skill premium.

Motivated by the above reasoning, I argue that the link between employment scale choice and employment composition choice is important for understanding observed patterns of worker skill allocation and premia across firms, and within firms over time. I begin by documenting the relationship between scale and skill composition in the cross-section of Australian firms in the period $2011-2020$: larger firms are more skilled. Next, I show that the cross-sectional patterns do not hold within a firm over time: as a firm becomes larger, it deskills. To understand the mechanisms behind these outcomes, I propose a simple model in which firms choose scale and skill composition. I prove the model can explain both the cross-sectional and within-firm facts, and argue that Hicks-neutral productivity differences between firms alone are unable to simultaneously explain the cross-sectional and within-firm patterns. To validate the model mechanisms, I provide estimates of labor supply elasticities, and examine a tax policy change in the state of South Australia where I predict how firms of different sizes should have adjusted their scale, skill composition, and skill premia. I show these predictions are supported empirically. Finally, I develop and calibrate a quantitative model of the Australian labor market to examine how a change in the skill composition of the aggregate labor force would affect the scale and skill distribution of labor across firms. I consider an increase in the educated share of the workforce from 39% to 50%, and find that this increases the share of employment in large firms ($\geq$ 1000 workers) by 1.5%. The increased skill share also leads to increased earnings in the largest and smallest firms and decreased skill premia within every firm, but an increase in the aggregate skill premium.

First, I compare the patterns of scale and skill composition in the cross-section with the patterns within a firm over time. In the cross-section, medium-sized firms are less skilled, and small and large firms are more skilled. Skill premia increase monotonically in firm size. These results broadly hold even when controlling for industry or location, especially for the most skilled workers. These findings indicate that underlying differences between firms are not primarily explained by differences between industries or regions. However, looking within firms reveals that as a firm increases its employment, it tends to skew the composition of its employment toward low-skill workers. This suggests that the higher share of high-skill workers in large firms observed in the cross section arises due to persistent differences between small and large firms. Large firms tend to be more skilled, but they are not simply small firms that have grown. On the contrary, as small firms grow, they typically reduce the share of high-skill workers they employ. In the cross-section, the initial decrease in skill ratio against size may occur because there is little difference between small and medium-sized firms, so the within-firm decrease in skill ratio with size is observed. Large firms tend to be different from medium-sized firms, however, in that for any given scale they prefer to be much more skilled, and this effect dominates the deskilling effect within large firms as they become large, so the largest firms in the cross-section are the most highly skilled.

Second, I provide a simple model to understand the observed patterns of scale and skill. Firms differ in both a Hicks-neutral component of productivity (total factor productivity, TFP) and a component of productivity that makes high-skill workers more productive (skill-biased productivity, SBP), and view worker types as imperfect substitutes. The imperfect substitutability of worker skills means that the marginal productivity of each worker depends on the composition of the entire team within a firm: an additional high-skill worker increases productivity more in firms with fewer high-skill workers. Firms face upward-sloping labor supply curves, where low-skill labor is more elastic than high-skill. Firms facing positive TFP shocks desire to scale up their employment, and find it optimal to do so via hiring relatively more low-skill workers, who are easier to hire. Hiring more low-skill workers changes the team composition and increases the marginal productivity of high-skill workers, however, so firms also increase high-skill hiring, and raise their skill premia because high-skill labor supply is less elastic. Firms with higher SBP tend to be both larger and have a more skilled workforce, so in the aggregate it is possible for the difference between low and high SBP firms to generate a positive relationship between firm size and skill composition in the cross-section[1]. The simple model also clarifies that standard models of firm dynamics, which only allow for heterogeneity in total factor productivity, are insufficient to explain the observed patterns of both scale and skill composition. The reason is that, with only a single firm-level productivity shifter, the cross-sectional patterns of size and skill composition should

[1]This result may be understood as a version of Simpson's Paradox: trends within each group (e.g., firm) may reverse when the groups are aggregated. This may occur because the within-group differences (firms deskill as they scale up) are dominated by opposite between-group differences (firms with higher SBP tend to be larger and more skilled).

be in accordance with the firm-level patterns, as the only difference between a small firm and a large firm must be the same as the difference between a small firm and itself when it becomes large. The empirical contrast between the aggregate and firm-level patterns motivates allowing firms to differ in not only total factor productivity, but also a dimension of skill-biased productivity.

Third, I provide two pieces of empirical evidence to support the model assumptions and predictions. The first is a back-of-the-envelope calculation of labor supply elasticities for each education group, using the fact that employment-to-employment transitions provide information about the sensitivity of firm-level employment to wages offered. For the second piece of evidence, I study a tax policy change in South Australia. The policy altered the payroll tax schedule of firms, so that the marginal tax rate for firms with payrolls between 1.5 and 1.7 million increased from approximately 4% to approximately 27%. Incorporating this policy change into the model, I predict that firms directly affected by the payroll tax change should decrease employment, shift toward a higher-skilled workforce, and decrease their skill premia. Alternatively, firms away from the treated region should see either no effect, or perhaps small effects in the opposite direction. These model predictions are supported empirically.

Fourth, I provide a quantitative model that matches the empirical patterns of the firm size distribution, the firm employment distribution, the distribution of skill ratios, the distribution of skill premia, and the earnings of low-skill workers. I use it to answer what will happen as Australia becomes more educated, motivated by the increasing share of educated workers. I consider an 11 percentage point increase in the share of the educated population (0.5 compared to 0.39) and show that employment in large firms increases by 1.5%. Additionally, this educational shift leads to a decrease in skill premia for all firms, with an average (firm-weighted) decrease of 0.1%, but an increase in the aggregate (worker-weighted) skill premium of 0.1%. The aggregate skill premium rises because employment is reallocated from firms with small skill premia to firms with large skill premia, and firms with higher shares of high-skill workers see larger absolute gains in earnings for all workers.

Related Literature I contribute to four areas of literature. First, I speak to the empirical literature on how workers sort to firms. Following Abowd, Kramarz, and Margolis (1999), recent work from Song et al. (2019), Bonhomme, Lamadon, and Manresa (2019), and Bonhomme et al. (2023) has shown that more productive workers tend to sort to more productive firms, though these frameworks assume that firm productivity is constant over time. Additionally, work by Haltiwanger, Lane, and Spletzer (1999) and Haltiwanger, Hyatt, and McEntarfer (2018) has shown that larger firms tend to be more productive and have more productive workers. I study how the skill of workers in a firm changes as the firm's scale changes, and show that, perhaps surprisingly given the above context, scaling up is associated with skilling down, and that skill premia only modestly rise with employment size. Similar results for employment have been documented by Gulyas (2020) in Germany, though they find opposite predictions regarding skill premia.

Second, I contribute to the theoretical literature seeking to explain the sorting patterns documented above. In this vein, Shimer and Smith (2000) and Borovicková and Shimer (2024) argue that more productive workers sort to more productive firms due to complementarity between worker skill and firm technology following Becker (1973), and Lise, Meghir, and Robin (2016) and Lise and Robin (2017) allow for firm dynamics, but the notion of firm and job are undifferentiated in their models. Closest to this paper is Eeckhout and Kircher (2018), which integrates the choice of worker composition into a span-of-control theory of the firm a la Lucas (1978), and Deb et al. (2022), where firms vary in skill-specific productivity and labor supply elasticities vary by skill. I study the empirical relevance of the relationship between size and scale, and allow for dynamics in the context of firms choosing both scale and skill composition in the simple model. In an extension in Appendix D I integrate the compositional choice into a model where firms face TFP shocks and may exit, and thus provide a theory of firm dynamics in skill composition similar to Hopenhayn (1992), which is not provided by previous work.

Third, I contribute to the rapidly expanding literature viewing firms as "teams" of workers cooperating to produce output. The approach of Freund (2024) is similar in their consideration of worker composition affecting output, but there are no firm dynamics. Similarly Herkenhoff et al. (2024), which is primarily focused on the learning aspect of team production, has no theory of heterogeneity in firm technology. Importantly, the discussion of firm scale is limited to at most two workers in both works, so they cannot explain the joint patterns of firm scale and worker composition. In contrast, I fit a quantitative version of my model to exactly hit the firm size and employment distributions, as well as the patterns of skill ratio, skill premia, and earnings across firm sizes.

Fourth, I contribute to the literature seeking to understand the employment size and worker earnings distributions across firms. Theoretically, this discussion goes back to at least the span-of-control model of Lucas (1978), which posits that firms face convex marginal costs, and thus the most productive entrepreneurs run the largest firms. Rosen (1982) augments the framework to allow for complementarity between skilled workers and managers, so is able to explain why the earnings distribution is skewed not only for managers, but also for workers. More recently, Poschke (2018) builds on the span-of-control model and argues that skill-biased technological change contributed to rising dispersion in firm size via increasing dispersion in entrepreneurial productivity of workers. Though I also consider how skill bias may impact firm size, I instead focus on how differences in skill bias alter the employment composition choices of firms, and how these choices relate to firm size choice, in a setting where firm-level labor supply elasticities differ by worker skill. Empirically, Autor et al. (2020), Rinz (2022), Hsieh and Rossi-Hansberg (2023), and Autor, Patterson, and Van Reenen (2023) provide evidence that employment concentration has risen at the national level in the United States, and Bajgar et al. (2023) provide evidence of increasing concentration in Europe, albeit in sales rather than employment. To explain these trends, Autor et al. (2020) suggest there must be some forces increasing market "toughness", and Autor, Patterson, and Van Reenen (2023) show a shift toward services and away from manufacturing. Hsieh and Rossi-Hansberg (2023) argue that increases in the scalability of services contribute to the dispersion in employment scale across firms. Rather than focus on changes in employment concentration across industries, I document how employment scale and composition are related both in the cross-section and within firms in the aggregate economy, and argue that differences between firms in skill bias lead to dispersion in firm size, and therefore an increase in the high-skill labor share would increase the skill bias advantage, and increase employment concentration.

The organization of the paper is as follows. In Section 2 I describe the data. In Section 3

I document the relationship between firm size and skill composition in the cross-section and within firms. In Section 4 I provide support for the assumption that low-skill labor supply to firms is more elastic than high-skill. In Section 5 I provide a simple model to explain the facts. In Section 6 I provide additional empirical support for the model assumptions and predictions by considering the impact of a regional change in tax policy. In Section 7 I explain the quantitative model and consider the counterfactual implications of a rise in the high-skill share of the workforce. Section 8 concludes.

2 Data Description

I construct a linked employer-employee panel of Australian workers and firms for the years 2011 - 2020 using confidential tax data from the Australian Bureau of Statistics. The data include earnings from each firm for each worker in Australia. I link to each worker an education level from the 2016 Census, and use this as a measure of skill. I also link each worker to their location of residence. I link to each firm its industry and construct a measure of value-added[2]. I keep workers between the ages of 25 and 60, and firms with positive employment, nonnegative value-added, and nonnegative cost-of-goods-sold. I observe approximately 55 million worker-years, and approximately 4 million firm-years. In Appendix B.2 I provide a detailed description of the data sources and variable construction. In Appendix B.4 I provide summary statistics concerning worker education, earnings, skill ratios, and skill premia, by firm size and age.

3 Motivating Facts

I start by examining the patterns of skill composition and skill premia versus firm size. My first finding is a U-shaped relationship between firm size and skill, meaning the firms with the most skilled composition of workers tend to either be small or large, but not medium-sized. I show that this pattern holds even when controlling for detailed industrial classification, indicating it is not driven entirely by industrial differences between differently-sized firms. My second finding is that this pattern no longer holds when considering size changes within firms: as firms scale up, they skill down. I repeat these exercises instead considering within-firm skill premia to deliver my third finding: skill premia are much larger in larger firms, but as firms grow they only mildly increase their skill premia. These results suggest that firms differ in a time-invariant dimension which determines how they value high-skill workers, motivating the model in Section 5, which allows me to explain both the cross-sectional and within-firm findings.

3.1 Fact 1: The largest and smallest firms are most skilled

To assess the relationship between size and skill composition at the level of the aggregate economy, I bin employment and consider the least-squares regression of the (log) skill ratio

[2] The Australian goods and services tax (GST) is a value-added tax, thus firms must report a detailed breakdown of their input costs and output revenues.

on indicators for each bin, and include time fixed effects. I index bins by b and let B_b be the interval of employment for firms in bin b, and run the following regression.

$$\ln \frac{H_{jt}}{L_{jt}} = \delta_t + \sum_b \beta_b \mathbf{1}\{emp_{jt} \in B_b\} + \epsilon_{jt} \tag{1}$$

where j indexes firms, t is time, δ_t are time fixed effects, β_b is the increase in log skill ratio for employment in firm B_b, relative to the baseline bin of firms with employment $1-10$.

The results are shown in Table 4 and plotted against mean employment in each bin in Figure 1. The U-shaped relationship between firm size and firm skill ratio continues to hold, so that medium-size firms tend to have the lowest skill ratio, but that the smallest and largest firms are more skilled. It is also worth noting that the minimum skill ratio bin is $20-50$ for college-educated workers compared to $100-200$ for graduate-educated workers. Therefore, firms in bin $100-200$ tend to have a higher college ratio, compared to bin $20-50$, but a lower graduate ratio.

The differences in skill composition across the firm size distribution are quantitatively large. In small firms ($0-9$ workers) 45% of workers are high-skill; in medium-sized firms (20 - 99 workers) only 33% of workers are high-skill; in the largest firms (5000+ workers) 52% of workers are high-skill. The results are more stark when focusing on workers with a graduate degree. In firms with at least one graduate worker and $0-9$ total workers, 39% of workers have a graduate degree; in firms with $100-1000$ workers, only 7% of workers have a graduate degree; in firms with 5000+ workers, 14% of workers have a graduate degree.

The broad-level specification in equation 1 only controls for aggregate time trends, so the identifying variation for the β_b coefficients comes from both within-firm difference across time, and between-firm differences. Therefore, this specification alone does not inform about the degree to which each type of variation (within- versus between-firm) is generating the observed U-shape. For example, one explanation for the above relationship between size and skill composition is that all firms follow the U-shape as they expand. Another explanation is that firms maintain a constant skill ratio at all times, but that firms which choose to become medium-sized are the same ones that choose a low skill ratio. Analogously, the firm differences could be at the level of the industry or location of the firm. To discern between these myriad explanations for the cross-sectional patterns in the aggregate economy, I now repeat the above specification, but control for variation in industrial division, detailed industry, and geography. To this end, I check that the relationship in the aggregate holds when I control for industry by considering the following regression, where $a(j)$ is the ANZSIC industry of firm j

$$\ln \frac{H_{jt}}{L_{jt}} = \delta_t + \delta_{a(j)} + \sum_b \beta_b \mathbf{1}\{emp_{jt} \in B_b\} + \epsilon_{jt} \tag{2}$$

In Table 5 I show the results where I allow $a(j)$ to be the coarsest industry stratification, industry division, and in Table 6 I consider the more detailed definition at the level of 3-digit ANZSIC. In Figure 25 in Appendix B.4 I display the coefficients as a function of firm size for the coarse specification, and in Figure 2 I show the results for the detailed industry specification. The results are broadly similar between the two approaches. Although the estimates

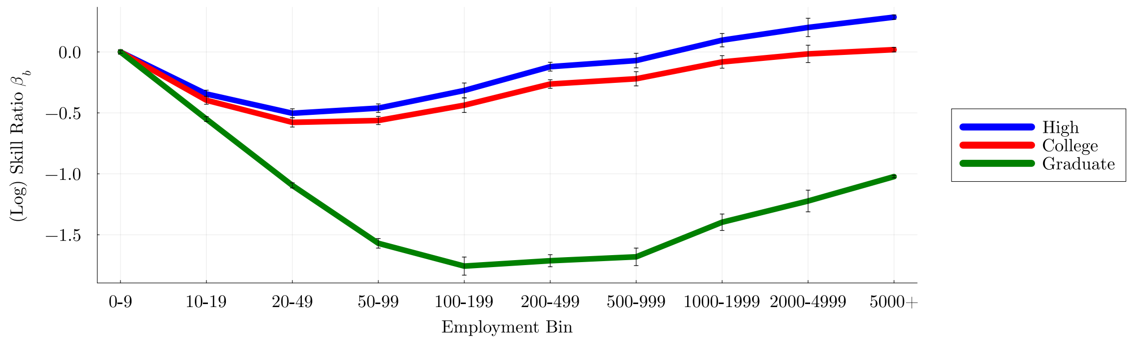


Figure 1: Coefficients from regressions of the log skill ratio on firm size, in the cross-section, from Table 4.

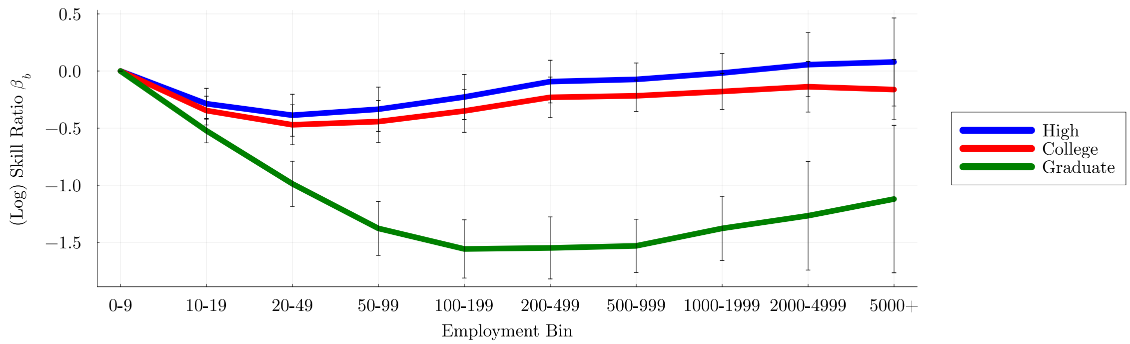


Figure 2: Coefficients from regressions of the log skill ratio on firm size, controlling for industry ANZSIC 3-digit, from Table 6

are much less precise, for firms with less than 1000 workers the point estimates are largely unchanged, indicating industrial differences are unable to explain the relationship between size and skill for most firms[3]. For the largest firms, controlling for industry reduces the coefficients, most notably for college-educated workers, indicating industrial differences are relevant for explaining differences in skill composition at the top of the firm size distribution.

Another potentially important difference between firms is location. To assess whether geographic differences might be driving the results, I control for a measure of firm location at the level of SA4+GCCSA, similarly to how I control for firm industry above. I present the results in Appendices B.3 and B.4, and find that controlling for location does not change the main results concerning the U-shape. Therefore the U-shaped relationship between size and skill composition cannot be explained by location differences alone, such as remoteness.

3.2 Fact 2: As firms become larger they become less skilled

Above, I showed that the cross-sectional patterns are broadly unchanged, albeit somewhat dampened, when controlling for differences in industry or location of firms. I now turn to the most detailed level possible, and repeat the above analyses, but control for permanent

[3]This finding is unsurprising in light of the fact that large heterogeneity in firm productivity has been documented to exist even within narrowly defined industries, e.g. by Syverson (2004).

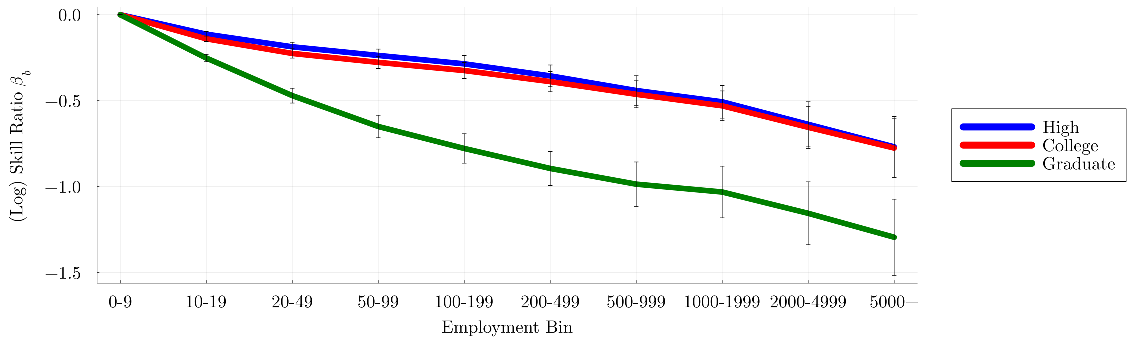


Figure 3: Coefficients from regressions of the log skill ratio on firm size, controlling for firm, from Table 8

heterogeneity at the firm level by including firm fixed effects. The motivation is that these specifications control for unobservable differences between firms that are time-invariant, such as a preference for high-skill workers independent of total employment. The specification is

$$\ln \frac{H_{jt}}{L_{jt}} = \delta_t + \delta_j + \sum_b \beta_b \mathbf{1}\{emp_{jt} \in B_b\} + \epsilon_{jt} \tag{3}$$

The results are in Table 8 and Figure 3. Surprisingly, the relationship between size and skill ratio is uniformly decreasing: when firms scale up, they skill down. The estimates are precise, and on the same order of magnitude as the cross-sectional results. To get a concrete sense of the difference between the cross-sectional and within-firm results, compare the coefficients for employment bins $50-99$ and $2000-4999$. In the cross-section, the log skill ratio for $50-99$ worker firms is -0.46, so 39% of the workforce is high-skill, but for $2000-4999$ worker firms the skill ratio is 0.2, so 55% of the workforce is high-skill. In contrast, when a firm grows from having $50-99$ workers to $2000-4999$ workers, on average they reduce their share of high-skill workers from 44% to only 35%.

The coexistence of Facts 1 and 2 implies that upward-sloping part of Fact 1 is entirely due to a composition effect: the firms that are larger tend to be more skilled, but they are not simply small firms with more employment, because when firms increase their employment they become less skilled. Explaining these two facts simultaneously requires a theory of firms that are heterogeneous in at least two dimensions. With only one dimension of heterogeneity, cross-sectional differences between firms and differences within a firm over time must be the same; there is only one possible way that firms are different. With at least two dimensions, it is possible for within-firm differences over time to be different from cross-sectional differences between firms, as one dimension of heterogeneity can drive firm dynamics, and another dimension may contribute to persistent (or permanent) differences between firms.

3.3 Fact 3: Skill premia increase more rapidly with size in the cross-section than within firms

I repeat the above analyses considering skill premia within a firm, rather than the skill ratio. The cross-section results relating firm size to (log) skill premium are in Table 10 and Figure

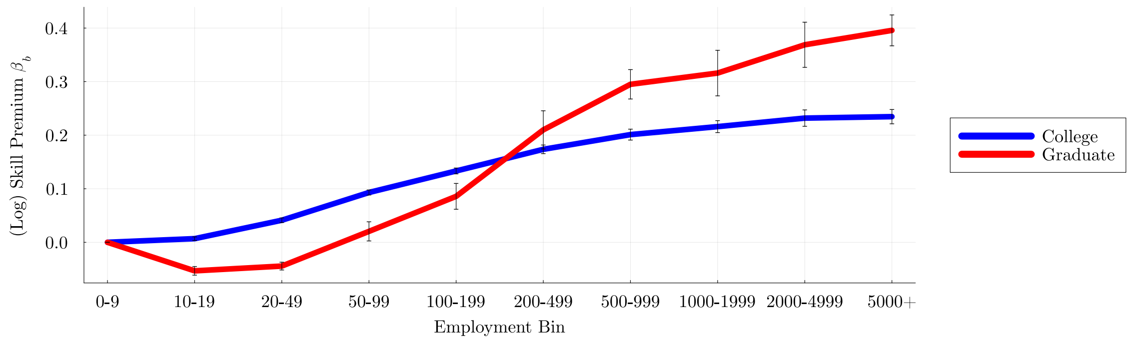


Figure 4: Coefficients from regressions of the log skill premium on firm size, in the cross-section, from Table 10

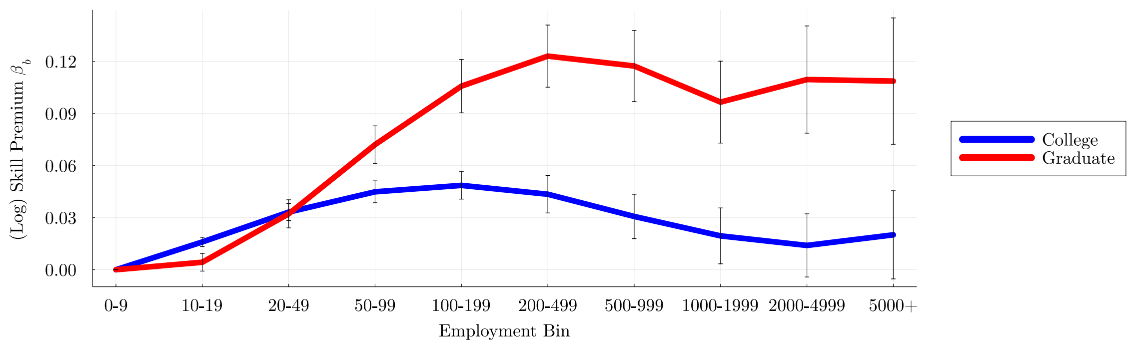


Figure 5: Coefficients from regressions of the log skill premium on firm size, controlling for firm, from Table 11

4, and the within-firm results are in Table 11 and Figure 5. In the cross-section, the log skill premium is monotonically increasing firm size, but within firms there is a plateau around the $200 - 499$ bin, and even a slight decrease afterwards. Furthermore, the increases in the within-firm specification are only about 25% of the full increase in the cross-section. Therefore, although large firms tend to have a college skill premium which is 20% higher than small firms, when those same small firms grow to the size of the large firms, they tend to only increase their skill premia by about 2%. For graduate workers, the largest firms have skill premia which are 40% higher than the smallest firms, but when these smallest firms grow to the same size of the large firms, they only increase their skill premia by about 11%. These results again point toward differences between small and large firms being driven by time-invariant differences in how firms are able to utilize workers of each skill type.

4 Estimating Firm-Level Labor Supply by Worker Skill

Below, I explain the observed patterns of skill composition in a simple model, and a crucial assumption is that labor supply is more elastic for low-skill workers. Estimating labor supply to individual firms is notoriously difficult, as the decision to move between firms, and the origin and destination wages, are all endogenous objects. One way to obtain a measure of how sensitive workers are to wage differences is to consider the separation or quit elasticity, as in Bassier, Dube, and Naidu (2022) and Autor, Dube, and McGrew (2023)[4]. The motivating observation is that workers are more likely to make an employment-to-employment transition when they currently have a lower wage. Most models of on-the-job search along the lines of Burdett and Mortensen (1998) will generate this pattern, as a lower wage will mean that a higher share of potential job offers dominate the current wage and lead to job transition. To answer the question of *how* responsive worker quit rates are to wages, I estimate the following specification.

$$EE_{it} = \delta_t + \delta_{j(i,t)} + \beta \ln w_{i,t-1} + \epsilon_{it} \tag{4}$$

where EE_{it} is an indicator for an employment-to-employment transition between $t-1$ and t, $w_{i,t-1}$ are worker earnings in $t-1$, and δ_t and δ_j are time and firm fixed effects, which I include to control for the fact that some firms $j(i,t)$ may generally have higher poaching rates. The results are in Table 9. The interpretation of the coefficient is that a 1% increase in a worker's earnings will decrease by $|\frac{\beta}{100}|$ their probability of quitting to another firm. To convert these measures to labor supply elasticities, I divide by the EE rate for each type of worker and multiply by -2, and report the results in Table 1[5]. The estimates appear reasonable, though are lower than the findings for the United States pre-pandemic reported by Autor, Dube, and McGrew (2023).

Education	**Quit** β	**EE rate**	**Labor Supply Elas** $(-2 \times \frac{\beta}{EE})$
Less High School	-0.040	0.174	0.460
High School	-0.038	0.190	0.400
College	-0.036	0.185	0.389
Graduate	-0.035	0.180	0.389

Table 1: Construction of a measure of labor supply elasticity for each skill type. "Quit β" is the coefficient from equation 4 estimating the responsiveness of worker quits to log earnings. "EE rate" is the employment-to-employment transition rate for each worker type. "Labor Supply Elas" uses the two previous measures to construct a measure of labor supply elasticity in accordance with models of dynamic monopsony a la Manning (2013).

5 Framework to Rationalize Empirical Findings

I start with a simple model able to qualitatively explain the observed patterns documented above. I follow Lucas (1978) in assuming firms are heterogeneous, but depart in two ways. First, instead of assuming firm technology is subject to decreasing returns to scale and the labor market is competitive, I assume constant returns to scale production, but that firm-level labor supply curves are upward-sloping, and that there are two types of labor. Second, firms are heterogeneous, not only in total factor productivity (TFP) z_{jt}, which is dynamic, but also

[4] Another approach is to consider the recruiting elasticity as in Hirsch et al. (2022).

[5] I review a simple model of dynamic monopsony following Manning (2013) in Appendix B.5 to show why this is an appropriate construction.

in skill-biased productivity (SBP) κ_j, which is fixed over time. These minimal assumptions (two labor types and two productivity margins) generate an equilibrium distribution in both firm size *and* worker composition. In order to be in line with the above empirical regularities, I make one key assumption: labor supply is more elastic for low-skill workers, $\eta_L > \eta_H$. The labor supply assumption will induce firms facing positive TFP shocks to scale up production, but they will do so by shifting the composition of their labor force toward the relatively easier to hire low-skill workers, so they "scale up by skilling down". Within a firm, this force generates the negative relationship between firm high-skill share and size over time[6]. The distribution of TFP and SBP across firms will generate the aggregate relationship between firm size and skill share, as high SBP firms will tend to be larger and more-skilled. When the SBP distribution is such that the between-firm difference in skill-share and size dominates the within-firm negative relationship, the cross-sectional outcome above is generated: larger firms tend to be more skilled. I relegate the algebra and computational details for this section to Appendix A.

Technology A unit measure of firms produce with the following constant elasticity of substitution technology, where both z_{jt} and κ_j are firm-specific, but $\kappa_j \in \{\kappa_L, \kappa_H\}$ is fixed over time, and $z_{jt} \in \{z_L, z_H\}$ evolves stochastically, transitioning from low state to high state at rate λ_H and high state to low state at rate λ_L. The share of firms with κ_L is s_L.

$$Y(z, \kappa, n_L, n_H) = zN(\kappa, n_L, n_H) \tag{5}$$

$$N(\kappa, n_L, n_H) \equiv \left(n_L^{\frac{\sigma-1}{\sigma}} + (\kappa n_H)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{6}$$

Labor Supply Firms face constant elasticity of substitution labor supplies with elasticities η_L and η_H and total market labor $\overline{n}_L$ and $\overline{n}_H = 1-\overline{n}_L$ for low and high skill labor, respectively,

$$n_L(w) = \left(\frac{w}{W_L}\right)^{\eta_L} \overline{n}_L \tag{7}$$

$$n_H(w) = \left(\frac{w}{W_H}\right)^{\eta_H} \overline{n}_H \tag{8}$$

The W_i are the market wage indices for each labor type i, and given by

$$W_i = \left(\int w_i^{\eta_i} \mathrm{d}F_i(w_i)\right)^{\frac{1}{\eta_i}} \tag{9}$$

where F_i is the equilibrium distribution of w_i in the economy. I provide a microfoundation for these labor supply curves in Appendix A.1.

[6]A variation of this mechanism is key in Mertens and Schoefer (2024), where they show that firms scale up production by substituting toward inputs, which are more elastically supplied than labor. I instead focus on substitution between labor types and abstract from intermediates.

Firm Optimality The problem of firm j with (z_{jt}, κ_j) at each instant[7] is to maximize revenues net of labor costs, where the homogeneous good produced is the numeraire. To do so, firms equate the marginal revenue product of labor with the marginal cost for each worker type, accounting for the monopsony power afforded by the upward-sloping labor supply curves. The optimality conditions may be found in Appendix A.3.

Equilibrium The market equilibrium is a set of wages $w_i(z,\kappa)$, employment levels $n_i(z,\kappa)$, and distribution of firms over (z,κ) such that

(i) Given the wage indices implied by the distribution of firms and $w_i(z,\kappa)$, wage $w_i(z,\kappa)$ implies labor quantity supplied $n_i(z,\kappa)$ for all i, z, κ.

(ii) Given the wage indices implies by the distribution of firms and $w_i(z,\kappa)$ and the implied labor supply schedules, firm (z,κ) optimally hires $n_i(z,\kappa)$ at $w_i(z,\kappa)$ for all i

(iii) The distribution of firms across (z,κ) is consistent with the share of firms for each κ and the dynamics of z

The above definition is one of partial equilibrium, but I explain in A.8 how to make additional assumptions so that the goods market clears but does not distort labor decisions away from the partial equilibrium outcome. The equilibrium of the economy cannot generally be given in closed-form, but I detail how to solve the equilibrium numerically in A.4 and prove the following proposition regarding qualitative comparative statics in A.5.

Proposition 1.

(i) In the limit as $\sigma \to \infty$

  (a) Fixing z, firms with larger κ are larger, i.e. N is larger

  (b) Fixing z, firms with larger κ have higher skill shares, i.e. $\frac{n_H}{n_L}$ is greater

  (c) Fixing z, firms with larger κ have higher skill premia

  (d) Fixing κ, as a firm's z increases, they become larger

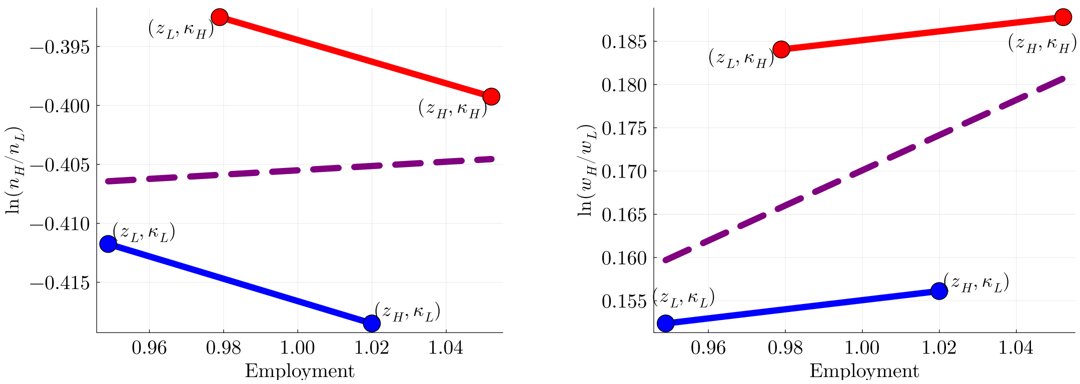
  (e) Fixing κ, as a firm's z increases, the share of their workforce that is high-skill declines

(ii) There exists $\sigma > 0$ such that the above statements hold and, additionally,

  (f) Fixing κ, as a firm's z increases, their skill premium increases

The first part of the proposition regards the limiting case when labor types are perfect substitutes. In this case, skill-biased productivity κ is irrelevant for a firm's hiring of low-skill labor, but increases the productivity of high-skill workers, incentivizing increased hiring.

[7] There are no choices being made that depend on the expected future path of (z,κ), so the dynamic problem of the firm is simply a sequence of static problems. This result does not occur if it is costly to hire and maintain labor, or if firms make entry and exit decisions, as in the quantitative model of Appendix E.



(a) Skill Ratios

(b) Skill Premia

Figure 6: Skill Ratios and Premia: the circles are the different firm types, with the solid lines connecting the two z states a firm with a fixed κ may move between. The dashed line is a least-squares fit of the relationship, using the cross-section of all firms.

This increase in hiring requires moving up the labor supply curve of high-skill workers, increasing their wage, thus the skill premium. Total factor productivity incentivizes increased hiring of both skill types, but particularly the more elastic low-skill, which lowers the ratio $\frac{n_H}{n_L}$ within the firm. Intuitively, as TFP rises firms want to scale up production, but it is relatively more difficult to hire high-skill workers, so fewer are hired. In the case of perfect substitutes, the skill premium is invariant to TFP, however, as the relative MRPL between worker types only depends on SBP. This knife-edge result motivates the final part of the proposition, which says that allowing for some amount of imperfect substituability between worker types will mean that firms prefer to scale up by increasing high-skill labor more, relative to the perfect substitutes case. This occurs because firms recognize that changing the composition of labor in the firm effects the scale of production, due to imperfect substitutability, and therefore they are slower to shift their composition toward low-skill workers. The relatively slower shift in skill composition means firms must move up the high-skill labor supply curve more, so skill premia rise. Firms scale up partly by deskilling, and they do so because they must tradeoff their desire to maintain a given skill composition, which increases the scale of output, with the increased labor costs associated with scaling up any given labor composition, given that differing elasticities between high- and low-skill labor supplies.

Note that the proposition is entirely concerned with the comparisons between firms with different (z, κ), but silent on aggregate relationships. The reason is that the distribution of firms across κ, as well as the transition rates for z, determine the firm (z, κ) distribution, and in general aggregate relationships between size and high-skill share and skill premia are not possible to find without a computer. To speak to the macro relationships, I solve the parameterization of the model given in Table 3 numerically and present the results in Figure 6.

As expected, the relationship between high-skill share and size is negative within a firm, but the macro relationship is positive, because the firms with larger κ tend to be larger and have a higher skill share. The positive relationship is then driven by comparison *between* firms with different κ, and in fact within-firm variation is a force in the opposite direction. A naive researcher attempting to understand how a firm changes the composition of its labor force as it grows would find a result of the wrong *sign* if they merely used a cross-section, and assumed that large firms are a proxy for small firms that increased in size.

The skill premia result is less stark, because the signs match at the firm level and in the cross-section, but the magnitudes are completely different. While firms with higher κ tend to be larger and pay a higher skill premium, each firm's skill premium only slightly increases as the firm grows. Again, a researcher seeking to understand the relationship between skill premium and size over a firm's lifecycle would wrongly conclude that firms dramatically increase their skill premia as they scale up, if they only looked at a cross-section, whereas in reality skill premia only slightly increase as each firms grows.

I can also directly consider what the above regression specifications will recover in the model, using the distribution of firms and the firm optimality conditions. The skill ratio of each firm is not a simple function of total employment and parameters alone, but the optimality conditions can be rearranged to yield the following relationship between log total employment and log skill ratio for each firm,

$$\underbrace{\ln \frac{n_{Hjt}}{n_{Ljt}}}_{\text{skill ratio}} = \underbrace{\left(\frac{\sigma}{\eta_L} - \frac{\sigma}{\eta_H}\right) \ln(n_{Ljt} + n_{Hjt})}_{\text{scale}} + \underbrace{(\sigma - 1) \ln \kappa_j}_{\text{skill-bias}} \tag{10}$$

$$+ \underbrace{\frac{\sigma}{\eta_L} \ln \frac{n_{Ljt}}{n_{Ljt} + n_{Hjt}} - \frac{\sigma}{\eta_H} \ln \frac{n_{Hjt}}{n_{Ljt} + n_{Hjt}}}_{\text{substitution}} + \text{constant} \tag{11}$$

where the constant does not vary across firms. This form reveals that a firm's log skill ratio may be decomposed into the sum of four terms. The constant term captures aggregates that affect all firms in the same manner, such as the total supplies of each skill, the wage indices, and the markdowns[8] [9]. The scale term captures the fact that scaling firms will change their skill composition, and that this component decreases the skill share when $\eta_L > \eta_H$, and the effect is larger when the skills are more substitutable. The skill-bias term captures the fact that firms with higher skill-biased productivity choose higher skill ratios. The skill substitution term is the residual that captures how skill shares affect the skill ratio, accounting for the scale and skill bias terms.

Before considering the relevant case of $\eta_L > \eta_H$, it is instructive to consider the simpler case of $\eta_L = \eta_H \equiv \eta$. Then the decomposition is

[8] The constant is $\sigma \left(\ln \frac{\eta_H+1}{\eta_H} - \ln \frac{\eta_L+1}{\eta_L} + \frac{1}{\eta_L} \ln \overline{n}_L - \frac{1}{\eta_H} \ln \overline{n}_H + \ln W_H - \ln W_L \right)$

[9] It is worth noting that the fact that all firms operate with the same markdown for each labor type is an endogenous outcome due to the constant elasticity supplies. In a more general setting where labor supplies are not CES, the markdowns would also vary across firms and be another term in the decomposition.

$$\underbrace{\ln\frac{n_{Hjt}}{n_{Ljt}}}_{\text{skill ratio}} = \underbrace{0}_{\text{scale}} + \underbrace{(\sigma-1)\ln\kappa_j}_{\text{skill-bias}} \tag{12}$$

$$+ \underbrace{\frac{\sigma}{\eta}\ln\frac{n_{Ljt}}{n_{Hjt}}}_{\text{substitution}} + \text{constant} \tag{13}$$

The optimal skill ratio is independent of the scale of employment, because when scaling up it is equally costly to move up each supply curve[10]. In this case, as firms face TFP shocks, they scale their workforce up or down accordingly, but always maintain the same composition, because the cost-minimizing bundle of skills is scale-invariant. This is no longer true when $\eta_L > \eta_H$, because as a firm scales production up, the least-cost bundle of labor skews towards the more elastic low-skill type. This result may be derived from the optimality conditions, but an intuitive way to understand this point is to imagine a firm is optimally producing, then is tasked with doubling output. When $\eta_L = \eta_H \equiv \eta$, the firm will double both types of labor, and costs will rise by $2^{\frac{\eta+1}{\eta}}$. Any other choice which skews the composition towards low- or high-skill labor will be more costly due to Jensen's inequality, as total costs for each skill type are convex, so the cost increase from hiring relatively more of one type of skill are greater than the cost decrease from hiring relatively less of the other. When $\eta_L > \eta_H$, doubling both labor inputs would lead to high-skill labor costs increasing more quickly than low-skill, so instead the firm *will* skew their hiring towards low-skill workers to minimize costs.

The decomposition is most useful for understanding the recovered relationships between size and skill ratio in the cross-section and within firms in Section 3. In a cross-sectional regression of firm size on skill ratio, the skill-bias and skill substitution terms are omitted variables which may statistically bias[11] the coefficient on log employment up or down, depending on the joint distribution of TFP and SBP. If firms which have high TFP also tend to have high SBP, then log employment and κ will be positively correlated, and the coefficient on log employment will be statistically biased upward. More generally, for employment levels where the average firm tends to have have higher SBP, the cross-sectional regression will attribute the higher skill ratio to the employment level[12]. In particular, consider the following cross-sectional regression of skill ratio on firm size, the analogue of equation 1.

$$\ln\frac{n_{Hjt}}{n_{Ljt}} = \alpha + f(\ln(n_{Ljt} + n_{Hjt})) + \epsilon_{jt} \tag{14}$$

where f is some function to be fitted, and in the empirical specifications above it is a piecewise constant function with constants that vary by firm size bin. Within each bin, this regression

[10]This result can also be observed directly in the optimality conditions in A.3.

[11]Unfortunately, "bias" has two meanings in this setting. I will say *skill bias* when referring to the economic force that causes firms to prefer high-skill workers, and *statistical bias* when referring to econometric estimators which are biased.

[12]It is easy to check that, for any fixed total employment level, higher κ must generate higher skill ratios, as the only way the equation can hold for a higher κ is if n_H increases and n_L decreases.

will attribute all differences in skill ratios to differences in total employment. Therefore, if the smallest and largest firms tend to have higher κ than medium-sized firms, then the U-shape above would be recovered, as the specification infers that the bins with more high κ firms and a higher skill ratio are due to the employment level. I show this pattern of κ is exactly what is recovered in the quantitative model below.

The statistical bias from the skill bias term also motivates the inclusion of firm-level fixed effects. Firm κ_j is time-invariant, so by controlling for each firm the skill bias term is absorbed into the firm fixed effect, and the statistical bias introduced by the composition of κ varying over the employment distribution is removed. Then the following regression recovers the relationship between employment and skill ratio on average across firms, controlling for differences in SBP, where f is as above.

$$\ln \frac{n_{Hjt}}{n_{Ljt}} = \alpha_j + f(\ln(n_{Ljt} + n_{Hjt})) + \epsilon_{jt} \tag{15}$$

The recovered relationship between log employment and the log skill ratio will now be unambiguously and uniformly negative. This can be seen by considering the above decomposition and holding fixed any given κ_j. In this case, when employment increases within a firm, the only way for the equation to hold is if the share of low-skill workers increases, which is not true if differences in the distribution of κ across the employment distribution are not controlled for, as in the specification of equation 14.

It is worth clarifying that having at least two degrees of firm heterogeneity is essential to obtain differing signs between micro and macro relationships between size and skill composition. The reason is that, with only one dimension of heterogeneity, comparing large firms to small firms is the same as comparing a firm to itself over time, because the cross-sectional heterogeneity between firms is the same as the temporal heterogeneity within a firm. As an example, suppose some exogenous process causes firms to vary their employment n over time, and that they maintain skill composition sc as a function of firm size n according to $sc = \frac{1}{1+n}$. Then the relationship is negative within each firm, but also must be negative in the cross-section, and in fact will be the exact same function. Therefore to explain the observed sign difference in the within-firm and cross-sectional relationships between size and skill composition, it is not merely enough to embed a firm technology that combines skills, such as Eeckhout and Pinheiro (2014), into a workhorse model of firm dynamics, such as Hopenhayn (1992); there must also be firm heterogeneity in the skill bias of technology across firms.

Related to the previous point, in order to generate the observed negative relationship between size and high-skill share at the firm level, it is essential in the above model that the firm dynamics are driven by TFP, the Hicks-neutral productivity shifter, because this leads to firms skilling down as they scale up, due to the relatively more elastic low-skill labor supply. If the model above instead stipulated that SBP κ were dynamic and z were fixed, then firms would increase their high-skill share as they grow, yielding the opposite prediction from what is observed. That a dynamic z is more in line with what I observe than a dynamic κ is not surprising, as this is what would occur if firms tend to become more efficient in general as they innovate, rather than simply more efficient in using skilled workers.

While the above model relies heavily on the differing labor supply elasticities between worker types to obtain that scaling and skilling are related, this result can also be achieved with alternative technological assumptions. Even in the case where $\eta_L = \eta_H$, so there is no supply-side motivation to skew towards low-skill workers, if the technology embeds some scaling dependence into skill-bias, for example by replacing κ with a function $\kappa(N)$, where $\kappa'(N) < 0$, then the results above will be recovered. In this case, firms still skew toward low-skill workers as they grow, but the reason is no longer simply that it becomes relatively harder to move up the high-skill supply curve, but rather that, at any given skill ratio, as the firm scales up, low-skill workers become relatively more marginally productive, compared to high-skill workers. This factor likely also plays a role in reality, for example in scaling a manufacturing operation. The first handful of PhD operations researchers are incredibly helpful in determining how to organize production, but as production is scaled up, more researchers add little to improving efficiency, whereas additional line workers remain approximately equally marginally productive.

6 The South Australian Payroll Tax Cut

The interdependence of scale and skill decisions for firms will mean that policies which cause a firm to alter its scale of employment may also cause firms to alter their skill composition. A recent policy change in South Australia provides a setting for assessing this prediction of the above model. I below explain the details of the policy change, augment the simple model to allow for taxes, use the augmented model to predict how firms should have reacted, and validate the model predictions empirically.

In 2018, the newly elected Liberal Party of South Australia introduced a policy which altered the payroll tax schedule for firms. In particular, prior to the January 1, 2019 implementation of the change, firms with a payroll exceeding \$650K faced a marginal tax rate of approximately 5% uniformly. After the policy went into effect, firms faced a 0% marginal rate up to \$1.5 million, at which point they faced a whopping 27% marginal tax rate up to a payroll of \$1.7 million, at which point the 5% rate resumed. I reproduce the figure of Andrews, Buckley, and Lee (2024) illustrating the change in the tax schedule in Figure 7.

While the ostensible purpose of the reform was to protect small businesses from prohibitive tax burdens, the policy provided a small incentive for growth of firms that would previously have operated with payrolls between \$650 thousand and \$1.5 million (where the marginal tax rate fell from 5% to 0%), and a large disincentive for growth for firms that would previously have operated with payrolls between \$1.5 million and \$1.7 million (where the marginal tax rate increased from 5% to 27%). In line with these predictions, Andrews, Buckley, and Lee (2024) find bunching around the \$1.5 million threshold, payroll increases for firms that saw a decrease in their marginal tax rate, and payroll decreases for firms that saw an increase in their marginal tax rate. They also find evidence that the policy reduced worker earnings, in line with the predictions and findings of Garicano, Lelarge, and Van Reenen (2016), which considered a similar policy in France.

The policy should also have implications for firm skill composition and skill premia. To make this point, I extend the above model by assuming that firms additionally pay taxes on their payroll. The payroll tax schedule is described by the function $\mathcal{T}$, which maps from the

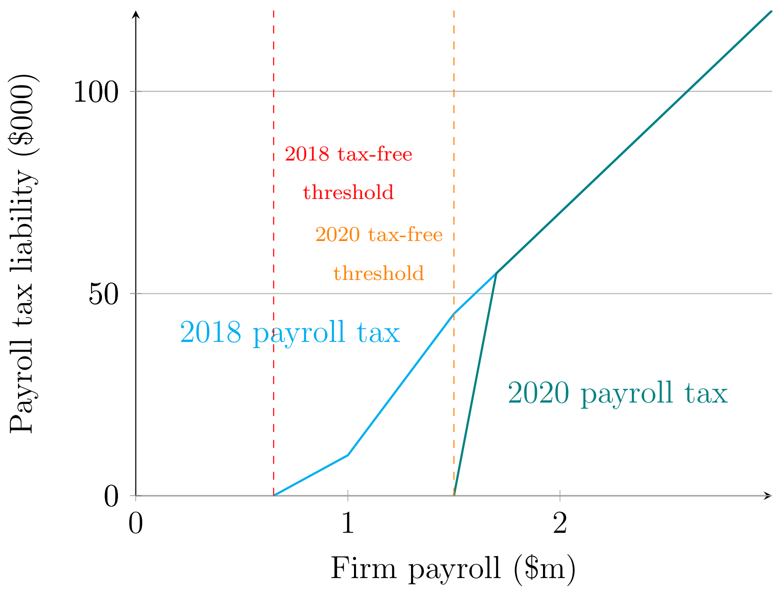


Figure 7: Illustration of the payroll tax schedules faced by firms in South Australia in 2018 and 2020. Reproduced from Andrews, Buckley, and Lee (2024).

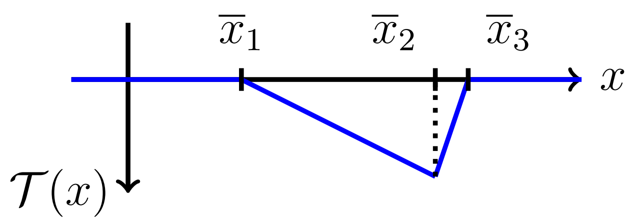


Figure 8: Tax schedule gives tax breaks to medium-sized firms. The schedule is regressive over a large set of payrolls, then sharply progressive over a small set of payrolls.

pre-tax payroll of a firm to the payroll tax, so that if a firm's total pre-tax wage bill were x, then their total post-tax bill is $x + \mathcal{T}(x)$. I consider the following tax schedule, which is visualized in Figure 8.

$$\mathcal{T}(x) = \begin{cases} 0 & x \leq \overline{x}_1 \\ a_1(x - \overline{x}_1) & \overline{x}_1 < x \leq \overline{x}_2 \\ a_1(\overline{x}_2 - \overline{x}_1) + a_2(x - \overline{x}_2) & \overline{x}_2 < x \leq \overline{x}_3 \\ a_1(\overline{x}_2 - \overline{x}_1) + a_2(\overline{x}_3 - \overline{x}_2) & x > \overline{x}_3 \end{cases} \tag{16}$$

I additionally impose $x_2 = \frac{a_2\overline{x}_3 - a_1\overline{x}_1}{a_2 - a_1}$. The policy $\mathcal{T}$ subsidizes small-to-medium-sized firms, but does so by lowering the marginal tax rate over a large region, then increasing the marginal tax rate over a small region, exactly analogous to the tax change in Australia, as all that matters for firm decisions are the changes in the marginal tax rates. To maintain simplicity, I assume the policy is funded by lump-sum taxes, as discussed in Appendix A.8.

This policy change will have non-uniform effects in altering firm employment decisions.

For some small firms, the tax break makes operating at a higher scale profitable, where it previously was not. For some large firms, it is now more profitable to descale to take advantage of the sharp tax break. The following proposition, proven in Appendix A.7, makes these points precise in partial equilibrium.

Proposition 2. Following a policy change from $\mathcal{T}_0 = 0$ to the $\mathcal{T}$ specified above, the partial equilibrium response of firms, that is, the response that keeps wage indices at pre-policy levels, may be characterized as follows. I use N as the measure of employment within a firm.

(i) Firms with payrolls less than $\overline{x}_1$ will either increase or not change their employment

(ii) Firms with payrolls in $[\overline{x}_1, \overline{x}_2)$ will increase their employment

(iii) Firms with payroll at $\overline{x}_2$ will not change their employment

(iv) Firms with employment in $(\overline{x}_2, \overline{x}_3]$ will decrease their employment

(v) Firms with payrolls greater than $\overline{x}_3$ will either decrease or not change their employment

The proposition only speaks to partial equilibrium outcomes, as the general equilibrium effects on wage indices are not simple to characterize, because they depend on the distribution of firms, in addition to the parameters governing the tax policy, and the elasticities of labor supply. Even absent monopsonistic forces (i.e., if $\eta_L, \eta_H \to \infty$), the effects are unclear, as labor demand from small firms will increase, and labor demand from large firms will decrease, so equilibrium wages for each worker type could rise or fall[13].

The partial equilibrium results are still useful empirically. Firms in $[\overline{x}_2, \overline{x}_3]$ should reduce their employment, and the only possible concern is that the change in relative wages for skill types could push in the opposite direction, for example if in the new $\frac{W_H}{W_L}$ is much lower. This seems unlikely to occur, because small firms scaling up will increase relative demand for high-skill labor (recall that skill premia rises with scale), so for some large firms to also scale up would need to mean that intermediate large firms scaled down to such an extent as to reduce aggregate demand for high-skill labor. This does not seem plausible, and I suspect it is impossible in this setting, though I have not been able to prove so.

Given the scaling results from above, and the relationships between scale, skill composition, and skill premia, I make the following predictions about the impact of the South Australian tax policy change:

(i) Firms facing marginal tax rate increases should decrease their total employment, increase their skill ratio, and decrease their skill premia.

(ii) Firms facing marginal tax rate decreases should increase their total employment, decrease their skill ratio and increase their skill premia.

[13] Garicano, Lelarge, and Van Reenen (2016) are able to obtain tighter results about GE, because they consider a tax change that only penalizes firms becoming larger. The $\mathcal{T}$ above, on the other hand, rewards some firms for becoming larger, while punishing others, relatively speaking. Their focus is also entirely on the scale margin.

All other firms will be affected by the policy as well, however, via the general equilibrium net increase or decrease in labor demands. If sufficiently many firms are directly affected by the tax decrease, then aggregate labor demand should increase, driving up wages for all workers. If sufficiently many firms are directly affected by the tax increase, then aggregate labor demand should decrease, driving wages down. Also note that the convexity in the tax change will be relevant: if equal shares of firms saw the tax increase and tax decrease, a decrease in labor demand would be expected, as the increase is much more extreme than the decrease. Nonetheless, for the purpose of making predictions I simply use the partial equilibrium results, operating under the assumption that any countervailing general equilibrium force should not dominate in this setting, an assumption which I above argued is plausible.

To test these hypotheses, I run the following specification, where y is skill ratio, skill premium, or total employment at the firm level, Δ is a two-year difference, $s(j)$ is the state of firm j, SA is the state of South Australia, and t is time.

$$\Delta \ln y_{jt} = \delta_{s(j)} + \delta_t + \beta \mathbf{1}\{s(j) = SA, t = 2020\} + \epsilon_{jt} \tag{17}$$

The results are in Table 2. The identifying assumption is that there may be permanent differences between South Australia and the rest of the Australian states in y, and there may be differences over time in y, but the only difference between South Australia and the rest of Australia which changed in fiscal year[14] $2019-2020$ was the change in payroll tax policy. Under this assumption, the coefficient β is the impact on the growth of y due to the tax change. A two-year difference is used, and the treatment period is fiscal year $2019-2020$. Both of these assumptions stem from the assumption that some degree of adjustment frictions exist in reality, therefore directly comparing $2018-2019$ (when the policy was only in effect for 6 months) to $2017-2018$ might attenuate the results, but comparing $2019-2020$ to $2017-2018$ allows for sufficient adjustment. I estimate the equation separately for the subset of firms with payroll between 1.5 and 1.7 million in 2018, and firms with payroll less than 1 million or more than 2 million. The first subset is designed to capture the set of firms affected by the increase in marginal tax rates, under the assumption that these firms would mostly operate around the same employment in 2020 if not for the policy change. The second subset is designed to capture the set of firms that should not be affected by the marginal tax increase, but which could be affected by the marginal tax decrease. I exclude the region of firms with payrolls from 1 to 1.5 million or 1.7 to 2 million from either sample because it is unclear how the policy might affect this subset, which teeters on the edge of being affected by the marginal tax hike.

I start by considering the effect on the set of firms with $2017-2018$ payrolls between \$1.5 and \$1.7 million, which saw an increase in their marginal tax rate. The results are reported in columns 1, 3, and 5 in Table 2. In the first column of Table 2, I find an increase in the average log skill ratio, in line with the prediction (i) above. In the third column, I verify that these firms did decrease their total payroll. In the fifth column, I find that the skill premium declined, also in line with prediction (i). Taken together, these results are in line with the interpretation given by the model, that these firms chose to adjust their payroll downward in response to the marginal tax hike, and in doing so they also adjusted their workforce

[14] Recall that an Australian fiscal year is July - July.

Dependent Variables:	Δ ln(High Skill Ratio)	Δ ln(High Skill Ratio)	Δ ln(Employment)	Δ ln(Employment)	Δ ln(Skill Premium)	Δ ln(Skill Premium)
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
$\mathbf{1}\{\text{state} = SA, \text{time} = 2020\}$	0.1483***	-0.0010	-0.0339***	0.0175***	-0.0673***	0.0038
	(0.0124)	(0.0044)	(0.0063)	(0.0039)	(0.0057)	(0.0035)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
state	Yes	Yes	Yes	Yes	Yes	Yes
Sample	$[1.5, 1.7]$	$[0, 1] \cup [2, \infty)$	$[1.5, 1.7]$	$[0, 1] \cup [2, \infty)$	$[1.5, 1.7]$	$[0, 1] \cup [2, \infty)$
Fit statistics						
Observations	13,788	720,494	15,772	2,017,076	13,623	673,834
R^2	0.01594	0.00302	0.00289	0.00026	0.00148	3.61×10^{-5}
Within R^2	0.00045	1.3×10^{-8}	3.16×10^{-5}	6.75×10^{-6}	0.00011	1.41×10^{-7}

Clustered (state) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2: Results from difference-in-difference specification specified in equation 17. The Sample row indicates the firm's payroll in 2018, in millions of dollars, and the operator Δ is a two-year difference.

composition, substituting towards a more skilled workforce, which is more cost-effective at smaller scales.

Now I consider the effect of the policy on the firms with payrolls less than \$1 million or greater than \$2 million. These effects of the tax change should be much weaker for this group for two reasons. First, firms with payrolls of less than \$650 thousand or more than \$1.7 million see no policy change, so the only effect on these firms comes from changes in their production process which would otherwise move them into a treated region, and general equilibrium effects, both of which should not be as strong as the direct effects for the above subset of exclusively treated firms. Second, even for the firms with payrolls between \$650 thousand and \$1 million, the marginal tax decrease (5% to 0%) is much less significant than then increases for the treated subset above (5% to 27%). As expected, in columns 2 and 4 I find null results for the skill ratio and skill premium, respectively. Additionally, in column 4, I find a muted but positive effect for total employment, in line with the prediction (ii).

One potential concern with the specification in equation 17 is the COVID pandemic affecting Australia in early 2020. The time fixed effects control for uniform effects of the pandemic nationally, so the only way for the pandemic to affect β would be if it differentially affected South Australia, or if the pandemic interacted with the tax change. In these cases, the coefficient would pick up the combined effect of the tax policy change, the region-specific effects of COVID, and the interaction effect of COVID differentially affecting South Australia due to the tax change. While further analysis into how the varied sectoral incidence of the pandemic potentially translated into varied geographic incidence is warranted, it is not obvious that the industrial composition of South Australia is different enough from the rest of Australia to generate large effects. Thus, I do not view the pandemic as a significant threat to the identification strategy.

While not the focus of this paper, these results speak to the concern that the distortive effect of income taxation may be amplified in environments featuring monopsony, raised by Berger et al. (2024). They point out that a progressive tax schedule causes firms to face a more inelastic effective labor supply, and thus reduce employment. In an environment where firms choose choose not only the scale of their workforce, but also the composition, changes in the progressivity of the payroll tax schedule will have uneven impacts across worker types. When scale and skill are inversely related at the firm level, as I find empirically, a more progressive payroll tax schedule will decrease the wages of all workers, but high-skill workers will see a more dramatic decline, as the within-firm skill premia will fall. The typical argument for using progressive taxes is to reduce inequality between workers, and Storesletten, Heathcote, and Violante (2020) use a framework with heterogeneous workers but a representative firm to clarify that a planner should account for whether income differences are due to skill or residual inequality when determining the progressivity of a tax. Accounting for differences in firm hiring across skills reveals an additional motivation for progressive taxes for a utilitarian planner, in that the progressivity will induce firms to stay smaller and lower their internal skill premia. I see better understanding the effect of payroll taxes on earnings between workers as an interesting avenue for future research.

7 Quantitative Model

Above, I established the empirical link between firm scale and skill composition, provided a model to explain the cross-sectional and within-firm patterns, and provided evidence in support of the model's key assumptions. I now want to understand how the rising share of educated workers in the Australian economy will affect the distributions of employment and skill premia across firms. To do so, I now build a richer version of the model by allowing firms to differ in the labor supply curves they face, in addition to TFP and SBP, and allowing for an arbitrary distribution over firm-level parameters. First, I explain the model and define equilibrium. Second, I show that within a restricted subset of the space of distributions, I can invert the model to obtain the unique distribution which exactly matches the above patterns of employment, skill ratio, and skill premia, by firm size in Australia. Finally, I consider a counterfactual increase in the share of skilled labor in the aggregate economy and examine the impact on the distributions of employment and earnings across firms.

7.1 Setup

Firms A measure M of firms operate. Each firm has total factor productivity parameter z, skill-biased productivity parameter κ, and faces a labor supply shifters A_L and A_H. The parameters are jointly distributed according to measure G, so that $\int_{(z,\kappa,A_L,A_H)} \mathrm{d}G(z,\kappa,A_L,A_H) = M$.

Workers A unit measure of workers earn labor income and profit income from owning the firms. Share $\overline{n}_L$ of workers are low skill, and $\overline{n}_H = 1 - \overline{n}_L$ are high-skill. Each worker supplies one unit of labor inelastically and has CES preferences over jobs, so solves the following problem, where F is the distribution over wages and supply shifters A.

$$\max_{n(w,A)} \left(\int (wAn(w,A))^{\frac{\eta}{\eta+1}} \, \mathrm{d}F(w,A) \right)^{\frac{\eta+1}{\eta}}$$
$$1 = \int n(w,A) \mathrm{d}F(w,A)$$

All workers earn the same profit income from an equal ownership share in all firms[15]. Worker utility is increasing in goods, and there is no savings technology, so workers spend their entire income on goods.

Technology Firms operate the same technology as in the simple model, in particular a constant returns to scale technology which aggregates labor bundles[16] according to an aggregator with constant elasticity of substitution σ.

$$Y(z, \kappa, n_L, n_H) = zN(\kappa, n_L, n_H) \tag{18}$$
$$N(\kappa, n_L, n_H) \equiv \left(n_L^{\frac{\sigma-1}{\sigma}} + (\kappa n_H)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} \tag{19}$$

All firms produce a homogeneous good which serves as the numeraire.

Labor Supply A firm with supply shifter A_i faces the following constant elasticity of substitution labor supply curve for type i.

$$n_i(w) = \left(\frac{A_i w}{W_i} \right)^{\eta_i} \overline{n}_i \tag{20}$$
$$\tag{21}$$

The appropriate wage indices are

$$W_i = \left(\int (A_i w)^{\eta_i} \mathrm{d}F_i(w, A_i) \right)^{\frac{1}{\eta_i}} \tag{22}$$

where F_i is the joint distribution over wages and supply shifters A. In C.1 I derive the above supply schedule from the worker problem, and in C.2 I explain how to augment the microfoundations from A.2 to generate these supply curves in another way. The A shifters are a means of explaining residual labor supplied to firms, and are useful for quantification because they will allow an exact fit of the model to the data along the key dimensions of

[15]It does not end up mattering how the profits are distributed across workers, as long as the profit distribution does not affect labor supply decisions.

[16]I abstract from capital for simplicity, so capital effects will show up through firm TFP, SBP, or labor supply shifter in the model inversion. Integrating capital into this framework is interesting future work, as the skill bias of capital is not clear. For example, Krusell et al. (2000) argue that capital is complementary to skill, but Bloom et al. (2014) clarify that information tech improvements will increase the demand for low-skill workers, whereas communication tech improvements will increase the demand for high-skill workers.

firm scale, skill ratio, and skill premia. The economic motivation for including the A shifters is that firms may differ in how they attract workers, holding constant the offered wage. One reason could be amenities which are not captured in the wage, as discussed in detail by Sorkin (2018). Another potential reason is that firms differ in their search costs, so firms which may post vacancies at a lower cost post more, and obtain more workers for any offered wage[17]. I do not take a stand on the exact cause of the variation in A in the Australian context, but believe unpacking this dimension of heterogeneity is interesting future work.

Optimality Firms maximize profits, so choose wages for each worker type to be a markdown $\frac{\eta_i}{\eta_i+1}$ from marginal revenue productivity of labor. The optimality conditions are in C.3.

Equilibrium The market equilibrium is the pair of functions of wages $w_i(z, \kappa, A_L, A_H)$ and employment levels $n_i(z, \kappa, A_L, A_H)$ such that

(i) Given the wage indices implied by the wage functions and distribution of firms G, wage $w_i(z, \kappa, A_L, A_H)$ implies labor quantity supplied $n_i(z, \kappa, A_L, A_H)$ for all firm types and skill types

(ii) Given the wage indices implied by the wage functions and distribution of firms G, firm (z, κ, A_L, A_H) optimally chooses wages $w_i(z, \kappa, A_L, A_H)$.

(iii) The goods market clears: Demand for goods by workers equals supply of goods by firms

The last condition will clear by Walras' Law, and is trivial to see will hold when the labor market clears: all firm income goes to either labor or profits, and this is exactly worker income, so they will demand the quantity of goods which is supplied. Proving the existence and uniqueness of an equilibrium is currently out of reach, as the optimality conditions for the firm do not yield closed-form results which could potentially be used to reduce the equilibrium to a pair of Hammerstein integral equations for the wage index, similarly to the proof approach in Allen and Arkolakis (2014). Nonetheless, every numerical simulation has converged to a solution via a tatonnement process, and the solution has remained invariant to the start value. I explain in more detail how to solve the model, below.

7.2 Calibration

Given parameters $\{\eta_L, \eta_H, \overline{n}_L, \overline{n}_H, M, \sigma\}$ and distribution $G(z, \kappa, A_L, A_H)$ the equilibrium of the model will imply a distribution of firm sizes, distribution of employment over firms, distributions of skill ratios, distribution of earnings, and thus distribution of skill premia over firms. The inverse problem is to find parameters and a distribution G which yield the above outcomes that are observed in the data, in particular the distributions of firms and employment in Figure 16, the college skill ratio in Figure 17, the high school and college earnings in Figure 18, and the skill premia in Figure 19. The model is not non-parametrically identified, as the dimension of G is larger than the dimension of the observed outcomes, so I impose sufficiently restrictive parametric assumptions on G to recover identification.

[17] I posit a theory along these lines in C.4.

There are potentially many ways to restrict the state space of G, so I suggest one which admits simple inversion of model parameters[18] [19]. In particular, I restrict the support of G to only have n points, where n is the dimension of each set of statistics I wish to match. Then each point in the support corresponds to a given firm measure, employment measure, skill ratio, earnings for low-skill workers, and skill premium. I now detail the inversion.

The first step is to calibrate the parameters for labor supply elasticities, firm measure, and elasticity of substitution between skill types in production[20]. I set $\eta_L = 0.40$ and $\eta_H = 0.39$ to match my estimates from above, and set $M = 0.69$ so that the average number of workers per firm is ≈ 15, matching the data. I set $\sigma = 2.2$ as the midpoint of the range suggested by Jerzmanowski and Tamura (2023).

I will use $n = 10$, so each of the employment bins from the empirical sections above correspond to one of the mass points. Within each bin, I calculate the earnings of high-skill workers using the earnings of low-skill workers and scaling by the skill premium. I could directly use the earnings of high-skill workers instead, and infer the skill premium in each bin, but I want to target the average skill premium for a firm in each bin, not just the average skill premium in the bin. In practice these two approaches do not differ much. I calculate the number of workers of each type employed using the total number of workers employed, $n_L + n_H$, and the skill ratio, $\frac{n_H}{n_L}$, as

$$n_L = \frac{n_L + n_H}{1 + \frac{n_H}{n_L}} \tag{23}$$

$$n_H = \frac{(n_L + n_H)\frac{n_H}{n_L}}{1 + \frac{n_H}{n_L}} \tag{24}$$

In this step I also calculate the skill share using the employment levels in each bin and the share of firms in each bin: $\overline{n}_i = \int n_i(z, \kappa, A_L, A_H) \mathrm{d}G(z, \kappa, A_L, A_H)$. Note that, because I have restricted the state space and the data give the firm distribution, I do not need to

[18] In this empirical approach, of selecting and inverting a distribution of parameters to exactly match a set of empirical moments of interest, I take motivation from the quantitative spatial literature, e.g. see Ahlfeldt et al. (2015) and Rosenthal-Kay (2024) for urban settings, Allen and Arkolakis (2014) for a trade setting, Monte, Redding, and Rossi-Hansberg (2018), Sarte et al. (2014) for a regional setting, Monte, Redding, and Rossi-Hansberg (2018) for a setting with regional trade and commuting, Allen and Arkolakis (2022) for a setting with transportation infrastructure, and Redding and Rossi-Hansberg (2017) for a review of this class of models. The unifying theme is that these models contain enough free parameters to exactly match a large number of moments across space. Similarly, I include enough points in the support of G, and include the supply shifters A, to exactly match moments across the firm size distribution.

[19] In Appendix D I consider a variation of the present model allowing for firm dynamics in TFP, and entry and exit. This variation of the model then requires not directly calibrating the distribution of firms, but calibrating the distribution of entering firms, and the stochastic process firms face. In Appendix E I additionally consider microfounding the upward-sloping labor supplies with worker search. This variation then requires additionally calibrating the parameters governing worker and firm search. In ongoing work I am utilizing these extended models to think about how worker and firm dynamics are impacted by changes in the distributions of worker skills or firm skill-biased productivities.

[20] The skill shares $\overline{n}_i$ will be pinned down by the distribution of total employment and skill ratios below. This approach for calculating the skill shares yields $\overline{n}_L = 0.39$, which is slightly different than the 0.41 measure from directly looking at the share of workers in the sample, due to some bias from averaging across logs.

know what the (z, κ, A_L, A_H) are at each point for the previous calculation, as I only need to know the distribution and the employment levels.

I can now find the labor supply shifters, A_i. The supply equations link earnings, employment levels, the distribution of employment, all pieces that have already been computed, with A, so I find the A for each firm that satisfies the supply equations. In principle, this would require solving a nonlinear system of equations, as the wage index depends on the A choice, but A is only determined up to scale, so I pick the scale of A that makes the wage index 1, and solve for A as

$$A = \left(\frac{n}{\overline{n}}\right)^{\frac{1}{\eta}} \frac{1}{w} \tag{25}$$

The final part of the inversion is the only step where the parameters cannot be recovered in closed-form from the data. To obtain TFP and SBP at each point, I solve for the (z, κ) which are consistent with the firm's optimality conditions, given the levels of employment and A calculated above. The equations at each point are[21]

$$z\left(\frac{n_L}{N}\right)^{-\frac{1}{\sigma}} = \frac{\eta_L + 1}{\eta_L}\left(\frac{n_L}{\overline{n}_L}\right)^{\frac{1}{\eta_L}} \frac{W_L}{A_L} \tag{26}$$

$$z\kappa^{\frac{\sigma-1}{\sigma}}\left(\frac{n_H}{N}\right)^{-\frac{1}{\sigma}} = \frac{\eta_H + 1}{\eta_H}\left(\frac{n_H}{\overline{n}_H}\right)^{\frac{1}{\eta_H}} \frac{W_H}{A_H} \tag{27}$$

$$N = \left(n_L^{\frac{\sigma-1}{\sigma}} + (\kappa n_H)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \tag{28}$$

7.3 Inversion Results

I present the results of the inversion. In Figure 9 I show the recovered labor supply shifters for each skill type. The A_i are monotonically increasing and fairly similar between types, though larger firms see slightly higher levels of relative high-skill supply than small firms. The reason for this positive relationship is that the labor supply elasticities and the difference in earnings between large and small firms is not enough to justify their massive differences in employment, alone. By imposing that labor supply for both worker types is highly inelastic, I am forced to infer that the supply shifters are largely responsible for the differences in employment between firms. Assuming larger values for η_i dampens the gradient in $\ln A_i$, because then the supply elasticity is able to generate more responsive employment to the earnings differences.

More interesting to note than the A patterns are the z and κ patterns in Figures 10 and 11. Small firms tend to have low TFP but high SBP, but medium-sized firms have higher TFP and lower SBP. Therefore, while medium-sized firms are more productive in general, which allows them to become larger, the smallest firms are better at using high-skill labor, so they employ a higher skill ratio. The high-skill workers at small firms are not paid larger skill premia, however, as the TFP in these firms is small, so the difference in productivity

[21] I include the W_i for completeness, but recall the normalization of A above implies $W_i = 1$.

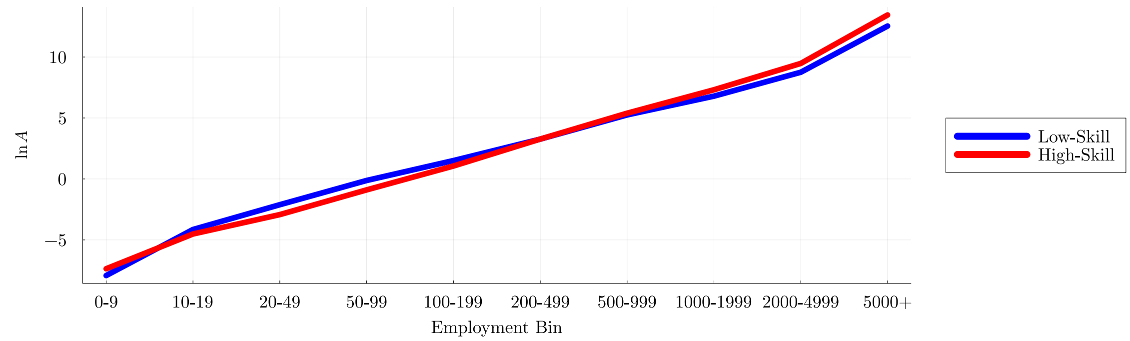


Figure 9: Recovered (log) labor supply shifters A by equilibrium firm size and worker type

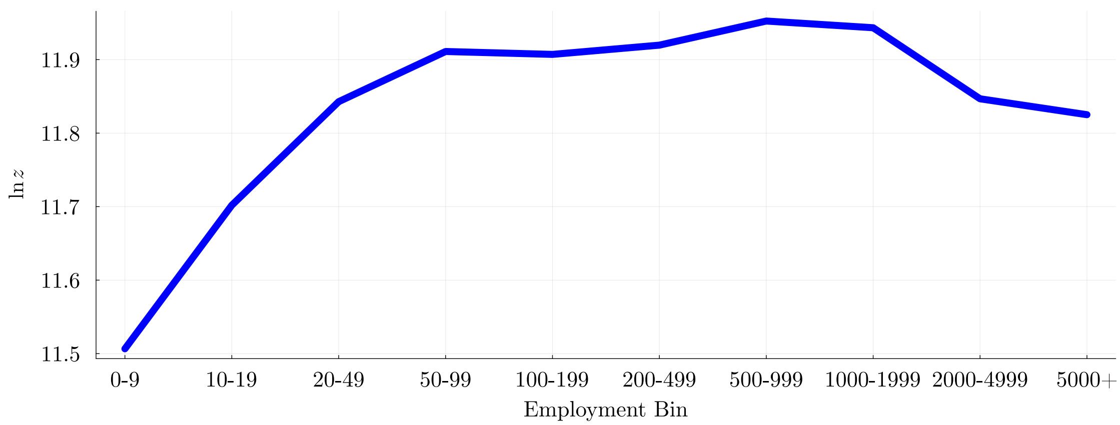


Figure 10: Recovered (log) total factor productivity z by equilibrium firm size and worker type

between low- and high-skill workers is not as amplified. While the medium-sized firms have lower SBP, their higher TFP is enough to amplify worker differences and lead to a higher skill premium than in the small firms.

The comparison between medium-sized firms and large firms is most striking, and is consistent with the notion that the largest firms are not simply the most productive, but the most productive with high-skill labor. While the largest firms do not have noticeably higher TFP, and in fact the absolute largest firms with 2000+ workers have lower TFP than slightly small firms, they do have higher SBP. This higher SBP is a force toward employing a higher skill ratio, and for the large firms this force dominates the force toward a lower skill ratio which comes from increases in TFP.

The above insights paint a novel picture of how firm size and productivity are related: medium-sized firms are more productive than small firms in a general sense, but large firms are not more productive than medium-sized firms in a general sense, but only more productive in using high-skill labor. An implication of this result is that it is potentially possible to understand the heterogeneity in firm growth rates noted by Luttmer (2011) in the context of firms having permanent differences in SBP, but facing stochastic TFP. Firms with small SBP may only choose to enter if they have sufficiently high TFP, and these firms will tend to stay medium-to-small for their entire lifetime. In contrast, firms with high SBP may be

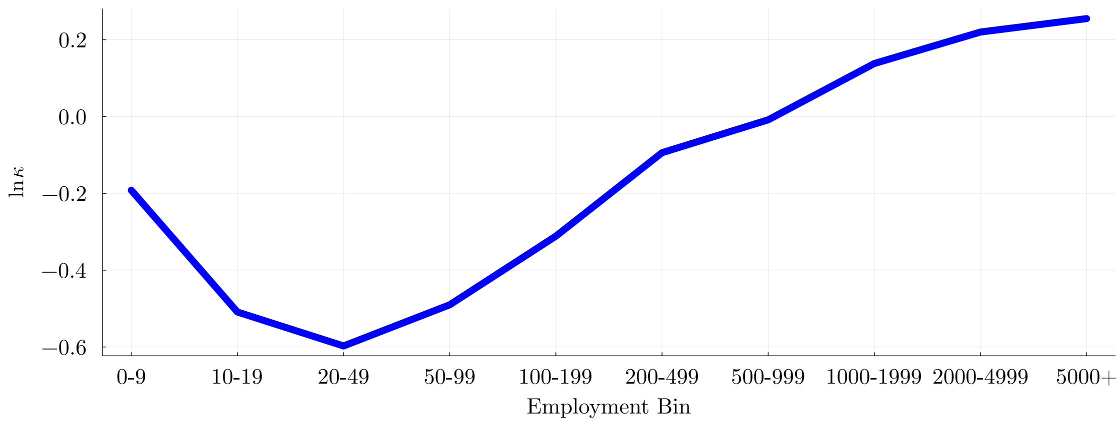


Figure 11: Recovered (log) skill-biased productivity κ by equilibrium firm size and worker type

willing to enter even with low TFP, and when their TFP grows they also grow to become large firms, as their high level of SBP supercharges the TFP growth. This also gives some potential context to the findings of Hurst and Pugsley (2011) that many small businesses never grow. If an entrepreneur wishes to enter a market, but they are not effective at using high-skilled labor, then they may choose to never scale beyond being medium-sized.

7.4 Counterfactual: Skilling Up the Workforce

The share of the Australian population with a bachelor's degree or higher increased from 24% in 2011 to 30% in 2020. This rise in high-skill supply, holding fixed technological changes, should lead to changes in the distribution of employment across firms, as firms which are particularly adept at utilizing high-skill labor will hire more, and also choose to scale up their operations by hiring more low-skill labor. To quantify the magnitude of these effects, I consider an increase in the share of high-skill workers from 39% to 50%, a not unreasonable projection for what the educated share of the workforce might be in Australia the next few decades.

Mechanically, I raise $\overline{n}_L$ from 0.39 to 0.5, then solve the new parameterization of the model, not changing any of the other parameters nor the firm distribution across (z, κ, A_L, A_H) from the inverted values found above. This change in $\overline{n}_i$ means that the supply of high-skill workers has increased, and the supply of low skill workers has decreased. As explained in the introduction, these forces have competing effects. On one hand, the adundance of high-skill labor will make firms want to scale up total employment, but on the other hand the scarcity of low-skill labor will make firms want to scale down. A firm's SBP will determine whether the combined effects lead to a net reduction in marginal costs, so the firm scales up, or net increase in marginal costs, which will cause the firm to descale. The results in Figures 12 and 13 are then unsurprising: employment moves toward the firms with the relatively larger κ, and those are the smallest and largest firms. The results are quite large within firms, as documented in 12 with the smallest firms increasing employment by 2%, the largest firms increasing employment by 3%, and some medium-sized firms seeing employment declines of 5%. In Figure 13 I instead consider how total employment in each bin changes, and find

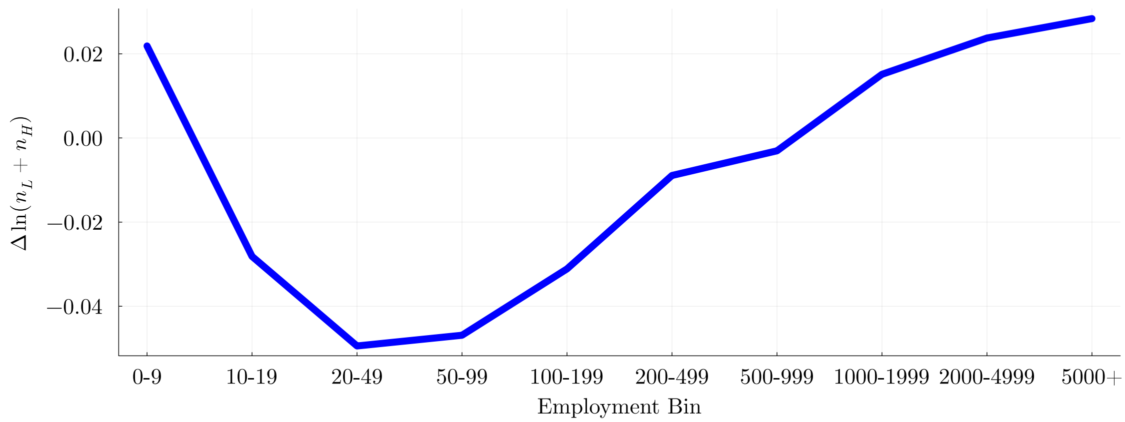


Figure 12: Change in log employment within each firm by baseline firm size, following increase in high-skill share of labor force

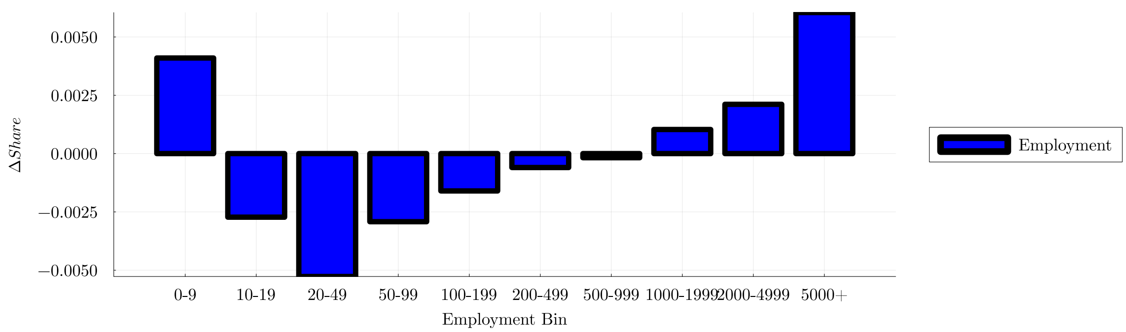


Figure 13: Change in share of total employment within each firm size bin, following increase in high-skill share of labor force

that 0.4% of employment moves toward the smallest firms, and 0.9% of employment moves toward the largest firms (1000+ employees).

The skill shift increases average earnings for both skill types, decreases skill premia within all firms, but increases the aggregate skill premium. The first result occurs because the addition of high-skill workers and removal of low-skill workers is not zero-sum; with more high-skilled workers, the productive capacity of the economy has increased. Some of these gains accrue to high-skill labor, and some accrue to low-skill labor who are now more marginally productive, and the remainder of the gains are firm profits. The decrease in skill premia within every firm is also not surprising, and follows from the simple economic logic that the supply curves facing the firm for low- and high-skill labor shift in opposite directions, so the relative cost of high skill labor falls, as in Katz and Murphy (1992). I display the changes in earnings by firm size, broken by skill, in Figure 14.

The rise in the aggregate skill premium is surprising, given the fall in skill premia within every firm. The key is that the distributions of low- and high-skill labor across firms are not the same, and the magnitudes of earnings vary across firms. Firms which tend to employ more high-skill workers in the baseline see larger absolute gains in earnings in the counterfactual. Additionally, firms which see the largest rise in employment (by all skill types) have the largest difference between low- and high-skill wages in absolute terms in the

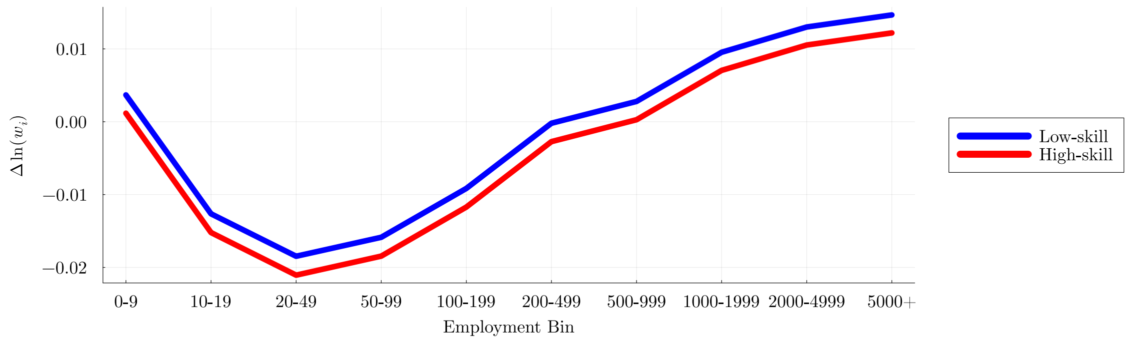


Figure 14: Change in log earnings within each firm by baseline firm size and skill type, following increase in high-skill share of labor force.

baseline. Lastly, there is an interaction term, in that the firms which see the largest increases in employment also see the largest increases in wages. These statements can be made precise in the following decomposition of changes in the aggregate wage for skill type i

$$\underbrace{\sum w_i^1 n_i^1 g_i - \sum w_i^0 n_i^0 g_i}_{\text{agg}} = \underbrace{\sum (w_i^1 - w_i^0) n_i^0 g_i}_{\text{wage}} + \underbrace{\sum w_i^0 (n_i^1 - n_i^0) g_i}_{\text{distribution}} + \underbrace{\sum (w_i^1 - w_i^0)(n_i^1 - n_i^0) g_i}_{\text{interaction}} \tag{29}$$

where i indexes a firm type. The decompositions for change in earnings by each labor type are as follows, in annual Australian dollars

$$\underbrace{101}_{\text{agg } L(100\%)} = \underbrace{71}_{\text{wage } L(70\%)} + \underbrace{27}_{\text{distribution } L(27\%)} + \underbrace{4}_{\text{interaction } L(4\%)} \tag{30}$$

$$\underbrace{187}_{\text{agg } H(100\%)} = \underbrace{133}_{\text{wage } H(71\%)} + \underbrace{51}_{\text{distribution } H(27\%)} + \underbrace{4}_{\text{interaction } H(2\%)} \tag{31}$$

The decomposition indicates that the majority of earnings gains comes from increases in worker earnings at firms which employ more workers, one-quarter of the gains come from workers reallocating towards firms which pay more, and a tiny contribution is due to the firms with the largest earnings increases gaining workers. The aggregate growth in earnings for high-skill workers is 0.3%, and the growth for low-skill workers is 0.2%, hence the skill premium rises by 0.1%.

The rise in aggregate skill premium in response to an increase in high-skill share of workers means the economy exhibits a notion of what Acemoglu (2007) terms "strong absolute equilibrium" skill bias. The original idea is that productivity of factors responds to changes in the supplies of factors, through changes in technological innovation decisions, and that if the responses are strong enough, an increase in supply of a factor leads to an increase in its price. In the present model, the increase in the relative average wage of high-skill workers in response to their increase in supply occurs because supply changes affect labor demand differently across the distribution of firms. Firms which have higher skill-biased productivity react by scaling up their employment of all worker types, and this increases the average earnings of high-skill workers more than low-skill workers, because these same firms tend to pay higher skill premia. The theoretical result that an increase in skill may induce enough reallocation of labor to generate an increase in the skill premium is interesting in its own right, but the above empirical exercise documenting that such an outcome could plausibly occur in Australia in coming decades provides an even more pressing motivation for the present study.

8 Conclusion

I have argued empirically and theoretically that the distribution of worker skill impacts the distribution of firm sizes. Empirically, small and large firms tend to be the most skilled, but this is entirely due to fundamental firm differences: within a firm, growth is associated with deskilling. I provided a simple model to make sense of these findings, wherein firms differ in both total factor productivity (TFP) and skill-biased productivity (SBP), and dynamics in TFP generate the within-firm negative relationship between size and skill, but the distribution of SBP across firms may generate a positive relationship. I supported the model's assumptions by estimating labor supply elasticities, which I found to be qualitatively in line with what the theory required. I made use of a natural policy experiment to support the predictions of the model. Finally, I quantified an augmented version of the model to infer the distribution of TFP and SBP across firm sizes, and found that the key difference between small and medium-sized firms is TFP, but the key difference between medium-sized and large firms is SBP. I used the quantitative model to predict what the effect of rising education shares in the Australian workforce may mean for the distribution of employment and earnings across firms. I found that employment and earnings increase for the firms with highest SBP, leading to 1% of employment shifting toward large firms and an aggregate increase in the skill premium, despite decreases in skill premia within each firm.

The labor market is rapidly evolving to adapt to technological change. Workers are becoming more educated, and innovations in information, communications, and technology (ICT), including the ascent of artifical intelligence (AI) are changing the productive potential of workers and firms. These advancements will no doubt alter both the skill supply of workers and the labor demand for workers across firms. By demonstrating the link between firm size and skill composition, I have clarified that we should expect the impact of these technologies on the productive potential of workers to impact the concentration of workers across firms. But I have only scratched the surface, and I look forward to continued work improving our understanding of how changes in the skill composition of workers, and the technological capacities of firms, will impact the distributions of employment and earnings.

References

Abowd, John M, Francis Kramarz, and David N Margolis (1999). "High wage workers and high wage firms". In: *Econometrica* 67.2, pp. 251–333.

Acemoglu, Daron (2001). "Good jobs versus bad jobs". In: *Journal of labor Economics* 19.1, pp. 1–21.

— (2002). "Technical change, inequality, and the labor market". In: *Journal of economic literature* 40.1, pp. 7–72.

— (2007). "Equilibrium bias of technology". In: *Econometrica* 75.5, pp. 1371–1409.

Ahlfeldt, Gabriel M et al. (2015). "The economics of density: Evidence from the Berlin Wall". In: *Econometrica* 83.6, pp. 2127–2189.

Allen, Treb and Costas Arkolakis (2014). "Trade and the Topography of the Spatial Economy". In: *The Quarterly Journal of Economics* 129.3, pp. 1085–1140.

— (2022). "The welfare effects of transportation infrastructure improvements". In: *The Review of Economic Studies* 89.6, pp. 2911–2957.

Anderson, Simon P, André De Palma, and Jacques-Francois Thisse (1987). "The CES is a discrete choice model?" In: *Economics Letters* 24.2, pp. 139–140.

Andrews, Dan, Jack Buckley, and Rachel Lee (Aug. 2024). *A Counterproductive Tax Cut? How Size-Based Payroll Taxes Can Create a Roadblock to Firm Growth.* Micro Note 23. Accessed: 2024-10-10. e61 Institute. URL: https://e61.in/wp-content/uploads/2024/08/A_Counterproductive_Tax_Cut_.pdf.

Andrews, Dan et al. (2019). *Wage growth in Australia: Lessons from longitudinal microdata.* Tech. rep. Treasury Working Paper.

Australian Bureau of Statistics (2006). *Australian and New Zealand Standard Industrial Classification (ANZSIC), 2006 (Revision 2.0).* Accessed: 2024-10-10. URL: https://www.abs.gov.au/statistics/classifications/australian-and-new-zealand-standard-industrial-classification-anzsic/2006-revision-2-0/nature-and-objectives-classification.

— (2024a). *2016 Census, Pay-As-You-Go Summaries, Personal Income Tax statements.* Tech. rep. Findings based on use of PLIDA data 2011 - 2020, accessed 2024.

— (2024b). *Business Activity Statements, Business Income Taxation forms, Pay-As-You-Go statements, Australian Business Register.* Tech. rep. Findings based on use of BLADE data 2011 - 2020, accessed 2024.

Autor, David, Arindrajit Dube, and Annie McGrew (2023). *The unexpected compression: Competition at work in the low wage labor market.* Tech. rep. National Bureau of Economic Research.

Autor, David, Christina Patterson, and John Van Reenen (2023). *Local and national concentration trends in jobs and sales: The role of structural transformation.* Tech. rep. National Bureau of Economic Research.

Autor, David et al. (2020). "The fall of the labor share and the rise of superstar firms". In: *The Quarterly Journal of Economics* 135.2, pp. 645–709.

Bajgar, Matej et al. (2023). "Industry concentration in europe and north america". In: *Industrial and Corporate Change*, dtac059.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu (2022). "Monopsony in movers: The elasticity of labor supply to firm wage policies". In: *Journal of Human Resources* 57.S, S50–s86.
Becker, Gary S (1973). "A theory of marriage: Part I". In: *Journal of Political economy* 81.4, pp. 813–846.
Berger, David, Kyle Herkenhoff, and Simon Mongey (2022). "Labor market power". In: *American Economic Review* 112.4, pp. 1147–1193.
Berger, David et al. (2024). "Monopsony Amplifies Distortions from Progressive Taxes". In: *AEA Papers and Proceedings*. Vol. 114. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 555–560.
Bilal, Adrien et al. (2022). "Firm and worker dynamics in a frictional labor market". In: *Econometrica* 90.4, pp. 1425–1462.
Bloom, Nicholas et al. (2014). "The distinct effects of information technology and communication technology on firm organization". In: *Management Science* 60.12, pp. 2859–2885.
Bloom, Nicholas et al. (2018). "The disappearing large-firm wage premium". In: *AEA Papers and Proceedings*. Vol. 108. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203, pp. 317–322.
Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa (2019). "A Distributional Framework for Matched Employer Employee Data". In: *Econometrica* 87.3, pp. 699–739.
Bonhomme, Stephane et al. (2023). "How Much Should We Trust Estimates of Firm Effects and Worker Sorting?" In: *Journal of Labor Economics* 41.2, pp. 699–739.
Borovicková, Katarına and Robert Shimer (2024). "Assortative Matching and Wages: The Role of Selection". In.
Brown, Charles and James Medoff (1989). "The employer size-wage effect". In: *Journal of political Economy* 97.5, pp. 1027–1059.
Buckley, Jack (Nov. 2023). *Productivity in Motion: The Role of Job Switching*. Micro Note 14. Accessed: 2024-10-10. e61 Institute. URL: https://e61.in/wp-content/uploads/2023/11/Productivity-in-motion-the-role-of-job-switching.pdf.
Burdett, Kenneth and Dale Mortensen (1998). "Wage differentials, employer size, and unemployment". In: *International Economic Review*, pp. 257–273.
Caliendo, Lorenzo, Ferdinando Monte, and Esteban Rossi-Hansberg (2015). "The anatomy of French production hierarchies". In: *Journal of Political Economy* 123.4, pp. 809–852.
Caliendo, Lorenzo et al. (2020). "Productivity and organization in Portuguese firms". In: *Journal of Political Economy* 128.11, pp. 4211–4257.
Deb, Shubhdeep et al. (2022). *Market power and wage inequality*. Tech. rep. IFS Working Papers.
Eeckhout, Jan and Philipp Kircher (2018). "Assortative matching with large firms". In: *Econometrica* 86.1, pp. 85–132.
Eeckhout, Jan and Roberto Pinheiro (2014). "Diverse organizations and the competition for talent". In: *International Economic Review* 55.3, pp. 625–664.
Freund, Lukas (2024). *Superstar Teams*. Tech. rep. Tech. rep., Cambridge University.
Garicano, Luis, Claire Lelarge, and John Van Reenen (2016). "Firm size distortions and the productivity distribution: Evidence from France". In: *American Economic Review* 106.11, pp. 3439–3479.

Gibson, John and Steven Stillman (2009). “Why do big firms pay higher wages? Evidence from an international database”. In: *The Review of Economics and Statistics* 91.1, pp. 213–218.

Gulyas, Andreas (2020). *Firm dynamics with labor market sorting.* Tech. rep. University of Bonn and University of Mannheim, Germany.

Haltiwanger, John, Henry Hyatt, and Erika McEntarfer (2018). “Who moves up the job ladder?” In: *Journal of Labor Economics* 36.S1, S301–S336.

Haltiwanger, John, Julia Lane, and James Spletzer (1999). “Productivity differences across employers: The roles of employer size, age, and human capital”. In: *American Economic Review* 89.2, pp. 94–98.

Hambur, Jonathan (2023). *Did labour market concentration lower wages growth pre-COVID?* 2023-01. Treasury Working Paper.

Herkenhoff, Kyle et al. (2024). “Production and learning in teams”. In: *Econometrica* 92.2, pp. 467–504.

Hirsch, Boris et al. (2022). “The wage elasticity of recruitment”. In.

Hopenhayn, Hugo A (1992). “Entry, exit, and firm dynamics in long run equilibrium”. In: *Econometrica: Journal of the Econometric Society*, pp. 1127–1150.

Hosios, Arthur J (1990). “On the efficiency of matching and related models of search and unemployment”. In: *The Review of Economic Studies* 57.2, pp. 279–298.

Hsieh, Chang-Tai and Esteban Rossi-Hansberg (2023). “The industrial revolution in services”. In: *Journal of Political Economy Macroeconomics* 1.1, pp. 3–42.

Hurst, Erik and Benjamin Wild Pugsley (2011). *What do small businesses do?* Tech. rep. National Bureau of Economic Research.

Jarosch, Gregor, Jan Sebastian Nimczik, and Isaac Sorkin (2024). “Granular search, market structure, and wages”. In: *Review of Economic Studies*, rdae004.

Jerzmanowski, Michal and Robert Tamura (2023). “Aggregate elasticity of substitution between skills: estimates from a macroeconomic approach”. In: *Macroeconomic Dynamics* 27.6, pp. 1597–1627.

Katz, Lawrence F and Kevin M Murphy (1992). “Changes in relative wages, 1963–1987: supply and demand factors”. In: *The quarterly journal of economics* 107.1, pp. 35–78.

Kiyotaki, Nobuhiro and Ricardo Lagos (2007). “A model of job and worker flows”. In: *Journal of political Economy* 115.5, pp. 770–819.

Kleinman, Benny (2024). “Wage Inequality and the Spatial Expansion of Firms”. In.

Krusell, Per et al. (2000). “Capital-skill complementarity and inequality: A macroeconomic analysis”. In: *Econometrica* 68.5, pp. 1029–1053.

Lise, Jeremy, Costas Meghir, and Jean-Marc Robin (2016). “Matching, sorting and wages”. In: *Review of Economic Dynamics* 19, pp. 63–87.

Lise, Jeremy and Jean-Marc Robin (2017). “The Macrodynamics of Sorting between Workers and Firms”. In: *American Economic Review* 107.4, pp. 1104–1135.

Lucas, Robert (1978). “On the size distribution of business firms”. In: *The Bell Journal of Economics*, pp. 508–523.

Luttmer, Erzo GJ (2011). “On the mechanics of firm growth”. In: *The Review of Economic Studies* 78.3, pp. 1042–1068.

Manning, Alan (2013). *Monopsony in motion: Imperfect competition in labor markets.* Princeton University Press.

Melitz, Marc J (2003). "The impact of trade on intra-industry reallocations and aggregate industry productivity". In: *econometrica* 71.6, pp. 1695–1725.

Melitz, Marc J and Gianmarco IP Ottaviano (2008). "Market size, trade, and productivity". In: *The review of economic studies* 75.1, pp. 295–316.

Mertens, Matthias and Bernardo Mottironi (2023). *Do larger firms exert more market power? Markups and markdowns along the size distribution*. Tech. rep. IWH-CompNet Discussion Papers.

Mertens, Matthias and Benjamin Schoefer (2024). "From Labor to Intermediates: Firm Growth, Input Substitution, and Monopsony". In: *Working Paper*.

Moll, Benjamin (2023). *Hopenhayn Model in Continuous Time*. Tech. rep. This note translates the Hopenhayn (1992) model of firm dynamics, exit, and entry into continuous time. Princeton University. URL: http://www.princeton.edu/~moll/HACTproject/option_simple.pdf.

Monte, Ferdinando, Stephen J Redding, and Esteban Rossi-Hansberg (2018). "Commuting, migration, and local employment elasticities". In: *American Economic Review* 108.12, pp. 3855–3890.

Oi, Walter Y and Todd L Idson (1999). "Firm size and wages". In: *Handbook of labor economics* 3, pp. 2165–2214.

Poschke, Markus (2018). "The firm size distribution across countries and skill-biased change in entrepreneurial technology". In: *American Economic Journal: Macroeconomics* 10.3, pp. 1–41.

Postel–Vinay, Fabien and Jean-Marc Robin (2002). "Equilibrium wage dispersion with worker and employer heterogeneity". In: *Econometrica* 70.6, pp. 2295–2350.

Postel-Vinay, Fabien and Helene Turon (2010). "On-the-job search, productivity shocks, and the individual earnings process". In: *International Economic Review* 51.3, pp. 599–629.

Redding, Stephen J and Esteban Rossi-Hansberg (2017). "Quantitative spatial economics". In: *Annual Review of Economics* 9.1, pp. 21–58.

Rinz, Kevin (2022). "Labor market concentration, earnings, and inequality". In: *Journal of Human Resources* 57.S, S251–S283.

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

Rosenthal-Kay, Jordan (2024). "Urban costs around the world". In.

Sarte, Pierre-Daniel et al. (2014). "The Impact of Regional and Sectoral Productivity Changes on the US Economy". In: *2014 Meeting Papers*. 426. Society for Economic Dynamics.

Shimer, Robert and Lones Smith (2000). "Assortative matching and search". In: *Econometrica* 68.2, pp. 343–369.

Song, Jae et al. (2019). "Firming up inequality". In: *The Quarterly journal of economics* 134.1, pp. 1–50.

Sorkin, Isaac (2018). "Ranking firms using revealed preference". In: *The quarterly journal of economics* 133.3, pp. 1331–1393.

Storesletten, Kjetil, Jonathan Heathcote, and Giovanni Violante (2020). *How Should Tax Progressivity Respond to Rising Income Inequality*. Tech. rep. CEPR Discussion Papers.

Syverson, Chad (2004). "Market structure and productivity: A concrete example". In: *Journal of political Economy* 112.6, pp. 1181–1222.

Traina, James (2018). "Is aggregate market power increasing? production trends using financial statements". In: *Production Trends Using Financial Statements (February 8, 2018)*.

A Details for Simple Model

As its name suggests, solving the simple model does not require complex machinery mathematically nor computationally. Nonetheless, I explain the full details required to solve for an equilibrium in this section, and prove Proposition 1.

A.1 Labor Supply Microfoundation with CES

One microfoundation for the constant elasticity of substitution labor supply curves given is that each worker has a unit of labor they spread across firms in order to maximize a CES aggregate of earnings. Workers then take the distribution of wages F as given[22] and solve[23]

$$W \equiv \max_{n(w)} \left(\int_0^\infty (wn(w))^{\frac{\eta}{\eta+1}} \, \mathrm{d}F(w) \right)^{\frac{\eta+1}{\eta}}$$
$$1 = \int_0^\infty n(w) \mathrm{d}F(w)$$

The first-order condition for the worker for wage w is

$$W^{\frac{1}{\eta+1}} w^{\frac{\eta}{\eta+1}} n(w)^{-\frac{1}{\eta+1}} = \lambda$$

where λ is the Lagrange multiplier for the total labor constraint. Multiplying both sides by $n(w)$ and integrating yields $W = \lambda$, so each worker's labor supply curve is

$$n(w) = \left(\frac{w}{W} \right)^\eta$$

Using this condition in the definition of W yields

$$W = \left(\int_0^\infty w^\eta \mathrm{d}F(w) \right)^{\frac{1}{\eta}}$$

There are $\overline{n}$ total workers, hence total labor supplied to a firm offering w is

$$n(w) = \left(\frac{w}{W} \right)^\eta \overline{n}$$

[22] I drop type subscripts here for clarity of notation.

[23] Note that this form is different than the one used by Berger, Herkenhoff, and Mongey (2022), which assume that workers have CES disutility across firms. In contrast, I am assuming workers have CES utility across earnings at firms.

A.2 Labor Supply Microfoundation with Continuous Choice

Along the same logic of Anderson, De Palma, and Thisse (1987), which showed that that CES and one form of discrete choice problem yield the same demand, it is possible to view choosing an employer as a discrete choice problem and yield the same CES supplies as above. Rather than assuming a measure of $\overline{n}$ of representative workers that each supply a distribution labor across all firms, assume each worker selects a single firm to supply their unit of labor, and has idiosyncratic preferences across firms. In particular, suppose a worker receives utility $w\epsilon$ from working at firm offering w, where ϵ is independent and identically distributed Frechet(η). Then the probability of selecting any given firm offering w is

$$n(w) = \frac{w^{\eta}}{\int (w')^{\eta} \mathrm{d}F(w')}$$

This is the same as the labor supply curves in the previous section[24].

A.3 Firm Optimality

The problem of firm (z, κ) is

$$\max_{n_i} z \left(n_L^{\frac{\sigma-1}{\sigma}} + (\kappa n_H)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} - \sum_{i \in \{L,H\}} \left(\frac{n_i}{\overline{n}_i} \right)^{\frac{1}{\eta_i}} W_i n_i$$

The first-order condition for employment of each worker type is

$$zN^{\frac{1}{\sigma}} n_L^{-\frac{1}{\sigma}} = \frac{\eta_L + 1}{\eta_L} \left(\frac{n_L}{\overline{n}_L} \right)^{\frac{1}{\eta_L}} W_L \tag{32}$$

$$zN^{\frac{1}{\sigma}} \kappa^{\frac{\sigma-1}{\sigma}} n_H^{-\frac{1}{\sigma}} = \frac{\eta_H + 1}{\eta_H} \left(\frac{n_H}{\overline{n}_H} \right)^{\frac{1}{\eta_H}} W_H \tag{33}$$

These conditions may be combined with the definition of the labor supply curves to relate firm-level employment to skill premia, as

$$\frac{w_H(z,\kappa)}{w_L(z,\kappa)} = \frac{\frac{\eta_H}{\eta_H+1}}{\frac{\eta_H}{\eta_H+1}} \kappa^{\frac{\sigma-1}{\sigma}} \left(\frac{n_H(z,\kappa)}{n_L(z,\kappa)} \right)^{-\frac{1}{\sigma}} \tag{34}$$

This condition says that optimizing firms' skill premia are entirely determined by their skill ratio and SBP κ. Note, however, that when $\eta_L > \eta_H$, firms will decrease their skill ratio as they change their scale of operation (say due to TFP shocks), and thus increase their skill premia.

[24] Again, this form is different than the one offered by Berger, Herkenhoff, and Mongey (2022). They assume that workers face a choice problem across firms, but conditional on a firm work more or less to obtain a given level of earnings. Instead, I assume workers face a choice problem across firms, but always provide a single unit of labor, conditional on selecting a firm.

A.4 Solving for Equilibrium

Finding the equilibrium requires finding the set of wages, employment levels, and distribution of firms that is consistent with the labor supply schedules, labor demand induced by firm optimality, the given shares of firms across κ, and the flow of firms between z.

The distribution of firms is exogenous, so easiest to start with. The measure of firms in state (z, κ) is the product of the share of firms in z and κ in the stationary distribution. The share of firms with κ_L is s_L and therefore the share of firms with κ_H is $s_H = 1 - s_L$. The stationary distribution of firms across z, call it h, satisfies the stationary KF equation and the requirement of being a proper distribution

$$0 = \dot{h} = \begin{bmatrix} -\lambda_H & \lambda_L \\ \lambda_H & -\lambda_L \end{bmatrix} h$$

$$1 = \sum_{i \in \{L,H\}} h_i$$

The solution is

$$\begin{bmatrix} h_L \\ h_H \end{bmatrix} = \begin{bmatrix} \frac{\lambda_L}{\lambda_L + \lambda_H} \\ \frac{\lambda_H}{\lambda_L + \lambda_H} \end{bmatrix}$$

Next, the labor supply curves provide closed-form mappings from wages to labor quantities, given the firm distributions from above, hence for any guess of wages I can calculate the implied labor supplied.

Lastly, I check that the labor supplied is consistent with the labor demanded given by the firm optimality conditions.

The full algorithm for solving the model is

(I) Calculate the exogenous distribution of firms over (z, κ), where state (z_i, κ_j) has measure $s_i \frac{\lambda_j}{\lambda_L + \lambda_H}$.

(II) Using a nonlinear equation solver, find the set of wages $w(z, \kappa)$ that clear all labor markets

   (i) Guess $w_i(z, \kappa)$

   (ii) Compute the wage indices W_i

   (iii) Compute labor quantities supplied $n_i(z, \kappa) = \left(\frac{w_i(z,\kappa)}{W_i}\right)^{\eta_i}$.

   (iv) Compute the residual demand equations by using $n_i(z, \kappa)$ in the firm optimality conditions (37) and (38).

A.5 Proof of Proposition 1

Proof. (i) Note that in the limiting case the optimality conditions are

$$z = \frac{\eta_L + 1}{\eta_L} \left(\frac{n_L}{\overline{n}_L}\right)^{\frac{1}{\eta_L}} W_L$$
$$z\kappa = \frac{\eta_H + 1}{\eta_H} \left(\frac{n_H}{\overline{n}_H}\right)^{\frac{1}{\eta_H}} W_H$$

(a) The optimality conditions imply n_L is invariant to κ, and n_H is increasing in κ, therefore $N = n_L + \kappa n_H$ increases.

(b) From above, n_H is increasing but n_L is constant, hence $\frac{n_H}{n_L}$ is increasing.

(c) The skill premium is

$$\frac{w_H}{w_L} = \frac{\frac{\eta_H}{\eta_H+1}}{\frac{\eta_L}{\eta_L+1}}\kappa$$

Therefore increases in κ increase the skill premium (note that z does nothing in this limit case).

(d) The optimality conditions imply that both labor types should increase with z, thus N increases.

(e) The optimality conditions imply the following relationships

$$\frac{\mathrm{d}\ln n_L}{\mathrm{d}\ln z} = \eta_L > \eta_H = \frac{\mathrm{d}\ln n_H}{\mathrm{d}\ln z}$$

Therefore $\frac{n_H}{n_L}$ is decreasing in z.

(ii) The optimality conditions are compositions of functions which are continuous in σ, provided the indeterminate case $\sigma = 1$ is allowed to be the natural Cobb-Douglas limit. Therefore the equilibrium n_i are continuous functions of σ, and the comparative static results are also continuous in σ. Therefore, given the results for the $\sigma \to \infty$ case, it must be that there is a σ sufficiently large that the above results hold, as they all refer to signs of comparative statics, so breaking any of the results would require discontinuity.

(f) The skill premium within a firm is

$$\frac{w_H}{w_L} = \frac{\frac{\gamma_H}{\eta_H+1}}{\frac{\eta_L}{\eta_L+1}}\kappa^{\frac{\sigma-1}{\sigma}} \left(\frac{n_H}{n_L}\right)^{-\frac{1}{\sigma}}$$

From above, the skill ratio is falling in z, thus the skill premium is rising. Note that this result holds for any $\sigma \in (0, \infty)$, thus it is possible to maintain a large enough σ that all the above results go through, as argued above, but still generate this result.

$\square$

A.6 Adding Payroll Taxes

Accounting for the payroll tax, the problem of firm (z, κ) is

$$\max_{n_i} z \left(n_L^{\frac{\sigma-1}{\sigma}} + (\kappa n_H)^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}} - \sum_{i \in \{L,H\}} \left(\frac{n_i}{\overline{n}_i} \right)^{\frac{1}{\eta_i}} W_i n_i - \mathcal{T} \left(\sum_{i \in \{L,H\}} \left(\frac{n_i}{\overline{n}_i} \right)^{\frac{1}{\eta_i}} W_i n_i \right)$$

The first-order condition for employment of each worker type is

$$z N^{\frac{1}{\sigma}} n_L^{-\frac{1}{\sigma}} = \frac{\eta_L + 1}{\eta_L} \left(\frac{n_L}{\overline{n}_L} \right)^{\frac{1}{\eta_L}} W_L \left[1 + \mathcal{T}' \left(\sum_{i \in \{L,H\}} \left(\frac{n_i}{\overline{n}_i} \right)^{\frac{1}{\eta_i}} W_i n_i \right) \right] \tag{35}$$

$$z N^{\frac{1}{\sigma}} \kappa^{\frac{\sigma-1}{\sigma}} n_H^{-\frac{1}{\sigma}} = \frac{\eta_H + 1}{\eta_H} \left(\frac{n_H}{\overline{n}_H} \right)^{\frac{1}{\eta_H}} W_H \left[1 + \mathcal{T}' \left(\sum_{i \in \{L,H\}} \left(\frac{n_i}{\overline{n}_i} \right)^{\frac{1}{\eta_i}} W_i n_i \right) \right] \tag{36}$$

When the first-order conditions are sufficient for optimality, equation 34 from above holds as is, because the payroll tax is on the entire wage bill, therefore, conditional on a given set of relative wages, the payroll tax does nothing to skew hiring toward either skill type. However, the payroll does increase the cost of operating at a larger scale, assuming $\mathcal{T}' \geq 0$ globally, and firms make scale and composition decisions jointly, so firms will adjust their skill composition. Intuitively, suppose an economy where $\mathcal{T} = 0$, but then a new tax schedule is introduced satisfying $\mathcal{T}' > 0, \mathcal{T}'' > 0$. For any given set of wage indices W_i, all firms will see their marginal costs increase and scale back hiring of both labor types. The higher elasticity of low-skill labor supply will mean that it takes relatively larger reductions in low-skill employment than high-skill employment to decrease the wage bill[25], so in descaling the relative share of high-skill workers will increase.

The caveat about first-order conditions being sufficient above is critical. If the tax schedule is not convex, then second-order conditions may not hold. I now consider a particular tax schedule that is similar to the one I study empirically, but which is not convex. Suppose

[25] Why not simply reduce the wage bill entirely by decreasing high-skill employment, which is least elastic? Pursuing this policy would also reduce output at a higher rate. Reducing low-skill labor is less detrimental to output.

$$\mathcal{T}(x) = \begin{cases} 0 & x \leq \overline{x}_1 \\ a_1(x - \overline{x}_1) & \overline{x}_1 < x \leq \overline{x}_2 \\ a_1(\overline{x}_2 - \overline{x}_1) + a_2(x - \overline{x}_2) & \overline{x}_2 < x \leq \overline{x}_3 \\ a_1(\overline{x}_2 - \overline{x}_1) + a_2(\overline{x}_3 - \overline{x}_2) & x > \overline{x}_3 \end{cases}$$

This tax schedule imposes that firms below threshold $\overline{x}_1$ pay zero taxes, then face a marginal tax rate a_1 up to $\overline{x}_2$, then face marginal tax rate a_2 up to $\overline{x}_3$, beyond which they pay no additional taxes. To determine their optimal employment, firms may solve the problem piecewise as follows. First, find the optimal choice that would be made if $\mathcal{T}' = 0$ globally, and compute the implied profits under that choice, but using the tax schedule. Next, do the same for $\mathcal{T}' = a_1$ and $\mathcal{T}' = a_2$. Last, check the three corners $\overline{x}_i$. The optimal decision is the one that delivers the maximal profits out of these five potential candidates.

A.7 Proof of Proposition 2

Proof.

(i) For firms below $\overline{x}_1$ originally, the tax does not affect the optimality conditions, hence there is no reason why a firm in this region would decrease employment, because if that policy were optimal, they would have pursued in under $\mathcal{T}_0$.

(ii) These firms' optimality conditions have will no longer hold, because the payroll subsidy has lowered the right side of each equation. Therefore payroll must rise to equalize, which means N must rise, because if it were possible for N to fall and the payroll rise, the firm would have chosen to operate at a lower payroll with higher N originally.

(iii) If x maximizes two different functions, then x maximizes their sum. In this case, the two different functions are the profit function of the firm before the tax policy change, and the negative of the total tax burden schedule, $-\mathcal{T}$.

(iv) As above, the optimality conditions no longer hold, but now the right side is too high. The firm must lower payroll, which means decreasing N, as argued already.

(v) As in the first point, if it were optimal for these firms to increase employment, then they would have done so under the $\mathcal{T}_0$.

□

A.8 General Equilibrium in the Simple Model

The primary purpose of the simple model is for understanding the distribution of firms and the labor markets of workers, therefore abstracting from the product market clearing does not seem unreasonable. Nonetheless, it is easy to add assumptions about how profits are disbursed which allow the goods market to clear due to Walras' Law. The only requirement is that the profit disbursement is nondistortive. To this end, suppose all workers own a portfolio of the firms which pays constant profits. Importantly, where a worker works, per the microfoundations in A.1 or A.2, does not affect their profit earnings. Then assume that each worker lives hand-to-mouth and uses the entirety of their earnings to buy goods. Then none of the decisions or outcomes in the main paper change, and the goods market will clear because the total income of all firms is exactly equal to the total income of all workers, accounting for both labor and profits.

It is also possible to maintain goods market clearing once a payroll tax or subsidy is added. One way to do so is for the government to levy a lump-sum tax on (or pay a lump-sum transfer to) all workers. As above, this lump-sum addition will do nothing to distort worker decisions, so all of the analysis in the main text goes through, and the goods market will clear because the sum of all worker and firm taxes and subsidies is zero, and the remaining income earned by firms exactly equals the income of workers.

A.9 Simple Model Parametrization

Parameter	**Description**	**Value**
σ	Elasticity of substitution between n_L and n_H in production	1.8
κ_L	Low type skill-biased productivity	1
κ_H	High type skill-biased productivity	1.1
z_L	Low type total factor productivity	1
z_H	High type skill-biased productivity	1.1
λ_L	Transition rate to z_L	0.5
λ_H	Transition rate to z_H	0.5
s_L	Share of firms with κ_L	0.5
$\overline{n}_L$	Share of low-skill workers	0.6
η_L	Elasticity of labor supply for low-skill workers	0.8
η_H	Elasticity of labor supply for high-skill workers	0.7

Table 3: Parameters in simple model

B Additional Empirical Details

B.1 Disclaimer

The results of these studies are based, in part, on data supplied to the ABS under the Taxation Administration Act 1953, A New Tax System (Australian Business Number) Act 1999, Australian Border Force Act 2015, Social Security (Administration) Act 1999, A New Tax System (Family Assistance) (Administration) Act 1999, Paid Parental Leave Act 2010 and/or the Student Assistance Act 1973. Such data may only used for the purpose of administering the Census and Statistics Act 1905 or performance of functions of the ABS as set out in section 6 of the Australian Bureau of Statistics Act 1975. No individual information collected under the Census and Statistics Act 1905 is provided back to custodians for administrative or regulatory purposes. Any discussion of data limitations or weaknesses is in the context of using the data for statistical purposes and is not related to the ability of the data to support the Australian Taxation Office, Australian Business Register, Department of Social Services and/or Department of Home Affairs' core operational requirements.

Legislative requirements to ensure privacy and secrecy of these data have been followed. For access to PLIDA and/or BLADE data under Section 16A of the ABS Act 1975 or enabled by section 15 of the Census and Statistics (Information Release and Access) Determination 2018, source data are de-identified and so data about specific individuals has not been viewed in conducting this analysis. In accordance with the Census and Statistics Act 1905, results have been treated where necessary to ensure that they are not likely to enable identification of a particular person or organisation.

B.2 Data Construction

I use data from administrative tax records for workers and firms and the Australian Census, all provided by the Australian Bureau of Statistics and accessed in the secure DataLab environment, for fiscal years[26] $2011 - 2020$. The firm data comes from the Business Longitudinal Analysis Data Environment (BLADE). The subset of BLADE I employ in this project sources data from Australian Business Register and the Australian Tax Office. The worker data comes from the Person Level Integrated Data Asset (PLIDA)[27]. The subset of PLIDA I employ sources data from the Australian Bureau of Statistics (ABS), the Australian Tax Office (ATO), and the 2016 Census[28].

I construct firm-level variables from BLADE (Australian Bureau of Statistics (2024b)). This dataset covers the universe of businesses registered for the goods-and-services tax (GST), indexed by Australian Business Number (BN), but does not include sole traders nor partnerships that submit personal income tax returns instead of business income tax returns. All firms with total sales exceeding \$75,000 dollars must register. To maintain consistency with the ABS, Department of the Treasury, and previous work, such as Andrews et al. (2019), Buckley (2023), and Hambur (2023), I use an enterprise group (BG) as the definition of a firm, which may include multiple BNs, and do so because this aggregation is meant to capture the set of legal entities under common control, including conglomerates that may include multiple subsidiaries, but where decisions are made jointly, in line with the notion of a firm I use throughout this paper. Firm turnover (total sales), wage bills, and other expenses come from Business Activity Statements that firms are legally required to file with the ATO. Additionally, firms file Business Income Taxation forms which report more detailed breakdowns of taxable income or loss. The ATO Pay-As-You-Go (PAYG) statements from each firm for each employee are also aggregated by the ATO to construct a measure of total headcount for each firm. The ABS also calculates a "full-time equivalent" (FTE) measure of employment, using a combination of headcount, firm wage bill, and

[26] An Australian fiscal year is July to July. To be in line with the time indexing of the ABS, I refer to the year $2010 - 2011$ as fiscal year 2011.

[27] The worker data sources were referred to as the Multi-Agency Data Integration Project (MADIP) until 2023, when the rename to PLIDA occurred.

[28] It is also possible to access the 2011 Census instead of the 2016 Census, but no project may access both simultaneously.

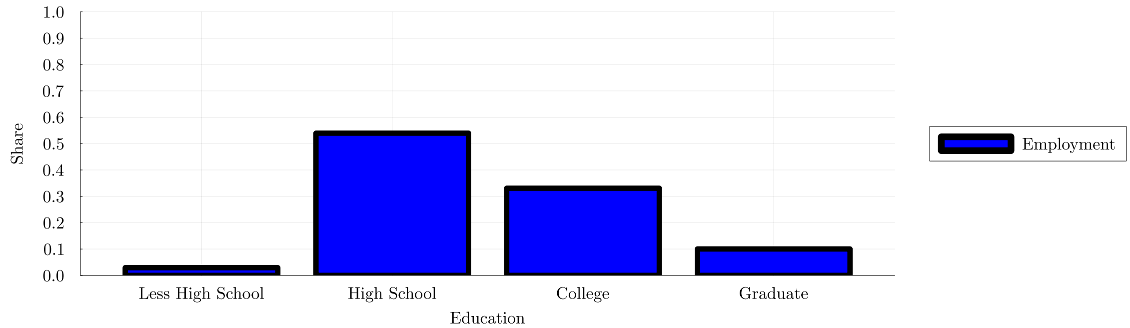


Figure 15: Distribution of education for employed Australians, 2011 - 2020

firm industry, which is meant to capture the fact some workers are part-time, so comparing headcounts may not be the best way to compare effective employment. I use this measure of FTE for computing labor productivity within a firm, but the choice between headcount and FTE appears to be inconsequential. Firm industry comes from Cross-Sectional Indicative data from the Australian Business Register. Industry is determined by Australian and New Zealand Standard Industrial Classification (ANZSIC) codes. These industries are partitioned according to "supply-side based industry definitions and groupings", and thus are "business units in a particular class will use similar inputs and apply similar transformation processes to produce similar outputs" (Australian Bureau of Statistics (2006)).

I construct worker-level variables from PLIDA (Australian Bureau of Statistics (2024a)). As mentioned above, each firm provides a payment summary for each worker each year to the ATO for use in calculating PAYG tax contributions[29]. These statements includes the gross payments to each worker, which are summed to calculate annual earnings, and are used to link workers to ABNs. I use Personal Income Tax (PIT) statements for determining worker location and occupation. Worker location is determined by residence[30], and at the level of Mesh Code, which I then aggregate into a geography of Statistical Area Level 4 (SA4) + Greater Capital City Statistical Areas (GCCSAs). SA4s are meant to capture local labor markets of most cities, but in the largest cities (e.g. Sydney) it may be the case that a labor market is more accurately a GCCSA, so in the Greater Capital Cities I aggregate SA4s up to GCCSAs. The 2016 Census provides individual birth dates, and thus ages, as well as level of highest educational attainment, which I group into four exhaustive categories: Less than High School, High School, College, and Graduate. I show the share of workers in each category in Figure 15.

Lastly, I drop all observations where: firm FTE is less than 1, total sales are negative, labor productivity (value-added/worker) is negative, cost of goods sold is negative, workers are younger than 25 or older than 60. I observe approximately 55 million worker-years, and approximately 4 million firm-years.

Firm ages are constructed using firm birth years derived from a combination of Business

[29]These are analogous to W-2 forms in the United States.

[30]Some analyses use job-level location data, such as Hambur (2023), but only residential data is available to researchers outside the Treasury, and in any case it seem unlikely that the difference should matter, as SA4+GCCSA is a sufficiently coarse geographic measure. The distinction would be essential for understanding urban patterns (e.g. commuting), but that is not my focus in this paper.

Activity Statements and the ABS Businesss Registers. The birth date data has a few issues, however, which I detail in Appendix B.6. Fortunately, for my sample period of 2011 - 2019, these issues are not a problem when focusing on firms less than 10 years old, so for each year I can simply assign firms with age greater than 10 to a "10+" bin and obtain a useful and accurate measure of age.

Firm location is not simple to measure, as discussed in Hambur (2023), where a notion of plant is used instead, because around half of firms operate in multiple regions. I am concerned with the skill composition within a firm, which may vary widely between firm headquarters and plants, as argued by Kleinman (2024), so in fact "the" location of many firms is not even well-defined. Nonetheless, for my purposes I assign each firm to the location of a randomly chosen worker in the firm, which I view as a fairly agnostic way of deciding where to allocate firms.

To construct a measure for value-added, I subtract the cost of goods sold (COGS) from the total income of the firm. To construct COGS, I consider total expenses net of depreciation, rent and leasing expenses, external labor and contractors, interest payments, bad debts, royalties, wages for internal labor, and supperannuation. This approach is meant to be as close as possible to capturing COGS in accordance with accounting for Selling and General Administration costs, following Traina (2018). Lastly, all values are deflated according to industry-level price deflators constructed by the ABS.

In the knowledge hierarchy extension of the quantitative model, firms must decide how to organize workers into layers of expertise[31]. I follow previous empirical work of Caliendo, Monte, and Rossi-Hansberg (2015) and Caliendo et al. (2020) to assign workers within a firm to a layer of the organization according to their occupation. Worker occupations are determined at the worker-year level, though workers may hold multiple occupations in a given year. Workers report the occupation in which they earned the most income. Occupations are determined by the Australian and New Zealand Standard Classification of Occupations (ANZSCO), which defines an occupation as a set of jobs that require the performance of similar or identical sets of tasks. I then classify each occupation into its hierarchy layer, which is meant to capture the degree to which each occupation requires expertise and/or capacity to manage workers. Workers in occupations with ANZSCO less then 120000 are in level 4, which is the subset of managerial occupations that are top managers and experts in a firm, such as CEOs. The next layer is the remaining managerial occupations, such as sales managers and CFOs. The third layer includes all professionals, such as software programmers. The final layer is all residual occupations, so is fairly broad but generally meant to capture jobs that not entail management or expert advising, such as trade workers and labourers. I first classify each worker according to their layer per the above scheme, but then within each firm reclassify the layer of the worker according to the total number of layers with a firm. For example, a firm containing 10 construction workers, 3 architects, and a CEO, contains layers 1,2, and 4, so only three unique layers, and the CEO will be reclassified as layer 3 worker, and the firm will be considered to only have 3 layers.

[31] I prefer to think of upper layer workers as experts rather than managers, as it seems more natural to me that when less skilled workers ask for help from more-skilled workers, this is because they are experts, not necessarily superior at management.

Dependent Variables:	ln(High Skill Ratio)		ln(College Skill Ratio)		ln(Graduate Skill Ratio)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
employment ∈ 10-19	-0.3449***		-0.3963***		-0.5506***	
	(0.0092)		(0.0097)		(0.0053)	
employment ∈ 20-49	-0.5030***		-0.5777***		-1.095***	
	(0.0159)		(0.0167)		(0.0102)	
employment ∈ 50-99	-0.4611***		-0.5626***		-1.570***	
	(0.0186)		(0.0196)		(0.0107)	
employment ∈ 100-199	-0.3162***		-0.4370***		-1.757***	
	(0.0176)		(0.0170)		(0.0200)	
employment ∈ 200-499	-0.1209***		-0.2633***		-1.713***	
	(0.0312)		(0.0304)		(0.0379)	
employment ∈ 500-999	-0.0708***		-0.2202***		-1.681***	
	(0.0185)		(0.0179)		(0.0253)	
employment ∈ 1000-1999	0.0965***		-0.0819**		-1.397***	
	(0.0304)		(0.0297)		(0.0366)	
employment ∈ 2000-4999	0.2009***		-0.0161		-1.223***	
	(0.0278)		(0.0262)		(0.0344)	
employment ∈ 5000+	0.2855***		0.0187		-1.023***	
	(0.0385)		(0.0364)		(0.0454)	
ln(employment)		-0.1660***		-0.2053***		-0.4459***
		(0.0066)		(0.0066)		(0.0049)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,557,864	1,557,864	1,485,797	1,485,797	485,895	485,895
R^2	0.03678	0.02886	0.05353	0.04736	0.23191	0.21326
Within R^2	0.03574	0.02781	0.05255	0.04638	0.23107	0.21240

Clustered (time) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4: Regressions of firm log skill ratio on firm size bins (columns 1,3,5) and log employment (columns 2,4,6), including time fixed effects. High-Skill workers are college or graduate educated.

B.3 Tables

I display below the regression tables referenced in the main text, and used to construct the main text figures and tables.

B.4 Additional Summary Statistics

I summarize a variety of cross-sectional patterns present in the data. First, I bin firm size and display the distribution of firms and employment in Figure 16. As expected, I find that most firms are small, with 81% of firms employing 10 or less workers, but that employment is much more dispersed, with 42% of workers employed in firms with 500 or more workers. The measure of skill used throughout is worker education, where

I consider the measure of skill composition within a firm that I will use throughout the remainder of the paper, which is the log of the skill ratio, defined as the number of workers of a given skill (college, graduate, or college+graduate) over the number of workers with less than a college education, within a firm. This measure is appealing because it can be related

Dependent Variables:	ln(High Skill Ratio)	High Skill Ratio	ln(College Skill Ratio)		ln(Graduate Skill Ratio)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
employment ∈ 10-19	-0.3122*		-0.3699**		-0.5328***	
	(0.1411)		(0.1317)		(0.1115)	
employment ∈ 20-49	-0.4416**		-0.5228**		-1.016***	
	(0.1899)		(0.1805)		(0.2210)	
employment ∈ 50-99	-0.3991*		-0.5053**		-1.433***	
	(0.1867)		(0.1775)		(0.2604)	
employment ∈ 100-199	-0.2763		-0.3994**		-1.613***	
	(0.1775)		(0.1686)		(0.2723)	
employment ∈ 200-499	-0.1099		-0.2515		-1.586***	
	(0.1491)		(0.1400)		(0.2559)	
employment ∈ 500-999	-0.0770		-0.2253*		-1.568***	
	(0.1158)		(0.1097)		(0.2197)	
employment ∈ 1000-1999	0.0165		-0.1545		-1.363***	
	(0.1377)		(0.1284)		(0.2463)	
employment ∈ 2000-4999	0.0652		-0.1384		-1.254***	
	(0.1908)		(0.1575)		(0.3327)	
employment ∈ 5000+	0.0652		-0.1833		-1.137**	
	(0.2485)		(0.1794)		(0.4124)	
ln(employment)		0.1879		-0.1934**		-0.4149***
		(0.1734)		(0.0742)		(0.0729)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
industry	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,557,864	1,557,864	1,485,797	1,485,797	485,895	485,895
R^2	0.19546	0.11983	0.20084	0.19787	0.35683	0.34480
Within R^2	0.03288	0.01120	0.05095	0.04743	0.22558	0.21110

Clustered (time & industry) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 5: Regressions of firm log skill ratio on firm size bins (columns 1,3,5) and log employment (columns 2,4,6), including time and ANZSIC division fixed effects. High-Skill workers are college or graduate educated.

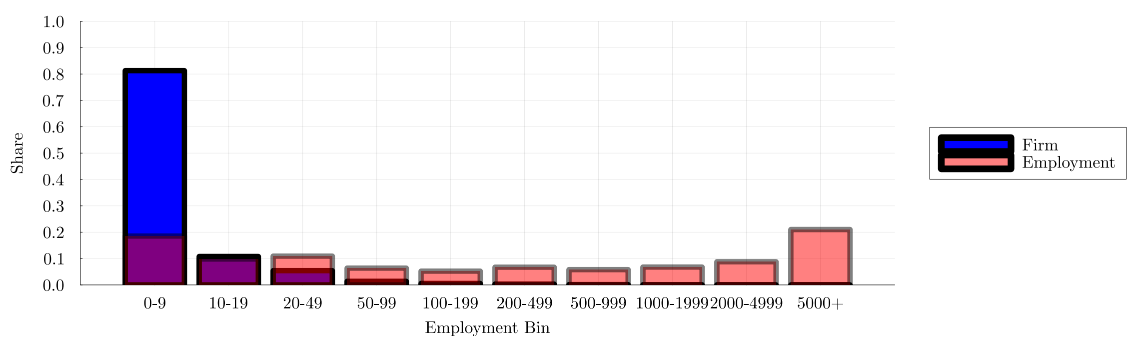


Figure 16: Distribution of firms and employment across employment size bins

Dependent Variables:	ln(High Skill Ratio)		ln(College Skill Ratio)		ln(Graduate Skill Ratio)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
employment ∈ 10-19	-0.2864***		-0.3463***		-0.5226***	
	(0.0688)		(0.0643)		(0.0544)	
employment ∈ 20-49	-0.3871***		-0.4708***		-0.9874***	
	(0.0937)		(0.0893)		(0.1008)	
employment ∈ 50-99	-0.3346***		-0.4428***		-1.378***	
	(0.0988)		(0.0942)		(0.1205)	
employment ∈ 100-199	-0.2274**		-0.3497***		-1.558***	
	(0.1006)		(0.0953)		(0.1301)	
employment ∈ 200-499	-0.0925		-0.2304**		-1.549***	
	(0.0953)		(0.0907)		(0.1385)	
employment ∈ 500-999	-0.0733		-0.2171**		-1.531***	
	(0.0733)		(0.0704)		(0.1191)	
employment ∈ 1000-1999	-0.0172		-0.1792*		-1.378***	
	(0.0869)		(0.0808)		(0.1434)	
employment ∈ 2000-4999	0.0558		-0.1376		-1.267***	
	(0.1433)		(0.1131)		(0.2429)	
employment ∈ 5000+	0.0793		-0.1624		-1.121***	
	(0.1967)		(0.1349)		(0.3294)	
ln(employment)		-0.1365***		-0.1805***		-0.4052***
		(0.0401)		(0.0372)		(0.0353)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
industry	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,557,864	1,557,864	1,485,797	1,485,797	485,895	485,895
R^2	0.24325	0.23969	0.24667	0.24527	0.39815	0.38876
Within R^2	0.02660	0.02202	0.04315	0.04137	0.21382	0.20154

Clustered (time & industry) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 6: Regressions of firm log skill ratio on firm size bins (columns 1,3,5) and log employment (columns 2,4,6), including time and ANZSIC 3-digit fixed effects. High-Skill workers are college or graduate educated.

Dependent Variables:	ln(High Skill Ratio)		ln(College Skill Ratio)		ln(Graduate Skill Ratio)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
employment ∈ 10-19	-0.3401***		-0.3944***		-0.5548***	
	(0.0692)		(0.0616)		(0.0456)	
employment ∈ 20-49	-0.5092***		-0.5860***		-1.107***	
	(0.1102)		(0.1015)		(0.0837)	
employment ∈ 50-99	-0.4904***		-0.5913***		-1.588***	
	(0.1457)		(0.1360)		(0.1451)	
employment ∈ 100-199	-0.3766**		-0.4937***		-1.789***	
	(0.1544)		(0.1445)		(0.1740)	
employment ∈ 200-499	-0.2097		-0.3461**		-1.757***	
	(0.1430)		(0.1356)		(0.1808)	
employment ∈ 500-999	-0.1852		-0.3260**		-1.744***	
	(0.1113)		(0.1063)		(0.1575)	
employment ∈ 1000-1999	-0.0489		-0.2150***		-1.498***	
	(0.0565)		(0.0549)		(0.0926)	
employment ∈ 2000-4999	0.0571		-0.1488**		-1.317***	
	(0.0695)		(0.0554)		(0.1117)	
employment ∈ 5000+	0.1214*		-0.1335*		-1.141***	
	(0.0549)		(0.0669)		(0.0583)	
ln(employment)		-0.1751***		-0.2152***		-0.4572***
		(0.0537)		(0.0508)		(0.0477)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
SA4+GCCSA	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,542,376	1,542,376	1,471,037	1,471,037	481,071	481,071
R^2	0.09083	0.08524	0.10139	0.09778	0.26744	0.25109
Within R^2	0.03841	0.03250	0.05689	0.05309	0.24458	0.22772

Clustered (time & SA4+GCCSA) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 7: Regressions of firm log skill ratio on firm size bins (columns 1,3,5) and log employment (columns 2,4,6), including time and SA4++GCCSA fixed effects. High-Skill workers are college or graduate educated.

Dependent Variables:	ln(High Skill Ratio)		ln(College Skill Ratio)		ln(Graduate Skill Ratio)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
employment ∈ 10-19	-0.1132***		-0.1401***		-0.2519***	
	(0.0082)		(0.0082)		(0.0110)	
employment ∈ 20-49	-0.1867***		-0.2252***		-0.4705***	
	(0.0136)		(0.0137)		(0.0220)	
employment ∈ 50-99	-0.2368***		-0.2772***		-0.6501***	
	(0.0189)		(0.0182)		(0.0335)	
employment ∈ 100-199	-0.2857***		-0.3247***		-0.7778***	
	(0.0249)		(0.0235)		(0.0435)	
employment ∈ 200-499	-0.3557***		-0.3890***		-0.8941***	
	(0.0323)		(0.0302)		(0.0503)	
employment ∈ 500-999	-0.4409***		-0.4628***		-0.9854***	
	(0.0438)		(0.0400)		(0.0658)	
employment ∈ 1000-1999	-0.5066***		-0.5298***		-1.031***	
	(0.0483)		(0.0441)		(0.0768)	
employment ∈ 2000-4999	-0.6375***		-0.6549***		-1.155***	
	(0.0669)		(0.0624)		(0.0933)	
employment ∈ 5000+	-0.7684***		-0.7754***		-1.294***	
	(0.0902)		(0.0868)		(0.1130)	
ln(employment)		-0.1862***		-0.2368***		-0.3885***
		(0.0070)		(0.0058)		(0.0115)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
firm	Yes	Yes	Yes	Yes	Yes	Yes
Fit statistics						
Observations	1,557,864	1,557,864	1,485,797	1,485,797	485,895	485,895
R^2	0.84931	0.85086	0.83871	0.84146	0.90089	0.90376
Within R^2	0.00518	0.01536	0.00759	0.02455	0.03401	0.06205

Clustered (time & firm) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 8: Regressions of firm log skill ratio on firm size bins (columns 1,3,5) and log employment (columns 2,4,6), including time and firm fixed effects. High-Skill workers are college or graduate educated.

Dependent Variables:	EE		EN		EE + EN	
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
ln(earnings) $\times$ $\mathbf{1}\{LessHighSchool\}$	-0.0479***	-0.0400***	-0.0650***	-0.0712***	-0.1129***	-0.1112***
	(0.0046)	(0.0038)	(0.0063)	(0.0068)	(0.0108)	(0.0102)
ln(earnings) $\times$ $\mathbf{1}\{HighSchool\}$	-0.0455***	-0.0377***	-0.0657***	-0.0717***	-0.1112***	-0.1094***
	(0.0044)	(0.0036)	(0.0064)	(0.0068)	(0.0106)	(0.0101)
ln(earnings) $\times$ $\mathbf{1}\{College\}$	-0.0447***	-0.0363***	-0.0651***	-0.0707***	-0.1098***	-0.1070***
	(0.0043)	(0.0035)	(0.0063)	(0.0067)	(0.0105)	(0.0100)
ln(earnings) $\times$ $\mathbf{1}\{Graduate\}$	-0.0442***	-0.0347***	-0.0645***	-0.0698***	-0.1087***	-0.1046***
	(0.0043)	(0.0034)	(0.0063)	(0.0067)	(0.0105)	(0.0098)
Fixed-effects						
time	Yes	Yes	Yes	Yes	Yes	Yes
firm		Yes		Yes		Yes
Fit statistics						
Observations	42,286,829	42,286,829	42,286,829	42,286,829	42,286,829	42,286,829
R^2	0.03875	0.14785	0.61283	0.64901	0.26790	0.36411
Within R^2	0.01533	0.00851	0.08489	0.07709	0.07543	0.05888

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 9: Regressions of employment-to-employment, employment-to-nonemployment, and combined transitions on log earnings, including time and firm fixed effects. High-Skill workers are college or graduate educated.

to skill premia in my simple model [32], and aids in comparing my results with the frictionless frameworks of Katz and Murphy (1992) and Acemoglu (2002)[33].

In Figure 17 I document the relationship between skill ratio and firm size. For both college and graduate workers, the patterns is U-shaped, but the trough occurs for larger firms for the most skilled workers, so that firms in the $20-49$ bin have the lowest skill ratio for the college measure, but firms in the $100-199$ have the lowest skill ratio for the graduate measure. These patterns remain if I instead weight the log skill ratios by firm employment, though the largest bin exhibits a marked decrease. As I argue below, these nonmonotonic patterns may be explained by a combination of a within-firm factor, which pushes the skill ratio downward with size, and a between-firm factor, which pushes the skill ratio upward with size.

I compute worker earnings by skill and firm size. In Figure 18, I compute the average earnings for each worker type in each firm, then take the average across firms. The patterns exhibit the large firm premium discussed by Brown and Medoff (1989) and Oi and Idson (1999) for each worker type, implying that the entirety of the large firm premium cannot be explained by differences in educational composition[34].

[32] See Appendix A.3.

[33] There is also an appealing notion of symmetry in terms of measuring skill composition with the log skill ratio for any aggregation of firms, which occurs because logs transform ratios into differences. For example, if I wished to measure the skill composition of 1000 firms, and 500 firms had 2 high-skill workers and 1 low-skill, while the remaining 500 had 1 high-skill and 2 low-skill, I would like for my measure to report that the aggregate skill ratio is 1. With simple weighting of the ratios I would measure $\frac{500}{1000} \cdot 2 + \frac{500}{1000} \cdot \frac{1}{2} = 1.25$, but with logs I have $\frac{500}{1000} \cdot \ln 2 + \frac{500}{1000} \cdot \ln \frac{1}{2} = 0 = \ln 1$.

[34] For more on the large firm premium internationally, see Gibson and Stillman (2009) and Bloom et al. (2018).

Dependent Variables:	log(Skill Premium College)		log(Skill Premium Graduate)	
Model:	(1)	(2)	(3)	(4)
Variables				
employment $\in$ 10-19	0.0068***		-0.0533***	
	(0.0017)		(0.0041)	
employment $\in$ 20-49	0.0412***		-0.0444***	
	(0.0022)		(0.0037)	
employment $\in$ 50-99	0.0930***		0.0204**	
	(0.0024)		(0.0091)	
employment $\in$ 100-199	0.1332***		0.0858***	
	(0.0028)		(0.0123)	
employment $\in$ 200-499	0.1736***		0.2099***	
	(0.0041)		(0.0182)	
employment $\in$ 500-999	0.2011***		0.2950***	
	(0.0052)		(0.0140)	
employment $\in$ 1000-1999	0.2160***		0.3160***	
	(0.0057)		(0.0217)	
employment $\in$ 2000-4999	0.2319***		0.3688***	
	(0.0078)		(0.0215)	
employment $\in$ 5000+	0.2346***		0.3957***	
	(0.0068)		(0.0147)	
log(employment)		0.0271***		0.0273***
		(0.0008)		(0.0029)
Fixed-effects				
time	Yes	Yes	Yes	Yes
Fit statistics				
Observations	1,469,215	1,469,215	481,912	481,912
R^2	0.00204	0.00173	0.00992	0.00685
Within R^2	0.00132	0.00102	0.00424	0.00116

Clustered (time) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 10: Regressions of firm log skill premia on firm size bins (columns 1,3) and log employment (columns 2,4), including time fixed effects.

Dependent Variables:	log(Skill Premium College)		log(Skill Premium Graduate)	
Model:	(1)	(2)	(3)	(4)
Variables				
employment $\in$ 10-19	0.0160***		0.0043	
	(0.0027)		(0.0051)	
employment $\in$ 20-49	0.0332***		0.0322***	
	(0.0049)		(0.0081)	
employment $\in$ 50-99	0.0449***		0.0721***	
	(0.0063)		(0.0108)	
employment $\in$ 100-199	0.0486***		0.1058***	
	(0.0079)		(0.0154)	
employment $\in$ 200-499	0.0435***		0.1231***	
	(0.0108)		(0.0179)	
employment $\in$ 500-999	0.0307**		0.1174***	
	(0.0128)		(0.0205)	
employment $\in$ 1000-1999	0.0195		0.0966***	
	(0.0161)		(0.0236)	
employment $\in$ 2000-4999	0.0140		0.1096***	
	(0.0182)		(0.0309)	
employment $\in$ 5000+	0.0201		0.1087**	
	(0.0254)		(0.0364)	
log(employment)		0.0204***		0.0374***
		(0.0028)		(0.0051)
Fixed-effects				
time	Yes	Yes	Yes	Yes
firm	Yes	Yes	Yes	Yes
Fit statistics				
Observations	1,469,215	1,469,215	481,912	481,912
R^2	0.66476	0.66477	0.71178	0.71180
Within R^2	8.8×10^{-5}	0.00011	0.00025	0.00030

Clustered (time & firm) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 11: Regressions of firm log skill premia on firm size bins (columns 1,3) and log employment (columns 2,4), including time and firm fixed effects.

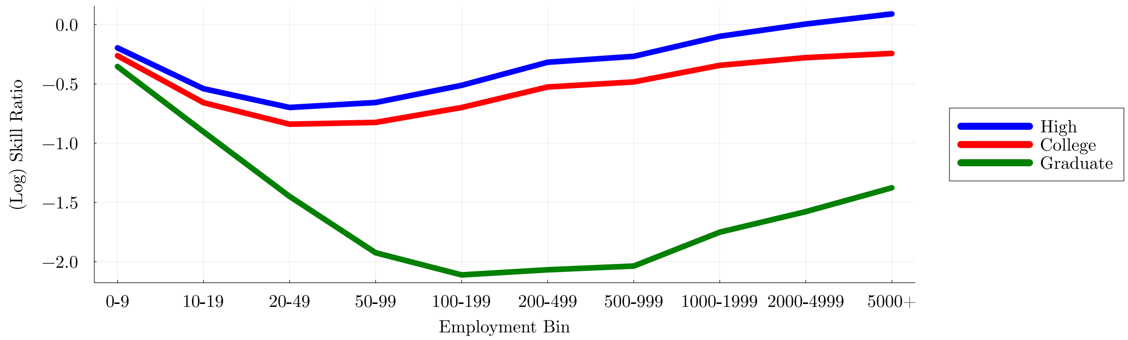


Figure 17: Average skill ratios for each firm size bin, broken by skill type. The skill ratio for skill type X is defined as the number of workers will skill X divided by the number of workers with high school or less education. High-skilled workers have a college or graduate education.

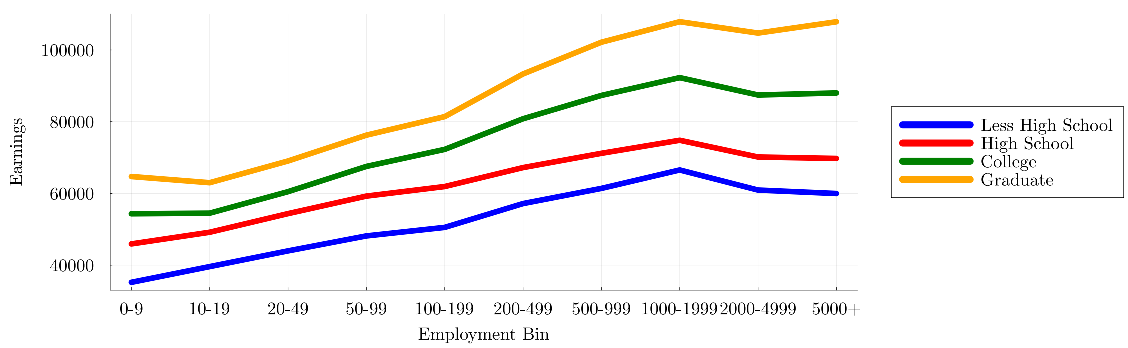


Figure 18: Average earnings within each firm size bin, broken by skill

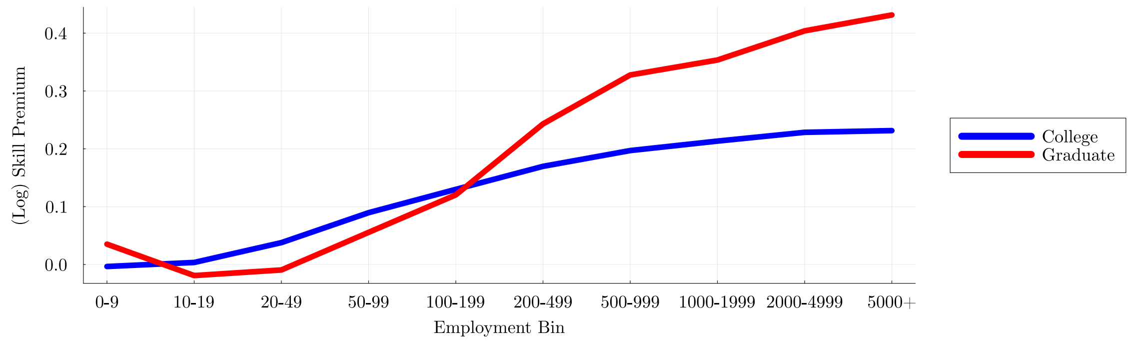


Figure 19: Average skill premia within each firm size bin, broken by skill. The X measure of skill premia within a firm is the ratio of the average earnings of workers with education X to the average earnings of a high school worker within the firm.

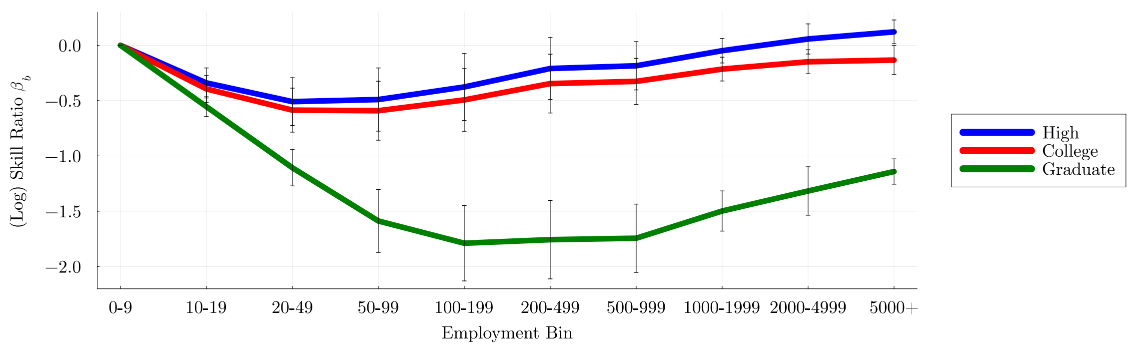


Figure 20: Coefficients from regressions of the log skill ratio on firm size, controlling for SA4+GCCSA, from 7

In Figure 19 I document how log skill premia vary with firm size. Whereas the above trends in earnings revealed that all worker types tend to earn more at larger firms, the fact that skill premia are rising with firm size reveals that high-skill workers see their earnings rise relatively more than low-skill workers. This pattern is consistent with larger firms operating more skill-biased technology, as I posit below.

I now summarize by firm age. In Figure 21 I show the distribution of firms and employment by age, where I bin all firms with age ≥ 11 together, for the reasons explained in Section 2. Most firms are at least a decade old, and most workers are employed in these same firms.

In Figure 22 I display the log skill ratios of firms by firm age. The measure is fairly flat across young ages, then declines more steeply for older ages. Similarly to the size patterns, I argue below this aggregate pattern is consistent with a combination of two forces: firms tend to decrease their skill ratio as they age, and firms with an initially higher skill ratio tend to survive longer. For young ages, these two forces approximately cancel out, with exiting firms raising the skill ratio, but surviving firms counteracting by shifting toward lower skill ratios. For older ages, the exit effect is less prevalent, so the deskilling effect within firms dominates.

In Figure 23 I consider worker earnings by firm age, broken by skill. Workers tend to

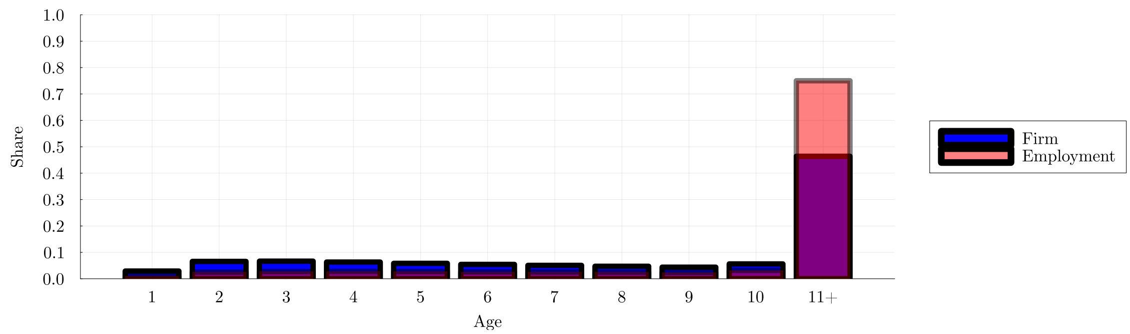


Figure 21: Distribution of firms and employment by firm age

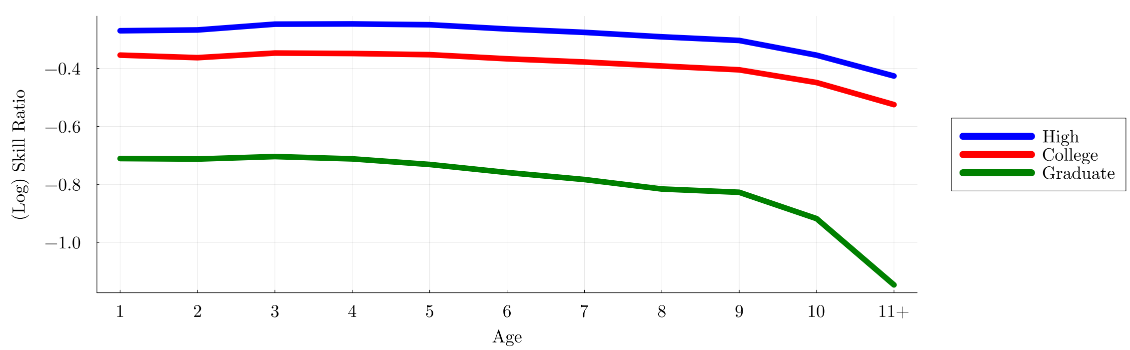


Figure 22: Average skill ratios for each firm age, broken by skill type. The skill ratio for skill type X is defined as the number of workers will skill X divided by the number of workers with high school or less education. High-skilled workers are simply workers that have a college or graduate education.

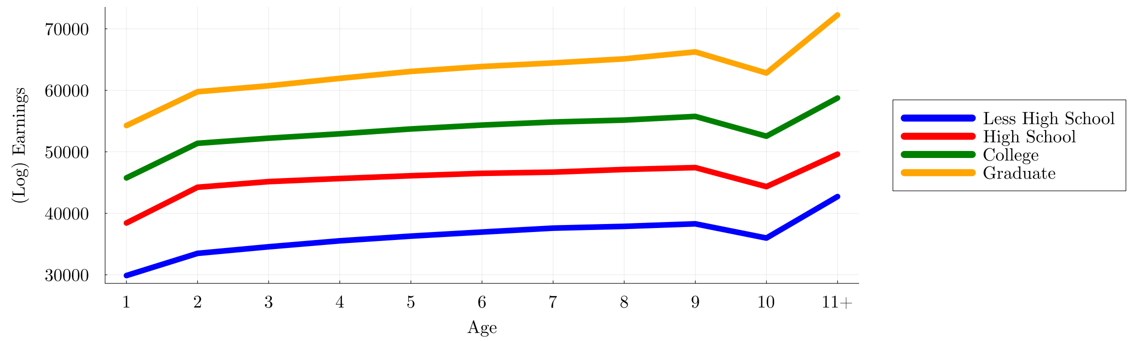


Figure 23: Average earnings for each firm age, broken by skill

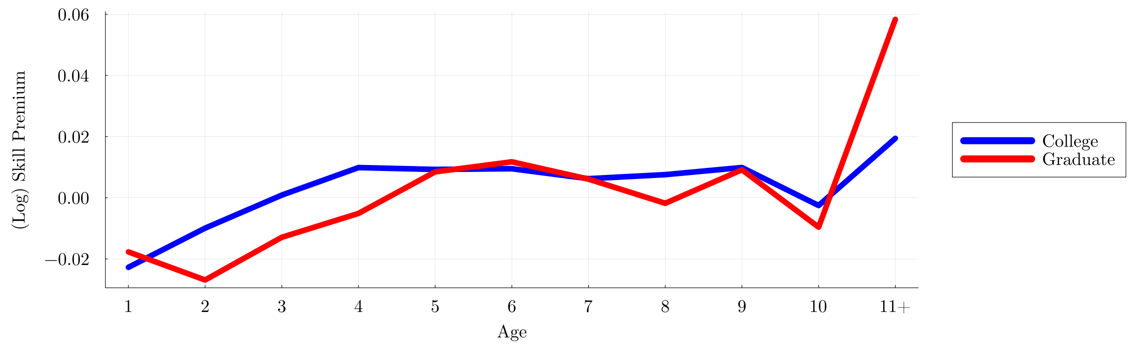


Figure 24: Average skill premia within each firm size bin, broken by skill. The X measure of skill premia within a firm is the ratio of the average earnings of workers with education X to the average earnings of a high school worker within the firm.

earn more in older firms, however, turning to Figure 24 I find that skill premia are fairly constant across the age distribution.

As discussed in B.2, the availability of worker occupation allows me to sort workers into a hierarchy within each firm. Given these hierarchies, it is then possible to consider how measures of skill ratio and skill premia vary with firm organizational structure. This work is ongoing, and will be a separate project.

B.5 Motivating the Separation Elasticity

The baseline simple and quantitative models assume upward-sloping labor supply curves, which yield tractability, but are static models of monopsonistic competition. In contrast, there is evidence that in reality at least some share of labor market monopsonistic (or oligopsonistic, as in Jarosch, Nimczik, and Sorkin (2024)) forces are dynamic, following the terminology in Manning (2013) , in the sense that worker movement between firms is frictional, and this is what allows firms to markdown wages. To understand this, consider a simple model[35] where firms post wages, and homogeneous workers search for jobs. In the stationary distribution, the measure of recruits, $R(w)$, at each firm offering w each instant must exactly offset the measure separating, $n(w)s(w)$, so $R(w) = n(w)s(w)$. Then defining

[35] I follow Hirsch et al. (2022) in making these points.

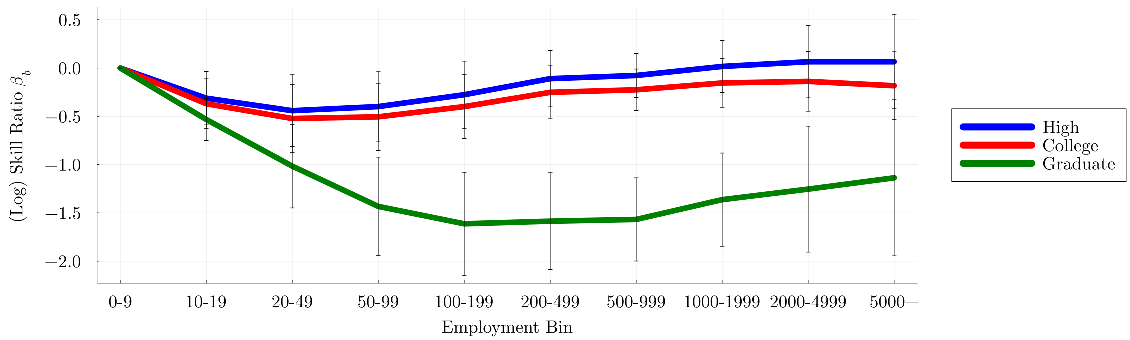


Figure 25: Coefficients from regressions of the log skill ratio on firm size, controlling for industry division, from Table 5

$\epsilon_x(w) \equiv \frac{\mathrm{d}\ln x(w)}{\mathrm{d}\ln w}$, it must be that $\epsilon_n(w) = \epsilon_R(w) - \epsilon_s(w)$, so the labor supply elasticity is the difference between the recruiting and separation elasticities. Furthermore, if it is assumed that every recruit at one firm is a separation from another firm (so ignoring recruitment from nonemployment), then the average recruit elasticity must equal the average separation elasticity. If it is further assumed that these elasticities are constant across firms, then it must be the case that $\epsilon_R = \epsilon_s$, so $\epsilon_n = -2\epsilon_s$. This result motivates estimating the labor supply elasticity via estimating the separation (or quit) elasticity as I do in Section 4. The assumptions made to arrive at this conclusion may not be benign, however, and thus I view this exercise as closer to a back-of-the-envelope calculation than a robust measure of the labor supply elasticities firms face.

B.6 Measured firm age distribution

Firm ages are constructed using firm birth years derived from a combination of Business Activity Statements and the ABS Businesss Registers, but the measure is imperfect. First, the earliest birth date recorded in the data is the $1992 - 1993$ fiscal year, leading to a winsorizing for older firms at that birth year. Second, and more problematic, is that the introduction of the goods and services tax (GST) in the year 2000 changed what qualified as a business entity, and thus many firms that previously did not file taxes were required to start that year, and in turn many firms were "born" that year. These two issues combine to yield an "age" distribution with two discrete jumps.

To understand how these measurement issues should affect the observed age distribution for my sample period, consider a simple model of firm entry and exit, where a fixed measure m_t of firms enter each year t, then exit at rate δ. Then in year t, there are $(1-\delta)m_{t-1}$ firms of age 1, and $(1-\delta)^a m_{t-a}$ firms of age a. The observed firm age distribution can then be roughly approximated by assuming m_t is constant in every year except for FY 1993, where m_t is larger, and FY 2001, where m_t is much larger. With these assumptions I generate the age distribution in Figure 26.

The two bumps in mass can be observed, and by selecting an appropriate δ the model does surprisingly well at replicating the patterns overall. The only feature missed is the tail of super young firms. The reason for this is that the exogenous exit pattern assumed is invalid for the youngest firms, which often enter and then promptly exit, so it is as if δ is

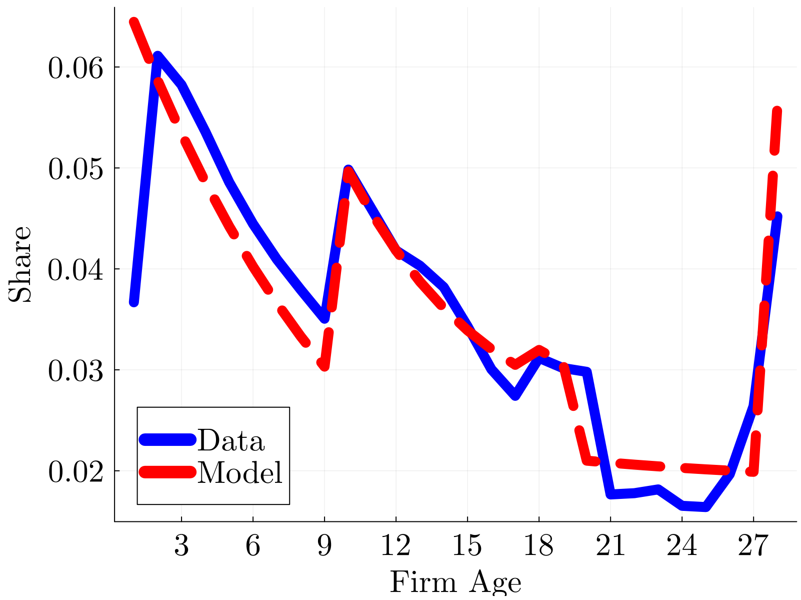


Figure 26: Observed age distribution and predicted age distribution from model assuming firms are "born" due to measurement error in FY 1993 and 2001.

quite high in the first year.

C Additional Details for Quantitative Model

C.1 Augmented Labor Supply Microfoundation with CES

It is possible to generate the CES supply curves in Section 7 using the microfoundation in A.1, but additionally allowing for differences in supply shifter A differences across firms. The relevant problem for workers is then

$$W \equiv \max_{n(w,A)} \left(\int (wAn(w,A))^{\frac{\eta}{\eta+1}} \, \mathrm{d}F(w,A) \right)^{\frac{\eta+1}{\eta}}$$
$$1 = \int n(w,A) \mathrm{d}F(w,A)$$

The solution is the following labor supply schedule and appropriate wage index, where the integral now must account for the joint distribution $F(w, A)$.

$$n(w,A) = \left(\frac{Aw}{W}\right)^{\eta} \overline{n}$$
$$W = \left(\int (Aw)^{\eta} \mathrm{d}F(w,A) \right)^{\frac{1}{\eta}}$$

There are two ways to think about the exact way the A shifters are incorporated to generate the workers' preferences above. One way is to assume that A are firm amenities, so instead of valuing only the earnings w at a firm, workers value the product of the earnings and the shifter, Aw, then have love-of-variety. In this interpretation, firms with better A face greater supplies of workers purely because workers are more willing to work at these firms.

The second interpretation is to assume workers continue to value all jobs equally, but that firms post A^η unique vacancies. Then firms recognize that posting wage w will result in $\left(\frac{w}{W}\right)^\eta \overline{n}$ hires per vacancy, so total supply is $\left(\frac{Aw}{W}\right)^\eta \overline{n}$. In this interpretation, firms with better A are able to offer more job variety to workers, so may post a lower wage to obtain any desired labor quantity.

It is worth clarifying that, under both interpretations presented here, the advantages conferred to firms with higher A are exogenous and costless to the firm. However, it is possible to endogenize these outcomes, say due to heterogeneous vacancy-posting, and I detail this mechanic in C.4 below.

C.2 Augmented Labor Supply Microfoundation with Continuous Choice

The continuous choice interpretation of the labor supply curves from A.2 may also be extended to include the A shifters, and again there are two possible interpretations. In the amenity interpretation, workers receive utility $Aw\epsilon$, where ϵ is iid Frechet(η), so more workers choose to work at firms with higher A. In the second interpretation, firms post A^η vacancies, so for a given wage workers are no more likely to value a job at a high A firm versus a low A firm, but high A firms have more jobs.

C.3 Firm Optimality Conditions

Similarly to the A.3, the problem of firm (z, κ, A_L, A_H) is

$$\max_{n_i} z\left(n_L^{\frac{\sigma-1}{\sigma}} + (\kappa n_H)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} - \sum_{i\in\{L,H\}} \left(\frac{n_i}{\overline{n}_i}\right)^{\frac{1}{\eta_i}} \frac{W_i}{A_i} n_i$$

The first-order condition for employment of each worker type is

$$zN^{\frac{1}{\sigma}} n_L^{-\frac{1}{\sigma}} = \frac{\eta_L + 1}{\eta_L}\left(\frac{n_L}{\overline{n}_L}\right)^{\frac{1}{\eta_L}} \frac{W_L}{A_L} \tag{37}$$

$$zN^{\frac{1}{\sigma}} \kappa^{\frac{\sigma-1}{\sigma}} n_H^{-\frac{1}{\sigma}} = \frac{\eta_H + 1}{\eta_H}\left(\frac{n_H}{\overline{n}_H}\right)^{\frac{1}{\eta_H}} \frac{W_H}{A_H} \tag{38}$$

In this form, it is clear that having a higher A_i is equivalent to observing a lower wage index W_i, a notion which makes precise the intuition that higher A firms face less competition.

C.4 Generating heterogeneous A with heterogeneous vacancy costs

When firms face heterogeneous vacancy-posting costs, it is as if the firms with lower costs face increased supply curves. I clarify this notion.

Suppose the following timing for firms making employment decisions

(i) Firms learn their vacancy cost parameters $\tilde{A}_i$, but do not know (z, κ)

(ii) Firms choose how many vacancies to post, subject to cost $c(v, \tilde{A})$

(iii) Firms learn (z, κ)

(iv) Firms decide wages w_i for their posted vacancies

The problem of choosing vacancies may be solved by backward induction. The last step is to make the employment decisions discussed in the main text. Let the value of these outcomes, given (z, κ) draw and vacancies (v_L, v_H), be $V(z, \kappa, v_L, v_H)$. Then let $\mathbf{V}(v_L, v_H) \equiv \mathbb{E}_F\left[V(z, \kappa, v_L, v_H)\right]$ be the ex ante expected value of posting vacancies (v_L, v_H), given distribution $F(z, \kappa)$. The vacancy-posting problem of the firm is then

$$\max_{v_L, v_H} \mathbf{V}(v_L, v_H) - c(v_L, \tilde{A}_L) - c(v_H, \tilde{A}_H)$$

As $\mathbf{V}$ is convex[36], assume that c is sufficiently convex so that the problem has a well-defined solution. That solution satisfies

$$\begin{aligned} \mathbf{V}_L &= c_2(\tilde{A}_L, v_L) \\ \mathbf{V}_H &= c_2(\tilde{A}_H, v_H) \end{aligned}$$

If higher $\tilde{A}$ lowers the marginal cost of vacancies, i.e. $c_{12} < 0$, then firms with higher $\tilde{A}_i$ will post higher v_i. Then the result is achieved, that heterogeneity in firm vacancy costs $\tilde{A}$ generates endogenous differences in posted vacancies v. As this justification is not required for any of the results in the main text, and including it would distract from the main points, I simply say $A_i \equiv v_i(\tilde{A}_L, \tilde{A}_H)^{\eta_i}$ and avoid discussing vacancies altogether.

D Model with Firm Dynamics: Entry and Exit

In the main text, my quantitative model is static, and simply a distribution of firms over parameters governing productivity and labor supply. I now build a richer version of the model which allows for more general distributions of TFP and SBP, and accounts for firm entry and exit decisions.

[36] Consider scaling both v_i by $\lambda \in \mathbb{R}$. Then a firm can post the same wages as before and scale their revenues and costs by λ, but in general it will not be optimal to do so, and they will adjust their wages, so $\mathbf{V}$ will scale by more than λ.

I opt for constant elasticity of substitution labor supplies, which lends tractability to the analysis. These forms allow for relatively simple computation of the profit function during solving for the equilibrium, and generate simple to compute wage distributions across and within firms. Nonetheless, the CES assumption is rather strong, so in Appendix D.5 I discuss alternative assumptions which generate labor supply curves with variable markdowns. It is also possible to microfound the upward-sloping labor supply curves firms face via worker search, and I explain how this may be done in E. The fact that a firm's hiring decisions will depend on the firm's current skill composition means that changes in the distribution of skill will impact worker dynamics, such as EE flows, and job creation and destruction. These worker dynamics are not the focus of this paper, but I am considering these questions in other work.

I now explain the model.

D.1 Setup and Equilibrium

Overview Time is continuous. Workers have preferences across firms which generate monopsonistic competition in each labor market. An infinite pool of potential firms exists, which may enter by paying entry cost c_e, and all firms are risk-neutral. Upon entering, firms draw $(z, \kappa) \sim H$, then face an exogenous and idiosyncratic stochastic process for z, which follows a geometric Brownian motion with drift, and κ is permanent. Firms may exit at will.

Workers Each worker has a permanent type $i \in \{L, H\}$ and is endowed with one unit of labor. There are $\overline{\nu}_i$ workers of type i, so the labor supply curves faced by each firm are[37]

$$\nu_i(w) = \left(\frac{w}{W_i}\right)^{\eta_i} \overline{\nu}_i \tag{39}$$

$$W_i = \left(\int_0^\infty w^{\eta_i} \mathrm{d}G_i(w)\right)^{\frac{1}{\eta_i}} \tag{40}$$

where G_i is the equilibrium distribution of wages for type i. Workers own the firms, and in particular each worker owns the same index of all firms, so that profits are the same for all workers[38]. This assumption is a simple way to make sure that profit incentives do not distort labor supply decisions. There is no saving technology, so workers spend the entirety of their combined labor and profit income on goods each instant.

Firms Every instant, firms must decide whether to continue operating, and if they continue operating they must decide how much of each type of labor to employ. Conditional on operating, a firm maximizes profits, accounting for the monopsonistic power afforded by the upward sloping labor supply curves from workers.

[37] I offer two possible microfoundations for these supply curves in Appendices A.1 (CES love-of-variety in earnings) and A.2 (continuous choice with preference shocks).

[38] Some firms are unprofitable, and in these cases the workers are making investments that never earn a positive return. For example, when firms enter, workers invest c_e to start the firm, but some of these firms immediately realize they do not expect to turn a profit, so exit. Other firms operate because they initially expect to turn a profit, but after a series of unlucky TFP draws no longer expect profits, so exit.

$$\pi(z,\kappa)=\max_{\nu_i} z\left(\nu_L^{\frac{\sigma-1}{\sigma}}+(\kappa\nu_H)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}\alpha}-\sum_{i\in\{L,H\}} w_i(\nu_i)\nu_i-c_f$$

where $w_i(\nu_i)$ are the inverse labor supply curves, c_f is the fixed operating cost, and α governs the returns to scale for the labor bundle in producing output. The firm optimality conditions for choosing labor are derived in A.3, and imply that each firm will hire each worker type up to the point where the marginal revenue product of labor equals the marked down wage, where the markdowns are $\frac{\eta_i}{\eta_i+1}$. In the relevant case where $\eta_L \neq \eta_H$, there is no closed-form solution for (ν_L, ν_H).

To determine whether or not to exit, firms must evaluate the expected present discounted value of continuing to operate, accounting for the fact that future exit decisions will be optimally made, given the exogenous stochastic process for z[39]. They discount the future at rate ρ, therefore the problem in sequence form is

$$v(z_0,\kappa)=\max_{\tau}\mathbb{E}\left\{\int_0^{\tau} e^{-\rho t}\pi(z_t,\kappa)\mathrm{d}t\right\} \tag{41}$$

$$\mathrm{d}\ln z_t=\mu\mathrm{d}t+\sigma\mathrm{d}W_t \tag{42}$$

In D.3 I show how to derive the recursive form of this problem, which is the following Hamilton-Jacobi-Bellman-Variational-Inequality.

$$0=\min\{v,\rho v(z,\kappa)-\pi(z,\kappa)-\mu z v_z(z,\kappa)-\frac{\sigma^2}{2}z^2 v_{zz}(z,\kappa)\} \tag{43}$$

Firms deciding whether to enter must assess whether the expected profitability of entrance covers the entry cost. In other words, firms will continue to enter as long as $\int v(z,\kappa)\mathrm{d}H(z,\kappa) > c_e$. When the inequality goes the other direction, no firms will enter. The expected value of entering goes to infinity as the measure of operating firms goes to zero, as there is no competition for workers in that case, and the expected value of entering goes to zero as the measure of operating firms goes to infinity, as all profits are competed away in that case. Therefore, in equilibrium the expected value of entrance will exactly offset the entry cost.

Given the entry and exit decisions of firms, it is possible to calculate the implied distribution of operating firms via the following stationary Kolmogorov Forward equation, which holds for $(z,\kappa)\in\mathcal{O}$ where firms do not exit.

$$0=-\partial_z\left[\mu z g(z,\kappa)\right]+\partial_{zz}\left[\frac{\sigma^2}{2}z^2 g(z,\kappa)\right]+mh(z,\kappa) \tag{44}$$

The measure of firms entering each instant is m[40]. Outside the region $\mathcal{O}$, where firms operate, it must be the case that $g(z,\kappa)=0$.

[39] My formulation and solution of this augmented Hopenhayn (1992) model of firm dynamics follows the continuous-time approach suggested in Moll (2023).

[40] Note that I am using the convention that m only includes firms that enter after observing their initial draw. This terminology is not always clear and consistent in work building on Hopenhayn (1992) and Melitz (2003).

Equilibrium A stationary equilibrium of the model is a set $\{\nu_i(z,\kappa), w_i(z,\kappa), v(z,\kappa), g(z,\kappa), \mathcal{O}, m\}$, where ν_i are labor quantities, w_i are wages, v are firm valuations, g is the distribution of operating firms, $\mathcal{O}$ is the subset of (z,κ) space such that firms continue operating, and m is the measure of entering firms, such that

(i) Given the wages and distribution of firms over (z,κ), workers supply the labor quantities $\nu_i(z,\kappa)$

(ii) Given the wages and distribution of firms over (z,κ) for all other firms, the firm at (z,κ) offers wages $w_i(z,\kappa)$ an employs labor $\nu_i(z,\kappa)$, provided they are operating

(iii) Given the exogenous process for TFP and the profits from optimally demanded labor as specific above, firm valuation satisfies the optimality condition in equation 43, and yields operating subset $\mathcal{O}$

(iv) Given the operating region $\mathcal{O}$ and entry rate m, the distribution of firms satisfies the Kolmogorov Forward equation 44.

(v) Potential firms are indifferent between entering and not entering: $\int v(z,\kappa)\mathrm{d}H(z,\kappa) = c_e$

(vi) The goods market clears, meaning that every good produced is purchased by some worker.

The stationary equilibrium is block recursive in the sense that the only aggregates workers and firms care about are the wage indices. So it is possible to make a guess of the W_i, then in the first "block" calculate the implied employment and entry and exit decisions of firms, and in the second "block" use those decisions to calculate the implied wage indices. I detail how to compute the equilibrium in **??**.

D.2 Markdowns

The model also generates predictions for markdowns for every worker type at every firm type. The CES specification further simplifies calculation, as a worker of type i will earn a constant markdown of $\frac{\eta_i}{\eta_i+1}$ from their marginal revenue productivity of labor[41]. Then the average markdown within a firm is simply the weighted average of the markdowns of the worker types, and the cross-sectional and firm-level markdown patterns mirror the skill ratio results. Higher skill ratios mean more weight is placed on the larger markdown, therefore as firms grow they tend to mark down average wages less, but in the cross-section the average markdown within a firm is inverse U-shaped with respect to firm employment.

The interaction of scale and skill consideration in theory, and the observed empirical patterns, add two new considerations to the recent discussion concerning the markdown distribution across firms. First, firm-level markdowns depend on the composition of workers, and the markdowns paid to each worker type. If high-skill workers tend to accept larger markdowns, then firms which employ a higher share of high-skill workers will tend to have

[41] I review and prove this simple fact in Appendix D.4.

larger markdowns. Second, the skill composition of a firm is not, in general, orthogonal to the scale choice of a firm. Therefore at least part of the larger markdown in larger firms documented in Mertens and Mottironi (2023) may be due not to falling labor elasticity as firms are observed further up the supply curve for each worker type, but to falling average labor elasticity in larger firms as they tend to employ a higher share of inelastic labor.

Understanding the difference between the cross-sectional and firm-level results is also crucial for policy. Consider a policymaker with the objective of reducing wage markdowns. Based on the cross-sectional observation that larger firms pay lower markdowns, they implement a progressive payroll tax, with the thought the firms paying the largest markdowns will be most incentivized to descale, and thus decrease the gap between wages and MRPL. Beyond the fact that the presence of monopsonistic competition will amplify tax distortions, as argued in Berger et al. (2024), the policy will do the exact opposite of its intended purpose, as all firms will reduce their scale, but in doing so raise the share of their workforce which is high-skill, and thus marked down most aggressively, raising their firm-level markdown. This is exactly the opposite prediction from what would be obtained in models similar to Melitz and Ottaviano (2008) with falling labor supply elasticities, because in those models the cross-sectional differences between firms are identically the differences within a firm over time, or "large firms are just small firms that get big". Instead, once I account for the differences in skill-biased productivity across firms, large and small firms are fundamentally different in how productive they find each skill, and so policies that incentivize descaling will not necessarily cause large firms to look like small firms.

The above argument points to a potentially troubling tradeoff for policymakers: when firms scale up, between-worker inequality within the firm rises, but so does the worker share of earnings. The manager of a firm only cares about total profits, so after a positive TFP shock they raise profits by scaling up production, raising the wage bill even more than profits, and within that wage bill the relative increase in high-skill payments rises more than the payments to low-skill workers.

It is worth noting that the CES assumption means the entirety of the differences in skill premia across firms come from (equilibrium) differences in MRPL between worker types. Firms with higher TFP will tend to skew the employment toward low-skill workers, where supply is more elastic, and in doing so will lower the MRPL of these workers relatively more than the MRPL of high-skill workers, but the contribution from markdown differences to skill premia is unaffected. This strong result is easily broken, for example by any supply curve with variable elasticity, and in Appendix D.5 I explain how this would then allow a portion of skill premia differences between firms to come from changes in markdowns within firm type.

D.3 Sequence to recursive form of firm problem

Recall the sequence formulation of the problem

$$v(z_0, \kappa) = \max_{\tau} \mathbb{E}_0 \left\{ \int_0^{\tau} e^{-\rho t} \pi(z_t, \kappa) \mathrm{d}t \right\} \tag{45}$$

$$\mathrm{d} \ln z_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t \tag{46}$$

In order to derive the recursive form, first generalize the problem to allow for time dependence in the value function, as

$$v(z_t, \kappa, t) = \max_{\tau} \mathbb{E}_t \left\{ \int_t^{\tau} e^{-\rho(s-t)} \pi(z_s, \kappa) \mathrm{d}s \right\} \tag{47}$$

$$\mathrm{d} \ln z_s = \mu \mathrm{d}s + \sigma \mathrm{d}W_s \tag{48}$$

Differentiate both sides with respect to t

$$v_t = -\pi(z_t, \kappa) + \rho v(z_t, \kappa, t) - \mu z v_z(z_t, \kappa, t) - \frac{\sigma^2}{2} z^2 v_{zz}(z_t, \kappa, t)$$

The left side is straightforward. The first term on the right side comes from the change in the lower bound of integration, and the second term comes from the change in the discounting term. The third and fourth terms account for the change in expectation of z which occurs in with marginal change in time, according to the geometric Brownian motion process. In the stationary equilibrium $v_t = 0$ and v does not depend on t, and rearranging the equation yields the recursive form for firms that continue to operate.

$$\rho v(z, \kappa) = \pi(z, \kappa) + \mu z v_z(z, \kappa) + \frac{\sigma^2}{2} z^2 v_{zz}(z, \kappa)$$

As it stands, the above equation simply values a firm that operates forever, as the option value of exiting has been ignored. To account for the exit option, I also require $v(z, \kappa) \geq 0$ everywhere. Then the full description of firm value is that it satisfies the above equation except if that would imply $v < 0$, in which case $v = 0$ instead. The flow value for a given state (z, κ) is $\rho v(z, \kappa)$, and when $v \geq 0$ this equals the right side of the above equation, but when $v = 0$ the value provided by the right side of the equation must be less than or equal to zero, else the firm would not exit. Then the value satisfies one of the following complementarity conditions.

$$\rho v(z, \kappa) = \pi(z, \kappa) + \mu z v_z(z, \kappa) + \frac{\sigma^2}{2} z^2 v_{zz}(z, \kappa), \quad v(z, \kappa) \geq 0$$
$$\rho v(z, \kappa) \geq \pi(z, \kappa) + \mu z v_z(z, \kappa) + \frac{\sigma^2}{2} z^2 v_{zz}(z, \kappa), \quad v(z, \kappa) = 0$$

The above form reveals that finding v is a linear complementarity problem (LCP). Fortunately, LCPs are well understood and the solution is not difficult to find computationally[42]. A compact way to write the condition is

$$0 = \min\{v, \rho v(z, \kappa) - \pi(z, \kappa) - \mu z v_z(z, \kappa) - \frac{\sigma^2}{2} z^2 v_{zz}(z, \kappa)\}$$

This is the form of a Hamilton-Jacobi-Bellman-Variational-Inequality.

[42] I use the `Julia` function `mcpsolve` from the `NLsolve` library.

D.4 CES implies constant markdown

Consider a production function $y(n)$ and a labor supply curve $n(w)$. The problem of a firm is

$$\max_{w} y(n(w)) - wn(w)$$

The optimality condition is

$$y'(n(w))n'(w) = n(w) + wn'(w)$$

Recognizing that $y'(n)$ is the marginal revenue productivity of labor (MRPL) and defining $\eta(w) \equiv \frac{\mathrm{d}\ln n(w)}{\mathrm{d}\ln w} = \frac{n'(w)w}{n(w)}$ as the elasticity of labor supply, the above expression may be rearranged to yield

$$w = \frac{\eta(w)}{\eta(w)+1} MRPL$$

This fairly general result, that the wage markdown may be written as a function of the labor supply elasticity alone, only requires the assumption that firms are profit maximizing. The CES result is then trivial, as $n(w) = \left(\frac{w}{W}\right)^{\eta} \overline{n}$ implies the elasticity is constant for any point on the supply curve: $\frac{\mathrm{d}\ln n(w)}{\mathrm{d}\ln w} = \eta$.

D.5 Variable Markdowns

In this section, I discuss the implications of allowing firms to face supply curves with variables elasticities of substitution, a la Melitz and Ottaviano (2008) and discussed for goods demand in the appendix of Autor et al. (2020). This flexibility creates an additional force through which markdowns may differ across firms, as the average markdown within a firm will not depend not only on the composition of employment, as in the CES case, but also on the scale of employment, as firms must move up labor supply curves and the elasticities for each worker type vary.

Consider a worker seeking to maximize their earnings subject to a linear-quadratic constraint on their total amount of labor. In particular, their problem is

$$\begin{aligned}
\max_{n(w)} \int wn(w)\mathrm{d}F(w) \\
1 = \alpha\mathbb{E}[n(w)] + \frac{\beta}{2}\mathbb{V}[n(w)] \\
\mathbb{E}[n(w)] \equiv \int n(w)\mathrm{d}F(w) \\
\mathbb{V}[n(w)] \equiv \int (n(w))^2 \,\mathrm{d}F(w)
\end{aligned}$$

The optimality conditions are

$$w - \lambda\left(\alpha + \beta n(w)\right) = 0$$

where λ is the Lagrange multiplier for the total labor constraint. Rearranging reveals linear supply curves

$$w = \tilde{\alpha} + \tilde{\beta} n(w)$$

where the tildes are to indicate the changes in the constants due to the Lagrange multiplier. The elasticity of supply on these curves is

$$\frac{\mathrm{d}\ln n(w)}{\mathrm{d}\ln w} = \frac{n'(w)w}{n(w)} = \frac{\frac{1}{\tilde{\beta}}w}{\frac{1}{\tilde{\beta}}\left(w - \tilde{\alpha}\right)} = \frac{w}{w - \tilde{\alpha}}$$

Unless $\tilde{\alpha} = 0$, the elasticity varies as firms move along the curve. When $\tilde{\alpha} > 0$, the elasticity approaches 1 from above, and when $\tilde{\alpha} < 0$, the elasticity approaches 1 from below. By the reasoning in the previous section, that the wage markdown depends on the labor supply elasticities alone, it is the case that markdowns also vary along the supply curve. With supply curves like these, firms make scale and composition decisions accounting for the fact that moving up supply curves may allow for increased or decreased optimal markdowns. For example, in the CES case firms would never choose to scale down in order to obtain more surplus per labor input, but with variable supply elasticities firms may consider this option.

E Model with Firm Dynamics and Worker Dynamics

E.1

I now build a model of the aggregate labor market of Australia, able to speak to patterns of worker job and earnings dynamics, as well as firm dynamics in scale and skill composition. The most novel aspect of the model is that the firm-level labor supply elasticity differences between skill types arise endogenously, due to the interaction between worker search and the equilibrium distribution of labor demands from other firms. Intuitively, all firms prefer to employ high-skill workers, so it is, on average, more difficult to poach a high-skill worker than a low-skill worker. Growing firms then skew their hiring toward low-skill workers.

Overview Time is continuous. Workers and firms are risk-neutral. Workers search when unemployed and on-the-job for wage contracts that maximize their expected present discounted value (EPDV) of earnings. Firms search for workers and offer contracts in order to maximize their EPDV of profits. Potential firms enter whenever the EPDV of profits is positive, and incumbent firms exit when their EPDV of continuing is negative.

Workers There is a unit mass of workers. The shares of low and high skill workers are $\overline{\nu}_L$ and $\overline{\nu}_H$. I will use ν to denote the vector of labor within a firm, with the convention $\nu \equiv (\nu_L, \nu_H)$.

Firms There is an infinite mass of potential firms, all ex-ante identical. Upon choosing to enter, a firm pays entry cost c_e to draw initial skill-biased productivity (SBP) $\kappa \sim H_\kappa$ and initial total factor productivity (TFP) $z \sim H_z$, and start with $\nu^0 > 0$ workers[43]. The firm may then continue operating indefinitely or exit at will, and at rate δ_f will exogenously dissolve, sending their entire labor force to unemployment. The draw for SBP κ is permanent, and the stochastic process for z is idiosyncratic across firms and obeys the infinitesimal generator $\mathcal{A}$. All firms produce an identical good, which is the numeraire, according to

$$Y(\nu, z, \kappa) = z\left(\nu_L^{\frac{\sigma-1}{\sigma}} + (\kappa\nu_H)^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}\alpha}$$

where $\sigma \in (1, \infty)$ is the elasticity of substitution between worker skill types and $\alpha \in (0, 1)$ governs the returns to scale. Instantaneous profits are

$$\pi(\nu, z, \kappa) = Y(\nu, z, \kappa) - c_f - W(\cdot)$$

where c_f are fixed operating costs. The wage bill W is not generally a function of (κ, ν) alone, as there will be path dependence in wage determination, but I defer discussion of the determination of wages to below.

Operating firms are continuously deciding who to hire, by determining vacancy-posting toward each skill type of worker, and whether it makes sense to exit. In general, firms must account for how the wage distribution of their labor force will evolve in making hiring, firing, and exit decisions. However, the wage contracting structure I impose below allows me to solve the model in allocations, and thus sidestep having to keep track of wage distributions.

Search and Wage Contracting I now specify the mechanics of the labor market. First, I specify how workers and firms search and meet. Second, I describe the space of feasible wage contracts, and how this restriction simplifies computing the equilibrium. Third, I lay out the relevant Hamilton-Jacobi-Bellman and Kolmogorov Forward equations sufficient for determining in equilibrium in allocations. Fourth, I show that the specified contract space allows for simple simulation of wage paths.

Workers search for firms when unemployed and on-the-job. Unemployed workers search with a unit of search intensity, and employed workers search with $\xi < 1$ units of search intensity, so that total search unit of type i are $\mathbf{s} = u_i + (1 - u_i)\xi$, where u_i is the unemployment rate. Firms determine the number of vacancies to post in each labor market $i \in \{L, H\}$, so aggregate vacancies are $\mathbf{v}_i = \int v(x)\mathrm{d}H(x)$, where H is the measure of firms with state x, which I leave unspecified at present, but will below argue $x = (\nu, z, \kappa)$ under the my specified

[43] In principle, there is no reason why firms could not start with $\nu^0 = 0$, but the vacancy cost formulation I use below will imply that firms with zero workers face an infinite vacancy-posting cost, so it is more tractable to assume firms start with some workers to keep these costs finite.

wage contracting environment. The rate at which matches occur is determined by constant returns to scale matching technology $m(\mathbf{v}, \mathbf{s}) = A\mathbf{v}^{\beta}\mathbf{s}^{1-\beta}$. Therefore, search is random within each skill market, but semi-directed in the sense that firms may choose the number of vacancies for each skill. Thus, defining $\theta_i = \frac{\mathbf{v}_i}{\mathbf{s}_i}$ as the tightness in a market, workers find firms at rate $p(\theta) = \frac{m(\mathbf{v},\mathbf{s})}{\mathbf{s}} = A\theta^{\beta}$ per unit of search and firms find $q(\theta) = \frac{m(\mathbf{v},\mathbf{s})}{\mathbf{v}} = A\theta^{\beta-1}$ workers per vacancy posted, and for each match the probability that the worker is unemployed is $\phi_i \equiv \frac{u_i}{\mathbf{s}_i}$. At rate δ, workers exogenously separate to unemployment.

Upon meeting, a Bertrand competition for the worker occurs, in the tradition of Postel–Vinay and Robin (2002), where the incumbent firm and challenging firm may each offer the worker a take-it-or-leave-it wage contract (if the worker is unemployed, unemployment "offers" the wage contract of continued unemployment). I restrict the space of permissible wage contracts as follows: A wage contract (V, ν, z, κ) is a legally binding agreement between a worker and firm with the following conditions:

(i) The firm may fire the worker at will.

(ii) The worker may quit at will, including quitting to unemployment or to another firm.

(iii) Workers may not enter into any contractual agreements with other workers[44].

(iv) As the state of the firm (ν, z, κ) evolves, the firm will update the worker's wage so the job continues to deliver value V to the worker, with the caveat

- If the firm can ever credibly show that firing the worker is preferred to maintaining V, but that there exists $V' \in (0, V)$ such that the firm would maintain the worker, then the contract is updated to (V', ν, z, κ), where V' the maximal value for which the firing threat is credible.

(v) The firm will post vacancies in a "privately efficient" manner, meaning that the vacancy-posting policy will maximize the sum of the surpluses of the firm and all of its workers.

This contract structure satisfies the assumptions prescribed in Bilal et al. (2022), in particular limited commitment, mutual consent, external negotiation, internal negotiation, and privately efficient vacancy-posting. It is similar in spirit to the contracts permitted in Postel-Vinay and Turon (2010) and Lise, Meghir, and Robin (2016), except in those settings all "firms" are simply jobs, and thus the complexities that come from dealing with multiworkers firms are sidestepped.

Generally, solving for the equilibrium wage contract (or lack thereof) between an arbitrary worker and firm meeting is not possible in this class of models, as the wage contract may depend on the entire wage distribution of workers within and between all firms. However, the fact that the allocations of workers may be solved in terms of the joint surplus alone allows for also solving for the equilibrium contracts between workers and firms at any given

[44] This condition rules out workers forming legally binding contracts which then raise credible threats. For example, worker A could contract with worker B that if A does not receive an improved contract in horizon Δt, then A must quit. After Δt has elapsed, the quitting threat is real, due to the contract, and if the marginal value of the worker to the firm exceeds V', the worker will then receive the improved contract, even though there is no real outside threat other than the arbitary internal contract.

meeting. The logic is to view the firm problem of hiring as a mean field game with HJB and KF equations, and then realize that the KF equation is already solved, and the HJB is trivial to solve given the joint surplus solution. For the HJB, at each meeting the firm and worker take as given the stochastic process for (ν, z, κ) within the firm, as well as the joint marginal surplus from matching, and use these objects to value wage contracts and worker flows. Then wage contracts may be found by solving the HJBs. For the contract and allocation decisions to be optimal and feasible, the implied decisions from all meetings must generate worker flows consistent with the assumed process for (κ, ν). As the wage contract structure is already known to be consistent with the allocations from the solution of the joint surplus alone, the KF must hold. This logic, of breaking the problem into choosing aggregate flows and individual contracts, then checking they are consistent, is particular to the limitations put on the wage contract space here, because the surplus problem is comparatively easy to solve, whereas in general the surplus problem required to be solved for the KF will require the distribution of wages as part of the state space, and thus be intractable.

A wage contract is a promise to update worker wages constantly so that worker value of the job remains at a constant value, with renegotiation off the table unless either party can credibly threaten to dissolve the match otherwise. Generally, solving for the equilibrium wage contract (or lack thereof) between an arbitrary worker and firm meeting is not possible in this class of models, as the wage contract may depend on the entire wage distribution of workers within and between all firms. However, the fact that the allocations of workers may be solved in terms of the joint surplus alone allows for also solving for the equilibrium contracts between workers and firms at any given meeting. The logic is to view the firm problem of hiring as a mean field game with HJB and KF equations, and then realize that the KF equation is already solved, and the HJB is trivial to solve given the joint surplus solution. For the HJB, at each meeting the firm and worker take as given the stochastic process for (ν, z, κ) within the firm, as well as the joint marginal surplus from matching, and use these objects to value wage contracts and worker flows. Then wage contracts may be found by solving the HJBs. For the contract and allocation decisions to be optimal and feasible, the implied decisions from all meetings must generate worker flows consistent with the assumed process for (κ, ν). As the wage contract structure is already known to be consistent with the allocations from the solution of the joint surplus alone, the KF must hold. This logic, of breaking the problem into choosing aggregate flows and individual contracts, then checking they are consistent, is particular to the limitations put on the wage contract space here, because the surplus problem is comparatively easy to solve, whereas in general the surplus problem required to be solved for the KF will require the distribution of wages as part of the state space, and thus be intractable.

It is instructive to work through the situations in which a wage contract will be renegotiated, before viewing the relevant value equations which determine how wages are set.

(i) Suppose a worker meets a challenging firm.

  (a) If the challenger cannot profitably offer a contract which dominates their current contract, no renegotiation will occur.

  (b) If the challenger can profitably offer a dominating contract, but not enough so to poach the worker, renegotiation will occur and yield an improve contract that is minimally acceptable to keep the worker.

(c) If the challenger can profitably offer a dominating contract which the incumbent cannot match, the worker will move to the challenging firm, and receive a contract that is as valuable to them as the best possible contract the incumbent could have offered.

(ii) Suppose (ν, z, κ) changes.

(a) If the marginal surplus from the match is negative, the match will dissolve, as no wage contract can satisfy both the firm and the worker.

(b) If the marginal surplus share of the worker falls below zero, because future opportunities for better wage contracts have worsened, but the total marginal surplus of the match remains positive, the worker can credibly threaten to quit, so renegotiation occurs, minimally improving the contract for the worker to prevent quitting.

(c) If the marginal surplus share of the firm falls below zero, because future production opportunities with the worker have worsened, but the total marginal surplus of the match remains positive, the firm can credibly threaten to fire, so renegotiation occurs, minimally improving the contract for the firm to prevent firing.

As the marginal surplus of a match is equal to the sum of the marginal worker value and the marginal firm value (including all incumbent workers at the firm), it is sufficient to consider the worker values for determining when and how contracts are updated. The value of a contract V at firm (ν, z, κ) to a worker[45] satisfies

$$\begin{aligned}\rho V &= w(V, \nu, z, \kappa) - b + (\delta + \delta_f)[0 - V] \\ &\quad + \xi p(\theta) \int_V^\infty [\min\{S_\nu(\nu, z, \kappa), S'_\nu\} - V] \mathrm{d}H(S'_\nu) \\ &\quad + \mathcal{B}(\nu, z, \kappa)[V]\end{aligned}$$

Discounting is at rate ρ. The first term is the flow surplus the worker earns from the job, which is the wage net of the unemployment benefit they would earn if unemployed. The second term accounts for exogenous separation by workers and exogenous dissolution of firms. The third term accounts for the increases in value a worker may earn from meeting other firms, where the maximal gain is the total marginal surplus of the current firm, but other offers may induce a wage contract negotiation that raises the value at the current job between the current value and maximal value of incumbent marginal surplus. The final term accounts for flows in (ν, z, κ) that may affect the marginal value of the worker, accounting for the asymmetry in how V may change so that when S_ν increases V is unchanged, but when S_ν falls below V, then V will be updated to S_ν. For a concrete example, suppose there are only 3 points in the state space, with values $a < b < c$, and transitions occur according to the following transition rate matrix.

[45] I drop type dependence here, to simplify notation, with the understanding that the logic applies to each worker type.

$$B = \begin{bmatrix} -0.1 & 0.05 & 0.05 \\ 0.2 & -0.3 & 0.1 \\ 0.1 & 0.4 & -0.5 \end{bmatrix}$$

If $V \in (a, b)$ and at state b, then V will transition to value a at rate 0.2 as $b \to a$, but will remain unchanged if $b \to c$. If $V \in (a, b)$ and at state c, then V will transition to a at rate 0.1 but otherwise remaning unchanged if $c \to b$. If $V \in (b, c)$ and at state c, then V will transition to b at rate 0.4 and to a at rate 0.1.

The above value equation determines wages. Upon any meeting or state change at a firm for a worker, the new value for the worker is determined (i.e. when a challenger approaches, the new value for the worker is either the current value, the full challenger surplus, or the full incumbent surplus, when the firm state change, the new value is the minimum of the previous value and marginal surplus of the worker). This value can then be used in the above equation (along with the relevant (ν, z, κ), depending on where the worker works after the negotiations are complete) to find the wage $w(V, \nu, z, \kappa)$ that satisfies the equation. This approach allows for simple wage path simulation: draw worker meetings, separations, and state changes, then determine how worker values are updated, then find the wages that yield the values. The surplus equation is already solved, which allows for easy comparison of worker value and marginal surplus for any challenger or state change.

Externalities It is worth clarifying all the externalities that exist in the quantitative model. The first and most obvious externality arises due to congestion in vacancy-posting, as discussed by Hosios (1990). Firms do not account for the fact that their vacancy-posting decisions contribute to aggregate vacancies, and tighten labor markets, thus resources are wasted by overposting[46]. This first externality is concerned with inefficiently allocating workers between employment and unemployment.

The second externality is that firms do not internalize that their vacancy-posting choices affect the type of firm a worker finds, as in Acemoglu (2001), and therefore low-productivity firms do not internalize that their vacancies cause workers to find high-productivity firms at a lower rate. This second externality is concerned with inefficiently allocating employment across heterogeneous firms.

The third and final externality is the most subtle, and is that firms and workers do not internalize that their decision to consummate a match affects the distributions of workers and firms searching. The key idea raised by Kiyotaki and Lagos (2007) is that, conditional on a meeting, a match will be consummated if positive surplus is generated for the meeting parties, but this will impact the pool of worker and firm matches in the economy, which affects the matching opportunities of all other workers and firms. In the present setting, the concern is that some workers meet firms and, given the pool of workers and firms available, it is optimal to consummate, but in fact a utilitarian planner would prefer that the match not occur, as this will alter the distribution of matches in a way that generates better matches, and to a large enough degree to justify missing out on the private gains. It is important to

[46] There could also be underposting, if workers received some share of the surplus when poached, but in the present case the entirety of the surplus accrues to the firm.

note, however, that in contrast to the above two externalities regarding inefficient vacancy-posting, this third externality concerns inefficient consummating of matches, conditional on meeting, and would be present even with exogenous contact rates, as in Herkenhoff et al. (2024).

The above discussion clarifies that there is potential for policy to improve upon the decentralized allocation. Policies which alter the surplus share between firms and workers would alter the incentives for vacancy-posting, and so could be useful in addressing the first and second externalities, and a program aimed at coordinating matching between workers and firms, such as through subsidizing firm search for workers that will generate the most social surplus, could address the second and third externalities. The focus of this paper is to document and explain the patterns of skill composition across firms, not to answer what policies could reduce the inefficiencies in the labor market. I see further exploration on the policy front as a fruitful for avenue for future work.

F Knowledge hierarchy extension

In the baseline model a firm's workers are imperfectly substitutable by assumption, but it is possible to provide a microfoundation for a firm to value teams of workers via production in a knowledge hierarchy. Beyond providing a more satisfactory theoretical explanation for why firms require teams of workers, as opposed to just collections of parallel jobs, this addition brings predictions for the organizational structure of firms, which I can observe in the microdata. This extension requires a fair amount of machinery not present in the main text, and I am in the process of moving this work into a completely separate project.