Internal Pay Equity and the Quantity-Quality Trade-Off in Hiring*

Michael Amior† and Shmuel San‡

September 10, 2025

Abstract

Firms face significant constraints in their ability to differentiate pay by worker productivity. We show how these internal equity constraints generate a quantity-quality trade-off in hiring: firms which offer higher wages attract higher skilled workers, but cannot profitably employ lower skilled workers. In equilibrium, this mechanism leads to workplace segregation and pay dispersion even among ex-ante identical firms. Unlike in a conventional monopsony model, firms use higher pay to improve hiring quality, even at the cost of lower quantity. Our framework provides a novel interpretation of the (empirically successful) log additive AKM wage model, and shows how log additivity can be reconciled with sorting of high-skilled workers to high-paying firms. It can also rationalize a hump-shaped relationship between firm size and firm pay—and, by implication, the small wage return to firm size. Finally, our model provides new insights into aggregate-level and regional changes in worker-firm sorting and earnings inequality, and public-private sector wage differentials—which we explore empirically using Israeli administrative data.

*We are grateful to Momi Dahan, Oren Danieli, Simon Jäger, Pat Kline and Elio Nimier-David (discussant) for helpful comments, and to seminar participants at Ben Gurion, Tel Aviv, Hebrew University, Yale, NYU, Princeton and the NLS-E 2025 conference. We also thank Matan Goldman for excellent research assistance, and David Gordon, Anat Katz, Rebecca Krieger, Julia Vidar, and many other employees of the Israeli Central Bureau of Statistics, for their invaluable administrative support. This research was generously supported by Israel Science Foundation grants 1277/21 (Amior) and 1472/22 (San), and the Maurice Falk Institute.

†King's Business School; Hebrew University of Jerusalem; Centre for Economic Performance. Email: **michael.amior@kcl.ac.uk**.

‡Hebrew University of Jerusalem. Email: **muly.san@mail.huji.ac.il**.

1 Introduction

Firms face significant constraints in their ability to differentiate pay between workers, stemming from workers' equity concerns. These constraints manifest not only horizontally—between workers performing similar jobs—but also vertically—across different levels of a firm's hierarchy (Akerlof and Yellen, 1990; Romer, 1992; Manning, 1994; Bewley, 1999; Galuscak et al., 2012; Weil, 2014; Saez et al., 2019; Giupponi and Machin, 2022; Brochu et al., 2025). Empirical studies from diverse contexts show that perceived pay inequity or unfairness can harm effort, group morale and retention (Pfeffer and Langton, 1993; Card et al., 2012; Breza et al., 2018; Dube et al., 2019; Cullen, 2024). In this paper, we explore how equity constraints influence firms' pay and recruitment strategies in labor market equilibrium. We show that this generates a quantity-quality trade-off in hiring, and job rationing in high-paying firms—which sheds new light on numerous labor market phenomena.

Our point of departure is the monopsony model of Card et al. (2018). Within this setting, we impose a strict limit on the extent to which firms can differentiate pay by worker productivity (as in Akerlof and Yellen, 1990). This constraint is admittedly rather blunt, but it captures the central mechanism in the simplest possible way: that firms are compelled to compress internal pay differentials more than they otherwise would. When this constraint binds, firms must trade off quantity with quality in hiring: higher wages help attract higher-skilled workers, but make it unprofitable to employ lower-skilled workers. This trade-off sustains two distinct firm strategies in equilibrium: (i) a "selective" strategy, paying high wages to recruit high-skill workers, while rationing low-skill employment, and (ii) an "inclusive" strategy, paying lower wages to maintain a larger but lower-skilled workforce. The prevalence of selective firms is increasing in both the bite of the equity constraint and the abundance (and productivity) of high-skilled labor. This results in substantial workplace segregation and firm wage dispersion, even among ex-ante identical firms—with important implications for both the level and evolution of aggregate earnings inequality.

Our framework implies that wages are log additive in firm and worker fixed effects, in common with the "AKM" wage model (Abowd et al., 1999). Intuitively, a binding equity constraint compels firms to adopt a single proportional pay premium (or "company wage policy", in the language of Manning, 1994, or Giupponi and Machin, 2022), which they apply uniformly to their workforce. Though the AKM model is often chosen for econometric convenience, it happens to fit the data remarkably well in numerous settings (Card et al., 2013; Kline, 2024); and our framework provides a novel conceptual interpretation. As we describe below, our interpretation is also consistent with how managers and recruiters describe setting wages in practice (Bewley, 1999).

Our model can also account for the heavy sorting of high-skilled workers to high-paying firms, as documented by Card et al. (2018); Dauth et al. (2022); Bonhomme et al. (2023); Haanwinckel (2023); Babet et al. (2025). As Bonhomme et al. (2019) and Kline (2025) emphasize, it is difficult to reconcile this positive sorting with log additive wages, if sorting is driven by worker-firm complementarities in production (as in Becker, 1973). But in our model, firms use higher pay specifically to improve hiring *quality*, even at the cost of lower *quantity* (in stark contrast to conventional models): this generates sorting even in the absence of productive complementarities, and even with no ex-ante firm heterogeneity.

While the quantity-quality trade-off generates a positive relationship between firm pay and hiring quality, it also mutes its relationship with workforce size. Once we allow for (skill-neutral) variation in productivity across firms, our model implies a concave or even hump-shaped relationship between workforce size and firm pay. This is because the density of selective firms grows more quickly higher up the firm pay distribution, so the quantity-quality trade-off becomes more acute. This insight can help explain the surprisingly small wage return to firm size (Sokolova and Sorensen, 2021): conventional monopsony models would require implausibly elastic labor supply to individual firms to explain this finding (Bloesch and Larsen, 2023), but small firm size premia are a natural consequence of binding equity constraints in our framework.

The equity constraint can also help address conceptual challenges, highlighted by Card (2022), in the interpretation of firm pay dispersion. The AKM literature suggests that firms can recruit otherwise identical workers at very different wages. In modern labor models, this is often attributed to idiosyncratic job preferences: i.e., employees of low-wage firms are *uninterested* in alternative jobs which pay more. But this makes it difficult to explain evidence on turnover, job ladders and sluggish recovery from job displacement. In contrast, older job search models embed a role for *luck* in wage determination and can therefore explain these phenomena better, but they presume that workers are *unaware* of more lucrative outside job options—an assumption which may be difficult to justify in practice (Caldwell et al., 2025). We argue that an equity constraint can provide an alternative source of luck, by inducing selective firms to *ration* their demand for low-skill labor. In equilibrium, this generates a pool of "unfortunate" low-skill workers who would like to work for selective firms but are denied access—and instead end up in lower-paying inclusive firms.

Finally, these insights offer a synthesis between competing interpretations of growing earnings inequality. Autor et al. (2008) and Dustmann et al. (2009) have reasserted the role of skill-biased technical change, in the context of the US and Germany. However, Card et al. (2013), Song et al. (2019) and Sorkin and Wallskog (2023) show that much of the expansion in inequality—in both countries—was driven by larger sorting of high-skilled workers to high-paying firms, coupled with growing dispersion in firm pay. This process is also reflected in rising workplace segregation, with firms' workforces increasingly sharing similar wage premia (Song et al., 2019), occupational rank (Babet et al., 2025) and education (Dillon et al., 2025). We argue that these trends may themselves be a consequence of technical change: in the presence of a binding equity constraint, growth in high-skilled productivity makes the selective strategy more tempting for firms. And as these firms capture an ever larger share of high-skilled workers (and ration their low-skilled counterparts), this amplifies the impact on earnings inequality.

This hypothesis is closely related to Acemoglu et al. (2001), who argue that technical change was responsible for de-unionization: since unions compress wages within firms, improvements in skilled workers' outside options encouraged them to defect to non-union firms, hastening the demise of unions. But as Bewley (1999) emphasizes, wage compression is not the preserve of unionized firms; and we interpret de-unionization as one manifestation of a broader phenomenon of skilled workers defecting to *selective* firms. Similarly, Weil (2014) and Goldschmidt and Schmieder (2017) emphasize the growing prevalence of outsourcing, and Gola (2024) explicitly connects it with technical change; but again, we argue that outsourcing is one manifestation of a broader phenomenon.

Our model also shows how a growing *supply* of skills (and not just changes in demand) can make the selective pay strategy more tempting for firms—and thereby contribute to larger workplace segregation and earnings inequality. According to our estimates, this supply channel is crucial to explaining these aggregate trends. This prediction is reminiscent of the directed technical change model of Acemoglu (1998), but our story is very different—and centered around workforce segregation, rather than innovation.

It should be emphasized that workplace segregation is an important outcome in its own right: it has long-term implications for economic mobility which are not captured by our model. Though the literature has focused predominantly on residential segregation (by income or ethnicity), workplace segregation may be no less important for mapping social interactions and their consequences (Hellerstein and Neumark, 2008). In particular, firms' adoption of selective hiring strategies may deny low-income individuals access to learning opportunities, human capital externalities (Barza et al., 2024; Dillon et al., 2025) and job networks (San, 2023) associated with high-skilled peers.

Empirical implementation. We test the model's predictions using Israeli administrative data from 1990 to 2019, which provides detailed information on workers' education, wages, and employment histories. The Israeli context is particularly suitable for this analysis, as the contemporaneous tech boom provides valuable empirical variation. The period saw large growth in workforce education, particularly in STEM degrees—coinciding with a rapid increase in the wage returns to these degrees. Our core empirical analysis focuses on cross-sectional variation in wages, employment and skill shares across firms. But our model also makes predictions for market-level variation (both temporal and spatial), as the prevalence of highly productive labor affects the attractiveness of the selective hiring strategy.

The empirical evidence strongly supports our theoretical framework. First, we show that the relationship between firm size and pay follows an inverse-U shape, consistent with the quantity-quality trade-off in our model. This pattern is entirely attributable to low-educated workers: just as our model predicts, high-educated employment increases monotonically with firm pay. We also see the same patterns within two-digit industry categories. These results imply heavy sorting of high-skilled workers to high-paying firms; but despite this, we show that firm wage premia are remarkably similar across education groups—consistent with the log additive AKM wage model and previous empirical work. We are not the first to document non-monotonicities in the firm size-pay relationship: see Bloom et al. (2018) and Kline (2024), who focus on the reverse effect (from firm size to pay). But we reveal the central role of lower-skilled workers in generating this pattern—and offer a new interpretation. The hump-shape relationship appears to be a general phenomenon: we find similar patterns in Northern Italy in the Veneto Worker History file, a popular dataset in the labor literature—which Kline (2024) uses for his analysis.

These qualitative patterns offer compelling support for our interpretation of the data. But we also fit the data quantitatively to our very parsimonious model, using a three-group skill classification (non-graduates, non-STEM graduates, and STEM graduates) and skill-neutral firm heterogeneity. Despite its simplicity, our model is able to match the key results surprisingly well: (i) log additive wages, (ii) skill sorting patterns, and (iii) the hump-shaped size-pay relationship (and small firm size premium). Our estimates imply that the equity constraint compresses the STEM degree return by 69% within firms (relative to the productivity differential), and the non-STEM degree return by 45%. We then compare our model's performance against three alternative frameworks. First, a model with skill-neutral firm heterogeneity but no equity constraint can generate log additive wages, but fails to produce worker-firm sorting or the hump-shaped size-pay relationship. Second, a model with productive complementarities between workers and firms can generate strong sorting patterns, but necessarily violates log additivity by introducing worker-firm interactions in wages. Third, a model with skill-varying labor supply elasticities can produce worker-firm sorting (while preserving log additivity), but cannot generate the non-monotonic relationship between firm size and wages in the data. Only our equity constraint framework can simultaneously accommodate all three empirical regularities. Importantly, the equity constraint also has a strong basis in the theoretical and empirical literature, highlighted above: we have not merely designed it to fit these empirical facts.

We also use our model to assess the welfare implications of eliminating the internal equity constraint (and implicitly, the equity concerns which underpin it). This brings aggregate efficiency gains (via improved amenity matches), as low-skilled workers can now access the full set of firms.[1] But it also exacerbates inequality: STEM graduate welfare grows by 42%, whereas non-graduate welfare declines by 2%. However, an alternative "policy" which prohibits selective hiring strategies (akin to mandating uniform hiring practices across firms) would bring both greater equity *and* efficiency gains. Non-graduate welfare would increase by 15%, while STEM graduate welfare would decrease by 5%; and at the same time, the elimination of workplace segregation would improve the quality of amenity matches. Note this analysis neglects any longer-term externalities associated with persistent workplace segregation.

We then explore market-level variation in firms' pay and recruitment strategies, over time, regions and sectors: this analysis sheds new light on known empirical phenomena. First, at the aggregate level: given the growth in the relative supply and productivity of STEM workers (in the context of the Israeli tech boom), our model predicts greater adoption of selective hiring strategies. This should be reflected in greater pay dispersion across firms and heavier sorting of skilled workers to high-paying firms—and indeed, this is what the data show. This phenomenon may help explain similar trends in other countries, as documented by Card et al. (2013), Song et al. (2019), Bonhomme et al. (2023) and Babet et al. (2025).

Second, we explore spatial variation in these outcomes. According to the model, selective hiring strategies should be more pervasive in higher-skilled regions. And indeed, we show empirically that regions with larger graduate employment shares (and larger skill expansions over time) exhibit greater firm pay dispersion and worker sorting (and greater increases in these outcomes over time). These results speak to influential work by Dauth et al. (2022) and Card et al. (2025), who show that larger cities exhibit heavier sorting of workers to firms. Dauth et al. (2022) attribute this effect to scale economies in matching, but our model offers an alternative interpretation—arising from a quantity-quality trade-off in hiring. As an out-of-sample validation, we use our nationally calibrated model to predict regional variation in the extent of workplace sorting—based solely on regional skill composition. Despite its parsimony, the model performs surprisingly well in both the fit and magnitude of the effects.

Finally, we compare outcomes in the public sector—treating it as a "control" environment where administrative units cannot adopt independent pay strategies. In many countries, the

[1] In an alternative framework with job search frictions, this would manifest in reduced low-skilled unemployment: see Appendix F.

public sector offers lower returns to skill; and this is typically attributed to tighter internal equity constraints (Borjas, 2002; Mazar, 2011). However, we show that skill returns are no larger within *individual* private sector firms than in the public sector: i.e., equity constraints are similarly tight. Instead, what distinguishes the private sector is its fragmentation into many independent firms (or "fissuring", in the language of Weil, 2014). This fragmentation facilitates larger returns to skill at the *aggregate* level, as firms adopt differential pay strategies, and high-skilled workers sort into high-paying firms — consistent with our model.

Related literature. The fact that our model can explain numerous labor market phenomena provides compelling *indirect* evidence for our story. But the equity constraint at its heart is also consistent with how managers and recruiters describe wage-setting in practice. In a seminal study, Bewley (1999) conducted hundreds of interviews with business leaders to better understand the considerations underpinning compensation decisions. Just as in our model, respondents report considerable flexibility in choosing the overall wage level of a firm (as captured by the AKM firm effect in our model), but much less latitude in the choice of internal pay differentials (i.e., the relative worker effects). Galuscak et al. (2012) reaches similar conclusions from surveys of European firms. This is reportedly because employees have significantly more interaction with their co-workers (across the organizational hierarchy) than with their peers outside, and questions of fairness are therefore much more salient for internal than external comparisons. External pay differentials do not generate the same emotional impact, and matter much less for group cohesion and morale. Limited information on outside options may also be a factor, as emphasized by Jäger et al. (2024). At the same time, managers do report that external pay differentials matter for recruitment and retention of *high-quality* workers, just as in our model. We argue that the quality motive is a direct consequence of internal pay compression.

Employees expect vertical pay differentials to be "fair": this typically requires some kind of correspondence with productivity, though Meyer (1975) finds that workers systematically over-value their own contribution to output relative to their co-workers. Notably, employees appear to be more concerned about pay gaps relative to co-workers in *higher* pay grades than those below them: Akerlof and Yellen (1990) and Romer (1992) cite numerous experimental and field research studies on this theme. Some organizations try to suppress internal pay comparisons through social norms or explicit threats, but these practices are often illegal (Cullen, 2024) and their existence illuminates the genuine constraints on wage-setting imposed by equity concerns (Akerlof and Yellen, 1990). Managers often prefer to actively share information on vertical pay differentials (e.g. by publishing formal pay grade schemes), to eliminate suspicion of "unfair" compensation practices and sustain team cohesion and morale

(Bewley, 1999). Though pressures to compress pay may be stronger in unionized firms, they are very much present in non-union firms also (Bewley, 1999). Giupponi and Machin (2022) and Brochu et al. (2025) provide compelling evidence that wage spillovers from minimum wage increases can be attributed to rigidity in internal pay differentials.

A vertical equity constraint also offers a natural interpretation of domestic outsourcing, whose prevalence has grown in recent years. Outsourcing offers a means of escaping the equity constraint, by institutionally separating high and low-skilled employees (Weil, 2014; Goldschmidt and Schmieder, 2017): on average, Drenik et al. (2023) show that high-paying firms only share half their pay premia with outsourced labor. Other studies have focused on observable outsourcing events, to identify causal effects: Goldschmidt and Schmieder (2017) and Daruich et al. (2024) show that outsourced workers suffer large wage losses; and revealingly, Deibler (2022) finds large wage *gains* for workers who remain. Though these results offer support for our hypothesis, outsourcing is merely one manifestation of the quantity-quality trade-off—which we argue is a much broader phenomenon. Firms' rationing of low-skilled employees may also be absorbed through technological substitution in production, whether within defined roles (i.e., employing higher-quality workers to do given tasks) or through the adoption of alternative production processes.

These insights add to a growing body of work documenting constraints on firms' ability to differentiate pay between their employees. Several papers show that firms cannot perfectly discriminate on workers' outside options: see e.g., Caldwell and Harmon (2019), Lachowska et al. (2022) and Di Addario et al. (2023). Jäger et al. (2023) provide evidence of restricted wage differentiation in the context of job surplus shocks. Hazell et al. (2022) have explored constraints on pay discrimination by geography; and Amior and Manning (2020), Amior and Stuhler (2023) and Arellano-Bover and San (2023) study the implications of imperfect pay discrimination between natives and migrants.[2] Our focus here is pay compression among workers of different productivity.

We are not the first to explore the equilibrium implications of internal pay compression between skill groups: Romer (1984) and Akerlof and Yellen (1990) show how equity constraints can generate workplace segregation and unemployment of low-skilled workers. Our key conceptual departure from these studies is to introduce wage-setting power, i.e., an imperfectly elastic supply of labor to the firm. This ensures that inclusive firms can maintain at least some high-skilled employment, despite offering them low pay. This is crucial to the profitability of the inclusive strategy, and hence the existence of a quantity-quality trade-off

[2]Our model is closely related to Amior and Stuhler (2023): that paper shows how constrained pay differentiation between natives and migrants can generate workplace segregation and pay dispersion between ex-ante identical firms. We apply this same idea to skill groups.

in equilibrium. In this respect, our model is closely related to Manning (1994), who imposes a "company wage policy" (with firms constrained to paying a single wage to heterogeneous workers) on an equilibrium search model. Conceptually, our contribution is to partially relax this constraint, to allow for a limited degree of pay differentiation between workers within firms. This allows us to generate a log additive (AKM-type) wage structure, with distinct firm and worker effects. We then show how this framework can deliver a quantity-quality trade-off in hiring, which can help explain several empirical regularities in the literature. Another closely related study—again, from the same period—is Romer (1992), who permits firms to post a non-degenerate wage schedule with respect to skill (ex ante), also in the context of a search model; but since each firm in his model only hires a single worker, there is no quantity-quality trade-off. Our model also shares an internal equity constraint and wage-setting power with Saez et al. (2019), who explore pay differentials by age; but we allow for equilibrium wage competition between firms, which drives many of our results. Frank (1984a,b) and Gola (2024) offer an alternative explanation for internal wage compression and workplace sorting, driven by workers' heterogeneous status concerns; though again, this story does not deliver a quantity-quality trade-off.

Finally, our model builds on an older literature on segmented or dual labor markets: see e.g. Doeringer and Piore (1971); Gordon et al. (1982); Dickens and Lang (1985); Bulow and Summers (1986); Akerlof and Yellen (1990), as well as recent work by Jäger et al. (2024). Though the specifics differ, these theories typically envisage distinct "primary" and "secondary" sectors, where the former offers superior wages and conditions to productively similar workers (in opposition to the human capital paradigm). Crucially though, primary sector jobs are rationed—just as for our "selective" firms. The literature offers numerous hypotheses to explain this segmentation, such as divide-and-rule management strategies, efficiency wages, and poor information on outside options. Building on these ideas, we emphasize the role of internal equity constraints.

In the next section, we present our theoretical framework and derive its key predictions. Section 3 describes our data, and Section 4 offers a quantitative assessment of our model: we document employment and wage patterns across the firm pay distribution, and calibrate the model to match these patterns. We then compare our model's performance against alternative frameworks, and assess key counterfactuals. In Section 5, we explore applications to temporal and spatial variation, as well as public-private sector differences; and we conclude in Section 6.

2 Equilibrium wage-setting model

We begin by developing a simple equilibrium wage-setting model, where firms are constrained in their ability to differentiate pay between workers of heterogeneous quality. As we will show, this pay equity constraint generates a novel trade-off between workforce quantity and quality, which can help shed new light on numerous labor market phenomena.

The economy consists of a continuum of firms (of measure k) and workers (measure n), who are either high or low-skilled. In the baseline model, we assume firms are identical: they produce a homogeneous output good, whose price is normalized to 1, with labor the sole factor of production, and skill types perfectly substitutable. As in Card et al. (2018), firms choose skill-specific wages to maximize profit, and their wage-setting power derives from workers' idiosyncratic preferences over jobs. We deviate from Card et al. by imposing a pay equity constraint: a strict within-firm limit on the wage differential between skill types.

Though the baseline model is purposefully simple, it is straightforward to extend. In Appendix C, we incorporate skill-neutral heterogeneity in firm productivity; Appendix D applies CES technology over skill inputs; and Appendix E extends to N skill types. Finally, Appendix F shows the key results can equally be derived in an environment with job search frictions (instead of idiosyncratic preferences), building on Burdett and Mortensen (1998) and Manning (1994). We will refer to these extensions in the discussion below, as the need arises.

We begin by specifying labor supply. The utility of worker i of skill type $s = \{h, l\}$ in firm f takes the form:

$$u_{isf} = \varepsilon \log w_{sf} + a_{if} \tag{1}$$

where w_{sf} is the wage paid by firm f to type-s workers; and the a_{if} are idiosyncratic workplace amenity values, distributed type-1 extreme value. The supply of skill s labor to a firm offering wage w is then:

$$l_s(w) = \Omega_s w^{\varepsilon} \tag{2}$$

where ε is the elasticity of labor supply to individual firms (which is finite if firms have wage-setting power), and the intercept Ω_s depends on the aggregate skill s workforce, n_s, and competing wage offers:

$$\Omega_s = \left(\int_f w_{sf}^{\varepsilon} df\right)^{-1} n_s \tag{3}$$

We now turn to production. Like Card et al. (2018), we assume that h- and l-type workers are perfect substitutes, but differ in efficiency units: h-types have (fixed) marginal product p_h, and l-types have marginal product p_l, where $p_h > p_l$. Firms choose wages w_s

and employment l_s of each skill type $s = \{h, l\}$ to maximize profit π:

$$\max_{w_h, w_l, l_h, l_l} \pi\left(w_h, w_l, l_h, l_l\right) = \left(p_h - w_h\right) l_h + \left(p_l - w_l\right) l_l \tag{4}$$

subject to labor supply constraints:

$$l_h \leq l_h\left(w_h\right) \tag{5}$$

$$l_l \leq l_l\left(w_l\right) \tag{6}$$

and a pay equity constraint:

$$w_l \geq \phi w_h \tag{7}$$

Equations (5) and (6) ensure that employment is bounded above by the labor supply curves: i.e., firms can only hire willing workers. The equity constraint (7) is our point of departure from standard monopsony models: firms cannot pay l-types less than a fraction ϕ of the h-type wage, where $\phi \leq 1$. Using the terminology of Weil (2014), this constraint may be interpreted in two ways: (i) as a "horizontal" equity constraint (with $\phi = 1$), where h and l-types are workers of different quality performing similar tasks (to different abilities), but firms cannot pay discriminate between them (a case explored by Manning, 1994); or (ii) as a "vertical" equity constraint (with $\phi < 1$), which limits the extent of pay differentiation between workers across the firm's hierarchy. The pay equity constraint can be microfounded using the efficiency wage model of Akerlof and Yellen (1990).[3]

The nature of labor market equilibrium depends on whether the equity constraint (7) binds or not. We will begin with the non-binding case, and then turn to the binding case.

2.1 Equilibrium if equity constraint does not bind

If the pay equity constraint does not bind, the labor supply constraints (5) and (6) must bind:

$$l_h^* = l_h\left(w_h\right) \tag{8}$$

$$l_l^* = l_l\left(w_l\right) \tag{9}$$

[3]Suppose an l-type worker's effort is given by $e_l = \min\left(\frac{w_l}{\tilde{w}_l}, 1\right)$, where $\tilde{w}_l = \phi w_h < p_l$ is a "fair wage" norm, and the corresponding productivity is $e_l p_l$. That is, workers only supply maximum effort ($e_l = 1$) if offered a wage exceeding the norm $\tilde{w}_l$. Under these assumptions, firms will never offer a wage below $\tilde{w}_l$. If they do so, profit per worker will equal $e_l p_l - w_l = \left(\frac{p_l}{\tilde{w}_l} - 1\right) w_l$, which is *increasing* in the wage offer w_l; so an offer below the norm $\tilde{w}_l$ cannot be optimal. Intuitively, the savings on labor costs (from an offer below $\tilde{w}_l$) will not justify the associated productivity losses.

Intuitively, since firms set wages below marginal products, they will hire all workers who are willing to join them. For skill type $s \in \{h, l\}$, the optimal wage is then:

$$w_s^* = \frac{\varepsilon}{1+\varepsilon} p_s \tag{10}$$

which is a fixed mark-down on the marginal product p_s (determined by the labor supply elasticity ε). The wage differential will then equal the productivity differential:

$$\frac{w_l^*}{w_h^*} = \frac{p_l}{p_h} \tag{11}$$

From equation (11), the equity constraint will not bind if $\phi \leq \frac{p_l}{p_h}$.

2.2 Equilibrium if equity constraint binds

Let β denote the bite of the pay equity constraint:

$$\beta \equiv \phi \frac{p_h}{p_l} \tag{12}$$

i.e., the ratio of the pay constraint ϕ to the productivity differential. The constraint binds if $\beta > 1$. Wages will then take log additive form:

$$\log w_{sf} = \eta_f + \lambda_s \tag{13}$$

The common firm effect η_f is chosen by firms, and is equal to $\log w_{hf}$ in the model. The skill effect $\lambda_s = I[s = l] \cdot \log \phi$ represents the fixed internal pay differential, which firms take as given.

In equilibrium, firms will adopt one of two pay strategies:

1. **Inclusive strategy (I).** Inclusive firms hire all willing workers, so the labor supply constraints bind for both skill types: i.e., $l_h^I = l_h\left(w_h^I\right)$ and $l_l^I = l_l\left(w_l^I\right)$. To accommodate both types, firms compress pay internally to satisfy the equity constraint, redistributing wages between h- and l-types (relative to the unconstrained optimum):

$$w_h^I = \frac{1 + \frac{1}{\beta} \cdot \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}} \cdot w_h^* < w_h^* \tag{14}$$

$$w_l^I = \frac{\beta + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}} \cdot w_l^* > w_l^* \tag{15}$$

See Appendix B.1 for derivations.

2. **Selective strategy (S).** Selective firms hire all willing h-type workers, so the h-type labor supply constraint binds, i.e., $l_h^S = l_h\left(w_h^S\right)$; but they ration l-type employment. Since the l-type marginal product is fixed at p_l, rationing l-types only makes sense if the l-type wage w_l (which is fixed at ϕw_h if the equity constraint binds) exceeds p_l. If this is indeed the case, firms will optimally reject all l-type workers: i.e., $l_l^S = 0$.[4] And since selective firms hire only h-types, they will optimally offer them the unconstrained optimal wage: i.e., $w_h^S = w_h^*$. See Appendix B.2.

Though firms in this exposition are identical, they may choose different pay strategies in equilibrium. Let σ denote the equilibrium share of firms which choose the selective strategy. Equilibrium is uniquely determined, and can take one of two forms:

1. **Zero skill segregation.** The inclusive strategy yields strictly larger profit than the selective strategy: $\pi^I > \pi^S$. So all firms adopt the inclusive strategy, i.e., $\sigma = 0$. They pay the same wages, and hire equal shares of h- and l-type workers.

2. **Partial skill segregation.** Both strategies yield equal profit ($\pi^I = \pi^S$), so firms are indifferent between them. Since firms are identical, the adopted strategy of any given firm is undetermined; but the selective share σ is uniquely determined and lies between 0 and 1. Note that equal profits ($\pi^I = \pi^S$) is *not* a knife-edge case: it is an equilibrium outcome, maintained by the value of σ. Selective firms pay high wages and recruit only h-types, and inclusive firms pay lower wages and recruit both h- and l-types; so skill types are partially segregated across firms.

   Note that firms' wage-setting power (i.e., a finite labor supply elasticity ϵ) is crucial to sustaining a partially segregated equilibrium, where inclusive firms offer lower pay. If labor supply were perfectly elastic, inclusive firms would not be able to maintain any h-type employment.[5]

As we show in Appendix B.3, the equilibrium σ can be expressed as:

$$\sigma = \begin{cases} 0 & \text{if } \beta < \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha} \\ \tilde{\sigma}\left(\alpha, \beta, \varepsilon\right) & \text{if } \beta \geq \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha} \end{cases} \tag{16}$$

[4] More generally, if the marginal product p_l is decreasing in l_l (e.g., if skill types are imperfect substitutes or if there are diminishing returns to labor), optimal l-type employment for selective firms may strictly exceed zero, but still lie below the labor supply curve: i.e., $0 \leq l_l^S < l_l\left(\phi w_h^S\right)$. So even here, it remains true that selective firms ration l-type labor. See Appendix D for an exposition with CES technology.

[5] This explains why low-paying inclusive firms do not exist in the model of Akerlof and Yellen (1990): though they impose a similar internal equity constraint, they assume a competitive labor market.

where

$$\alpha \equiv \frac{p_h n_h}{p_h n_h + p_l n_l} \tag{17}$$

is the (exogenous) h-type aggregate output share, and the function $\tilde{\sigma}(\alpha, \beta, \varepsilon)$ solves the implicit equation:

$$\left(1 + \frac{1-\alpha}{\alpha - \tilde{\sigma}}\right)^{1+\varepsilon} = \left(1 + \beta \frac{1-\alpha}{\alpha - \tilde{\sigma}}\right)^{\varepsilon} \tag{18}$$

Equation (16) shows that the equilibrium selective share σ is uniquely determined by three parameters: the h-type output share α, the constraint bite β, and the labor supply elasticity ε. If the constraint bite is sufficiently weak, i.e., if $\beta < \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha}$, the selective share σ is fixed at zero—and invariant to α, β and ε. But if β exceeds this threshold, σ is strictly increasing in all three parameters, in line with equation (18).

To see how equilibrium works, note that increases in α, β and ε (all else equal) would make the selective strategy more profitable, such that $\pi^S > \pi^I$. This would encourage more firms to adopt the selective strategy (so σ increases), which reduces its relative profitability as competition for h-types intensifies—and this process continues until we return to equal profit, with $\pi^S = \pi^I$. Note that $\sigma = 1$ cannot be an equilibrium: all l-type workers would then be unemployed, so any given firm could therefore earn arbitrarily large profit by adopting the inclusive strategy and hiring l-types at arbitrarily low wages.

2.3 Comparative statics: Impact of equity constraint

If the equity constraint binds, firms face a trade-off between quantity and quality in hiring: by offering higher pay, selective firms can hire more h-type workers, but must ration l-types. This trade-off has important implications for pay dispersion, workplace segregation, earnings inequality and efficiency, which we now discuss:

Proposition 1. *An equity constraint with sufficient bite β generates:*

(a) Pay dispersion even among productively identical firms.

(b) Rationing of l-type workers by high-paying firms and hence workplace segregation.

(c) Compression of skill wage differentials, but no change in aggregate earnings.

(d) Reduction in expected amenity match quality, for both skill types. Since aggregate earnings and output are unaffected, this implies aggregate efficiency losses.

If the equity constraint has sufficient bite, and specifically if $\beta > \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha}$, the equilibrium selective share σ will exceed zero. Selective firms will offer a high wage w_h^S, and inclusive firms a low wage w_h^I, to identical workers: this is part (a) of the proposition. This equilibrium is sustained by a quantity-quality trade-off: high-paying (selective) firms recruit more h-types, but this strategy compels them to ration l-type labor; and the equilibrium σ ensures that firms are indifferent between strategies.[6] Therefore, h-types disproportionately concentrate in selective firms (which offer them higher pay), and l-types only work for inclusive firms (as selective firms deny them employment): this is part (b).

Next, consider the implications for wage equity. In an equilibrium with zero skill segregation, a binding equity constraint must compress wage differentials between skill types: firms simply respond by redistributing earnings between skill types, in line with (14) and (15). However, if β increases beyond the $\frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha}{1-\alpha}$ threshold, firms begin to adopt the selective strategy and ration l-type employment. Despite reduced *within*-firm wage differentials, growing segregation *between* firms gradually erodes the pay compression effect. In the limit, as the bite β becomes very large, the labor market converges to perfect skill segregation, with σ equal to the h-type output share α, and earnings differentials matching the unconstrained equilibrium.[7] Up to this limit however, we show in Appendix B.4 that expected wage differentials are always narrower than in a counterfactual with no binding equity constraint. We also show that β has no effect on aggregate earnings, so these equity effects involve redistribution of earnings between workers alone. This is part (c) of the proposition.

We now explore the welfare implications of the equity constraint—and implicitly, of the equity concerns which underpin it. In this model, welfare depends not only on wages, but also on workplace amenities. If β exceeds the $\frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha}{1-\alpha}$ threshold (so the selective share σ exceeds zero), Appendix B.5 shows that the equity constraint reduces the expected value of amenity matches, for both skill types. For l-types, this is because they are denied access to selective firms—and therefore have fewer firms to choose from.[8] The amenity loss may be so large that expected l-type utility decreases (despite the increase in earnings). For h-types, the amenity loss is a consequence of firm pay dispersion: h-types are willing to sacrifice amenity match quality to secure employment at high-paying selective firms. Since aggregate earnings, profit and output are unchanged[9], these amenity losses imply aggregate efficiency

[6]Equilibrium pay dispersion among identical firms is reminiscent of Burdett and Mortensen (1998). In both their model and ours, pay dispersion arises from a trade-off in the wage-setting decision, with different strategies yielding identical profit. For Burdett and Mortensen, this trade-off arises from the standard quantity motive of a non-discriminating monopsonist, in the context of on-the-job search: larger pay reduces profit per worker, but increases firm size. In our model, there is an additional quality motive in the trade-off, which arises from the binding pay constraint: firms use pay to shape their workforce composition, and not just workforce size. This quality motive delivers equilibrium pay dispersion even without on-the-job search.

[7]In this equilibrium, selective and inclusive firms exclusively employ h- and l-types respectively—the former because of rationing, and the latter because their h-type wage offer is arbitrarily low (as the bite β becomes large). The selective share σ is bounded above by the h-type output share α, since this ensures equal profits $\pi^S = \pi^I$ under perfect segregation.

[8]In an alternative job search framework, this loss of access would manifest in higher unemployment for low-skilled workers, rather than lower-quality amenity matches: see Appendix F

[9]Output in this model is fixed by assumption, as workers are equally productive in all firms. And since

losses: this is part (d). This result is in stark contrast to alternative models with firm-worker complementarities in production, where sorting of high-skilled workers to high-paying firms is associated with efficiency *gains*.

2.4 Implications for firm size

We next consider the implications for firm size:

Proposition 2. *An equity constraint with sufficient bite β generates:*

(a) A negative relationship between log firm size and pay, if firms are ex-ante identical.

(b) An initially positive and concave (and possibly hump-shaped) relationship, if there is skill-neutral heterogeneity in firm productivity.

We begin with part (a). In the baseline model with identical firms, if the equity constraint has sufficient bite (such that the selective share $\sigma > 0$), selective firms will offer higher pay, but will employ fewer workers overall. This is a necessary consequence of the quantity-quality trade-off. Since firms are identical, the selective and inclusive strategies must deliver equal profit in partially segregated equilibria. But selective firms employ more skilled workers, who individually generate larger profits; and therefore, to ensure equal profit, selective firms must employ fewer workers overall. See Appendix B.6 for a formal proof.

Of course, in practice, larger firms do typically pay higher wages; but the firm size premium is much smaller than what conventional monopsony models predict. To address this phenomenon, Appendix C extends the model to include skill-neutral heterogeneity in firm productivity. In a given firm f with firm-specific parameter x_f, suppose the h-type and l-type marginal products are $p_{hf} = x_f p_h$ and $p_{lf} = x_f p_l$ respectively, where x_f is distributed log normally across firms. This heterogeneity introduces an *orthogonal* source of variation, which generates a countervailing *positive* correlation between firm size and pay. This positive correlation arises from the standard quantity motive: productive firms benefit more on the margin from larger employment, so they offer higher pay. However, firm productivity x_f makes no difference to the relative value of the selective and inclusive strategies; and hence, the selective share σ is independent of x_f.

Therefore, firm wage premia may now vary for two (orthogonal) reasons: (i) the choice of hiring strategy (selective firms offer higher pay) and (ii) variation in productivity x_f (productive firms offer higher pay). Together, (i) and (ii) generate the concave relationship described by Proposition 1b. Since firms hire all willing h-type workers, the relationship between log h-type employment and log firm pay will simply trace the labor supply curve in

aggregate earnings are unchanged, the same must then be true of profit.

(2): it will be positive and linear, with elasticity ε. However, the same is not true for l-type employment. For sufficiently low pay, the standard quantity motive dominates, and the slope will equal ε: higher-paying firms are more productive and recruit more workers. But higher up the pay distribution, the density of selective firms rapidly expands, and the quality motive plays a more important role: l-types are increasingly rationed, and this may even cause the firm size-pay relationship to turn negative (producing a hump-shaped relationship).

2.5 Implications for aggregate-level earnings inequality

Our model also delivers new insights on the determinants of aggregate-level earnings inequality. Increases in the relative productivity of h-types, i.e., $\frac{p_h}{p_l}$, and in their relative labor supply, $\frac{n_h}{n_l}$, make the selective hiring strategy more attractive; and this yields testable implications for workplace segregation and earnings differentials.

To guide our conceptual discussion and the empirical analysis below, we will rely on a simple decomposition of skill wage differentials, derived from our model. Assuming the equity constraint binds (i.e., $\beta > 1$), Appendix B.7 shows that the skill differential in expected log wages can be expressed as:

$$E\left[\log w_h\right] - E\left[\log w_l\right] = \underbrace{\log \frac{1}{\phi}}_{\text{Within-firm}} + \underbrace{\frac{\sigma}{\alpha} \log \left(\frac{1-\sigma}{\alpha-\sigma}\right)^{\frac{1}{\varepsilon}}}_{\text{Between-firm}} \tag{19}$$

The first component on the right-hand side summarizes the contribution from within-firm pay differentials, i.e. the equity constraint ϕ. The second component summarizes the contribution from workplace segregation, i.e., the extent to which h-types are disproportionately employed by (high-paying) selective firms. Empirically, these components can be identified in two steps:

1. Estimate a log additive (AKM) model for wages, with worker and firm fixed effects.
2. Identify the first component using the mean differential in worker effects (between skill groups), and the second component by the mean differential in firm effects.

We now consider the determinants of these components. The within-firm component is exogenous in our model: it depends on the binding pay constraint ϕ, which we take as given. In practice though, internal pay differentials are likely to be responsive to the other parameters. First, changes in relative productivity $\frac{p_h}{p_l}$ may affect workers' concept of "fair" wages, and the equity constraint may adjust to some extent; in the extreme case, perfect pass-through would imply that the constraint bite β in equation (12) maintains its value. But internal pay differentials may also be increasing in the relative supply of skilled labor $\frac{n_h}{n_l}$:

as we will see next, the skill supply affects firms' incentives to adopt the selective strategy, and l-types may adjust their wage demands internally in response.

We next turn to the between-firm component. We have established above that workplace segregation requires a binding equity constraint. But the extent of segregation also depends on the relative productivity $\frac{p_h}{p_l}$ and labor supply $\frac{n_h}{n_l}$. In fact, holding the constraint bite β fixed, the impact of both can be summarized by a single parameter: the aggregate h-type output share α, as defined by (17). We make the following claim:

Proposition 3. *Assuming the equity constraint binds, and holding its bite β fixed, a larger h-type output share α increases (i) the equilibrium selective share σ and (ii) the between-firm component of the skill wage differential—as long as α is sufficiently large.*

See Appendix B.8 for a proof. If α is sufficiently small, such that $\frac{1-\alpha}{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha} < \beta$, it never makes sense for firms to adopt the selective strategy: there are not enough h-types (and/or they are not sufficiently productive) to justify rationing l-type employment. All firms will then offer the same wages to h- and l-types, defined by equations (14) and (15). Since the selective share σ is zero, there will be no workplace segregation and no between-firm component in the skill wage differential (19). In this scenario, the equity constraint compels firms to share any productive benefits of larger α equally between skill types.

But when the h-type output share α becomes sufficiently large, such that $\frac{1-\alpha}{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha} \geq \beta$, this sharing mechanism snaps: firms begin to adopt the selective strategy, and increasingly so as α grows. Since selective firms refuse to employ l-types, this expansion of σ ensures that only h-types capture the benefits from increases in α. This manifests through larger workplace segregation and a larger between-firm component in the skill wage differential. To see this effect in equation (19), note the selective share σ is bounded above by α.

2.6 Reinterpretation of firm pay dispersion

An important insight of the AKM literature is that firms offer differential pay to otherwise identical workers, an apparent violation of the "law of one price" in the labor market. Pay dispersion may partly reflect compensating non-wage amenities, though the evidence suggests that higher-paying firms typically offer *better* workplace amenities (Lamadon et al., 2022; Sockin, 2022; Caldwell et al., 2025). Much of the literature has instead focused on the role of wage-setting power, sustained either by workers' idiosyncratic preferences (as in Card et al. 2018) or search frictions (e.g. Hornstein et al., 2011). As Card (2022) argues, each approach offers strengths and weaknesses. In idiosyncratic preference models, workers choose the employer they most prefer; but this makes it difficult to explain evidence on turnover, job ladders and sluggish recovery from job displacement. Search models can explain these phenomena better, but they presume that workers are unaware of more lucrative outside job options—an assumption which may difficult to justify in practice (Caldwell et al., 2025).

An equity constraint offers an alternative interpretation of equilibrium wage dispersion, which may help resolve these challenges. In our model, l-types would prefer to work at selective firms; but given rigidity in internal pay structures, selective firms cannot profitably employ them. Of course, if selective firms *never* employ l-types (as in the exposition above), job rationing does not contribute to wage dispersion among l-types. But this extreme result is just an artifact of our (simplifying) assumption of linear technology: it is not true more generally. In Appendix D, we explore the case of CES technology[10]: selective firms do now employ l-types, but quantity is demand-rationed (relative to h-type employment). This generates a role for *luck* in wage determination, traditionally the preserve of search models: fortunate l-types secure employment in selective firms, whereas others are denied access and end up in lower-paying inclusive firms, for reasons unconnected with information. Like Akerlof (1980) and Romer (1984), we assume for simplicity that rationed jobs are allocated randomly.[11] But in an environment with explicit job queuing, a natural implication is longer queues for vacancies in high-paying firms—consistent with evidence from Caldwell et al. (2025). This insight shares intuition with models of directed search, such as Peters (2010); but we emphasize the essential role of internal equity constraints.

3 Data and descriptive statistics

3.1 Data sources

Our analysis draws on Israeli administrative data covering the period 1990-2019. The core dataset, provided by the Central Bureau of Statistics (CBS), contains detailed employment records that link workers to firms. We offer a detailed description of data definitions and processing in Appendix K, but summarize the main points here.

For each worker-firm match, we observe average monthly salary, industry classification, and an indicator for public sector employment. We restrict our main analysis to the private

[10] Saez et al. (2019) also explore a model with an internal equity constraint and CES technology. But our framework incorporates equilibrium wage competition between firms: among other implications, this means that workers denied access by selective firms find employment in inclusive firms.

[11] Note this random allocation rule would produce "inefficient rationing" (conditional on the equity constraint and equilibrium pay strategies), in the language of Lee and Saez (2012), since l-type workers with better amenity matches are no more likely to be selected. Efficient rationing, by contrast, would require firms to observe workers' amenity matches—and to prefer workers with better matches. This is a separate question from the efficiency of the equity constraint itself, which we explore in Proposition 1.

sector, where firms can plausibly adopt differential pay strategies (in line with our model). But in Section 5.3, we compare outcomes in the public sector—treating it as a "control" environment where pay-setting is more unified across administrative units.

We link these records to detailed information on worker demographics and education: we observe both highest degree completed and field of study. For our empirical analysis, we divide workers into three education groups: (i) no college degree, (ii) non-STEM graduate, and (iii) STEM graduate. The STEM/non-STEM graduate distinction has become increasingly salient in recent decades (Altonji et al., 2016; Kirkebøen et al., 2016), and especially in the context of Israel and its tech boom. Though our baseline model distinguishes between just two skill types, we show it is simple to extend to N types in Appendix E.

We also merge the employment records with workplace location data, borrowed from 20% samples of the Israeli census of 1995 and 2008. We group locations into 49 spatial units, based on Israel's "natural regions".[12] We exploit this spatial variation to test Proposition 3, on the market-level determinants of firm pay dispersion and workplace segregation.

3.2 Trends in earnings inequality and returns to education

Our sample period saw significant changes in the Israeli earnings distribution (Cornfeld and Danieli, 2015; Dahan, 2021). In this section, we highlight the large growth in returns to education and explore the contribution of sorting across firms. Most of the action occurred in the early part of the sample, before the mid-2000s, contemporaneous with a large expansion of workforce education—and was specific to STEM graduates.

Figure 1 shows that real earnings grew markedly across the full distribution. For men, the gap between the 90th and 10th percentiles changed little (Panel A). But these patterns mask important trends affecting key minority groups: the arrival of immigrants from the former Soviet Union (FSU) in the early 1990s and their subsequent labor market assimilation (Arellano-Bover and San, 2023), and the integration of Arab and ultra-orthodox Jewish women (Debowy et al., 2021). Once we exclude FSU immigrants, Arabs and ultra-orthodox Jews (who collectively account for 37% of our sample), Panels C and D show that the 90th percentile grew 20% more than both the 50th and 10th—for both genders, by the mid-2000s. Since then, men have seen some moderate compression of earnings differentials.

In Figure 2, we study changes in the standard deviation of residualized log earnings—similar to Card et al. (2013). Excluding minority groups, this grew about 0.07 for both genders by the mid-2000s (Panels C and D). As in the US (Autor et al., 2008), returns to

[12] These have been defined by the Central Bureau of Statistics to ensure a high degree of uniformity in the demographic, economic, and social characteristics of the constituent population. We have incorporated the three smallest regions into neighboring regions, to ensure sufficient sample size for all empirical analysis.

education play a crucial role: after residualizing by non-STEM and STEM degree effects (separately by year), about half this increase is eliminated (green line). And after conditioning on firm fixed effects (orange line), we see no increase at all: this highlights the importance of growing pay dispersion between firms, as in Card et al. (2013).

In Figure 3, we explore returns to education in greater detail. Unconditionally, the return to non-STEM degrees grew moderately from 0.2 to a little over 0.3 (blue line in Panel A), but the STEM degree return surged from 0.2 to 0.8 (Panel B). This reflects the importance of the local tech boom; note the temporary dip in the early 2000s coincided with the dot-com crash and an outbreak of local conflict. Conditioning on age, gender and minority groups (separately by year) makes little difference to these results (red line). But remarkably, firm fixed effects explain away *most* of the expansion of the STEM return (green line). That is, much of the change is driven by growing workplace segregation, with STEM graduates increasingly concentrating in high-paying firms. Interestingly, this result is not driven by *existing* firms changing their pay strategy: as the orange line shows, fixing the firm effects over time makes little difference.[13] This echoes Sorkin and Wallskog (2023), who find that growing dispersion in firm pay effects in the US was driven by the entry of new (more unequal) firm cohorts—rather than changes in pay policies of incumbent firms. See also Lachowska et al. (2023), who document the persistence of pay policies within firms over time.

Excluding minority groups in Panel D, the results look similar—but with two noticeable differences. First, the initial dip in STEM returns in the early 1990s (from Panel B) disappears: this dip was triggered by the arrival of FSU immigrants, many of whom had STEM degrees but were employed at unusually low wages. Second, the increase in STEM returns is now almost entirely concentrated in the 1990s. This is again due to the influence of the (now-excluded) FSU immigrants, whose wage assimilation contributed significantly to growth in aggregate STEM returns after 2000 (Arellano-Bover and San, 2023).

Notably, the rapid growth of STEM returns in the 1990s was accompanied by a large expansion of the STEM workforce. As Figure 4 shows, the STEM employment share grew from 2% to 9% by the early 2000s, and changed little thereafter. This growth was partly driven by FSU immigration, as well as college enrollment among native Israelis: the period saw a vast expansion of degree-granting academic colleges (Meltz, 2001). In contrast, the non-STEM degree share grew (proportionally) much less over the period, from 27% to 33% (Panel A). Motivated by our model, we will argue the growth in STEM employment in the 1990s was partly responsible for the increase in workplace segregation.

[13] For the orange line, we jointly estimate the degree returns for all years using the full worker panel. We interact education (and the age, gender and minority controls) with year effects, but we condition only on *time-invariant* firm effects (i.e., not interacted with year).

3.3 AKM variance decomposition

In line with our model, we identify a firm's pay policy using the firm fixed effect in a log additive AKM wage specification, across workers i and years t:

$$\log w_{it} = \eta_{f(i,t)} + \lambda_i + \delta_t + \gamma X_{it} + e_{it} \tag{20}$$

where $\eta_{f(i,t)}$ are firm effects (for the firm f employing worker i at time t), λ_i are worker effects, δ_t are year effects, and X_{it} includes time-varying controls.[14] The firm effects η_f are identified by worker mobility between firms, and the worker effects λ_i by pay differentiation within them. The firm effects η_f have a "causal" interpretation, under our model's assumptions: they summarize the wage effect of an amenity draw (or a "luck" shock, in the CES extension of Appendix D) which shifts a worker between firms with different productivity or pay-strategies—with the caveat that not all firms are viable counterfactuals for all workers (if l-types are rationed by selective firms).

Our model does not provide an interpretation of the error term e_{it}: any variation in e_{it} would violate the log additive specification implied by our assumptions. In this respect, the wage model's fit can be evaluated by its R-squared. Nevertheless, as long as workers do not sort into jobs according to their e_{it} realizations (e.g., if e_{it} reflects measurement error or transitory shocks), the firm fixed effects will still have a "causal" interpretation.

Table 1 presents summary statistics and an AKM variance decomposition, both for our full sample and separately by education group, for years between 2010 and 2019. To address measurement error in the estimated firm effects, we implement a split-sample correction.[15] Panel A shows that the AKM model fits the data remarkably well, explaining 91.8% of the overall variance in log wages—similar to estimates from other countries.[16] The worker fixed effects account for the largest share of wage variance (65.1%), while firm effects contribute 7.6%, and the covariance between worker and firm effects explains 18.4%. This indicates

[14] Following Card et al. (2018), we control for quadratic and cubic polynomials of age, centered around 40. Given our focus on education, we also interact both the age and year effects with education effects (non-graduate, non-STEM graduate and STEM graduate). These time and age-varying components are not explicit in the model we present above; but of course, they are important when moving to the data. To ease notation, we incorporate them within the worker effects in all the empirical analysis below.

[15] Specifically, we randomly assign workers to two equally sized samples, "A" and "B", and estimate separate AKM models for each sample. We then compute the firm effect variance as $\text{Cov}(\eta_f^A, \eta_f^B)$, where η_f^A and η_f^B are the firm effects estimated in the two samples. For the covariance between worker and firm effects, we use $\text{Cov}(\lambda_i^A, \eta_{f(i,t)}^B)$. And finally, for the worker effect variance, we first compute adjusted worker fixed effects using $\lambda_i^{adj} = \lambda_i^A + \eta_{f(i,t)}^B - \eta_{f(i,t)}^A$, and then use $\text{Cov}(\lambda_i^A, \lambda_i^{adj})$.

[16] Using German data, Card et al. (2013) estimate an R^2 of 90%–93% for the basic AKM model, compared to 92%–95% for the augmented model with match effects. In Portugal, the inclusion of match effects raises the R^2 from 93%–94% to 95% (Card et al., 2016). Finally, applying a correction for limited mobility bias, Kline et al. (2020) estimate an R^2 of 90% for northern Italy.

significant sorting of high-skilled workers to high-paying firms.

We also report results for an augmented specification with worker-firm interactions ("match effects"), which raises the R-squared by only 4pp (95.7% versus 91.8% for the AKM model). As in Card et al. (2013), this small improvement suggests that a log additive specification fits the data well, and match effects offer little additional explanatory power—consistent with our theoretical framework.

The next three columns show that the AKM model exhibits similarly high explanatory power across education groups, with R-squared around 90–92%. Notice also that there is significant firm-worker sorting even within education groups: this is consistent with our model's predictions, to the extent that education is an imperfect indicator of skill.

Panel B shows that average earnings are increasing in education, as in Figure 3. But now, using our AKM estimates, we can decompose these raw education differentials into within-firm and between-firm components—in line with equation (19). Firm effects account for $\frac{0.03-(-0.03)}{9.24-8.96} = 21\%$ of the wage differential between non-STEM graduates and non-graduates[17], and $\frac{0.18-(-0.03)}{9.70-8.96} = 28\%$ of the differential between STEM graduates and non-graduates. This illustrates the importance of worker-firm sorting in driving the return to education. Though substantial, the contribution of firm effects is proportionally smaller than in Figure 3: this is because we are now controlling for worker fixed effects, which partial out heterogeneity in worker quality *within* education groups.

4 Quantitative assessment of the model

In this section, we provide a quantitative assessment of our theoretical model. We first document key empirical patterns in the Israeli labor market: a hump-shaped relationship between firm size and wage premia, heterogeneity in this relationship by education (and heavy worker-firm sorting), and log additive wages. We then calibrate our model to match these patterns and compare its performance against alternative frameworks. We show that an internal equity constraint can simultaneously explain all three empirical regularities, while competing models cannot. Our counterfactual analyses further illustrate the distributional implications of internal pay constraints and quantify their welfare effects across worker types.

[17]Looking at the table, the worker and firm effect differentials do not perfectly sum to the raw wage gap—but they are very close. This reflects the strong fit of the AKM model.

4.1 Relationship between firm employment and pay premia

We begin in Figure 5 by plotting the relationship between log employment and wages across firms. We group firms into 20 bins according to their wage level, with each bin containing an equal number of firms. Panel A identifies firm pay with its average wage (which conflates firm premia with skill composition), and Panel B with its AKM fixed effect (i.e., η_f from equation (20), which conditions on worker effects). Both panels show a striking hump-shaped relationship: among low-paying firms, employment is steeply increasing in pay (consistent with an upward-sloping labor supply curve), but the relationship is reversed higher up the distribution.

As emphasized above, the AKM premia are subject to measurement error, and this may distort the estimates. In Panel A of Figure 6, we now apply a split-sample correction[18], but we still see a clear inverse-U shape. This pattern is indicative of a quantity-quality trade-off in hiring, and can be rationalized by Proposition 2. At the bottom of the pay distribution, firm size increases steeply in wage premia, consistent with the standard quantity motive: higher-paying firms are typically more productive and recruit more workers. However, the relationship is strongly concave—and even *decreasing* among the highest-paying firms. This is due to a rapidly expanding share of selective firms, which are prioritizing recruitment quality over quantity—and rationing lower-skilled employment.

We find similar patterns even within industries. In Panel B, we remove industry fixed effects from both the y-variable (log employment) and x-variable (firm premia); and the basic shape is preserved. This suggests it reflects fundamental trade-offs in firms' wage and hiring strategies, rather than simply sectoral differences in technology or skill requirements.

In Panels C and D, we exclude very small firms—with fewer than 5 employees. We continue to see a clear concave relationship in Panel C, but with no downward-sloping portion. However, the hump shape returns in Panel D when we remove industry effects.

The quantity-quality trade-off becomes more evident when we disaggregate employment by education. Figure 7 plots the relationship between firms' log education-specific employment (non-graduate, non-STEM graduate and STEM graduate) and their AKM premia. For non-graduates, we observe a strong hump-shaped pattern, with employment declining sharply at higher wage premia—and now in all four panels. Employment of non-STEM graduates is strongly concave (but less so than for non-graduates), but with no clear downward-sloping portion. In contrast, STEM employment increases close to linearly in firm wage premia, as

[18]To implement this correction, we randomly assign workers to two equally sized samples ("A" and "B"), and estimate the AKM model separately using each sample. We group firms into the 20 bins according to their sample A premia; and for each bin, we report the mean of the sample B premia on the x-axis. In Appendix Figure A1, we confirm that the sample B premia are monotonically increasing in the sample A premia, across the 20 firm bins.

predicted by our model: firms never ration high-skilled employment, so the green line simply traces out the isoelastic labor supply curve.

At least for the within-industry estimates in Panel D, notice also that the slopes (by education) are similar at the bottom of the firm pay distribution. This is consistent with the model, under the assumption of a common labor supply elasticity ε: see Section 2.4. Intuitively, among low-paying firms, the standard quantity motive dominates, and there is little skill rationing; so the employment slopes will be close to ε for all skill groups.

We are not the first to document non-monotonicities in the relationship between firm size and pay. Bloom et al. (2018) show that the *reverse* relationship (from firm size to pay) has become hump-shaped in the US in recent years, and Kline (2024) finds similar patterns in Northern Italy. These effects are plausibly a consequence of the non-monotonic relationship we document in Figure 6, and we offer a new interpretation of this finding. Note that our model guides us to study how firm size varies across the firm pay distribution (rather than the reverse relationship), because we assume firms can discriminate in hiring (on the y-axis) but not in pay premia (on the x-axis). This permits a meaningful disaggregation of employment (on the y-axis) by education, which speaks clearly to the quantity-quality trade-off.

Hump-shaped employment is not particular to the Israeli context. In Appendix L, we replicate our Figure 6 analysis using the Veneto Worker History (VWH) dataset (as used by Kline, 2024), which contains matched employer-employee records from Italy's Veneto region. As Figure A2 shows, we find a similar inverse-U relationship between firm size and wage premia. This suggests that the quantity-quality trade-off is a general phenomenon, arising from fundamental constraints on firms' wage-setting, rather than from country-specific institutions or policies.

Finally, it is worth emphasizing that concavity in the firm size-pay premium relationship is not unique to our model. As Kline (2025) emphasizes, its shape will depend on distributional assumptions on workers' outside options (or the specification of utility). For simplicity, we have assumed isoelastic labor supply elasticity in Section 2; but if workers' outside options are distributed according to a shifted power function (as in Card et al., 2018), the elasticity of labor supply will be decreasing in firm pay. However, while this can account for *concavity* in the relationship, it cannot rationalize a *hump shape*; and it cannot explain why it is specifically lower-skilled workers who drive this pattern. Our model can explain both these features, as high-paying selective firms ration their employment of low-skilled labor in a quantity-quality trade-off.[19]

[19]See Section 4.5 for further comparison between our model and alternatives from the literature.

4.2 Job mobility patterns

Our baseline model says nothing about job mobility: workers always choose their most preferred workplace, subject to the skill requirements of selective firms. For our empirical analysis, we have therefore focused on the distribution of employment *stocks* across firms. But under alternative assumptions, job rationing effects may also be visible in patterns of worker *flows*. In Appendix F, we explore an alternative environment where wage-setting power is derived from job search frictions—in the manner of Burdett and Mortensen (1998) or Manning (1994)—rather than idiosyncratic job preferences. Workers gradually work their way up a job ladder to ever higher-paying firms, but this process takes time and is subject to luck. In the presence of a binding equity constraint however, the job ladder will be "taller" for high-skilled workers. This is because high-paying firms will disproportionately adopt selective hiring strategies and deny employment to the low-skilled.

We empirically test this claim in Table A1. We begin by dividing firms into four quartiles, according to their AKM premia. For job movers initially employed in any given "origin" quartile, we compute the share who transition to each "destination" quartile—separately by education group. Looking first at STEM graduates (in Panel C), we see clear evidence of a job ladder: job movers from the bottom quartile regularly find work across the full pay distribution, from the bottom to the top; but workers initially in the top quartile rarely accept jobs at the bottom. In contrast, the job ladder appears significantly "shorter" for non-graduates (Panel A): very few job movers from the bottom quartile find employment at the top. This is consistent with the top-paying firms denying them jobs.

4.3 Log additivity of wages

If the equity constraint binds, firms in our model will share wage premia proportionally between skill types: i.e., wages will be log additive. We test this assumption by estimating the AKM model (20) separately by education group, and recovering group-specific firm premia. Figure 8 plots these group-specific premia against the aggregate (i.e., full sample) premia, across 20 bins (ordered by the aggregate premia). The bins are defined separately by education group, and contain equal numbers of group-specific workers; since STEM workers sort into higher-paying firms, the green bins are located more to the right.[20] Group-specific

[20] As before, we correct for measurement error using a split-sample method. We begin by randomly dividing workers into two samples: "A" and "B". For each sample, we estimate AKM firm premia using all workers ("aggregate premia") and separately by education group. For the non-graduate group (in blue), we split firms into 20 bins with equal numbers of non-graduate workers, according to their sample A aggregate premia. Along the x-axis, we report the mean sample B aggregate premia; and along the y-axis, we report the mean sample B non-graduate premia. The red and green dots repeat this exercise for non-STEM and STEM graduates, respectively.

and aggregate premia are normalized to zero for firms with mean (employment-weighted) aggregate premia. If wages are log additive, the firm premia should then be identical across groups: i.e., the group-specific premia should increase one-for-one with the aggregate premia, and should line up perfectly on the 45-degree (dashed) line. Looking at Figure 8, the data are remarkably close to the dashed line, for all three education groups. Panel B shows the same patterns manifest within industries. These results are consistent with Card et al. (2018), who find that relative pay premia (of graduates to non-graduates) are very similar in high and low value-added firms in Portugal.

4.4 Model quantification

The qualitative patterns above offer compelling support for our interpretation of the data. But we also fit the data quantitatively to our very parsimonious model. We study a specification with skill-neutral heterogeneity in firm productivity and three skill types, corresponding to non-graduates, non-STEM graduates and STEM graduates: we denote these l, m and h, respectively. The quantification exercise, detailed in Appendix G, identifies key parameters by matching observable moments in the data.

Table 2 summarizes the target moments and resulting parameter estimates. We identify the labor supply elasticity ($\varepsilon = 5.64$) using the average relationship between log firm size and AKM firm effects. And we calibrate the productivity variance ($\nu = 0.02$) to match the variance of AKM firm effects. To identify the (binding) equity constraint parameters (i.e., the ϕs), we use education differentials in the mean AKM worker effects (from Table 1). The aggregate wage differentials, i.e., $E[\log w_m] - E[\log w_l]$ and $E[\log w_h] - E[\log w_l]$, then pin down the extent of workplace segregation; and the model delivers the skill productivity differentials (and hence the bite of the constraints) which rationalize this segregation.

In this model with three skill types (a special case of the N-type model in Appendix E), firms pursue one of three strategies in equilibrium, which follow a hierarchical structure: (i) a fully inclusive L-strategy, where firms hire all willing workers; (ii) an intermediate M-strategy, where firms hire only m- and h-types; and (iii) a highly selective H-strategy, where firms hire only h-types. As before, these strategies differ in the optimal h-type wage: we denote these as w_h^L, w_h^M and w_h^H respectively. Wages of the other skill types are then fixed by the equity constraints, ϕ_l and ϕ_m. In our calibration, $\sigma^H = 8\%$ of firms pursue the H-strategy (and hire only STEM graduates), $\sigma^M = 17\%$ adopt the M-strategy (and hire both STEM and non-STEM graduates); and the remaining $\sigma^L = 75\%$ adopt the fully inclusive L-strategy and employ all willing workers.[21] Table 2 reports estimated h-type wage

[21]In principle, not all strategies need be active in equilibrium—though this happens to be the case in the national calibration. By contrast, in the regional calibration in Section 5.2 and Appendix J, our estimates

differentials between firms (of given productivity): H-strategy firms pay 0.41 log points more than L-strategy firms, and M-strategy firms pay 0.16 more.

The results point to substantial internal pay compression. Within firms, STEM graduates earn 0.54 log points more than non-graduates, compared to a productivity differential of 1.05. This tight constraint generates substantial job rationing of non-graduates by selective firms, which partially undoes the equalizing effects of the constraint: on aggregate, STEM workers earn 0.74 log points more than non-graduates.

Pay compression is milder for non-STEM graduates: within firms, they earn 0.24 log points more than non-graduates, compared to a productivity differential of 0.41 points. And hence, the sorting effects are correspondingly smaller.

4.5 Comparison with alternative models

To evaluate the performance of our framework (Model 1), we now compare it to three alternatives: an equivalent model with skill-neutral firm heterogeneity but no equity constraint (Model 2); a model with productive complementarities between worker skill and firm quality (Model 3); and one with skill-varying labor supply elasticities (Model 4). We describe how we quantify these alternative specifications in Appendix I. Only Model 1 can match all three empirical patterns documented in Figures 6–8.

Firm size-wage relationship. First, Figure 9 shows that only our model (Model 1) can successfully reproduce a hump-shaped relationship between firm size and wage premia. This pattern emerges through the quantity-quality trade-off, which originates from the equity constraint. The alternative models all predict monotonically increasing relationships, which trace out the (binding) labor supply functions: in the absence of an equity constraint, there is no reason for high-paying firms to ration low-skilled workers.

Employment by education and workplace sorting. It is well known that high-skilled workers sort into high-paying firms, but Figure 7 shows this sorting takes a very particular character—with low-skilled employment exhibiting a clear hump-shape in firm pay, and high-skilled employment increasing monotonically. Figure 10 shows that only Model 1 can replicate these patterns, as low-skilled workers are rationed by high-paying firms. Though Models 3–4 do generate positive sorting, they both predict monotonically increasing employment for all skill types—with no rationing of low-skilled workers at the top.

indicate that some strategies are inactive in particular regions and years.

Log additive wages. Empirically, high-paying firms offer approximately equal premia to all skill types (Figure 8), consistent with the log additive structure of AKM wage models. Model 1 delivers log additivity through the binding equity constraint, and Models 2 and 4 achieve this by assuming skill-neutral heterogeneity between firms: see Figure 11. Model 3 however violates log additivity, as the productive complementarities generate substantial worker-firm match effects.

These comparisons suggest that the equity constraint represents a fundamental feature of wage-setting, which can help explain multiple empirical regularities. While alternative models can match one or two of these results, only our framework can deliver all three. Aside from explaining these regularities, the equity constraint has intuitive appeal: as we argue in the introduction, it has a strong basis in both the theoretical and empirical literature.

4.6 Resolution of empirical puzzles

The results above offer a resolution to two important empirical puzzles in the literature on firm wage-setting: on (i) workplace sorting patterns and (ii) the firm size wage premium.

First, on sorting: our framework can reconcile the heavy sorting of high-quality workers to high-paying firms with the apparent log additivity of wages. The tension between the two has previously been highlighted by Bonhomme et al. (2019) and Kline (2025). The most natural explanation for sorting is productive complementarities (as in Model 3); but as Figure 11 shows, this structure fails to generate log additive wages in a vanilla monopsony framework. Intuitively, if the labor supply constraints bind, skill-differentiated sorting must originate from differential wage returns. Borovičková and Shimer (2024) argue that random match-specific productivity shocks can make wages appear log additive despite the presence of productive complementarities, in an alternative framework with match-level Nash bargaining. Lamadon et al. (2024) propose instead that differential valuations of workplace amenities can resolve this tension: if more productive firms have better amenities, and if high-skilled workers place greater value on these amenities, they will sort differentially into productive firms (even without differential wage returns). Finally, Kline (2025) proposes a very different story, where wages function as a screening device: if high-skilled workers have better outside options, and firms cannot condition wages on skill, higher wage offers may differentially attract high-skilled workers (see also Weiss, 1980). As Kline shows, this can yield comparable sorting patterns to Model 4. However, these alternative stories cannot reproduce the very particular character of sorting we identify in Figure 7, with the hump-shape in low-skilled employment and sharp decrease among the highest-paying firms. To account for this pattern, we require an explanation for why the highest-paying firms may choose to *ration* low-skilled employment: an equity constraint, a concept supported by a growing empirical literature, provides exactly this—through the quantity-quality trade-off.

Second, the model can help explain why wage returns to firm size are significantly smaller than standard monopsony models would predict—typically showing only a 0.05 log wage increase per log point in employment (Sokolova and Sorensen, 2021; Bloesch and Larsen, 2023). Conventional monopsony models require an implausibly elastic labor supply to individual firms to generate such small premia, with ε in the region of 20 (the inverse of 0.05). Our model, however, naturally produces a concave or even hump-shaped relationship between firm size and wages through the quantity-quality trade-off: selective firms offer higher pay, but ration their low-skilled employment; and this implies a much smaller wage return to firm size. A simulation of our very parsimonious model yields a firm size premium of 0.12 (from a regression of AKM firm premia on log employment): this is below the value implied by conventional monopsony models ($1/\varepsilon = 0.18$, for our ε of 5.64), and closer to our empirical estimate in Israeli data (0.052).[22]

It is notable that our estimate of the labor supply elasticity ($\varepsilon = 5.64$) aligns closely with recent estimates identified from within-firm variation. For example, tracing out the response to firm-level procurement auction shocks, Lamadon et al. (2022) and Kroft et al. (2020) find that employment grows 4-6 times as much as wages. This is consistent with our model: a skill-neutral productivity shock should not induce firms to adjust their hiring strategy (between selective or inclusive approaches), but only to adjust on the *quantity* margin (see Section 2.4). Hence, the response to such a shock should be fully captured by the ε parameter. In contrast, when studying the *cross-sectional* distribution of firms (as in Figures 6 and 7), variation in hiring strategy becomes much more salient—and employment may even be *decreasing* in wages among the highest-paying firms, as we show empirically.

4.7 Counterfactual analysis

Having established the empirical appeal of our model, we now quantify the welfare implications of internal equity constraints—and implicitly, of the equity concerns which underpin them. First, we examine the consequences of removing the equity constraint entirely, i.e., the reverse experiment of Proposition 1. Second, we consider a scenario where firms are prohibited from adopting the selective strategy. Table 3 presents changes in expected util-

[22] One can alternatively account for a small firm size premium by introducing a third factor which generates firm-level variation in employment independently of wages: in particular, Bloesch and Larsen (2023) propose a role for recruitment expenditures. However, our model makes a stronger prediction: that the relationship between firm size and pay is concave and potentially non-monotonic, and that this non-linearity is fully attributable to lower-skilled workers—just as we observe empirically.

ity by education in these counterfactuals, decomposed into contributions from expected log wages and amenity match quality. Note we weight utility and amenity effects by $\frac{1}{\varepsilon}$ for this exercise, to ensure they are in comparable log wage units: see equation (1). We leave technical derivations of these effects to Appendix H. It should be emphasized that this analysis neglects any longer-term externalities associated with workplace segregation.

Panel A shows the effects of removing the equity constraint. Consistent with Proposition 1c, we see greater wage inequality between skill groups: firms now set wages independently for each group, and no longer redistribute rents between them. STEM graduates enjoy the largest wage gains (0.19 log points), with a smaller increase for non-STEM graduates (0.04), and significant losses for non-graduates (-0.09). At the same time, all three groups benefit from improved amenity matches, consistent with Proposition 1d. Intuitively, in the counterfactual, high-skilled workers no longer need to sacrifice amenity match quality to secure employment at high-paying selective firms; and low-skilled workers are no longer rationed by selective firms, so they have more firms to choose from. Still, the amenity gain for non-graduates (0.05 log points) is insufficient to offset their wage losses; so their expected utility falls. In summary, eliminating the constraint brings efficiency gains (amenity matches improve, with no change in aggregate output), but exacerbates inequality.

However, an alternative "policy" which prohibits selective hiring (and the rationing of low-skilled labor) can bring both greater efficiency *and* equity. We explore this counterfactual in Panel B. In equilibrium, conditional on their productivity, all firms offer the same wages (in line with the inclusive strategy) and redistribute rents between their high and low-skilled employees. The wage (and welfare) effects are therefore reversed: non-graduates enjoy wage gains of 0.06 log points, non-STEM graduate wages are unchanged, and STEM wages contract by 0.18 log points. However, expected amenities still increase for all groups, just as in Panel A and for identical reasons: high-skilled workers benefit from reduced firm pay dispersion, and low-skilled workers from access to all firms. Since aggregate output is unchanged, we therefore have efficiency gains—alongside the improvement in equity.

The second counterfactual provides a useful theoretical benchmark for interpreting the public sector labor market—which we explore empirically in Section 5.3 below. Given its organizational unity, the public sector's various administrative units cannot easily adopt differential pay strategies; so effectively, they are compelled to pursue the inclusive strategy. As Panel B shows, this generates better outcomes for lower-skilled workers.

5 Applications

5.1 Decadal changes in model parameters

Until now, we have focused on empirical variation across the distribution of firms. But the model also has testable implications for *market-level* outcomes. In markets where high-skilled workers are more numerous and/or more productive (i.e., larger h-type output share α), Proposition 3 predicts that more firms will adopt the selective strategy in equilibrium—and compromise on hiring quantity in favor of quality. We should then expect greater dispersion in firm wage premia, with heavier sorting of high-skilled workers to high-paying firms.

The Israeli tech boom provides a natural setting to test these predictions. Recall that Figure 4 shows a large expansion of STEM graduate employment since 1990. At the same time, Figure 3 reveals large increases in the return to STEM degrees, much of which is driven by sorting of STEM workers to high-paying firms. In contrast, non-STEM graduate shares and wage returns have remained comparatively flat.

To interpret these changes, we replicate our analysis above (which focused on 2010-2019) for the two previous decades: 1990-1999 and 2000-2009. Separately by decade, we estimate the AKM wage equation (20) and calibrate a model with three skill types and heterogeneous firms, following the procedure of Section 4.4. We present the empirical moments and estimated parameters in Table A3. The final column, for the 2010s, is identical to Table 2.

Our model identifies a dramatic shift in the firm strategy mix. The share of firms pursuing the L-strategy (hiring all skill types) declined substantially from 94% in the 1990s to 75% in the 2010s. Conversely, the M-strategy share (selective hiring of m and h-types, i.e., non-STEM and STEM graduates) grew from 4% to 17% over the same period, and the H-strategy share (hiring only h-types) from 2% to 8%. This growing prevalence of selective hiring is reflected in the variance of firm pay premia, which grew from 0.023 to 0.032.

The model attributes these changes to a combination of (i) growing aggregate skill shares and (ii) growing skill productivity differentials—both of which make selective hiring more attractive. We have already seen this for the supply of skills (in Figure 4), whose growth mostly occurred in the early part of the sample. But Table A3 also shows increases in relative productivity, at least for STEM workers: relative to non-graduates, STEM productivity grew by 0.12 ($= 1.05 - 0.93$) log points.

At the same time, our estimates point to a relaxation of the internal equity constraints. Within firms, wage differentials between STEM graduates and non-graduates grew by 0.32 ($= 0.54 - 0.22$) log points—significantly more than the 0.12 increase in productivity differentials. Our model does not explain these changes, since the ps and ϕs are all exogenous parameters. But a natural interpretation is that a rapidly expanding supply of STEM labor generated intense pressure for firms to adopt selective hiring (and ration low-skilled employment), and this encouraged a relaxation of equity norms within firms.

In this way, our model offers a synthesis between competing interpretations of the growth in earnings inequality. While influential studies have reasserted the role of skill-biased technical change (Autor et al., 2008; Dustmann et al., 2009), others have highlighted the contribution of growing dispersion in firm pay—and the sorting of high-skilled workers to high-paying firms (Card et al., 2013; Song et al., 2019; Bonhomme et al., 2023). We argue that this sorting may itself be a consequence of technical change: facing a binding equity constraint, increases in both the productivity *and* abundance of high-skilled labor made the selective strategy more tempting for firms; and this amplified the impact on earnings inequality. Quantitatively, our model shows that skill wage differentials grew significantly more than productivity differentials over time. As we argue above, de-unionization (as in Acemoglu et al., 2001) and the expansion of domestic outsourcing (as in e.g., Goldschmidt and Schmieder, 2017; Gola, 2024) can be interpreted as manifestations of this broader phenomenon.

5.2 Spatial variation in quantity-quality trade-off

We next explore spatial variation in these outcomes. According to the model, selective hiring strategies should be more pervasive in higher-skilled regions. We test this claim empirically—both in the cross-section, and exploiting regional changes over time. We then quantitatively validate our estimates by extrapolating from our nationally calibrated model.

Empirical estimates

We rely on workplace location data from 20% samples of the Israel census of 1995 and 2008: note that much of the growth in STEM employment occurs between these years. We match these records with AKM firm wage premia estimated for two corresponding time intervals: 1993-1997 and 2006-2010. Appendix Table A4 documents regional variation in skill shares and wages in 1995 and 2008: mean graduate share grew from 0.39 to 0.49 between these years, and its standard deviation from 0.045 to 0.066. At the same time, regional dispersion in both the means and variances of firm wage premia grew significantly—as did dispersion in firm-worker sorting (as summarized by local correlations of firm and worker effects).

We estimate two specifications:

$$y_{rt} = \beta_0 + \beta_1 \text{GradShare}_{rt} + d_t + \varepsilon_{rt} \tag{21}$$

$$y_{rt} = \beta_0 + \beta_1 \text{GradShare}_{rt} + d_r + d_t + \varepsilon_{rt} \tag{22}$$

where y_{rt} is some outcome in region r at time t, GradShare_{rt} is the local graduate employment share, and d_r and d_t are region and year fixed effects respectively. The first specification leverages cross-sectional variation to compare regions with different graduate shares. The second exploits local changes in graduate shares within regions over time. We do not claim to be isolating "causal" variation in the local graduate share. Instead, we are using the graduate share as a proxy for the h-type output share α, which is increasing in both the relative employment and productivity of high-skilled workers. Proposition 3 makes predictions on how α affects the quantity-quality trade-off, and we seek to test these empirically.

We present our main results in Panel A of Table 4. First, columns 1-2 show that a larger regional graduate share is associated with significantly higher firm wage premia. This is consistent with more firms adopting a selective high-pay strategy. We find very similar coefficients using both between-region and within-region variation, suggesting that the relationship is very robust: a 10pp increase in local graduate share is associated with a 3-4% increase in average firm premia.

As more firms adopt the selective strategy, we also expect larger dispersion in firm pay premia—and greater sorting of high-skilled workers to high-paying firms. These predictions are validated by the remaining columns. A 10pp increase in local graduate share is associated with a 0.005 increase in the variance of firm wage premia (columns 3-4) and a 0.06-0.07 point increase in the correlation between worker and firm AKM effects. Again, the results are remarkably similar in the between-region and within-region specifications. These specifications also deliver consistently large "within" R-squared (i.e., after partialing out the fixed effects), ranging from 26% to 54%. In Appendix Table A5, we replace the graduate share with distinct regional STEM and non-STEM shares: the effects are mostly driven by the former. This is consistent with the Israeli tech boom playing an important role.

These results build on the influential work of Dauth et al. (2022), who find significantly higher correlation of firm and worker fixed effects in larger German cities. Card et al. (2025) replicate this result for the US, and show additionally that assortative matching is increasing in local graduate share—with a remarkably similar slope coefficient to ours.[23] Dauth et al. (2022) attribute the city size effect to increasing returns in the local matching technology. In this paper, we propose a complementary agglomeration story: larger regions typically have larger skill shares, and this encourages more firms to adopt selective hiring strategies if equity constraints bind. At least in our Israeli data, the empirical evidence appears to favor our interpretation. In column 5 of Panel B (Table 4), we estimate a positive effect of log regional employment on sorting in column 5, consistent with previous work.[24] But only the

[23]Card et al. (2025) estimate a slope of 0.85 in the US, very similar to our between-region estimate of 0.60 in column 5. We show additionally that the effect is robust to including region fixed effects (in column 6).

[24]Dauth et al. (2022) estimate a coefficient of 0.061 in German data, and Card et al. (2025) estimate 0.039

graduate share effect remains statistically significant in the presence of region fixed effects (see column 6); and Panel C shows additionally that graduate share captures the effect in a horse-race between the two variables. Still, it is worth emphasizing that our regions are significantly smaller than those used by Dauth et al. (2022) and Card et al. (2025), and this may influence the results.[25]

The sorting effects are also visible in local skill returns. We show this explicitly in Appendix Table A6, where we estimate effects of graduate share on wage returns to non-STEM degrees (Panel A) and STEM degrees (Panel B). Columns 1-2 show positive effects—significantly larger for STEM returns—both for between-region and within-region variation. We next disaggregate these effects using the decomposition of equation (19). Columns 3-4 show that much of the effect is driven by unobserved worker heterogeneity (associated with the AKM worker effects), especially in the between-region specification—a major theme of Card et al. (2025). But consistent with Proposition 3, differential sorting of graduates to high-paying firms also plays an important role (columns 5-6), and especially for STEM graduates—amounting to about half the overall effect in the fixed effect specification.[26]

Quantitative validation

The estimates above are *qualitatively* consistent with Proposition 3. But quantitatively, can the model rationalize effects of regional skill share on workplace sorting of this magnitude? To address this question, we now use our *nationally* calibrated model to predict the impact of observable spatial variation in skill shares. This exercise effectively serves as an out-of-sample validation of the national model.

As before, we rely on the three-type variant of the model. Using the (observed) local employment shares of each skill type, we solve for the equilibrium strategy shares and wage differentials in each region and census year (1995 and 2008). We fix all the remaining exogenous parameters (relative productivities, equity constraints and the labor supply elasticity) using the national calibration from Section 5.1 (for the 2000-9 interval).

Figure 12 shows the equilibrium strategy shares across regions and census years. In line with Proposition 3, higher-skilled regions have fewer inclusive L-strategy firms (which hire all skill types) and more selective firms—adopting either the M-strategy (hiring both m and h-types, i.e. non-STEM and STEM graduates) or H-strategy (hiring h-types only). This

in the US, which are larger than our 0.020 coefficient in column 5.

[25] Our regions have a mean population of 150,000 in 2008, compared to about 400,000 for German and American commuting zones; though Dauth et al. (2022) show their results are robust to using finer spatial variation across German counties, which have mean population of 250,000.

[26] Analogously, Card et al. (2025) show that college graduates sort disproportionately to high-wage industries in larger cities; and this contributes significantly to regional variation in college wage premia.

pattern becomes more pronounced in 2008, following the expansion of the STEM workforce.[27]

We next explore whether these selective hiring effects can account for the *empirical* variation we observe in Table 4—in mean firm premia, their variance, and workplace sorting.[28] We present our results in Panel D of the same table, separately for each of the three outcomes. Comparing Panels A and D, we observe that the model captures not only the direction of all the skill share slopes—but even performs well on *magnitude*, especially for the mean and variance outcomes. For the correlation of worker and firm effects, the model actually *overpredicts* the effect of graduate share: the predicted slope coefficients in Panel D are about double those in the data (Panel A). Finally, in Figure 13, we plot the same three outcomes ("data") directly on our predictions ("model"), across all region-year pairs. Despite its parsimony, the model fits the data surprisingly well—with R-squared between 0.3 and 0.7, and slope coefficients of about 0.5 to 0.7.

Beyond the parsimony of the model, and the neglect of other sources of regional variation, there are good conceptual reasons to expect the model to overpredict the impact of local skill share. When extrapolating from the national calibration (using variation in local skill share alone), we are implicitly treating regions as independent entities. But in practice, regions are tied through commuting and migration flows; and furthermore, half the workers in our sample are employed by multi-region firms. To the extent that firms struggle to spatially differentiate pay (as in Hazell et al. 2022), this should moderate local effects of skill shares—relative to the model's predictions.

5.3 Public sector wage returns

Until now, we have restricted our empirical analysis to the private sector. But the public sector offers an interesting "control" environment, where administrative units cannot adopt independent pay strategies. In this final section, we explore this comparison empirically.

It is commonly thought that public sector wage-setting is distinguished by tighter constraints on internal pay differentiation. This can explain why it offers relatively low returns to skill: see e.g. Borjas (2002) on the US, and Mazar (2011) on Israel. But our framework offers a very different interpretation. We argue that individual private sector firms are no better at differentiating pay than the public sector: rather, the key distinction lies in the private sector's fragmentation into many independent firms. This fragmentation facilitates

[27] In the national calibration, we observe a single equilibrium with all three strategies active (L, M, and H). In contrast, across individual regions, we see examples of four equilibrium configurations (L-strategy only, L+M, L+H, and L+M+H). See Appendix J for further details.

[28] To predict the local correlation of firm and worker effects in our model, we assign the national-average worker AKM effects to workers in each of the three education groups, and estimate their local correlation with simulated firm AKM premia by region-year pair.

larger returns to skill at the *aggregate* level, as firms adopt differential pay strategies, and high-skilled workers sort into high-paying firms. That is, the public sector is an empirical analogue of the counterfactual with no selective strategy in Table 3.

To test this interpretation, we estimate the AKM model of equation (20) on the full sample, including both private and public sector employment. In our data, "firms" in the public sector identify different administrative units with distinct tax codes. In Table 5, using our fixed effect estimates, we then decompose the wage returns to education into between-firm and within-firm components (in line with equation (19)), separately for each sector.

In Israel, the overall return to non-STEM degrees is slightly larger in the public sector—a consequence mostly of health and education professionals. But the return to STEM degrees is much larger in the private sector: 0.741 versus 0.499. Table 5 shows this difference is almost entirely attributable to the *between-firm* component. Within-firm wage differentials (as identified by the worker effects) are remarkably similar across sectors: this suggests that both sectors face similar internal constraints on pay differentiation. Instead, consistent with our hypothesis, the large private sector returns to STEM arise from the sorting of high-skilled workers into high-paying firms—a mechanism that is absent in the public sector, where administrative units cannot easily compete on pay.

Additional statistics in Table 5 provide further context. The public sector employs workers with higher average worker effects (0.162 versus -0.050 in the private sector), suggesting positive selection into public employment. Interestingly, the public and private sectors offer similar average firm effects, but the variance in the private sector is much larger (0.034 versus 0.017). Again, this reflects the fragmentation of the private sector—and the ability of firms to adopt distinct pay and hiring strategies in equilibrium.

6 Conclusion

It has long been argued that firms face significant constraints in their ability to differentiate pay internally, a claim supported by recent empirical work. In this paper, we show how these internal equity constraints generate a quantity-quality trade-off in hiring. Firms must choose between a selective hiring strategy—paying high wages to attract high-quality workers, while rationing lower-skilled employment—or an inclusive strategy—maintaining lower wages to employ a larger, more diverse workforce. Unlike in conventional monopsony models, firms use higher pay to improve hiring quality, even at the cost of lower quantity.

This insight can help resolve several empirical puzzles in the labor literature. It provides a novel interpretation of the (empirically successful) log additive AKM wage model, and shows how log additivity can be reconciled with sorting of high-skilled workers to high-paying firms. It can also help explain why firm size premia are surprisingly small, without requiring implausibly elastic labor supply. And it provides a new interpretation of systematic variation in workplace segregation between regions and sectors. Finally, it offers a synthesis between competing interpretations of the growth in earnings inequality: we propose that increases in firm pay dispersion and workplace skill segregation (as documented in several countries) is a *consequence* of technical change and the expansion of workforce education. This growing workplace segregation may have detrimental implications for economic mobility over the longer term—through externalities not captured by our model.

Using detailed administrative data from Israel, we find strong empirical support for our model's predictions. We show that the relationship between firm size and wage premia follows a striking inverse-U shape (both on aggregate and within industries), whose concavity can be attributed to high-paying firms rationing low-skilled employment. Our very parsimonious model can successfully reproduce these empirical patterns, in contrast to alternative monopsony models. Furthermore, a national calibration of this model can predict (out-of-sample) spatial variation in workplace segregation surprisingly well—when extrapolating on local skill shares alone. Finally, we use the model to explore key counterfactuals: eliminating the equity constraint (and implicitly, the equity concerns which underpin it) would improve equity, at the expense of efficiency; but an alternative "policy" which prohibits the selective hiring strategy would benefit both equity and efficiency. This latter counterfactual provides a novel interpretation of pay-setting in the public sector.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999). High Wage Workers and High Wage Firms. *Econometrica 67*(2), 251–333.

Acemoglu, D. (1998). Why do New Technologies Complement Skills? Directed Technical Change and Wage Inequality. *Quarterly Journal of Economics 113*(4), 1055–1089.

Acemoglu, D., P. Aghion, and G. L. Violante (2001). Deunionization, Technical Change and Inequality. In *Carnegie-Rochester Conference Series on Public Policy*, Volume 55(1), pp. 229–264. Elsevier.

Akerlof, G. A. (1980). A Theory of Social Custom, of Which Unemployment May be One Consequence. *Quarterly Journal of Economics 94*(4), 749–775.

Akerlof, G. A. and J. L. Yellen (1990). The Fair Wage-Effort Hypothesis and Unemployment. *Quarterly Journal of Economics 105*(2), 255–283.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2016). The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. In *Handbook of the Economics of Education*, Volume 5, pp. 305–396. Elsevier.

Amior, M. and A. Manning (2020). Monopsony and the Wage Effects of Migration. CEP Discussion Paper No. 1690.

Amior, M. and J. Stuhler (2023). Immigration, Monopsony and the Distribution of Firm Pay. CEP Discussion Paper No. 1690.

Arellano-Bover, J. and S. San (2023). The Role of Firms and Job Mobility in the Assimilation of Immigrants: Former Soviet Union Jews in Israel 1990-2019. IZA Discussion Paper No. 16389.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in US Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics 90*(2), 300–323.

Babet, D., O. Godechot, and M. G. Palladino (2025). In the Land of AKM: Explaining the Dynamics of Wage Inequality in France. Banque de France Working Paper No. 987.

Barza, R., E. L. Glaeser, C. A. Hidalgo, and M. Viarengo (2024). Cities as Engines of Opportunities: Evidence from Brazil. NBER Working Paper No. 32426.

Becker, G. S. (1973). A Theory of Marriage: Part I. *Journal of Political Economy 81*(4), 813–846.

Bewley, T. F. (1999). *Why Wages Don't Fall During a Recession*. Harvard University Press.

Bloesch, J. and B. Larsen (2023). When do Firms Profit from Wage Setting Power? Unpublished.

Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018). The Disappearing Large-Firm Wage Premium. *AEA Papers and Proceedings 108*, 317–322.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2023). How Much Should We Trust Estimates of Firm Effects and Worker Sorting? *Journal of Labor Economics 41*(2), 291–322.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A Distributional Framework for Matched Employer Employee Data. *Econometrica 87*(3), 699–739.

Borjas, G. J. (2002). The Wage Structure and the Sorting of Workers into the Public Sector. NBER Working Paper No. 9313.

Borovičková, K. and R. Shimer (2024). Assortative Matching and Wages: The Role of Selection. NBER Working Paper No. 33184.

Breza, E., S. Kaur, and Y. Shamdasani (2018). The Morale Effects of Pay Inequality. *Quarterly Journal of Economics 133*(2), 611–663.

Brochu, P. R., D. A. Green, T. Lemieux, and J. H. Townsend (2025). The Minimum Wage, Turnover, and the Shape of the Wage Distribution. NBER Working Paper No. 33479.

Bulow, J. I. and L. H. Summers (1986). A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment. *Journal of Labor Economics 4*(3), 376–414.

Burdett, K. and D. T. Mortensen (1998). Wage Differentials, Employer Size, and Unemployment. *International Economic Review 39*, 257–273.

Caldwell, S., I. Haegele, and J. Heining (2025). Firm Pay and Worker Search. NBER Working Paper No. 33445.

Caldwell, S. and N. Harmon (2019). Outside Options, Bargaining, and Wages: Evidence from Coworker Networks. Unpublished.

Card, D. (2022). Who Set Your Wage? *American Economic Review 112*(4), 1075–1090.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018). Firms and Labor Market Inequality: Evidence and Some Theory. *Journal of Labor Economics 36*(S1), S13–S70.

Card, D., A. R. Cardoso, and P. Kline (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. *Quarterly Journal of Economics 131*(2), 633–686.

Card, D., J. Heining, and P. Kline (2013). Workplace Heterogeneity and the Rise of West German Wage Inequality. *Quarterly Journal of Economics 128*(3), 967–1015.

Card, D., A. Mas, E. Moretti, and E. Saez (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review 102*(6), 2981–3003.

Card, D., J. Rothstein, and M. Yi (2025). Location, Location, Location. *American Economic Journal: Applied Economics 17*(1), 297–336.

Cornfeld, O. and O. Danieli (2015). The Origins Of Income Inequality In Israel—Trends And Policy. *Israel Economic Review 12*(2), 51–95.

Cullen, Z. (2024). Is Pay Transparency Good? *Journal of Economic Perspectives 38*(1), 153–180.

Dahan, M. (2021). Income Inequality in Israel: A Distinctive Evolution. In A. Ben-Bassat, R. Gronau, and A. Zussman (Eds.), *The Israeli Economy, 1995–2017: Light and Shadow in a Market Economy*, pp. 362–96. Cambridge: Cambridge University Press.

Daruich, D., M. Kuntze, P. Plotkin, and R. Saggio (2024). The Consequences of Domestic Outsourcing on Workers: New Evidence from Italian Administrative Data. Unpublished.

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum (2022). Matching in Cities. *Journal of the European Economic Association 20*(4), 1478–1521.

Debowy, M., G. Epstein, and A. Weiss (2021). Returns to Education and Labor Market Experience in Israel. In A. Weiss (Ed.), *State of the Nation Report: Society, Economy and Policy 2021*. Taub Center.

Deibler, D. (2022). The Effect of Outsourcing on Remaining Workers, Rent Distribution, and Inequality. Unpublished.

Di Addario, S., P. Kline, R. Saggio, and M. Sølvsten (2023). It Ain't Where You're From, It's Where You're At: Hiring Origins, Firm Heterogeneity, and Wages. *Journal of Econometrics 233*(2), 340–374.

Dickens, W. T. and K. Lang (1985). A Test of Dual Labor Market Theory. *American Economic Review 75*(4), 792–805.

Dillon, F. M., E. L. Glaeser, and W. R. Kerr (2025). Workplace Segregation between College and Non-college Workers. NBER Working Paper No. 33653.

Doeringer, P. and M. J. Piore (1971). *Internal Labor Markets and Manpower Adjustment*. New York: DC Heath and Company.

Drenik, A., S. Jäger, P. Plotkin, and B. Schoefer (2023). Paying Outsourced Labor: Direct Evidence from Linked Temp Agency-Worker-Client Data. *Review of Economics and Statistics 105*(1), 206–216.

Dube, A., L. Giuliano, and J. Leonard (2019). Fairness and Frictions: The Impact of Unequal Raises on Quit Behavior. *American Economic Review 109*(2), 620–663.

Dustmann, C., J. Ludsteck, and U. Schönberg (2009). Revisiting the German Wage Structure. *Quarterly Journal of Economics 124*(2), 843–881.

Frank, R. H. (1984a). Are Workers Paid Their Marginal Products? *American Economic Review 74*(4), 549–571.

Frank, R. H. (1984b). Interdependent Preferences and the Competitive Wage Structure. *RAND Journal of Economics 15*(4), 510–520.

Galuscak, K., M. Keeney, D. Nicolitsas, F. Smets, P. Strzelecki, and M. Vodopivec (2012). The Determination of Wages of Newly Hired Employees: Survey Evidence on Internal versus External Factors. *Labour Economics 19*(5), 802–812.

Giupponi, G. and S. Machin (2022). Company Wage Policy in a Low-Wage Labor Market. CEP Discussion Paper No. 1869.

Gola, P. (2024). The Pond Dilemma with Heterogeneous Relative Concerns. Unpublished.

Goldschmidt, D. and J. F. Schmieder (2017). The Rise of Domestic Outsourcing and the Evolution of the German Wage Structure. *Quarterly Journal of Economics 132*(3), 1165–1217.

Gordon, D. M., R. Edwards, and M. Reich (1982). *Segmented Work, Divided Workers: The Historical Transformation of Labor in the United States*. Cambridge University Press Cambridge.

Haanwinckel, D. (2023). Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution. NBER Working Paper No. 31318.

Hazell, J., C. Patterson, H. Sarsons, and B. Taska (2022). National Wage Setting. NBER Working Paper No. 30623.

Hellerstein, J. K. and D. Neumark (2008). Workplace Segregation in the United States: Race, Ethnicity, and Skill. *Review of Economics and Statistics 90*(3), 459–477. Publisher: The MIT Press.

Hornstein, A., P. Krusell, and G. L. Violante (2011). Frictional Wage Dispersion in Search Models: A Quantitative Assessment. *American Economic Review 101*(7), 2873–2898.

Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2024). Worker Beliefs about Outside Options. *Quarterly Journal of Economics 139*(3), 1505–1556.

Jäger, S., B. Schoefer, and J. Zweimüller (2023). Marginal Jobs and Job Surplus: A Test of the Efficiency of Separations. *Review of Economic Studies 90*(3), 1265–1303.

Kirkebøen, L. J., E. Leuven, and M. Mogstad (2016). Field of Study, Earnings, and Self-Selection. *Quarterly Journal of Economics 131*(3), 1057–1111.

Kline, P. (2024). Firm Wage Effects. In *Handbook of Labor Economics*, Volume 5, pp. 115–181. Elsevier.

Kline, P. (2025). Labor Market Monopsony: Fundamentals and Frontiers. NBER Working Paper No. 33467.

Kline, P., R. Saggio, and M. Sølvsten (2020). Leave-out Estimation of Variance Components. *Econometrica 88*(5), 1859–1898.

Kroft, K., Y. Luo, M. Mogstad, and B. Setzler (2020). Imperfect Competition and Rents in Labor and Product Markets: The Case of the Construction Industry. NBER Working Paper No. 27325.

Lachowska, M., A. Mas, R. Saggio, and S. A. Woodbury (2022). Wage Posting or Wage Bargaining? A Test Using Dual Jobholders. *Journal of Labor Economics 40*(S1), S469–S493.

Lachowska, M., A. Mas, R. Saggio, and S. A. Woodbury (2023). Do Firm Effects Drift? Evidence from Washington Administrative Data. *Journal of Econometrics 233*(2), 375–395.

Lamadon, T., J. Lise, C. Meghir, and J.-M. Robin (2024). Labor Market Matching, Wages, and Amenities. NBER Working Paper No. 32687.

Lamadon, T., M. Mogstad, and B. Setzler (2022). Imperfect Competition, Compensating Differentials, and Rent Sharing in the US Labor Market. *American Economic Review 112*(1), 169–212.

Lee, D. and E. Saez (2012). Optimal Minimum Wage Policy in Competitive Labor Markets. *Journal of Public Economics 96*(9-10), 739–749.

Manning, A. (1994). Labour Markets with Company Wage Policies. CEP Discussion Paper No. 214.

Mazar, Y. (2011). Self-Selection of Employees Moving Between the Public and Private Sectors. *Israel Economic Review 9*(1), 151–180.

Meltz, N. M. (2001). Academic Colleges: Transforming Higher Education in Israel. *Higher Education Policy 14*(4), 343–359.

Meyer, H. H. (1975). The Pay-for-Performance Dilemma. *Organizational Dynamics 3*(3), 39–50.

Peters, M. (2010). Noncontractible Heterogeneity in Directed Search. *Econometrica 78*(4), 1173–1200.

Pfeffer, J. and N. Langton (1993). The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively: Evidence from College and University Faculty. *Administrative Science Quarterly 38*(3), 382–407.

Romer, D. (1984). The Theory of Social Custom: A Modification and Some Extensions. *Quarterly Journal of Economics 99*(4), 717–727.

Romer, D. (1992). Why Do Firms Prefer More Able Workers? Unpublished.

Saez, E., B. Schoefer, and D. Seim (2019). Payroll Taxes, Firm Behavior, and Rent Sharing: Evidence from a Young Workers' Tax Cut in Sweden. *American Economic Review 109*(5), 1717–1763.

San, S. (2023). Who Works Where and Why: The Role of Social Connections in the Labor Market. Unpublished.

Sockin, J. (2022). Show Me the Amenity: Are Higher-Paying Firms Better All Around? CESifo Working Paper 9842.

Sokolova, A. and T. Sorensen (2021). Monopsony in Labor Markets: A Meta-Analysis. *ILR Review 74*(1), 27–55.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2019). Firming Up Inequality. *Quarterly Journal of Economics 134*(1), 1–50.

Sorkin, I. and M. Wallskog (2023). The Slow Diffusion of Earnings Inequality. *Journal of Labor Economics 41*(S1), S95–S127.

Weil, D. (2014). *The Fissured Workplace.* Harvard University Press.

Weiss, A. (1980). Job Queues and Layoffs in Labor Markets with Flexible Wages. *Journal of Political Economy 88*(3), 526–538.

Tables and figures

Table 1: Descriptive statistics and AKM decomposition by worker type

	All	Non-grads	Non-STEM grads	STEM grads
Panel A: AKM variance decomposition				
Var. log salary	0.42	0.32	0.43	0.48
AKM model (share):				
Var. worker effect	65.1	66.2	68.4	62.3
Var. firm effect	7.6	7.9	7.0	8.6
2 × Cov(worker, firm)	18.4	15.0	15.0	20.2
R-squared	91.8	90.0	91.1	92.0
Comparison match model (share:)				
R-squared	95.7	94.7	95.3	95.8
Panel B: Sample means and size				
Worker-years				
N.	16,388,759	9,617,127	5,273,282	1,498,350
Share N.	1.00	0.59	0.32	0.09
Av. log salary	9.12	8.96	9.24	9.70
Av. worker effect	-0.11	-0.24	-0.01	0.30
Av. firm effect	0.01	-0.03	0.03	0.18
Workers				
N.	2,964,754	1,744,388	980,462	239,904
Share N.	1.00	0.59	0.33	0.08
Firms				
N.	238,193			
Av. firm size	19.8			

Notes: Panel A presents variance decomposition results from an AKM model (one model for all worker types). We correct for measurement error using a split-sample procedure (see text). The final row reports the R-squared of an augmented model with interacted firm-worker fixed effects. Panel B presents the number of observations and averages of relevant variables for worker-years, workers, and firms. Sample consists of private sector firms between 2010 and 2019.

Table 2: Quantification of model parameters

Moments			Parameters		
Moment	Value	Explanation	Parameter	Value	Explanation
ε_{data}	2.66	Elasticity of firm size w.r.t firm effect	ε	5.64	Elasticity of labor supply w.r.t pay
V_{AKMf}	0.032	Variance of firm effect	ν	0.02	Variance of firm productivity
$\log \frac{\phi_m}{\phi_l}$	0.24	Relative m-type person effect	σ^M	0.17	Share firms with M-strategy
$\log \frac{1}{\phi_l}$	0.54	Relative h-type person effect	σ^H	0.08	Share firms with H-strategy
$E[\log w_m] - E[\log w_l]$	0.28	Relative expected m-type log wage	$\log \frac{w_h^M}{w_h^L}$	0.16	Relative M-strategy log wage
$E[\log w_h] - E[\log w_l]$	0.74	Relative expected h-type log wage	$\log \frac{w_h^H}{w_h^L}$	0.41	Relative H-strategy log wage
$\log \frac{n_m}{n_l}$	-0.60	Relative m-type employment	$\log \frac{\phi_m^\varepsilon \Omega_m}{\phi_l^\varepsilon \Omega_l}$	-1.05	Relative m-type labor-supply intercept
$\log \frac{n_h}{n_l}$	-1.86	Relative h-type employment	$\log \frac{\Omega_h}{\phi_l^\varepsilon \Omega_l}$	-2.85	Relative h-type labor-supply intercept
			Implied parameters		
			$\log \frac{p_m}{p_l}$	0.41	Relative m-type log productivity
			$\log \frac{p_h}{p_l}$	1.05	Relative h-type log productivity

Notes: This table shows the empirical moments used for model calibration (left columns) and the resulting parameter estimates (right columns). Sample consists of private sector firms in 2010-2019. See Appendix G for more details.

Table 3: Welfare effects of counterfactuals

Type	Exp log wage	Exp amenity	Exp utility
Panel A: No pay-equity constraint			
Non-graduates	-0.094	0.054	-0.040
Non-STEM graduates	0.038	0.032	0.071
STEM graduates	0.194	0.094	0.289
Panel B: No selective strategy			
Non-graduates	0.060	0.054	0.114
Non-STEM graduates	-0.001	0.032	0.031
STEM graduates	-0.175	0.094	-0.081

Notes: This table presents welfare changes from two counterfactual scenarios. Panel A shows what happens if we eliminate the pay equity constraint, allowing firms to set wages of skill types independently. Panel B shows what happens if we prohibit the selective pay strategy, requiring all firms to employ workers of all skill types. Worker types are defined by education: STEM graduates (type-h in the model), non-STEM graduates (type-m), and non-graduates (type-l). Changes in expected utility are decomposed into changes in expected log wages and expected amenity matches. Note we weight utility and amenity effects by $\frac{1}{\varepsilon}$ for this exercise, to ensure they are in log wage units: see equation (1).

Table 4: Regional effects on firm pay dispersion and sorting: data and model

	Mean: Firm AKM		Var: Firm AKM		Corr: Worker, Firm	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Base specifications						
Graduate share	0.304	0.365	0.048	0.052	0.595	0.684
	(0.053)	(0.067)	(0.005)	(0.019)	(0.095)	(0.229)
Within-R^2	0.391	0.536	0.515	0.256	0.337	0.271
Panel B: Effect of regional employment						
Log employment	0.011	0.008	0.002	0.001	0.020	0.020
	(0.003)	(0.015)	(0.000)	(0.003)	(0.005)	(0.036)
Within-R^2	0.194	0.011	0.212	0.001	0.163	0.009
Panel C: Controlling for both						
Graduate share	0.303	0.368	0.052	0.052	0.600	0.692
	(0.062)	(0.061)	(0.010)	(0.020)	(0.121)	(0.222)
Log employment	0.000	0.011	-0.000	0.001	-0.000	0.025
	(0.003)	(0.008)	(0.000)	(0.002)	(0.007)	(0.028)
Within-R^2	0.391	0.555	0.519	0.259	0.337	0.285
Panel D: Model-predicted outcomes						
Graduate share	0.326	0.371	0.047	0.071	1.326	1.667
	(0.027)	(0.030)	(0.012)	(0.014)	(0.250)	(0.228)
Within-R^2	0.855	0.899	0.366	0.593	0.508	0.707
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Sample	98	98	98	98	98	98

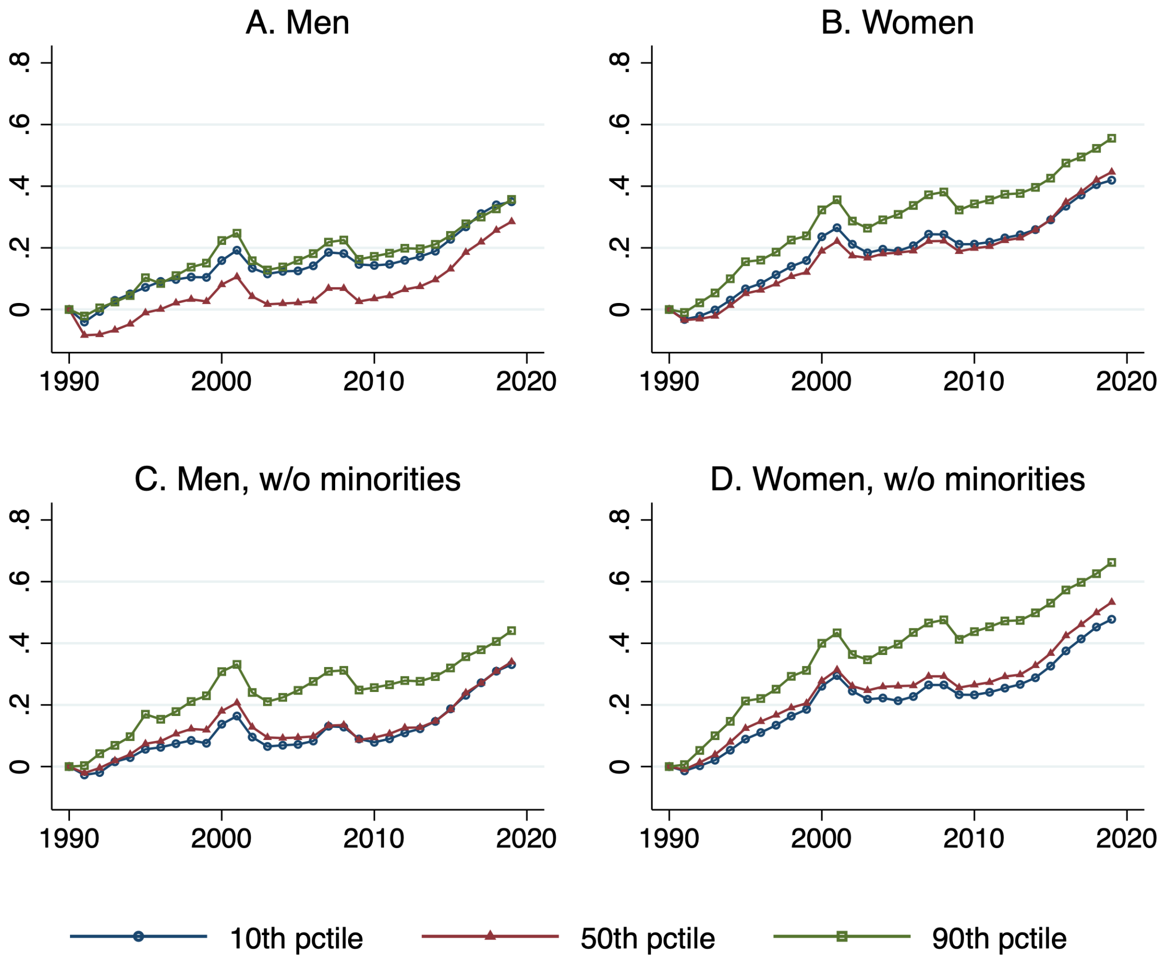
Notes: Panel A shows the relationship between regional graduate share and key local outcomes. Panel B reproduces these estimates, but with log regional employment instead of graduate share on the right-hand side. Panel C controls for both variables simultaneously. Panel D estimates effects on model-predicted outcomes, extrapolated from the national calibration. Odd-numbered columns exploit cross-sectional variation across regions, using equation (21). Even-numbered columns control for region fixed effects, as in equation (22), relying on within-region changes for identification. The dependent variables are the mean firm AKM premia (columns 1-2), the variance of firm AKM premia (columns 3-4), and the correlation between firm and worker AKM premia (columns 5-6). Sample consists of 49 regions observed in both 1995 and 2008 census years, for a total of 98 region-year observations. Observations are weighted by regional employment shares. Standard errors, clustered by region, in parentheses.

Table 5: Differences between sectors in return to education

	Private sector	Public sector
Panel A: Return to education		
Non-STEM graduates v non-graduates		
Log salary	0.282	0.307
Worker effect	0.221	0.316
Firm effect	0.060	-0.010
STEM graduates v non-graduates		
Log salary	0.741	0.499
Worker effect	0.520	0.486
Firm effect	0.221	0.018
PAnel B: Additional statistics		
Av. log wage	9.119	9.280
Av. worker effect	-0.050	0.162
Av. firm effect	0.012	-0.024
Var. firm effect	0.034	0.017

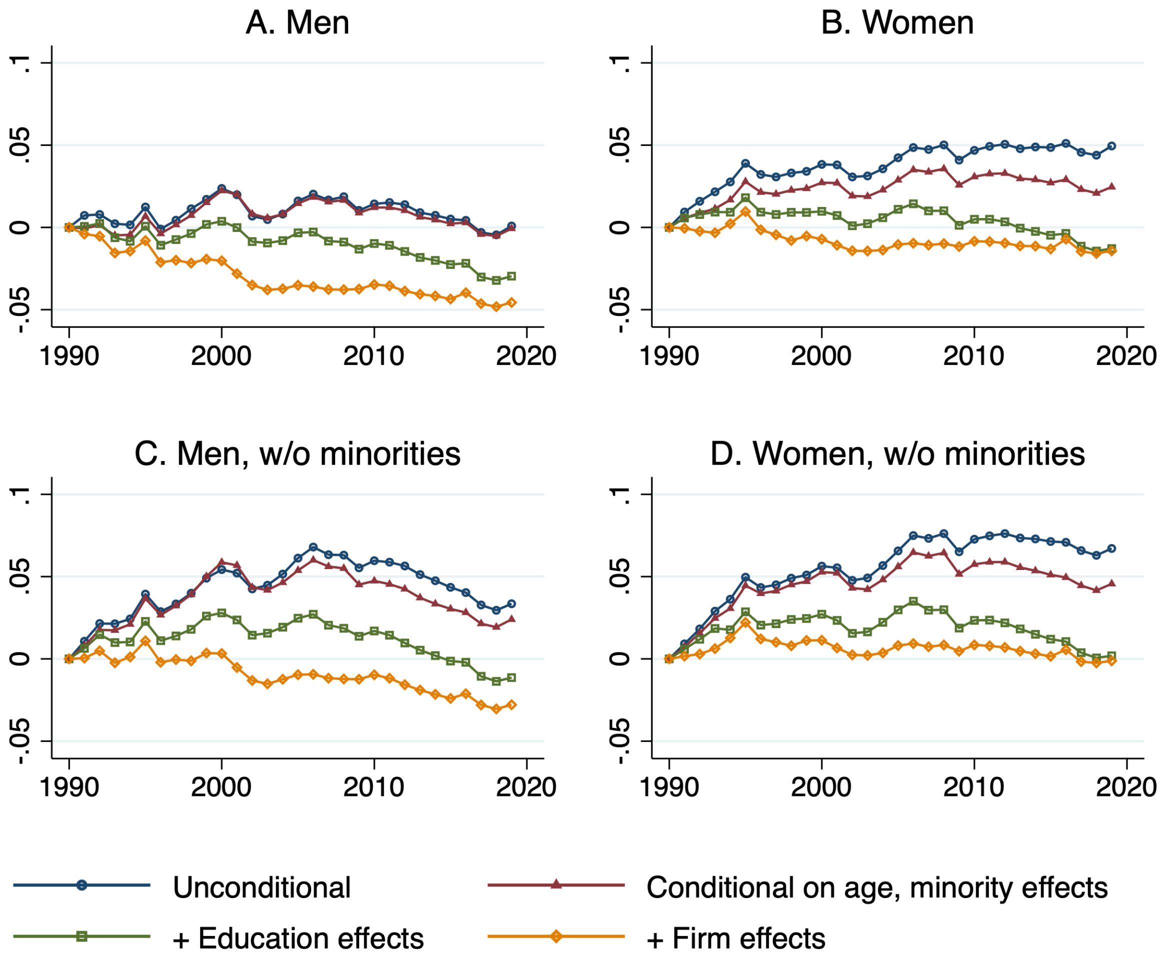
Notes: This table reports various outcomes separately for the private and public sectors, for the period 2010-2019. The first rows show aggregate employment shares of each education group, by sector. We then report mean wage differentials between STEM/non-STEM graduates and non-graduates, which we disaggregate into AKM worker and firm effects, in line with equation (19). For this exercise, we rely on a common AKM model estimated using data from both sectors. Finally, we report the variance of AKM firm effects by sector, corrected for measurement error using a split-sample procedure (as described in the text).

Figure 1: Percentiles of real earnings over time



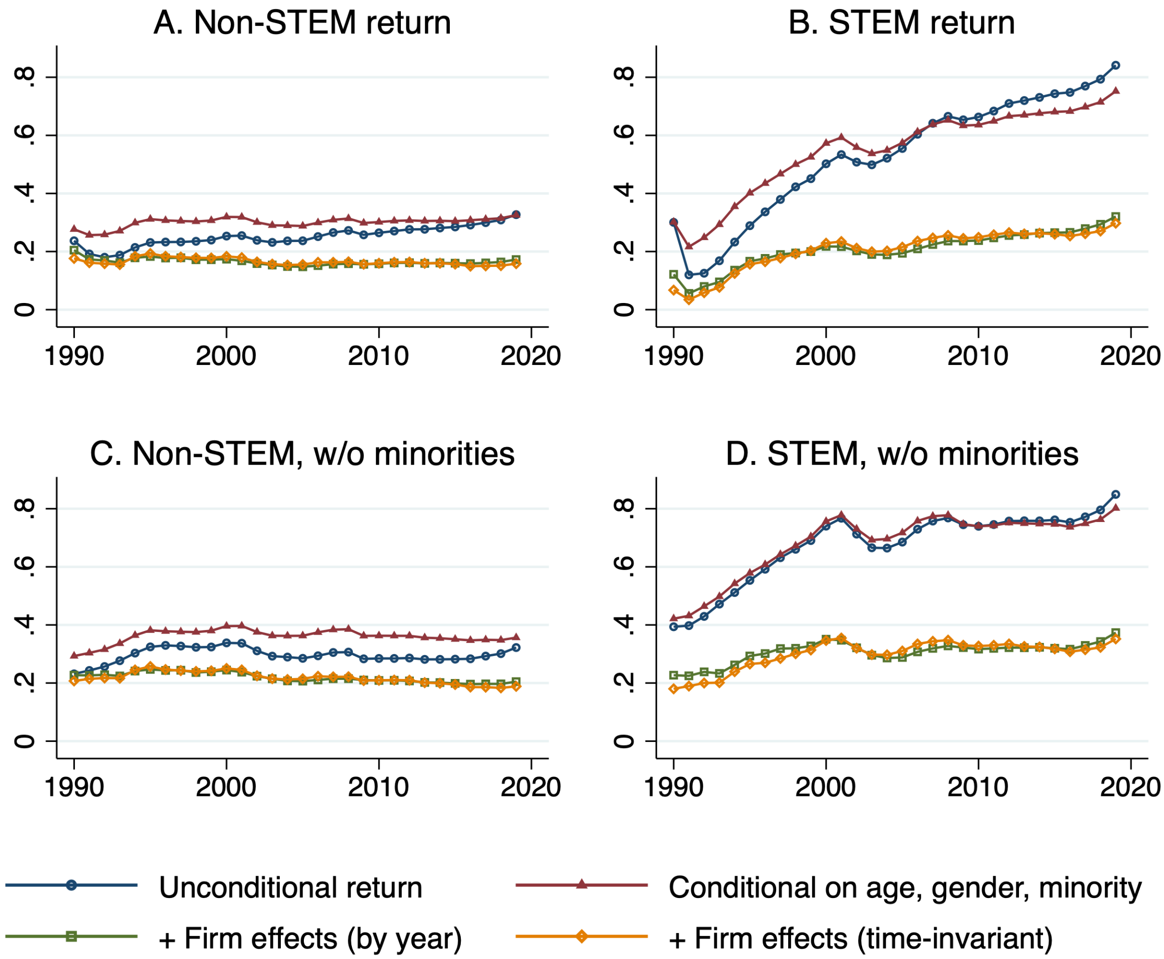
Notes: This figure shows the evolution of percentiles of real earnings from 1990 to 2019, separately for different sub-samples of the workforce. Panels A and B show trends for men and women respectively. Panels C and D exclude Arabs, ultra-orthodox Jews and FSU immigrants from the sample.

Figure 2: Standard deviations of residualized log earnings



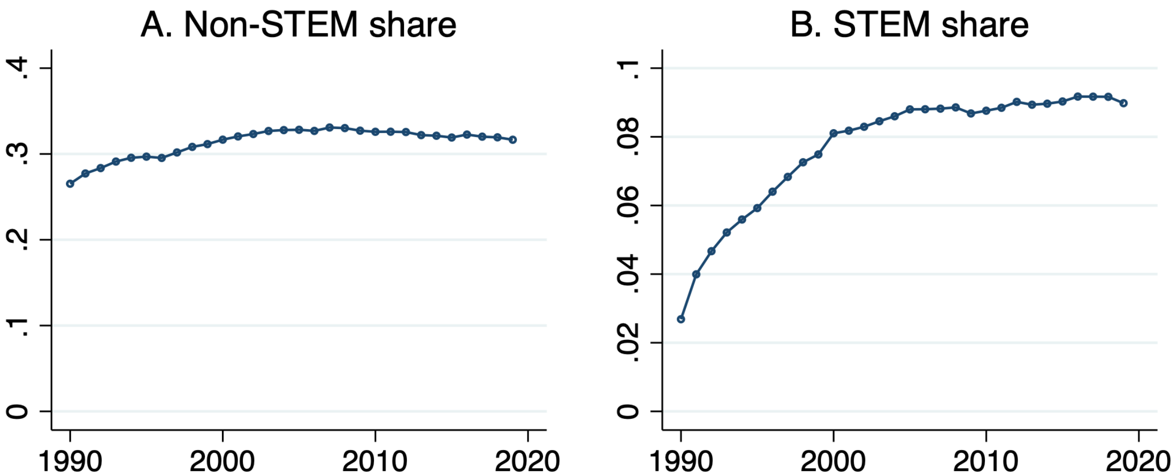
Notes: This figure shows the evolution of standard deviations of residualized log earnings from 1990 to 2019. Panel A and B show trends for men and women respectively, and Panels C and D exclude Arabs, ultra-orthodox Jews, and FSU immigrants from the sample. Each line shows standard deviations after residualizing log earnings by progressively more controls. The red line controls for a cubic in age, interacted with minority effects. The green line includes education effects (STEM and non-STEM degree), interacted with all previous variables. The orange line controls additionally for firm fixed effects.

Figure 3: Evolution of education wage premia



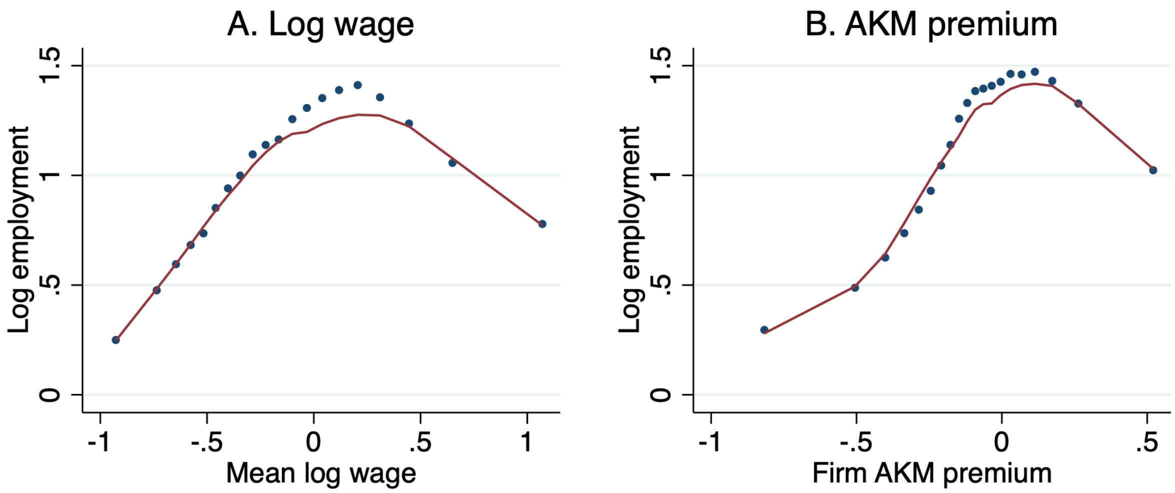
Notes: This figure shows the evolution of education wage premia from 1990 to 2019. Panel A shows earnings differentials between non-STEM graduates and non-graduates, and Panel B shows differentials between STEM graduates and non-graduates. Panels C and D show corresponding trends for a restricted sample which excludes Arabs, ultra-orthodox Jews, and FSU immigrants. Each panel plots four different specifications: the unconditional return (blue line), controlling for interactions between gender, minority effects (for Arabs, ultra-orthodox Jews and FSU immigrants) and an age cubic separately by year (red line), controlling additionally for firm effects separately by year (green line), and replacing these with time-invariant firm effects (orange line).

Figure 4: Evolution of graduate employment shares



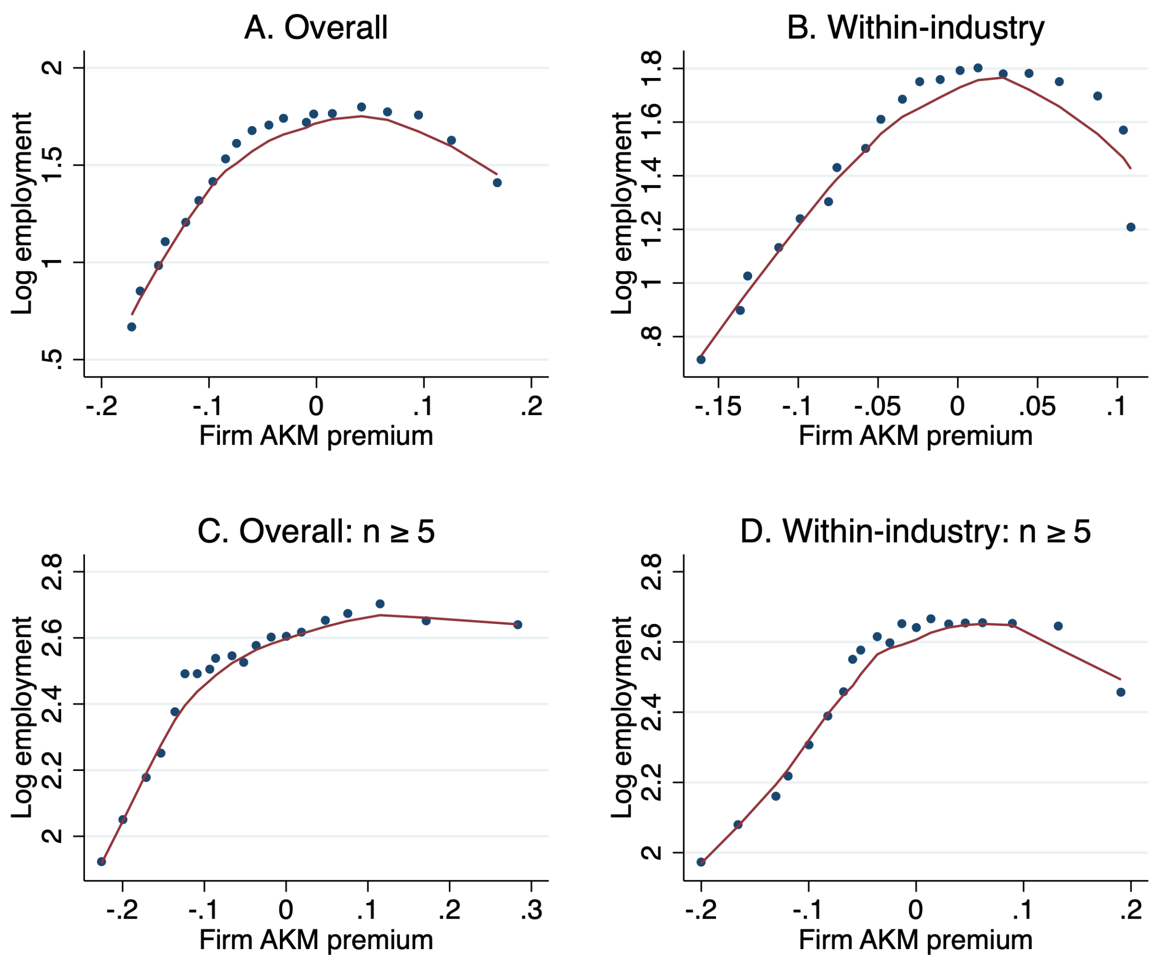
Notes: Panel A shows the aggregate employment share of non-STEM graduates, from 1990 to 2019. Panel B shows the aggregate employment share of STEM graduates.

Figure 5: Log employment by firm wage level and AKM premium



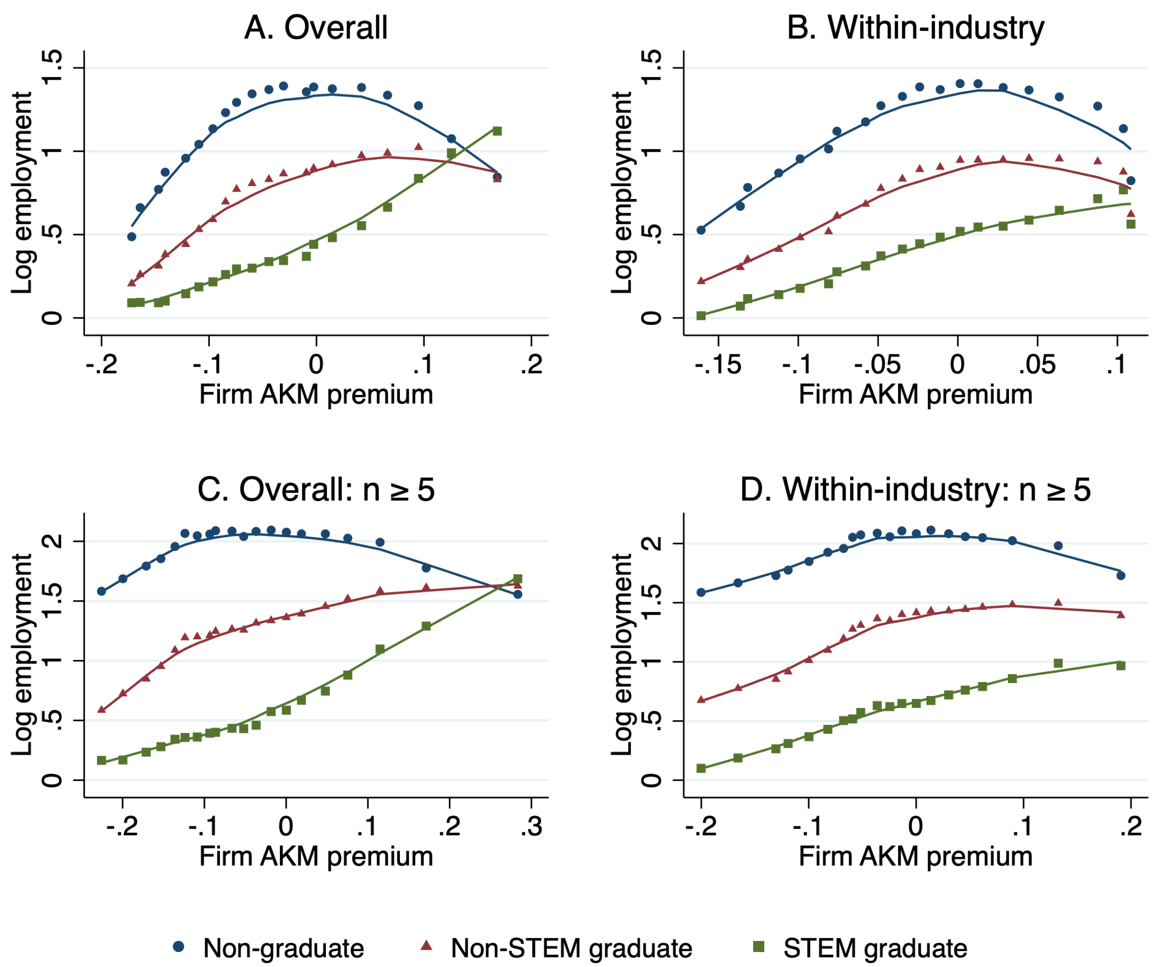
Notes: Panel A shows mean log firm employment across 20 bins (with equal numbers of firms), arranged by mean firm log wages. Panel B arranges firms instead by raw AKM firm premia, not adjusted for measurement error. Mean log wages and firm premia are normalized to their worker-weighted means. Sample consists of private sector firms in 2010-2019.

Figure 6: Employment by firm pay premium



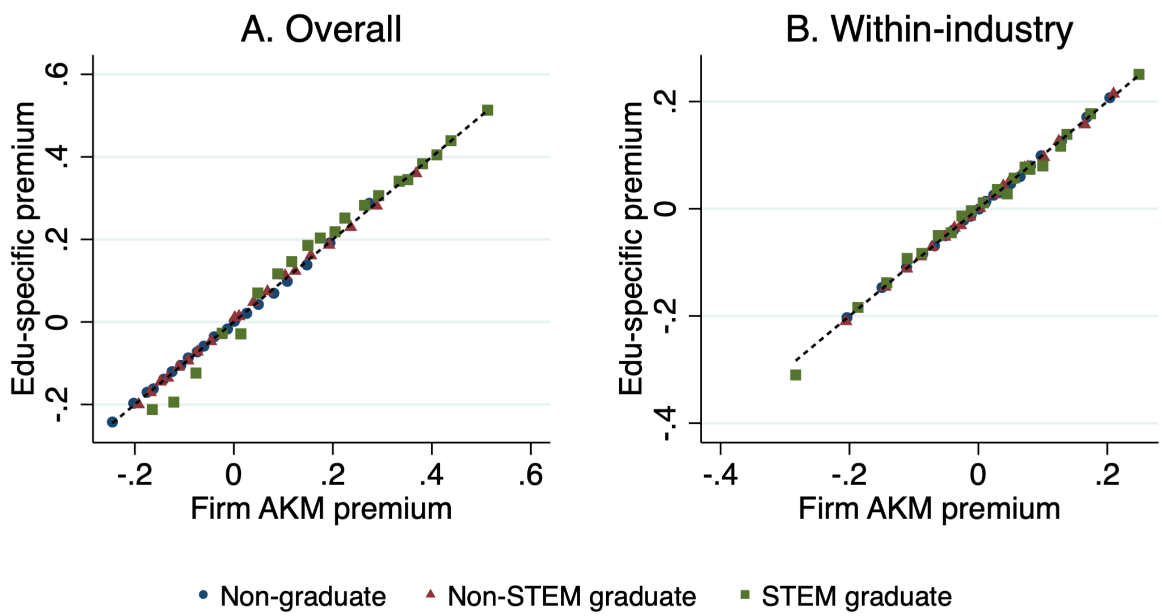
Notes: Panel A shows mean log firm employment across 20 bins (with equal numbers of firms), arranged by AKM firm premia. Firm premia are normalized to their worker-weighted mean. We implement a split-sample procedure to correct for measurement error in the firm premia, as described in Section 4.1. In Panel B, we remove industry fixed effects from both the y-variable (log firm employment) and the x-variable (firm premia). Panels C and D repeat this exercise after excluding firms with fewer than 5 employees. Sample consists of private sector firms in 2010-2019.

Figure 7: Employment by education and firm pay premium



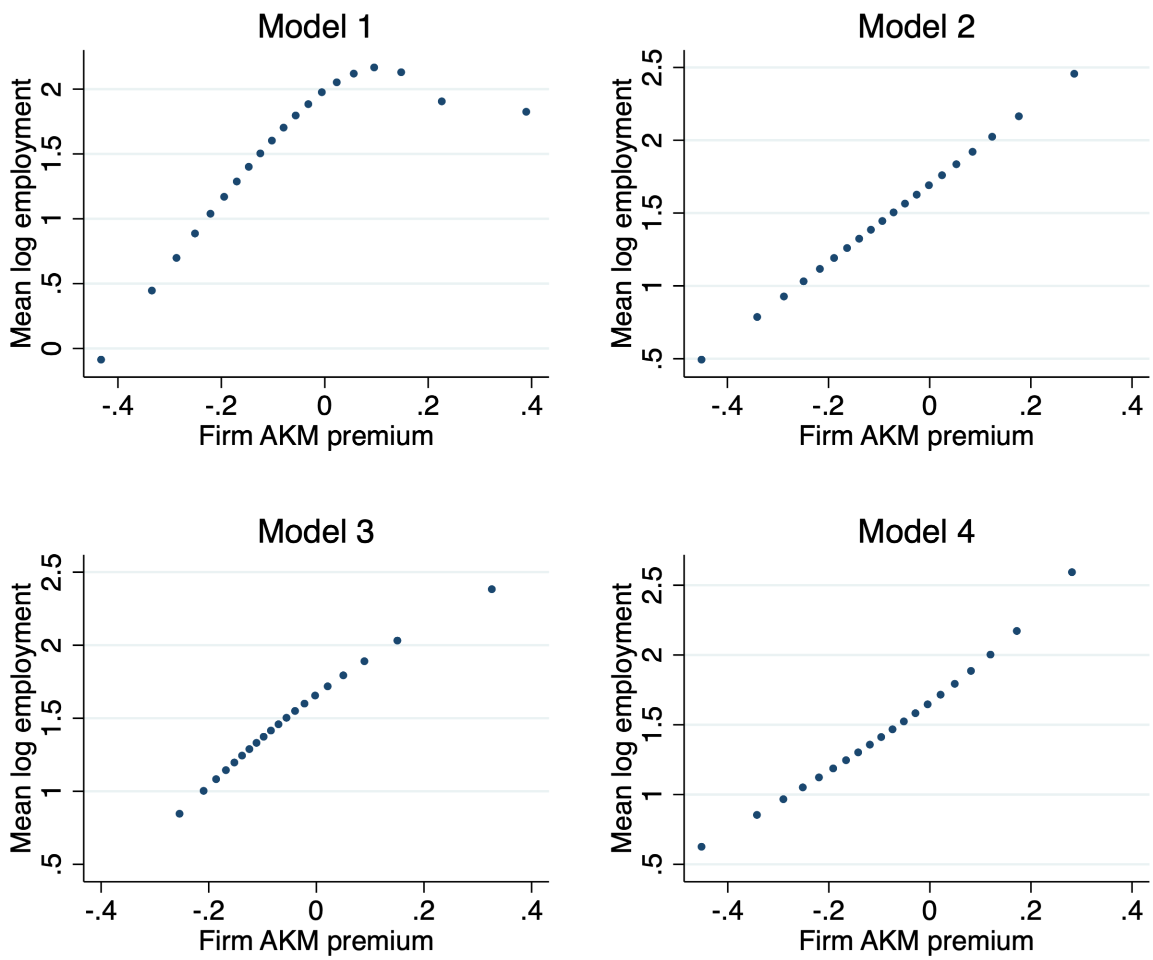
Notes: These plots repeat the exercise of Figure 6, but now showing mean log firm employment separately for three education groups: non-graduates, non-STEM graduates and STEM graduates. Sample consists of private sector firms in 2010-2019.

Figure 8: Education-specific AKM firm premia



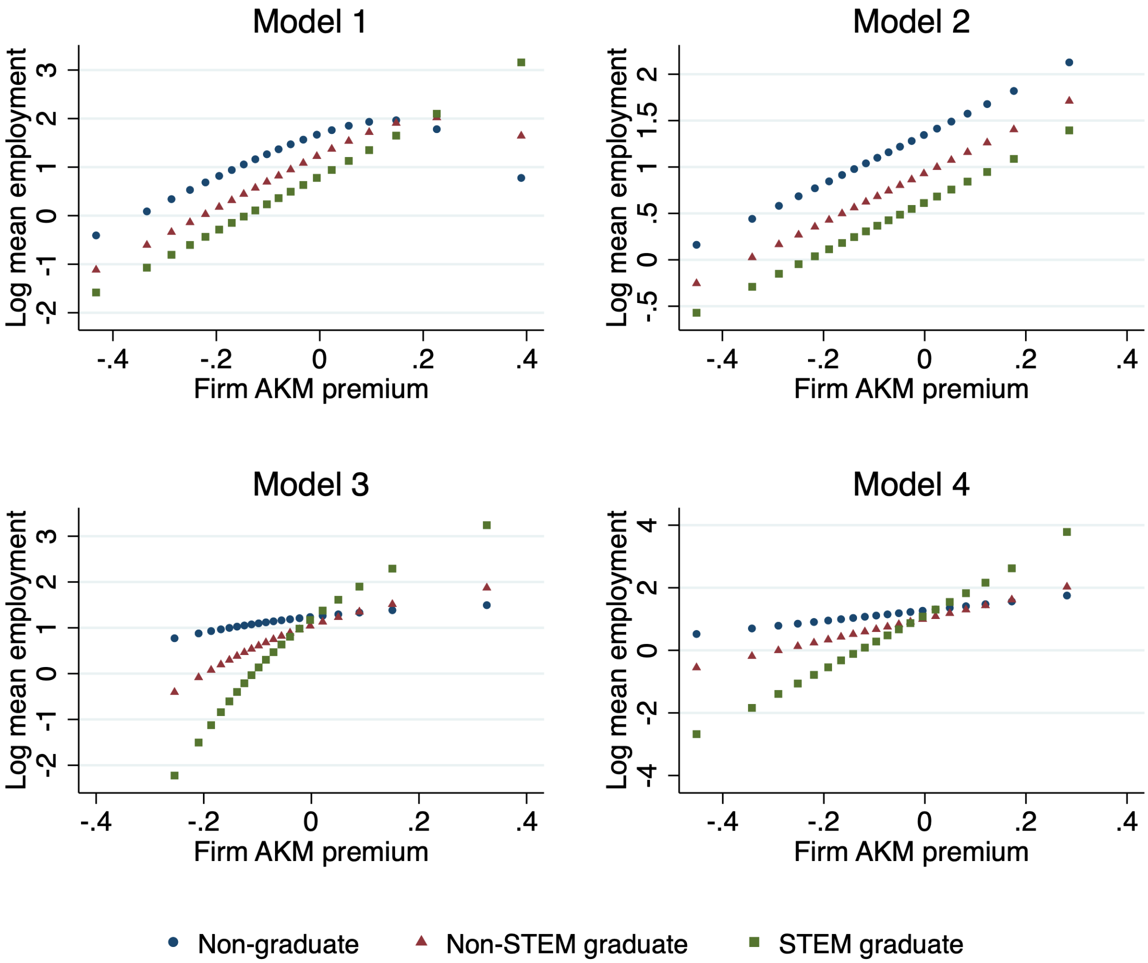
Notes: This figure estimates AKM firm premia separately by education group, and plots these group-specific premia against the aggregate (i.e., full sample) firm premia, across 20 bins (ordered by the aggregate premia). The bins are defined separately by education group, and contain equal numbers of group-specific workers. Group-specific and aggregate premia are normalized to zero for firms with mean (employment-weighted) aggregate premia. If wages are log-additive, the group-specific premia will line up perfectly on the 45 degree (dashed) line. Panel B repeats this exercise, after removing industry effects from the group-specific and aggregate premia. We implement a split-sample procedure to correct for measurement error in the firm premia, as described in Section 4.3. Sample includes private sector firms in 2010-2019.

Figure 9: Employment by firm pay premium: Models



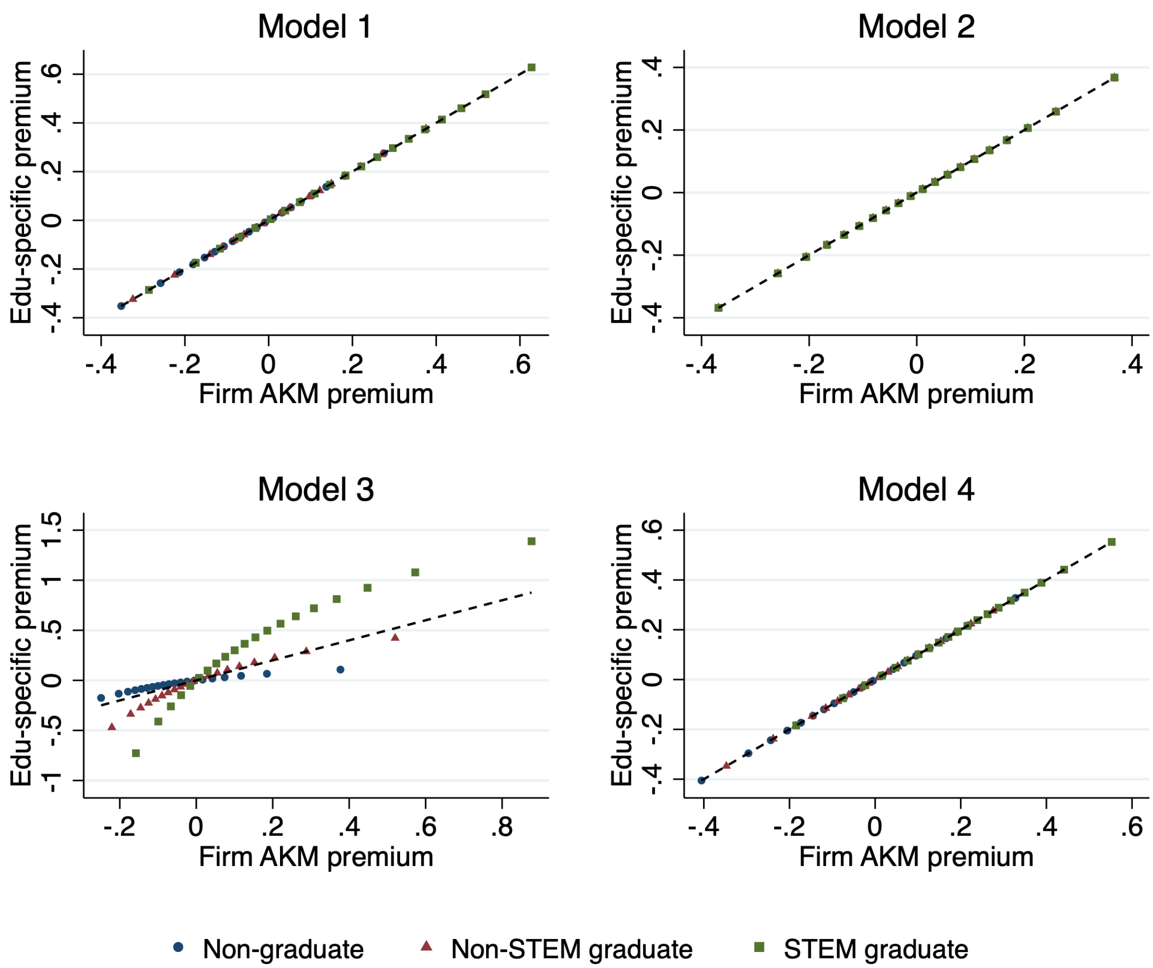
Notes: Figure shows mean log firm employment across 20 bins (each with equal numbers of firms), arranged by AKM firm premia, separately for the four models described in the text. Firm premia are normalized to their worker-weighted mean.

Figure 10: Employment by education and firm pay premium: Models



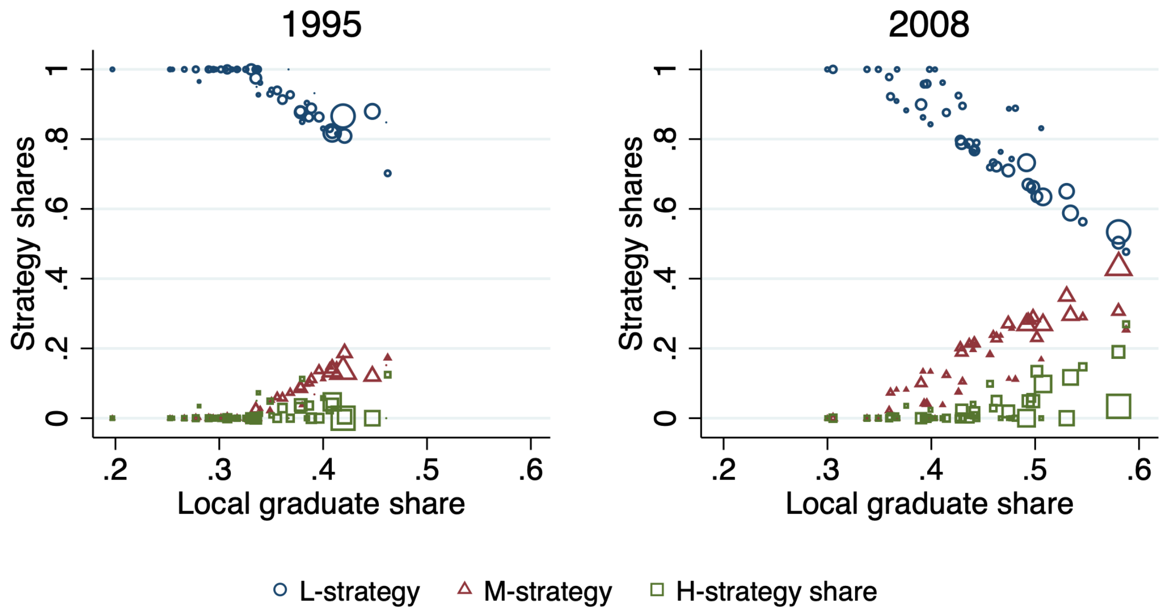
Notes: Figure shows log of mean firm employment by education group across 20 bins (each with equal numbers of firms), arranged by AKM firm premia, separately for the four models described in the text. Firm premia are normalized to their worker-weighted mean.

Figure 11: Education-specific AKM firm premia: Models



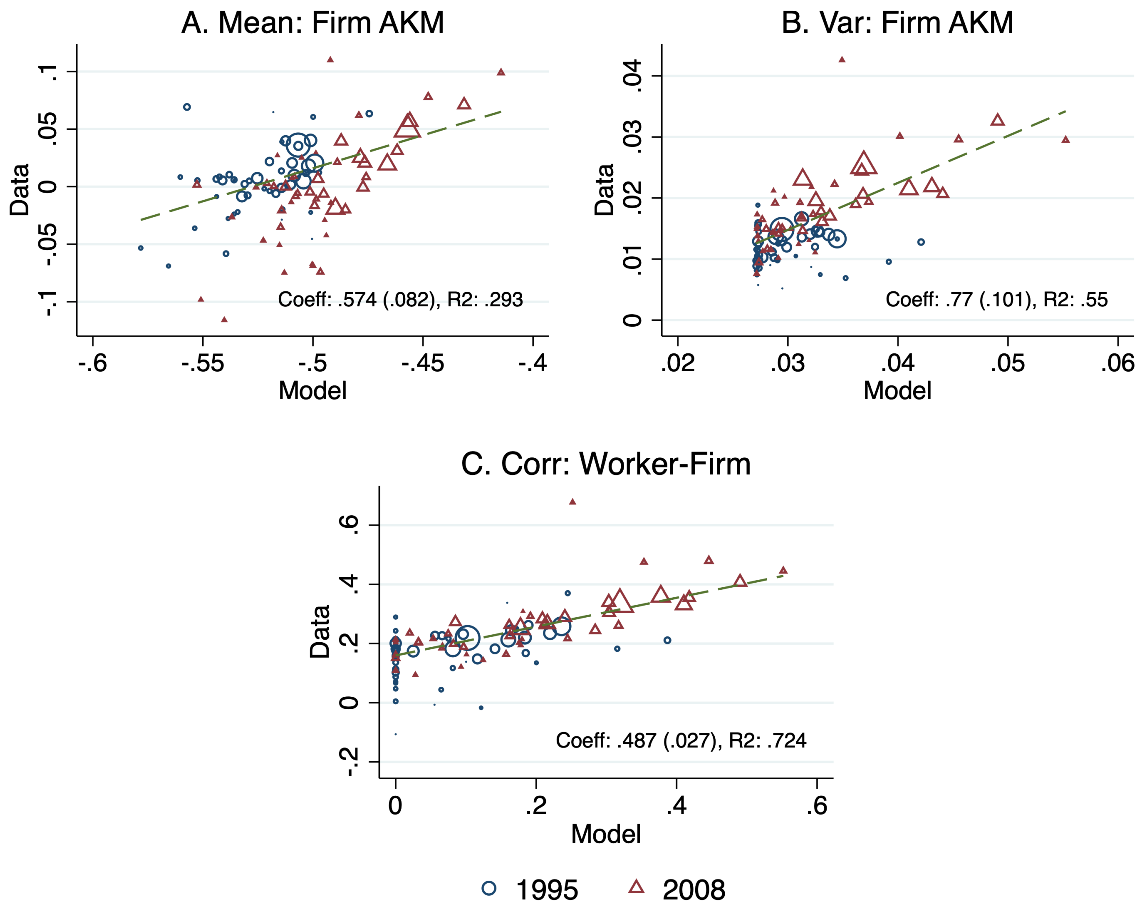
Notes: Figure shows education-specific firm wage premia across 20 bins (each with equal numbers of firms), arranged by aggregate AKM firm premia, separately for the four models described in the text. Firm premia are normalized to their worker-weighted mean.

Figure 12: Equilibrium strategy shares by region



Notes: Figure shows predicted shares of firms adopting L-strategy, M-strategy and H-strategy in each region, by regional graduate share, separately by census year. Marker size corresponds to regional employment. L-strategy firms hire all skill types, M-strategy firms hire only m and h-types, (i.e., non-STEM and STEM graduates), and H-strategy firms hire only h-types (i.e., STEM graduates).

Figure 13: Observed and predicted regional outcomes



Notes: Figure shows observed regional outcomes (on y-axis) against model predictions (x-axis), separately for (i) mean of firm AKM premia, (ii) variance of firm AKM premia, and (iii) correlation of firm and worker AKM premia, within each region-year pair. Each figure shows 98 markers, one for each region-pair: i.e., 49 regions in two census years (1995 and 2008). Marker size corresponds to regional share of total employment. Coefficient estimate and R-squared correspond to depicted OLS fit line.

Appendices: For Online Publication

A Appendix tables and figures

Table A1: Worker mobility patterns across firm pay quartiles

Origin quartile	Destination quartile			
	1 (lowest)	2	3	4 (highest)
Panel A: Non-graduates				
1 (lowest)	0.460	0.287	0.177	0.075
2	0.312	0.374	0.221	0.093
3	0.232	0.285	0.321	0.162
4 (highest)	0.156	0.174	0.258	0.412
Panel B: Non-STEM graduates				
1 (lowest)	0.362	0.294	0.209	0.135
2	0.252	0.338	0.252	0.158
3	0.165	0.232	0.326	0.277
4 (highest)	0.093	0.116	0.217	0.575
Panel C: STEM graduates				
1 (lowest)	0.239	0.204	0.215	0.342
2	0.169	0.246	0.254	0.331
3	0.087	0.133	0.281	0.499
4 (highest)	0.031	0.038	0.108	0.823

Notes: This table presents the share of job movers from each origin firm pay quartile (rows) who transition to each destination firm pay quartile (columns), separately by education group. Each row sums to 1. Firm quartiles are based on AKM firm fixed effects estimated for the period 2010-2019. Sample is restricted to private sector.

Table A2: Calibrated parameters across models

Parameter	Model 1 Baseline	Model 2 Skill-Neutral	Model 3 Skill-Biased	Model 4 Varying Elasticities
Labor supply parameters				
Labor supply elasticity (ε)	5.640	2.664	2.550	-
STEM-degree (ε_h)	-	-	-	8.674
Non-STEM (ε_m)	-	-	-	3.510
No degree (ε_l)	-	-	-	1.677
Productivity parameters				
Firm productivity variance (ν)	0.023	0.032	0.264	0.032
Log productivity (relative to STEM)				
Non-STEM intercept ($\log \frac{p_m}{p_h}$)	-0.634	-0.460	0.094	-0.093
No-degree intercept ($\log \frac{p_l}{p_h}$)	-1.047	-0.744	-0.083	-0.163
Non-STEM slope ($\theta_m - 1$)	-	-	-0.579	-
Non-degree slope ($\theta_l - 1$)	-	-	-0.866	-
Equity constraint parameters				
Non-STEM pay ratio ($\log \phi_m$)	-0.311	-	-	-
Non-degree pay ratio ($\log \phi_l$)	-0.537	-	-	-

Notes: This table presents the calibrated parameter values for each model variant. The baseline model ("Model 1") features equity constraints and skill neutral firm heterogeneity in productivity. For the remaining models, we dispose of the equity constraints: Model 2 features skill-neutral firm heterogeneity only, Model 3 allows for skill-biased firm heterogeneity in productivity, and Model 4 incorporates skill differences in labor supply elasticities (alongside skill-neutral firm heterogeneity).

Table A3: Quantification of model parameters by decade

Moments				Parameters			
Moment	1990s	2000s	2010s	Parameter	1990s	2000s	2010s
ε_{data}	2.74	2.62	2.66	ε	4.05	4.72	5.64
V_{AKMf}	0.023	0.035	0.032	ν	0.020	0.027	0.023
$\log \frac{\phi_m}{\phi_l}$	0.20	0.20	0.23	σ^M	0.04	0.14	0.17
$\log \frac{1}{\phi_l}$	0.22	0.40	0.54	σ^H	0.02	0.05	0.08
$E[\log w_m] - E[\log w_l]$	0.22	0.24	0.28	$\log \frac{w_h^M}{w_h^L}$	0.19	0.17	0.16
$E[\log w_h] - E[\log w_l]$	0.31	0.57	0.74	$\log \frac{w_h^H}{w_h^L}$	0.57	0.46	0.41
$\log \frac{n_m}{n_l}$	-0.77	-0.58	-0.60	$\log \frac{\phi_m^\varepsilon \Omega_m}{\phi_l^\varepsilon \Omega_l}$	-0.86	-0.91	-1.05
$\log \frac{n_h}{n_l}$	-2.39	-1.90	-1.86	$\log \frac{\Omega_h}{\phi_l^\varepsilon \Omega_l}$	-2.62	-2.59	-2.85
				Implied parameters			
				$\log \frac{p_m}{p_l}$	0.45	0.40	0.41
				$\log \frac{p_h}{p_l}$	0.93	0.98	1.05

Notes: This table extends the calibration exercise of Table 2 to previous decades. The left columns show the empirical moments used for model calibration, and the right columns the resulting parameter estimates, separately for three decadal intervals: 1990-1999, 2000-2009 and 2010-2019. The third column, corresponding to the 2010s, is identical to results reported in Table 2. Sample consists of private sector firms. See Appendix G for more details.

Table A4: Regional distribution of skill shares and wages

	1995		2008	
	Mean	SD	Mean	SD
Graduate share	0.389	0.045	0.494	0.066
Non-STEM graduate share	0.327	0.039	0.399	0.048
STEM graduate share	0.063	0.016	0.096	0.035
Mean: Log wage	8.979	0.112	9.086	0.137
Mean: Firm AKM	0.019	0.019	0.019	0.034
Mean: Worker AKM	8.977	0.098	9.076	0.107
Var: Firm AKM	0.013	0.002	0.021	0.005
Corr: Worker, firm	0.209	0.041	0.298	0.071

Notes: This table presents regional means and standard deviations of key variables in the 1995 and 2008 census years. Graduate share is the local fraction of workers with college degrees, which we disaggregate into non-STEM and STEM shares. AKM effects are estimated using employment records for the corresponding intervals: 1993-1997 for the 1995 census, and 2006-2010 for 2008. The sample consists of private sector firms across 49 regions in each census year.

Table A5: Regional effects of STEM and non-STEM employment shares

	Mean: Firm AKM		Var: Firm AKM		Corr: Worker, Firm	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Data						
Non-STEM grad share	0.155	0.133	0.034	-0.015	0.130	-0.128
	(0.103)	(0.088)	(0.006)	(0.023)	(0.089)	(0.181)
STEM grad share	0.622	0.628	0.079	0.128	1.585	1.606
	(0.109)	(0.109)	(0.016)	(0.017)	(0.152)	(0.259)
Within-R^2	0.511	0.657	0.572	0.493	0.599	0.484
Panel B: Model						
Non-STEM grad share	0.210	0.211	-0.007	-0.006	0.264	0.517
	(0.006)	(0.009)	(0.002)	(0.002)	(0.090)	(0.207)
STEM grad share	0.575	0.553	0.163	0.158	3.588	2.971
	(0.009)	(0.009)	(0.004)	(0.003)	(0.269)	(0.223)
Within-R^2	0.994	0.992	0.980	0.985	0.923	0.894
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Sample	98	98	98	98	98	98

Notes: This table replicates Panels A and D of Table 4, but replacing the regional graduate share on the right-hand side with distinct STEM and non-STEM shares. Odd-numbered columns exploit cross-sectional variation across regions, using equation (21). Even-numbered columns control for region fixed effects, as in equation (22), relying on within-region changes for identification. The dependent variables are the mean firm AKM premia (columns 1-2), the variance of firm AKM premia (columns 3-4), and the correlation between firm and worker AKM premia (columns 5-6). Sample consists of 49 regions observed in both 1995 and 2008 census years, for a total of 98 region-year observations. Observations are weighted by regional employment shares. Standard errors, clustered by region, in parentheses.

Table A6: Regional effects on local skill returns

	Log Wage		Worker AKM		Firm AKM	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Non-STEM graduates v non-graduates — data						
Graduate share	0.683	0.303	0.486	0.151	0.172	0.175
	(0.095)	(0.227)	(0.077)	(0.202)	(0.026)	(0.062)
Within-R^2	0.337	0.070	0.257	0.021	0.363	0.276
Panel B: STEM graduates v non-graduates — data						
Graduate share	1.665	0.901	1.136	0.451	0.512	0.439
	(0.235)	(0.581)	(0.208)	(0.461)	(0.050)	(0.131)
Within-R^2	0.440	0.108	0.343	0.042	0.547	0.273
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Sample	98	98	98	98	98	98

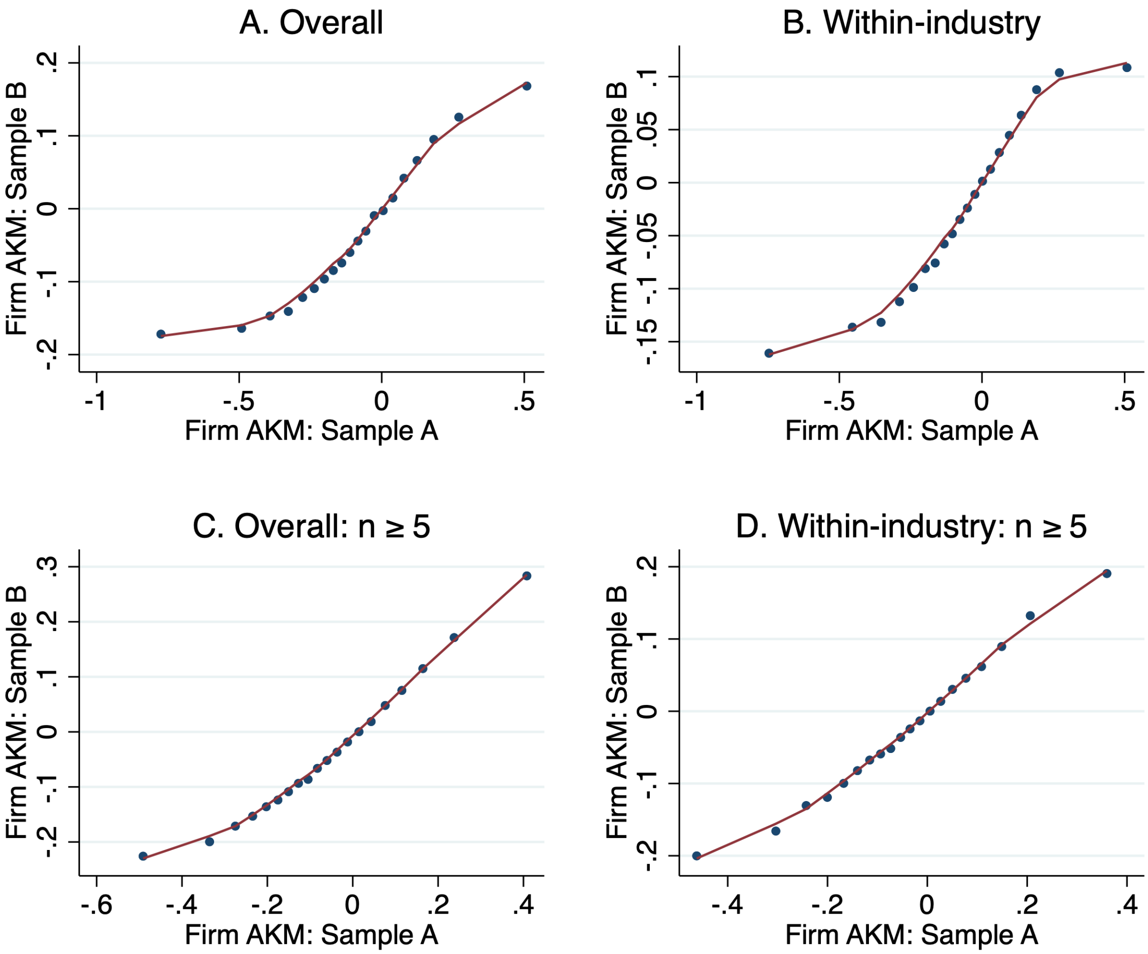
Notes: Table shows the relationship between regional graduate share and local skill returns. Odd-numbered columns exploit cross-sectional variation across regions, using equation (21). Even-numbered columns control for region fixed effects, as in equation (22), relying on within-region changes for identification. Panel A explores wage differentials between non-STEM graduates and non-graduates, and Panel B between STEM graduates and non-graduates. The dependent variables are the mean log wage differential (columns 1-2), the mean differential in AKM worker effects (columns 3-4), and the mean differential in AKM firm effects (columns 5-6). Sample consists of 49 regions observed in both 1995 and 2008 census years, for a total of 98 region-year observations. Observations are weighted by regional employment shares. Standard errors, clustered by region, in parentheses.

Table A7: Distribution of equilibrium configurations

	1995	2008
L-strategy only	0.429	0.122
$L + M$ strategies	0.204	0.367
$L + H$ strategies	0.061	0.000
$L + M + H$ strategies	0.306	0.510

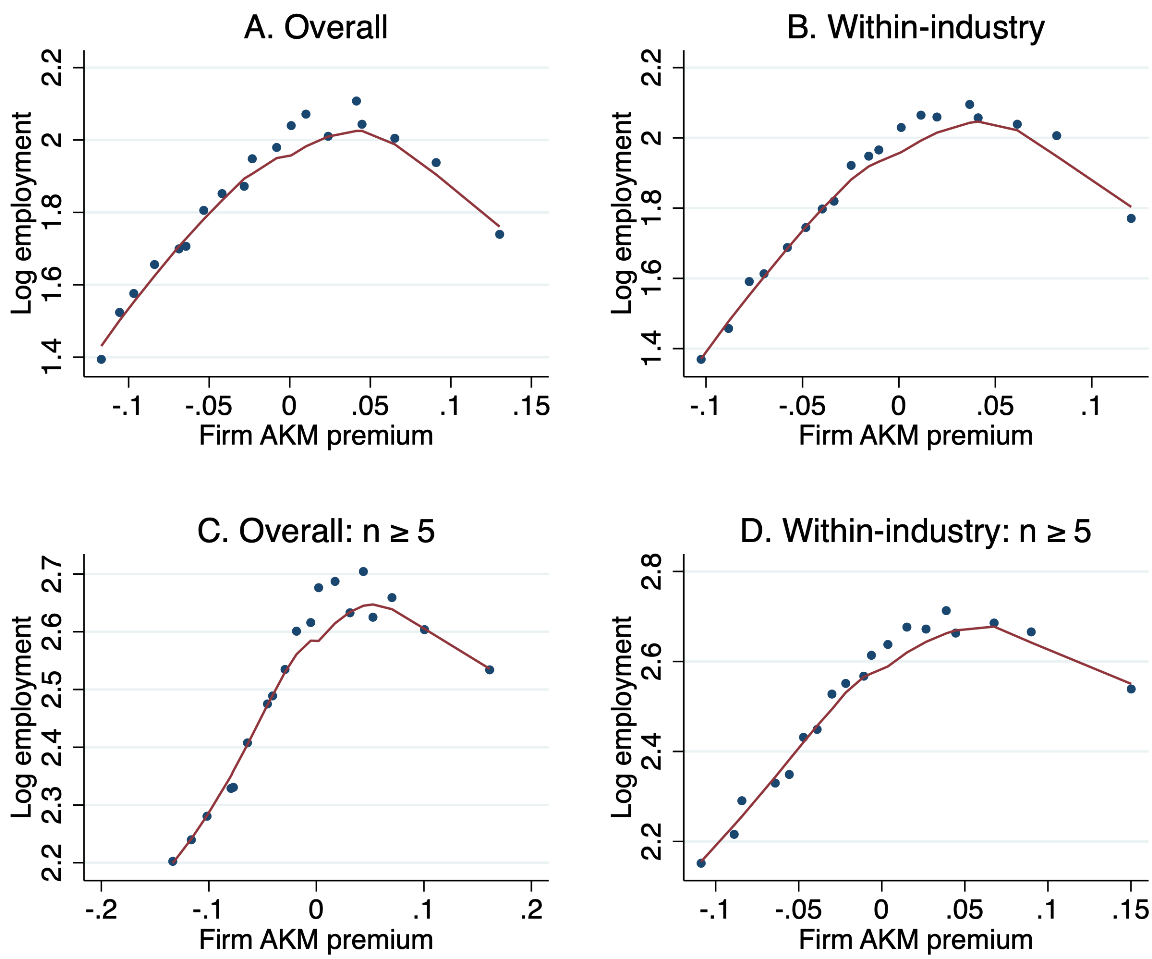
Notes: This table reports shares of regions with each equilibrium configuration, by census year. The first row shows the share of regions populated exclusively by inclusive L-strategy firms (i.e., hiring all skill types). Regions in the second row contain a mix of L- and M-strategy firms, and those in the third row a mix of L- and H-strategy firms. Regions in the final row feature firms with all three strategies in equilibrium.

Figure A1: Split sample: First stage



Notes: This figure shows the relationship between the AKM firm effects in the two random worker samples. This can be interpreted as a "first stage" relationship for our split-sample correction in the main text. In Panel B, we remove industry fixed effects from both the y-variable and the x-variable. Panels C and D repeat this exercise after excluding firms with fewer than 5 employees.

Figure A2: Employment by firm pay premium in Veneto



Notes: Panel A shows mean log firm employment across 20 bins (with equal numbers of firms) in the Veneto Worker History (VWH) database, arranged by AKM firm premia. Firm premia are normalized to the worker-weighted mean. We implement a split-sample procedure to correct for measurement error in the firm premia, as described in Section 4.1. In Panel B, we remove industry fixed effects from both the y-variable (log firm employment) and the x-variable (firm premia). Panels C and D repeat this exercise after excluding firms with fewer than 5 employees. Sample consists of private sector firms in 1992-2001.

B Theoretical proofs for baseline model

B.1 Derivation of optimal inclusive wages (14) and (15)

Suppose the equity constraint binds, i.e. $\phi > \frac{p_l}{p_h}$. For inclusive firms, the l-type wage w_l will then equal ϕw_h; and the labor supply constraints will bind: i.e. $l_s = l_s(w_s)$ for $s = \{h, l\}$. We can then re-write the firm's problem in (4) as:

$$\max_{w_h} \pi(w_h) = (p_h - w_h)\, l_h(w_h) + (p_l - \phi w_h)\, l_l(\phi w_h) \tag{B1}$$

The first order condition for the h-type wage w_h is then:

$$(p_h - w_h)\, l_h'(w_h) + \phi(p_l - \phi w_h)\, l_l'(\phi w_h) = l_h(w_h) + \phi l_l(\phi w_h) \tag{B2}$$

After replacing $l_s(w_s)$ with (2), and using $w_s^* = \frac{\varepsilon}{1+\varepsilon} p_h$ from (10), and $\beta = \phi \frac{p_h}{p_l}$ from (12), we have:

$$w_h^I = \frac{1 + \frac{1}{\beta} \cdot \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}} \cdot w_h^* \tag{B3}$$

which delivers (14); and (15) then follows from the binding pay constraint $w_l = \phi w_h$.

B.2 Derivation of optimal selective wage

For selective firms, only the h-type labor supply constraint binds, i.e. $l_h = l_h(w_h)$. We can then re-write the firm's problem in (4) as:

$$\max_{w_h, l_l} \pi(w_h) = (p_h - w_h)\, l_h(w_h) + (p_l - \phi w_h)\, l_l \tag{B4}$$

where l-type employment l_l is rationed, and must be strictly below the labor supply curve:

$$l_l < l_l(\phi w_h) \tag{B5}$$

Since marginal products are fixed by assumption, firms will only ration l_l if the l-type wage w_l (which is fixed by ϕw_h) exceeds their productivity p_l. But if this is indeed the case, firms will optimally reject all l-type workers: i.e., $l_l = 0$. Imposing $l_l = 0$, the first order condition for the h-type wage w_h is then:

$$(p_h - w_h)\, l_h'(w_h) = l_h(w_h) \tag{B6}$$

After replacing $l_h(w_h)$ with (2), we have:

$$w_h^S = \frac{\varepsilon}{1+\varepsilon} p_h = w_h^* \tag{B7}$$

where w_h^* is the optimal unconstrained wage.

B.3 Derivation of equilibrium equations (16) and (18)

Expressions for labor supply intercepts Ω_s

To solve for equilibrium, we first require expressions for the labor supply intercepts Ω_s, for $s = \{h, l\}$. Using equation (3), the intercept for h-type workers can be written as:

$$\Omega_h = \frac{n_h}{k} \left[\sigma \left(w_h^S\right)^{\varepsilon} + (1-\sigma)\left(w_h^I\right)^{\varepsilon}\right]^{-1} \tag{B8}$$

where n_h is the measure of h-type workers, and k is the measure of firms. The square brackets contain an average of the wages (with an ε exponent) of selective firms (weighted by the selective firm share σ) and inclusive firms (weighted $1-\sigma$). This weighted average represents the outside option of h-type workers.

Similarly, the intercept for l-type workers can be written as:

$$\Omega_l = \frac{n_l}{k} \left[(1-\sigma)\left(\phi w_h^I\right)^{\varepsilon}\right]^{-1} \tag{B9}$$

where n_l is the measure of l-type worker. Since l-type workers cannot access selective firms, the outside option in (B9) only accounts for inclusive firms.

Using the definitions of β and α in equations (12) and (17), the ratio of the two intercepts can be written as:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha} \cdot \frac{\beta}{\phi^{1+\varepsilon}} \left[1 + \frac{\sigma}{1-\sigma}\left(\frac{w_h^S}{w_h^I}\right)^{\varepsilon}\right] \tag{B10}$$

Finally, replacing w_h^I and w_h^S with (B3) and (B7), we have:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha} \cdot \frac{\beta}{\phi^{1+\varepsilon}} \left[1 + \frac{\sigma}{1-\sigma}\left(\frac{1+\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}{1+\frac{\phi^{1+\varepsilon}}{\beta}\cdot\frac{\Omega_l}{\Omega_h}}\right)^{\varepsilon}\right] \tag{B11}$$

which is an equilibrium relationship between the intercept ratio $\frac{\Omega_l}{\Omega_h}$ and selective share σ. To fix the equilibrium values of each, we need to assess the profits from the selective and inclusive strategies.

Expressions for inclusive and selective firm profits

Inserting the optimal inclusive wage (B3) into equation (4), and replacing $l_s(w_s)$ with (2), the inclusive profit can be written as:

$$\pi^I = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \cdot \frac{\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon}}{\left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^{\varepsilon}} \cdot \Omega_h p_h^{1+\varepsilon} \tag{B12}$$

Similarly, inserting the optimal selective wage (B7) into equation (B4), and replacing $l_h(w_h)$ with (2), the selective profit can be written as:

$$\pi^S = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \Omega_h p_h^{1+\varepsilon} \tag{B13}$$

Equilibrium with zero workplace segregation: $\sigma = 0$

For an equilibrium with zero workplace segregation ($\sigma = 0$), firms must strictly prefer the inclusive strategy: i.e. $\pi^I > \pi^S$. Using (B12) and (B13), this implies:

$$\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon} > \left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^{\varepsilon} \tag{B14}$$

But imposing $\sigma = 0$ on (B11), we have:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha} \cdot \frac{\beta}{\phi^{1+\varepsilon}} \tag{B15}$$

And after inserting this equation (B14) and rearranging:

$$\beta < \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha} \tag{B16}$$

which is the threshold condition for a $\sigma = 0$ equilibrium in equation (16).

Equilibrium with partial workplace segregation: $\sigma > 0$

For an equilibrium with partial workplace segregation ($\sigma > 0$), firms must be indifferent between the selective and inclusive strategies: i.e. $\pi^I = \pi^S$. Equating (B12) and (B13), this implies:

$$\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon} = \left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^{\varepsilon} \tag{B17}$$

Imposing this on (B11) yields:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha-\sigma} \cdot \frac{\beta}{\phi^{1+\varepsilon}} \tag{B18}$$

And replacing $\frac{\Omega_l}{\Omega_h}$ in equation (B17) with (B18):

$$\left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{1+\varepsilon} = \left(1 + \beta\frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon} \tag{B19}$$

This is an implicit equation which solves for $\tilde{\sigma}$ in equation (18), i.e. the value of σ in an equilibrium with partial workplace segregation.

B.4 Proof of Proposition 1c: Wage compression effects

Effect on expected log wages by skill type

In this appendix, we study the impact of the equity constraint on expected log wages by skill type. We begin with h-type wages. Let $E[\log w_h|\beta > 1]$ denote the expected log wage of h-types in an economy with a binding equity constraint. This is a weighted average of log wages paid by selective and inclusive firms, with weights equal to their shares of h-type employment:

$$E[\log w_h|\beta > 1] = \frac{(1-\sigma)\, l_h\left(w_h^I\right)\log w_h^I + \sigma l_h\left(w_h^S\right)\log w_h^S}{(1-\sigma)\, l_h\left(w_h^I\right) + \sigma l_h\left(w_h^S\right)} \tag{B20}$$

In a counterfactual unconstrained economy, all firms offer h-types the unconstrained optimum w_h^*, as defined by equation (10). As Appendix B.2 shows, the optimal selective wage w_h^S is equal to w_h^*. Using (B20), the impact of the equity constrained can then be written as:

$$E[\log w_h|\beta > 1] - \log w_h^* = \frac{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}}{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma} \log\frac{w_h^I}{w_h^*} \tag{B21}$$

Note the optimal inclusive wage w_h^I is less than the unconstrained optimum w_h^*: see equation (14). It follows that $E[\log w_h] - \log w_h^*$ is negative: i.e., the equity constraint reduces the expected log h-type wage.

We next turn to l-type wages. Let $E[\log w_l|\beta > 1]$ denote the expected log wage of l-types in an economy with a binding equity constraint. Since l-types are denied employment by selective firms, this is simply equal to the log inclusive wage. The impact of the pay constraint (relative to the unconstrained optimum) can then be written as:

$$E\left[\log w_l|\beta > 1\right] - \log w_l^* = \log \frac{w_l^I}{w_l^*} \tag{B22}$$

From equation (15), we know the inclusive wage w_l^I must exceed the unconstrained optimum w_l^*. So this expression must be positive: i.e., the equity constraint increases the expected log l-type wage.

Effect on aggregate earnings and profit

Here, we show that aggregate earnings are unaffected by the equity constraint. Since output in this model is fixed by assumption (workers are equally productive at all firms), it is sufficient to show that profit is unaffected by the equity constraint.

We begin by solving for profit π^* in an unconstrained economy. Applying the optimal wage (10) to (3), the labor supply intercepts for skill type s will equal:

$$\Omega_s^* = \left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon} \frac{n_s}{k} \cdot p_s^{-\varepsilon} \tag{B23}$$

Using this expression, and applying the binding labor supply curve (2) and optimal wage (10) to the profit function (4), we have:

$$\pi^* = \frac{1}{\alpha} \cdot \frac{n_h}{k} \cdot \frac{p_h}{1+\varepsilon} \tag{B24}$$

where the h-type output share α is defined by (17).

Next, we turn to profit under a binding equity constraint, which we denote $\pi|\beta > 1$. Since firms are identical (and earn equal profit), we can use the profit of inclusive firms from equation (B12):

$$\pi|\beta > 1 = \frac{\varepsilon^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}} \cdot \frac{\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon}}{\left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^{\varepsilon}} \cdot \Omega_h p_h^{1+\varepsilon} \tag{B25}$$

Inserting expressions for the h-type labor supply intercept Ω_h from (B8), the equilibrium intercept ratio $\frac{\Omega_l}{\Omega_h}$ from (B18), the optimal unconstrained wage w_h^* from (10), and the optimal inclusive wage w_h^I from (14), this can be written as:

$$\pi|\beta > 1 = \frac{\left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{1+\varepsilon}}{\sigma\left(1 + \beta\frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon} + (1-\sigma)\left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon}} \cdot \frac{n_h}{k} \cdot \frac{p_h}{1+\varepsilon} \tag{B26}$$

We now consider two cases. In a non-segregated equilibrium, the selective share σ will equal zero; and equation (B26) will collapse to the unconstrained profit π^* in (B24). In a partially segregated equilibrium, the equal profit condition in equation (B19) ensures that $\left(1+\beta\frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon} = \left(1+\frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon}$; so again, equation (B26) will collapse to the unconstrained profit π^*. Therefore, profit is unaffected by the equity constraint; so the same must be true of aggregate earnings.

B.5 Proof of Proposition 1d: Amenity and welfare effects

Let $\bar{u}_s$ denote the expected utility of skill type s workers, and let $\bar{a}_s$ denote their expected amenity match. Given the assumption that amenity effects are distributed type-1 extreme value, $\bar{u}_s$ will equal the log of the inclusive value:

$$\bar{u}_s = \log \int_f w_{sf}^{\varepsilon} df + \gamma \tag{B27}$$

where γ is Euler's constant. From equation (1), the expected amenity match $\bar{a}_s$ can then be imputed by subtracting ε times the expected log wage:

$$\bar{a}_s = \log \int_f w_{sf}^{\varepsilon} df - \varepsilon E\left[\log w_s\right] + \gamma \tag{B28}$$

Proposition 1d states that the equity constraint increases the expected match $\bar{a}_s$ for both skill types (relative to the unconstrained optimum), if the constraint has sufficient bite (such that the selective share σ exceeds zero). We prove this result for each skill type in turn.

Effect on expected amenity match for h-types

Let $\bar{a}_s^*$ denote the expected amenity match in an unconstrained economy. For h-types, wages are fixed at the unconstrained optimum w_h^* in all firms; so $\bar{a}_h^*$ is simply equal to Euler's constant γ.

Next, let $\bar{a}_h|\beta > 1$ denote the expected amenity match for h-types in an economy with a binding equity constraint. Using (B28) and (B21), this can be written as:

$$\bar{a}_h|\beta > 1 = \log\left[(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma\right] - \frac{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}}{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma} \log\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \gamma \tag{B29}$$

The impact of the equity constraint, compared to an unconstrained counterfactual, is therefore:

$$(\bar{a}_h|\beta > 1) - \bar{a}_h^* = \log\left[(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma\right] - \frac{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}}{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma} \log\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} \tag{B30}$$

Since inclusive firms pay less than the unconstrained optimum w_h^*, the term $\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}$ must lie between 0 and 1. Notice that for $\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} = 1$, $(\bar{a}_h|\beta > 1) - \bar{a}_h^*$ is equal to zero. But after differentiating (B30) with respect to $\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}$, it can be shown that $(\bar{a}_h|\beta > 1) - \bar{a}_h^*$ is strictly increasing in $\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}$ for $\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} < 1$, as long as the selective share σ exceeds zero. It follows that $(\bar{a}_h|\beta > 1) - \bar{a}_h^*$ must be less than zero, for $\sigma > 0$: i.e. an equity constraint with sufficient bite (such that $\sigma > 0$) reduces the expected amenity match for h-types.

Since the equity constraint also reduces the expected log wage of h-types (Proposition 1c), we can conclude that their expected utility (which is the sum of the two) must decrease.

Effect on expected amenity match for l-types

Let $\bar{a}_l|\beta > 1$ denote the expected amenity match for l-type in an economy with a binding equity constraint. Using (B28) and (B22), this can be written as:

$$\bar{a}_l|\beta > 1 = \log(1-\sigma)\, w_l^I - \log w_l^I + \gamma \tag{B31}$$

The impact of the equity constraint, compared to an unconstrained counterfactual, is therefore:

$$(\bar{a}_l|\beta > 1) - \bar{a}_l^* = \log(1-\sigma) \tag{B32}$$

which is less than zero, if the selective share σ exceeds zero. That is, the equity constraint with sufficient bite (such that $\sigma > 0$) reduces the expected amenity match for l-types.

This reduction in the expected amenity match offsets the increase in the expected log wage of l-types (from Proposition 1c). And it turns out the overall impact on their expected utility is ambiguous. To see this, consider two extreme cases. (i) Suppose the bite of the equity constraint delivers an equilibrium selective share σ of zero. In this case, there is no impact on the expected amenity match in (B32), but the expected log wage in (B32) does increase; so expected utility must increase also. (ii) Suppose the h-type output share is arbitrarily close to 1, so the same must be true of the equilibrium selective share σ. This ensures an arbitrarily large negative effect on the expected amenity match in (B32), which will dominate the impact on the expected log wage in (B32). In this case, the equity constraint must reduce expected utility.

B.6 Proof of Proposition 2a: Negative firm size premium

In the baseline model with productively identical firms, Proposition 3 states that high-pay (i.e., selective) firms will have lower employment overall. To prove this result, we derive firm size for the selective and inclusive strategies.

Selective firms only employ h-type workers, and they pay the unconstrained optimal wage: i.e., $w_h^S = w_h^*$. Therefore, using the labor supply function (2), their firm size is equal to:

$$l_h\left(w_h^S\right) = \Omega_h\left(w_h^*\right)^{\varepsilon} \tag{B33}$$

where Ω_h is the h-type labor supply intercept.

Inclusive firms employ both h- and l-type workers: they pay the former w_h^I and the latter $w_l^I = \phi w_h^I$. Therefore, using the labor supply functions in (2), their firm size is equal to:

$$\begin{aligned} l_h\left(w_h^I\right) + l_l\left(w_l^I\right) &= \Omega_h\left(w_h^I\right)^{\varepsilon} + \Omega_l\left(\phi w_h^I\right)^{\varepsilon} \\ &= \left(1 + \frac{\Omega_l}{\Omega_h}\phi^{\varepsilon}\right)\Omega_h\left(w_h^I\right)^{\varepsilon} \end{aligned} \tag{B34}$$

Replacing the optimal inclusive wage w_h^I with equation (14) gives:

$$l_h\left(w_h^I\right) + l_l\left(w_l^I\right) = \left(1 + \frac{\Omega_l}{\Omega_h}\phi^{\varepsilon}\right)\left(\frac{1 + \frac{1}{\beta}\cdot\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}\right)^{\varepsilon}\Omega_h\left(w_h^*\right)^{\varepsilon} \tag{B35}$$

Since we are comparing selective and inclusive firms, we must be in an equilibrium with positive selective share ($\sigma > 0$). Therefore, the equilibrium ratio $\frac{\Omega_l}{\Omega_h}$ of the labor supply intercepts can be summarized by equation (B18). After replacing $\frac{\Omega_l}{\Omega_h}$ with (B18) and rearranging, we have:

$$l_h\left(w_h^I\right) + l_l\left(w_l^I\right) = \frac{1-\sigma}{\alpha-\sigma}\cdot\frac{\alpha-\sigma+(1-\alpha)\frac{\beta}{\phi}}{\alpha-\sigma+(1-\alpha)\beta}\cdot\Omega_h\left(w_h^*\right)^{\varepsilon} \tag{B36}$$

Since $\sigma < \alpha$ in equilibrium and $\phi \leq 1$, this expression must exceed the selective firm size (B33). This confirms that selective firms do indeed have lower employment overall.

B.7 Derivation of equation (19): Decomposition of skill differential

We begin by deriving a simple expression for the ratio of the inclusive wage to the optimal unconstrained wage, i.e. $\frac{w_h^I}{w_h^*}$. Replacing the intercept ratio $\frac{\Omega_l}{\Omega_h}$ with (B18) in equation (14), we have:

$$\frac{w_h^I}{w_h^*} = \frac{1 + \frac{1-\alpha}{\alpha-\sigma}}{1 + \beta\frac{1-\alpha}{\alpha-\sigma}}\cdot w_h^* \tag{B37}$$

Using (18), this can be re-written as:

$$\frac{w_h^I}{w_h^*} = \left(\frac{\alpha-\sigma}{1-\sigma}\right)^{\frac{1}{\varepsilon}} \tag{B38}$$

Next, we turn to the skill differential. The expected log h-type wage is given by equation (B20). And since l-types are denied access to selective firms, they all receive the inclusive firm wage: i.e. $w_l^I = \phi w_h^I$. Subtracting one from the other, the expected skill differential is:

$$E\left[\log w_h\right] - E\left[\log w_l\right] = -\log \phi w_h^I + \frac{(1-\sigma)\, l_h\left(w_h^I\right)\log w_h^I + \sigma l_h\left(w_h^S\right)\log w_h^S}{(1-\sigma)\, l_h\left(w_h^I\right) + \sigma l_h\left(w_h^S\right)} \tag{B39}$$

Applying the labor supply function (2), and given that the selective wage w_h^S is equal to the optimal unconstrained wage w_h^*, we have:

$$E\left[\log w_h\right] - E\left[\log w_l\right] = -\log \phi + \frac{\sigma}{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma}\log\frac{w_h^*}{w_h^I} \tag{B40}$$

And after replacing $\frac{w_h^I}{w_h^*}$ with (B38), we reach equation (19) in the main text.

B.8 Proof of Proposition 3: Impact of h-type output share α

Suppose first that the h-type output share α is sufficiently large, such that $\frac{1-\alpha}{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha} > \beta$; so the selective share σ exceeds zero: see equation (16). For the purposes of this proof, it is useful to define the function $\Lambda(\beta,\varepsilon)$ as the solution of the implicit equation:

$$(1+\Lambda)^{1+\varepsilon} = (1+\beta\Lambda)^{\varepsilon} \tag{B41}$$

This is identical to the equilibrium equation (B19), except with $\frac{1-\alpha}{\alpha-\sigma}$ replaced by Λ, which exceeds zero if $\sigma > 0$. Using this definition, we can express equilibrium as:

$$\Lambda(\beta,\varepsilon) = \frac{1-\alpha}{\alpha-\sigma} \tag{B42}$$

From equation (B42), since Λ is fixed by the exogenous parameters β and ε (and invariant to α), the selective share σ must then be increasing in α.

Next, consider the between-firm component in equation (19), which is equal to $\frac{\sigma}{\alpha}\log\left(\frac{1-\sigma}{\alpha-\sigma}\right)^{\frac{1}{\varepsilon}}$.

Using (B42), this can be re-written as:

$$\text{Between-firm} = \left[1 - \frac{1-\alpha}{\alpha} \cdot \frac{1}{\Lambda(\beta,\varepsilon)}\right] \log\left(1+\Lambda(\beta,\varepsilon)\right)^{\frac{1}{\varepsilon}} \tag{B43}$$

Holding the exogenous parameters β and ε fixed, the between-firm component in (B43) must be increasing in α. This proves the first part of the proposition.

Now suppose that $\frac{1-\alpha}{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha} < \beta$. Equation (16) then shows that the selective share σ is fixed at zero and unaffected by α. We are therefore in a zero-segregation equilibrium, where all firms pursue the inclusive hiring strategy and offer identical wages: w_h^I for h-types, and w_l^I for l-types. Replacing the intercept ratio $\frac{\Omega_l}{\Omega_h}$ with (B15) in equation (14), the h-type inclusive wage in a zero-segregation equilibrium is equal to:

$$w_h^I = \frac{1}{\alpha + (1-\alpha)\beta} \cdot w_h^* \tag{B44}$$

which is increasing in α (as well as in h-type productivity p_h, via the optimal unconstrained wage term w_h^*). At the same time, the l-type wage is equal to ϕw_h^I (where ϕ is fixed by the equity constraint), so any productive benefits of larger α are shared equally with l-types.

C Extension with heterogeneous firms

In this appendix, we extend our baseline model to account for skill-neutral heterogeneity in firm productivity. In a given firm f with firm-specific parameter x_f, suppose the h-type and l-type marginal products are equal to $p_{hf} = x_f p_h$ and $p_{lf} = x_f p_l$ respectively, where $\tilde{x}_f \equiv \log x_f$ has distribution F across firms, where F is normal with mean 0 and variance ν. For the purposes of this analysis, suppose the equity constraint binds, and β exceeds $\frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}}-\alpha}{1-\alpha}$; so the equilibrium selective share σ exceeds 0.

C.1 Characterization of equilibrium with heterogeneous firms

We begin by characterizing the equilibrium in this extended model. Building from equation (10), for a firm with productivity x, the unconstrained optimum wage for skill type $s = \{h, l\}$ can be written as:

$$w_s^*(x) = \frac{\varepsilon}{1+\varepsilon} p_s x \tag{C1}$$

Selective firms with productivity x pay the unconstrained optimum to h-types:

$$w_h^S(x) = w_h^*(x) \tag{C2}$$

Replacing w_h^* with $w_h^*(x)$ in equation (14), inclusive firms with productivity x offer a wage equal to:

$$w_h^I(x) = \frac{1 + \frac{1}{\beta} \cdot \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}} w_h^*(x) \tag{C3}$$

Replacing p_h with $p_h x$ in equations (B12) and (B13), the profits associated with these strategies are:

$$\pi^S(x) = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \Omega_h (p_h x)^{1+\varepsilon} \tag{C4}$$

and

$$\pi^I(x) = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \cdot \frac{\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon}}{\left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^{\varepsilon}} \cdot \Omega_h (p_h x)^{1+\varepsilon} \tag{C5}$$

Comparing (B12) and (B13), it is clear that the productivity parameter x makes no difference to the relative profits of the two strategies; and hence, x does not affect the choice of strategy. It follows that selective and inclusive firms will be distributed identically in terms of x.

This allows us to characterize the pay distributions among selective and inclusive firms. Let F^s be the distribution of log h-type wages among selective firms, i.e. $\tilde{w}_h^S \sim F^S$, where the tilde indicates a log variable: $\tilde{w}_h^S \equiv \log w_h^S$. Similarly, let F^I be the distribution of log h-type wages among inclusive firms, i.e. $\tilde{w}_h^I \sim F^I$. Expressing these distributions in terms of log (rather than dollar) wages will simplify the proofs below. It then follows that:

$$F^S(\tilde{w}) = F^x\left(\tilde{w} - \tilde{w}_h^S(1)\right) \tag{C6}$$

$$F^I(\tilde{w}) = F^x\left(\tilde{w} - \tilde{w}_h^I(1)\right) \tag{C7}$$

where $\tilde{w}_h^S(1) = \log \frac{\varepsilon}{1+\varepsilon} p_h$ and $\tilde{w}_h^I(1) = \log \frac{1+\frac{1}{\beta}\cdot\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}{1+\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}} \frac{\varepsilon}{1+\varepsilon} p_h$. In summary, both the F^S and F^I distributions have identical variance (equal to ν, the same as for firm productivity x), but inclusive firms have a lower mean.

We now turn to the labor supply intercepts, Ω_h and Ω_l. Using equation (3), the intercept for h-type workers can be written as:

$$\begin{aligned}\Omega_h &= \frac{n_h}{k}\left[\sigma \int_{\tilde{w}} e^{\varepsilon \tilde{w}} dF^S(\tilde{w}) + (1-\sigma)\int_{\tilde{w}} e^{\varepsilon \tilde{w}} dF^I(\tilde{w})\right]^{-1} \\ &= \frac{n_h}{k}\left[\sigma\left(w_h^S(1)\right)^\varepsilon + (1-\sigma)\left(w_h^I(1)\right)^\varepsilon\right]^{-1}\left[\int_{\tilde{x}} e^{\varepsilon \tilde{x}} dF^x(\tilde{x})\right]^{-1}\end{aligned} \tag{C8}$$

where n_h is the measure of h-type workers, and k is the measure of firms. In the first line of (C8), the square brackets contain an average of the wages (with an ε exponent) of selective firms (weighted by the selective firm share σ) and inclusive firms (weighted $1-\sigma$). The second line follows from the definition of (C6) and (C7), as well as the fact that $\tilde{w}_h^S(x) = \tilde{w}_h^S(1) + \tilde{x}$ and $\tilde{w}_h^I(x) = \tilde{w}_h^I(1) + \tilde{x}$: this additive separability allows us to disentangle the x terms from the rest of the expression. Similarly, the labor supply intercept for l-type workers can be written as:

$$\begin{aligned}\Omega_l &= \frac{n_l}{k}\left[(1-\sigma)\int_{\tilde{w}} \phi^{\varepsilon} e^{\varepsilon\tilde{w}} dF^I(\tilde{w})\right]^{-1} \\ &= \frac{n_l}{k}\left[(1-\sigma)\left(w_h^I(1)\right)^{\varepsilon}\right]^{-1}\left[\int_{\tilde{x}} e^{\varepsilon\tilde{x}} dF^x(\tilde{x})\right]^{-1}\end{aligned} \tag{C9}$$

Putting these together, the intercept ratio is identical to equation (B18) in the baseline model:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha}\cdot\frac{\beta}{\phi^{1+\varepsilon}}\left[1+\frac{\sigma}{1-\sigma}\left(\frac{w_h^I(1)}{w_h^I(1)}\right)^{\varepsilon}\right] = \frac{1-\alpha}{\alpha-\sigma}\cdot\frac{\beta}{\phi^{1+\varepsilon}} \tag{C10}$$

And hence, the equilibrium selective share σ will take an identical form to the baseline model, as specified by equation (B19). Finally, inserting (C10) into (C3), the log wage of an inclusive firm with productivity x can be written as:

$$\tilde{w}_h^I(x) = \tilde{w}_h^S(x) - \kappa \tag{C11}$$

where k is the pay differential between equally productive inclusive and selective firms:

$$\kappa = \log\left[1+\frac{1-\alpha}{1-\sigma}(\beta-1)\right] \tag{C12}$$

C.2 Proof of Proposition 2b: Firm size effects

Proposition 2b states that log employment is initially positive and concave (and possibly hump-shaped) in log firm pay. The key insight underlying these results is that firm pay may vary for two reasons: (i) heterogeneity in productivity x and (ii) choice of selective or inclusive strategy. As we have shown above, these two sources of variation are orthogonal: a firm's productivity x has no effect on whether it adopts a selective strategy.

It is first useful to define the selective share $\sigma(\tilde{w}_h)$, among firms offering a log h-type wage equal to $\tilde{w}_h$:

$$\begin{aligned}\sigma\left(\tilde{w}_h\right) &= \frac{\sigma f^S\left(\tilde{w}_h\right)}{\left(1-\sigma\right) f^I\left(\tilde{w}_h\right)+\sigma f^S\left(\tilde{w}_h\right)} \\ &= \frac{\sigma f^S\left(\tilde{w}_h\right)}{\left(1-\sigma\right) f^S\left(\tilde{w}_h+\kappa\right)+\sigma f^S\left(\tilde{w}_h\right)} \\ &= \left[\frac{1-\sigma}{\sigma}\exp\left(-\frac{\kappa}{\nu^2}\left(\tilde{w}_h-\tilde{w}_h^S\left(1\right)+\frac{1}{2}\kappa\right)\right)+1\right]^{-1}\end{aligned} \tag{C13}$$

where σ is the unconditional selective share, and $\tilde{w}_h^S\left(1\right)=\log\frac{\varepsilon}{1+\varepsilon}p_h$. The second equality follows from the definitions of F^S and F^I in (C6) and (C7), and the definition of k in equation (C12): i.e., the pay differential between equally productive selective and inclusive firms. The final equality follows from the fact that F^S and F^I are normally distributed, with means $\tilde{w}_h^S\left(1\right)$ and $\tilde{w}_h^S\left(1\right)-k$ respectively, and variance ν^2. Equation (C13) shows that the selective share $\sigma\left(\tilde{w}_h\right)$ is increasing in firm pay, and varies from 0 (for very low $\tilde{w}_h$) to 1 (for very high $\tilde{w}_h$). Intuitively, selective firms pay higher wages (conditional on productivity x); so the higher up the pay distribution we move, the larger the representation of selective firms.

Next, we consider how log firm employment varies over the firm pay distribution. Let $E\left[\log l|\tilde{w}_h\right]$ denote the expectation of log firm employment, conditional on the firm offering a log h-type wage equal to $\tilde{w}_h$. This is a weighted average of the expected log employment of selective and inclusive firms, with weights equal to the selective and inclusive shares at $\tilde{w}_h$:

$$E\left[\log l|\tilde{w}_h\right]=\sigma\left(\tilde{w}_h\right)E\left[\log l^S|\tilde{w}_h\right]+\left[1-\sigma\left(\tilde{w}_h\right)\right]\log\left[\log l^I|\tilde{w}_h\right] \tag{C14}$$

where $\sigma\left(\tilde{w}_h\right)$ is defined by (C13). Since selective firms recruit all h-type workers who are willing to work at $\tilde{w}_h$, their expected employment is simply equal to the h-type labor supply curve. And since inclusive firms recruit all workers (both h- and l-type) who are willing to work at $\tilde{w}_h$, their expected employment is equal to the sum of the h- and l-type labor supply curves. So we have:

$$E\left[\log l|\tilde{w}_h\right]=\sigma\left(\tilde{w}_h\right)\log l_h\left(e^{\varepsilon\tilde{w}_h}\right)+\left[1-\sigma\left(\tilde{w}_h\right)\right]\log\left[l_h\left(e^{\varepsilon\tilde{w}_h}\right)+\log l_l\left(\phi^\varepsilon e^{\varepsilon\tilde{w}_h}\right)\right] \tag{C15}$$

Inserting the labor supply curve (2) and rearranging:

$$E\left[\log l|\tilde{w}_h\right]=\log\left(\Omega_h+\Omega_l\phi^\varepsilon\right)+\varepsilon\tilde{w}_h-\log\left(1+\frac{\Omega_l\phi^\varepsilon}{\Omega_h}\right)\sigma\left(\tilde{w}_h\right) \tag{C16}$$

The first term on the right-hand side is a constant. The second term is linear and increasing in $\tilde{w}_h$, with slope ε: this is the contribution of the upward-sloping labor supply curve (high-paying firms attract more workers). The final term is decreasing in the selective share $\sigma(\tilde{w}_h)$: at higher firm pay $\tilde{w}_h$, a larger share of firms are selective, so there is more rationing of l-types.

The first derivative of $E[\log l|\tilde{w}_h]$ can be written as:

$$\frac{d}{d\tilde{w}_h}E[\log l|\tilde{w}_h] = \varepsilon - \frac{k}{\nu^2}\log\left(1+\frac{\Omega_l\phi^\varepsilon}{\Omega_h}\right)\sigma(\tilde{w}_h)[1-\sigma(\tilde{w}_h)] \tag{C17}$$

As $\tilde{w}_h$ becomes small, the selective share $\sigma(\tilde{w}_h)$ goes to zero, and the derivative converges to the labor supply elasticity ε. But for larger $\tilde{w}_h$, the second term ensures that the derivative drops below ε. The second derivative can be written as:

$$\frac{d^2}{d\tilde{w}_h^2}E[\log l|\tilde{w}_h] = -\frac{k}{\nu^2}\log\left(1+\frac{\Omega_l\phi^\varepsilon}{\Omega_h}\right)\sigma(\tilde{w}_h)[1-\sigma(\tilde{w}_h)][1-2\sigma(\tilde{w}_h)] \tag{C18}$$

which is negative for sufficiently small $\tilde{w}_h$. This proves Proposition 2b: log employment is initially positive and concave (and possibly hump-shaped) in log firm pay.

Finally, notice the curvature of $E[\log l|\tilde{w}_h]$ is more substantial (and more likely to be hump-shaped) if the ratio $\frac{\kappa}{\nu^2}$ is larger. Recall that κ is the pay differential between inclusive and selective firms (for given firm productivity x), and ν^2 is the variance of log firm productivity. A larger κ indicates that the "quality motive" is more dominant for firms choosing high pay (firms seeking more h-type employment), and a larger ν indicates that the "quantity motive" is more important (firms seeking more workers of any type). Intuitively, the more important the quality motive, the stronger the quantity-quality trade-off, and the more substantial the curvature of $E[\log l|\tilde{w}_h]$.

D Extension with CES technology

In the baseline model, we assume that h-types and l-types are perfect substitutes in production. As we show in Section 2.2, it follows that selective firms do not hire any l-type workers: i.e. perfect rationing. More generally though, if h- and l-types are imperfect substitutes, selective firms may still employ some l-types—though less intensively than inclusive firms. As we explain in Section 2.6, this partial rationing implies a role for luck in wage determination, even in the absence of search friction: some l-types will be fortunate to find work in selective firms, but others not.

D.1 Firm's problem

In this appendix, we explore equilibrium in this more general setting, with imperfectly substitutable labor inputs. Specifically, suppose there is a measure k of identical firms which produce output y (with price normalized to 1) according to the following CES technology:

$$y\left(l_h, l_l\right) = p\left(\theta l_h^\gamma + \left(1-\theta\right) l_l^\gamma\right)^{\frac{1}{\gamma}} \tag{C19}$$

where p is a fixed productivity parameter, l_l is l-type employment, l_h is h-type employment, and $\frac{1}{1-\gamma}$ is the elasticity of substitution between skill types. In the baseline model, we have $\gamma = 1$, $p_h = \theta p$ and $p_l = \left(1-\theta\right) p$.

Firms choose wages w_s and employment l_s of each skill type $s = \{h, l\}$, to maximize profit π:

$$\max_{w_h, w_l, l_h, l_l} \pi\left(w_h, w_l, l_h, l_l\right) = y\left(l_h, l_l\right) - w_h l_h - w_l l_l \tag{C20}$$

subject to labor supply constraints:

$$l_h \leq l_h\left(w_h\right) \tag{C21}$$

$$l_l \leq l_l\left(w_l\right) \tag{C22}$$

and the pay equity constraint:

$$w_l \geq \phi w_h \tag{C23}$$

D.2 Equilibrium if equity constraint does not bind

If the equity constraint does not bind, the labor supply constraints (C21) and (C22) must bind:

$$l_h^* = l_h\left(w_h\right) \tag{C24}$$

$$l_l^* = l_l\left(w_l\right) \tag{C25}$$

Intuitively, since firms set wages below marginal products, they will optimally hire all workers who are willing to join them. For skill type $s \in \{h, l\}$, the optimal wages are then fixed markdowns $\frac{\varepsilon}{1+\varepsilon}$ on the marginal products:

$$w_h^* = \frac{\varepsilon}{1+\varepsilon} \cdot y_h\left(l_h^*, l_l^*\right) = \frac{\varepsilon}{1+\varepsilon}\left[\left(\frac{1-\theta}{\theta}\right)^{\frac{1+\varepsilon}{1+\varepsilon(1-\gamma)}}\left(\frac{\Omega_l^*}{\Omega_h^*}\right)^{\frac{\gamma}{1+\varepsilon(1-\gamma)}} + 1\right]^{\frac{1-\gamma}{\gamma}} \theta^{\frac{1}{\gamma}} p \tag{C26}$$

and

$$w_l^* = \frac{\varepsilon}{1+\varepsilon} \cdot y_l\left(l_h^*, l_l^*\right) = \frac{\varepsilon}{1+\varepsilon}\left[\left(\frac{\theta}{1-\theta}\right)^{\frac{1+\varepsilon}{1+\varepsilon(1-\gamma)}}\left(\frac{\Omega_h^*}{\Omega_l^*}\right)^{\frac{\gamma}{1+\varepsilon(1-\gamma)}}+1\right]^{\frac{1-\gamma}{\gamma}}(1-\theta)^{\frac{1}{\gamma}} p \quad \text{(C27)}$$

where $y_h\left(l_h, l_l\right) \equiv \frac{\partial y(l_h,l_l)}{\partial l_h}$ and $y_l\left(l_h, l_l\right) \equiv \frac{\partial y(l_h,l_l)}{\partial l_l}$ denote the h-type and l-type marginal products, and Ω_h^* and Ω_l^* are the labor supply intercepts in the unconstrained equilibrium. Since all firms offer the same wage, equation (3) implies these intercepts are equal to:

$$\Omega_s^* = \left(w_s^*\right)^{-\varepsilon} n_s \quad \text{(C28)}$$

for skill $s \in \{h, l\}$. The optimal wage differential will then equal:

$$\frac{w_l^*}{w_h^*} = \frac{y_l\left(l_h^*, l_l^*\right)}{y_h\left(l_h^*, l_l^*\right)} = \left[\frac{1-\theta}{\theta}\left(\frac{\Omega_h^*}{\Omega_l^*}\right)^{1-\gamma}\right]^{\frac{1}{1+\varepsilon(1-\gamma)}} = \frac{1-\theta}{\theta}\left(\frac{n_h}{n_l}\right)^{1-\gamma} \quad \text{(C29)}$$

where n_s is the aggregate measure of type-s workers. From equation (C29), the equity constraint will therefore not bind if $\phi \leq \frac{1-\theta}{\theta}\left(\frac{n_h}{n_l}\right)^{1-\gamma}$.

D.3 Equilibrium if equity constraint binds

Suppose instead that the equity constraint binds, i.e. $\phi > \frac{1-\theta}{\theta}\left(\frac{n_h}{n_l}\right)^{1-\gamma}$. Wages will then take log additive form, in line with equation (13). In equilibrium, firms will adopt one of two pay strategies, just as in the baseline model: inclusive or selective. We discuss each in turn, and then solve for the equilibrium share of firms which adopt each strategy.

Unlike in the baseline model, all firms hire at least some l-type workers. But they differ in their optimal skill hiring ratios: selective firms hire relatively more h-types, since they ration l-types. For the purposes of this analysis, it is useful to introduce new notation for these skill ratios. Let $\lambda^I \equiv \frac{l_l^I}{l_h^I}$ denote the optimal ratio of l-type to h-type employment for inclusive firms; and let $\lambda^S \equiv \frac{l_l^S}{l_h^S}$ denote the optimal skill ratio for selective firms.

Inclusive strategy (I)

Inclusive firms hire all willing workers, so the labor supply constraints bind for both skill types: i.e., $l_h^I = l_h\left(w_h^I\right)$ and $l_l^I = l_l\left(w_l^I\right)$. From equation (2), it follows that the optimal skill ratio is equal to:

$$\lambda^I \equiv \frac{l_l^I}{l_h^I} = \frac{l_l\left(\phi w_h^I\right)}{l_h\left(w_h^I\right)} = \frac{\phi^\varepsilon \Omega_l}{\Omega_h} \quad \text{(C30)}$$

where Ω_l and Ω_h are the labor supply intercepts. To accommodate both skill types, firms compress pay internally to satisfy the equity constraint, redistributing wages between h- and l-types (relative to the unconstrained optimum), just as in the baseline model. To solve for the inclusive strategy, we can replace the l-type wage w_l with ϕw_h in the firm's problem above (i.e. imposing that the equity constraint binds), and replace h- and l-type employment with the labor supply curves (i.e. imposing that the labor supply constraints bind). This simplifies the problem to:

$$\max_{w_h} \pi\left(w_h\right) = y\left(l_h\left(w_h\right), l_l\left(\phi w_h\right)\right) - w_h l_h\left(w_h\right) - \phi w_h l_l\left(\phi w_h\right) \tag{C31}$$

Solving this problem, the optimal inclusive h-type wage can be written as:

$$w_h^I = \frac{\left[(1-\theta)\left(\lambda^I\right)^\gamma + \theta\right]^{\frac{1}{\gamma}}}{1+\phi\lambda^I} \cdot \frac{\varepsilon}{1+\varepsilon} p \tag{C32}$$

And using the equations above, the associated profit is:

$$\pi^I = \frac{1}{\varepsilon}\left(1+\phi\lambda^I\right)\Omega_h\left(w_h^I\right)^{1+\varepsilon} \tag{C33}$$

Selective strategy (S)

The alternative "selective" strategy is to hire all willing h-type workers (so the h-type labor supply curve binds), but to freely choose l-type employment to maximize profit. This opens the possibility of rationing l-type labor, so the l-type labor supply constraint does not bind: i.e. $l_l^S < l_l\left(w_l^S\right)$. Intuitively, firms offer higher pay to compete more effectively for h-types; but the equity constraint forces them to share these wage rents with l-types—potentially to the detriment of profit. Rationing may then be an optimal response: by reducing l-type employment, firms can ensure that the l-type marginal product exceeds the wage they are compelled (by the equity constraint) to pay them. However, given diminishing returns to l-type labor (implied by the CES technology), selective firms need not ration *all* their l-type workers to ensure this condition is met—unlike in the baseline model.

Since selective firms may only partially ration their employment of l-types, we require some tie-break rule to determine which l-types are "fortunate" (and join a selective firm) and which are not. Like Akerlof (1980) and Romer (1984), we simply assume that rationed jobs are allocated randomly among willing workers.

To solve for the selective strategy, we can replace the l-type wage w_l with ϕw_h in the firm's problem above (i.e. imposing that the equity constraint binds), and h-type employment l_h with its labor supply curve (i.e. imposing that the h-type labor supply constraints binds);

but we allow firms to freely choose l-type employment l_l:

$$\max_{w_h, l_l} \pi\left(w_h, l_l\right) = y\left(l_h\left(w_h\right), l_l\right) - w_h l_h\left(w_h\right) - \phi w_h l_l \tag{C34}$$

This problem yields two first order conditions. The first order condition for w_h implies the following expression for the optimal selective wage:

$$w_h^S = \left[(1-\theta) + \theta\left(\lambda^S\right)^{-\gamma}\right]^{\frac{1-\gamma}{\gamma}} \frac{1-\theta}{\phi} \cdot p \tag{C35}$$

where

$$\lambda^S \equiv \frac{l_l^S}{l_h^S} = \frac{l_l^S}{l_h\left(w_h^S\right)} < \lambda^I \tag{C36}$$

is the optimal ratio of l-type to h-type employment in selective firms. Note that l-type employment is rationed, so $l_l^S < l_l\left(w_l^S\right)$; whereas all willing h-types are hired, so $l_h^S = l_h\left(w_h^S\right)$. This implies that $\lambda^S < \lambda^I$, since inclusive firms do not ration l-types. The λ^S ratio is pinned down by the first order condition for l_l:

$$\frac{\theta}{1-\theta}\varepsilon\phi\left(\lambda^S\right)^{1-\gamma} = (1+\varepsilon) + \phi\lambda^S \tag{C37}$$

i.e. λ^S is fully determined by the exogenous parameters θ, γ, ε and ϕ. Using the equations above, the associated profit is equal to:

$$\pi^S = \frac{1}{\varepsilon}\left(1+\phi\lambda^S\right)\Omega_h\left(w_h^S\right)^{1+\varepsilon} \tag{C38}$$

Labor supply intercepts and strategy shares

Since firms do not ration h-type employment, the labor supply intercept has identical form to the baseline model. Using equation (3), we have:

$$\Omega_h = \frac{n_h}{k}\left[(1-\sigma)\left(w_h^I\right)^{\varepsilon} + \sigma\left(w_h^S\right)^{\varepsilon}\right]^{-1} \tag{C39}$$

where n_h is aggregate h-type employment, k is the measure of firms, and σ is the share of firms which adopt the selective strategy. For l-types, we have:

$$\Omega_l = \frac{n_l}{k}\left[(1-\sigma)\left(\phi w_h^I\right)^{\varepsilon} + \frac{\lambda^S}{\lambda^I}\sigma\left(\phi w_h^S\right)^{\varepsilon}\right]^{-1} \tag{C40}$$

where $\frac{\lambda^S}{\lambda^I} = \frac{l_l^S}{l_l\left(\phi w_h^S\right)} < 1$ is the ratio of l-type employment to their potential supply, for selective firms: this represents the extent of rationing. Putting equations (C39) and (C40) together, we have:

$$\frac{\sigma}{1-\sigma} = \frac{\lambda^I - \frac{n_l}{n_h}}{\frac{n_l}{n_h} - \lambda^S} \left(\frac{w_h^I}{w_h^S}\right)^{\varepsilon} \tag{C41}$$

which pins down the equilibrium selective share σ. Since inclusive firms disproportionately hire l-types, we must have $\lambda^I > \frac{n_l}{n_h} > \lambda^S$: i.e. the skill ratio in inclusive firms exceeds the aggregate skill ratio, which in turn exceeds the skill ratio in selective firms.

Just as in the baseline model, equilibrium can take one of two forms: zero workplace segregation ($\sigma = 0$), with no rationing of l-type workers, or partial workplace segregation ($\sigma > 0$). We now assess each in turn.

Equilibrium with zero workplace segregation: $\sigma = 0$

In an equilibrium with zero workplace segregation ($\sigma = 0$), firms must strictly prefer the inclusive strategy: i.e. $\pi^I > \pi^S$. Using (C33) and (C38), this implies:

$$\left(\frac{w_h^I}{w_h^S}\right)^{1+\varepsilon} > \frac{1+\phi\lambda^S}{1+\phi\lambda^I} \tag{C42}$$

But imposing $\sigma = 0$ on (C41), we have:

$$\lambda^I = \frac{n_l}{n_h} \tag{C43}$$

Intuitively, since all firms are inclusive (and offer the same wages), their skill ratio λ^I must equal the aggregate ratio. Using equations (C32), (C35) and (C43), we can then re-write (C42) as:

$$\left(\frac{\left[(1-\theta)\left(\frac{n_l}{n_h}\right)^{\gamma} + \theta\right]^{\frac{1}{\gamma}}}{\left[(1-\theta) + \theta\left(\lambda^S\right)^{-\gamma}\right]^{\frac{1-\gamma}{\gamma}}} \cdot \frac{\phi}{1-\theta} \cdot \frac{\varepsilon}{1+\varepsilon}\right)^{1+\varepsilon} > \left(1+\phi\lambda^S\right)\left(1+\phi\frac{n_l}{n_h}\right)^{\varepsilon} \tag{C44}$$

where the selective firms' skill ratio λ^S is pinned down by the exogenous parameters in equation (C37). Therefore, both sides of this inequality are functions of the exogenous parameters; and if the inequality is satisfied, the selective share σ will indeed equal zero in equilibrium. The inclusive and selective h-type wages, w_h^I and w_h^S, can then be pinned down by (C32) and (C35). And since the equity constraint binds, the l-type wages can be computed as $w_l^I = \phi w_h^I$ and $w_l^S = \phi w_h^S$ respectively.

Equilibrium with partial workplace segregation: $\sigma > 0$

In an equilibrium with partial workplace segregation ($\sigma > 0$), firms must be indifferent between the selective and inclusive strategies: i.e. $\pi^I = \pi^S$. Equating (C33) and (C38), this implies:

$$\left(\frac{w_h^I}{w_h^S}\right)^{1+\varepsilon} = \frac{1+\phi\lambda^S}{1+\phi\lambda^I} \tag{C45}$$

Equilibrium can then be characterized by the following five equations: (C32), (C35), (C37), (C41) and (C45). These five equations determine five unknowns: the selective firm share σ; the inclusive and selective skill ratios, λ^I and λ^S; and the inclusive and selective h-type wages, w_h^I and w_h^S. Since the equity constraint binds, the l-type wages can then be computed as $w_l^I = \phi w_h^I$ and $w_l^S = \phi w_h^S$ respectively.

E Extension with N skill types

E.1 Description of framework

In this appendix, we generalize the baseline model from two to N skill types. Firms choose wages and employment, for every skill type s, to maximize profit:

$$\max_{\{w_s;l_s\}_{s=1}^N} \pi\left(w_1, ..., w_N; l_1, ..., l_N\right) = \sum_{s=1}^{N} \left(p_s - w_s\right) l_s \tag{C46}$$

where skill s productivity p_s is increasing in s, and skill types are perfect substitutes. Firms are subject to labor supply constraints:

$$l_s \leq l_s\left(w_s\right) \tag{C47}$$

where the labor supply curves $l_s\left(w_s\right)$ are defined by (2), and to pay equity constraints:

$$w_s \geq \phi_s w_N \tag{C48}$$

for every skill type s. The Nth equation of (7) is of course redundant, but it is useful for notation to normalize ϕ_N to 1. Analogous to the baseline model, we can also define the "bite" β_s of each equity constraint as:

$$\beta_s = \phi_s \frac{p_N}{p_s} \tag{C49}$$

where $\beta_N = 1$. We assume that β_s is strictly decreasing in s, so the equity constraints bind for all skill types s (since $\beta_s > 1$ for $s < N$), and the bite is stronger for less productive workers. This is necessarily the case if there is perfect pay equity ($\phi_s = 1$ for all s), or more generally if wages are compressed (within firms) relative to productivity differentials.

E.2 Equilibrium strategies

As in the baseline model, since the equity constraints bind, wages will take log additive form:

$$\log w_{sf} = \eta_f + \lambda_s \tag{C50}$$

where firms choose a common firm effect η_f (equal to w_{Nf} in the model, for the top skill type), and the skill effect $\lambda_s = \log \phi_s$ represents the fixed internal pay differential (which firms take as given).

Consider a firm which offers N-type workers a wage of w_N (which determines the common firm effect). Given the equity constraint, the profit from employing an s-type worker is then equal to:

$$p_s - \phi_s w_N = \left(\frac{1}{\beta_s} - \frac{w_N}{p_N} \right) \phi_s p_N \tag{C51}$$

using equation (C49). Firms will employ all willing s-type workers if $p_s \geq \phi_s w_N$ (so the s-type labor supply constraint will bind), and will employ none if $p_s < \phi_s w_N$. But since the constraint bite β_s is decreasing in s (by assumption), equation (C51) implies that if a firm employs s-type workers, it must also employ all workers with skill exceeding s.

It follows that there are N possible strategies in equilibrium (one corresponding to each skill type), which we index z. Firms adopting strategy z employ all workers with skill $s \geq z$, and reject all workers with skill $s < z$. More formally, let w_s^z denote the optimal wage paid by strategy-z firms to s-type workers, and let l_s^z denote the optimal employment of z-type workers. The labor supply constraints bind, i.e. $l_s^z = l_s(w_s^z)$, for all skill types $s \geq z$. And optimal employment $l_s^z = 0$ for all skill types $s < z$. Strategy z is internally consistent if hiring workers with skill $s < z$ is unprofitable at the chosen wage, i.e., if the s-type wage $w_s^z = \phi_s w_N^z$ (as fixed by the equity constraint) exceeds their productivity p_s.

As in the baseline model, though firms in the baseline model are identical, they may choose different pay strategies in equilibrium. Let σ^k denote the equilibrium share of firms which choose strategy z. Since all firms must choose one of these N strategies, these shares must sum to 1:

$$\sum_z \sigma^z = 1 \tag{C52}$$

E.3 Optimal wage of strategy-z firm

Strategy-z firms do not employ workers with skill $s < z$, so they are not subject to the equity constraint for these workers. But the labor supply constraints will bind for all skill types $s \geq z$. We can then re-write the firm's problem in (C46) as:

$$\max_{w_N} \pi^z (w_N) = \sum_{s \geq z}^{N} (p_s - w_s) l_s \tag{C53}$$

The first-order condition is then:

$$\sum_{s \geq z} \phi_s (p_s - \phi_s w_N) l_s' (\phi_s w_N) = \sum_{s \geq z} \phi_s l_s (\phi_s w_N) \tag{C54}$$

Using the labor supply constraint (2), this implies:

$$w_N^z = \frac{\sum_{s \geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}}{\sum_{s \geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}} \cdot \frac{\varepsilon}{1+\varepsilon} p_N \tag{C55}$$

Finally, using (C53), optimal profit of strategy-k firms is:

$$\pi^z = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \cdot \frac{\left(\sum_{s \geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N} \right)^{1+\varepsilon}}{\left(\sum_{s \geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N} \right)^{\varepsilon}} \cdot \Omega_N p_N^{1+\varepsilon} \tag{C56}$$

E.4 Labor supply intercepts

To solve for equilibrium, we next require expressions for the labor supply intercepts Ω_s. Since s-type workers are only employed by firms with strategy $z \leq s$, equation (3) implies:

$$\Omega_s = \frac{n_s}{k} \left[\sum_{z \leq s} \sigma^z (\phi_s w_N^z)^\varepsilon \right]^{-1} \tag{C57}$$

Taking the ratio relative to the top skill types ($S = N$), and weighting by ϕ_s^ε, we have:

$$\frac{\phi_s^\varepsilon \Omega_s}{\Omega_N} = \frac{\alpha_s}{\alpha_N} \cdot \frac{\beta_s}{\phi_s} \cdot \frac{\sum_z \sigma^z \left(\frac{w_N^z}{w_N^N} \right)^\varepsilon}{\sum_{z \leq s} \sigma^z \left(\frac{w_N^z}{w_N^N} \right)^\varepsilon} \tag{C58}$$

where

$$\alpha_s \equiv \frac{n_s p_s}{\sum_x n_x p_x} \tag{C59}$$

is the output share of s-type workers.

Also, from (C55), notice the optimal wage of strategy-N firms is:

$$w_N^N = \frac{\varepsilon}{1+\varepsilon} p_N \tag{C60}$$

So the relative wage $\frac{w_N^z}{w_N^N}$ in equation (C58) is equal to:

$$\frac{w_N^z}{w_N^N} = \frac{\sum_{s\geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}}{\sum_{s\geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}} \tag{C61}$$

for strategy $z < N$.

E.5 Equilibrium

In equilibrium, as long as $\alpha_1 > 0$, at least some firms must opt for strategy 1: i.e., $\sigma^1 > 0$. This is because type-1 workers are only employed by strategy-1 firms, and these workers cannot be left unemployed in equilibrium. Otherwise, the profit from strategy 1 would exceed all others, so at least some firms must adopt this strategy (a contradiction).

For all other strategies z, there are two possibilities. Either no firms adopt strategy z, so we have:

$$\sigma^z = 0 \tag{C62}$$

This requires that strategy z is less profitable than strategy 1 (i.e. $\pi^z < \pi^1$). Or alternatively, at least some firms adopt strategy z (i.e. $\sigma^z > 0$), which requires that strategies z and 1 are equally profitable (i.e. $\pi^z = \pi^1$). From equation (C56), equal profits implies:

$$\frac{\left(\sum_{s\geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{1+\varepsilon}}{\left(\sum_{s\geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{\varepsilon}} = \frac{\left(\sum_{s} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{1+\varepsilon}}{\left(\sum_{s} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{\varepsilon}} \tag{C63}$$

In equilibrium, we then have $3N-2$ unknowns: (i) the strategy shares σ^z for $z = 1, ..., N$; (ii) the optimal wages $\frac{w_N^z}{w_N^N}$ for strategies $z = 1, ..., N-1$ (relative to the strategy-N wage); and (iii) the relative labor supply intercepts $\frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}$ for skill types $s = 1, ..., N-1$. And we also have $3N-2$ equations: (i) the relative intercept equations (C58) for strategies $z = 1, ..., N-1$; (ii) the relative wage equations (C61) for strategies $z = 1, ..., N-1$; (iii) equation (C52), which ensures the strategy shares sum to 1; and finally, (iv) we have one equilibrium condition for every strategy $z = 2, ..., N$: either (C62) or (C63).

F Extension with job search frictions

In the baseline model, we attribute wage-setting power to workers' idiosyncratic preferences over firms. But we can derive similar results from an alternative framework with search frictions. In what follows, we extend the on-the-job search model of Burdett and Mortensen (1998), by allowing for two skill types and imposing an internal equity constraint—just as in our baseline model. This framework is similar to Manning (1994), except we allow for internal pay differentiation between skill types: i.e., we permit $\phi < 1$, rather than imposing $\phi = 1$. The introduction of search frictions also delivers additional testable predictions for worker mobility over the job ladder—which we test empirically in Section 4.2.

The firm's problem is identical to the baseline model: see equations (4)-(7). Firms choose wages and employment of h- and l-types, subject to the two labor supply constraints and the internal equity constraint. Production is linear in each skill type, with marginal products equal to p_h and p_l. What is new here is the form of the labor supply functions, $l_s(w_s)$ for skill type $s = \{h, l\}$: these functions depend on the nature of the job search process.

F.1 Derivation of labor supply functions

We begin by deriving the form of $l_s(w_s)$, the supply of type-s workers to firms paying wage w_s. For simplicity, we assume that all workers (whether employed or unemployed) randomly meet with firms at rate λ. Workers can leave a job for two reasons: either to move to a higher-paying firm, or due to separation to unemployment (at exogenous rate δ). To keep the exposition as simple as possible, we assume that workers receive a zero utility flow when unemployed. Since the offer rate λ does not vary by employment status, unemployed workers will accept any positive wage offer.

The labor supply functions depend on the unemployment rate and distribution of realized wages (which together determine recruitment), as well as the distribution of wage offers (which determines retainment). Let $F_s(w_s)$ be the equilibrium distribution of wage offers to type-s workers. And let $G_s(w_s)$ be the equilibrium distribution of wages among type-s workers. In steady-state, $G_s(w_s)$ will depend on the offer distribution $F_s(w_s)$. To see how, consider the group of firms offering wages below w_s to type-s workers. The inflow of type-s workers to this group must equal the outflow in equilibrium:

$$u_s \lambda F_s(w_s) n_s = \delta(1-u_s) G_s(w_s) n_s + \lambda(1-F_s(w_s))(1-u_s) G_s(w_s) n_s \tag{C64}$$

where n_s is the aggregate measure of type-s workers, and u_s is their unemployment rate. The left-hand side of equation (C64) shows the type-s inflow to this group of firms, which is composed exclusively of the unemployed. And the outflow on the right-hand side, from this same group of firms, consists of two components: (i) workers separating to unemployment (at rate δ), and (ii) workers moving to firms which pay above w_s. Note that (C64) is only defined for wages w_s below the marginal product (i.e. $w_s \leq p_s$), as firms will never employ workers at a loss. This is a non-trivial consideration for l-types when the equity constraint binds, as will be clear below.

The steady-state equilibrium unemployment rate of type-s workers is:

$$u_s = \frac{\delta}{\delta + \lambda F_s\left(p_s\right)} \tag{C65}$$

The job finding rate (out of unemployment) is equal to $\lambda F_s\left(p_s\right)$, because only those firms which offer wages below p_s will employ type-s workers.

Substituting (C65) into the equation above and rearranging gives:

$$G_s\left(w_s\right) = \frac{\delta}{\delta + \lambda\left[1 - F\left(w_s\right)\right]} \cdot \frac{F_s\left(w_s\right)}{F_s\left(p_s\right)} \tag{C66}$$

for $w_s \leq p_s$.

We can now derive the labor supply function $l_s\left(w_s\right)$ itself. This too is pinned down by a steady-state condition (equating inflows and outflows), but this time at the firm level:

$$\frac{\lambda}{k} u_s n_s + \frac{\lambda}{k}\left(1 - u_s\right) G_s\left(w_s\right) n_s = \left[\delta + \lambda\left(1 - F_s\left(w_s\right)\right)\right] l_s\left(w_s\right) \tag{C67}$$

The left-hand side is the flow of type-s workers recruited to a firm paying w_s: the first term is the inflow from unemployment (divided between the measure k of firms), and the second term is the inflow from firms paying below w_s. The right-hand side of (C67) is the outflow of workers from this firm: workers can exit either through separation to unemployment (at rate δ), or by meeting a firm offering a wage exceeding w. Using (C65) and (C66), this steady-state condition implies:

$$l_s\left(w_s\right) = \frac{\delta + \lambda}{\delta + \lambda F_s\left(p_s\right)} \cdot \frac{\delta\lambda}{\left[\delta + \lambda\left(1 - F_s\left(w_s\right)\right)\right]^2} \cdot \frac{n_s}{k} \tag{C68}$$

which is the type-s labor supply function.

F.2 Equilibrium if equity constraint does not bind

If the equity constraint does not bind, firms will earn a positive profit on the marginal hire; so the labor supply constraints must bind (i.e. firms hire all willing workers), just as in

Burdett and Mortensen (1998). That is, for each skill type s and for all firms, we have $l_s = l_s(w_s)$. The firm's problem can therefore be simplified to:

$$\max_{w_h, w_l} \pi(w_h, w_l) = \pi_h(w_h) + \pi_l(w_l) \tag{C69}$$

where $\pi_s(w_s)$ is the profit earned from skill type $s = \{h, l\}$:

$$\pi_s(w_s) = (p_s - w_s)\, l_s(w_s) \tag{C70}$$

We therefore effectively have a distinct Burdett-Mortensen model for each skill type.

Equilibrium offer distribution

As is well known, the equilibrium offer distribution $F_s(w_s)$ has no discrete mass point; otherwise, firms at mass points could profit by offering infinitesimally larger wages (this would increase employment discretely, as these firms poach workers from their original mass point). Additionally, the lowest offer in $F_s(w_s)$ must be zero, the reservation wage of unemployment workers; otherwise, the lowest-paying firm could profit by reducing its offer to zero (at no cost to employment).

Since firms are identical, all wage offers on the support of $F_s(w_s)$ must yield equal profit in equilibrium. And since the lowest-paying firm offers a zero wage, this implies:

$$\pi_s(w_s) = \pi_s(0) \tag{C71}$$

for all offers w_s on the equilibrium support. Using equations (C68), (C69) and (C70), this can be written as:

$$\frac{n_s}{k} \cdot \frac{\delta\lambda(p_s - w_s)}{[\delta + \lambda(1 - F_s(w_s))]^2} = \frac{n_s}{k} \cdot \frac{\delta\lambda p_s}{(\delta + \lambda)^2} \tag{C72}$$

Rearranging then yields the unconstrained type-s offer distribution:

$$F_s^*(w_s) = \frac{\delta + \lambda}{\lambda}\left[1 - \left(\frac{p_s - w_s}{p_s}\right)^{\frac{1}{2}}\right] \tag{C73}$$

Skill wage premium and firm-worker sorting

For convenience, suppose that firms are ranked identically in their wage offers to h- and l-types. That is, for any given firm f, $F_h(w_{hf}) = F_l(w_{lf})$. This is not true in general, since firms are indifferent between all wages in the offer distribution. But this outcome can be ensured by a negligible amount of imperfect substitutability (between worker types) in production.[1]

Rearranging (C73), the type-s wage of the percentile F firm will then be:

$$w_s^*(F) = \left[1 - \left(1 - \frac{\lambda}{\delta + \lambda} F\right)^2\right] p_s \tag{C74}$$

And therefore, the skill wage differential will simply equal the productivity differential in all firms:

$$\frac{w_l^*(F)}{w_h^*(F)} = \frac{p_l}{p_h} \tag{C75}$$

Just as in the baseline model, it follows that log wages will be additively separable in firm and worker effects. Finally, notice that equations (C68) and (C74) imply:

$$\frac{l_l^*(F)}{l_h^*(F)} = \frac{n_l}{n_h} \tag{C76}$$

i.e. the relative supply of skill types in all firms is identical to the relative aggregate supply, and independent of firm percentile F; so there is no assortative matching. Intuitively, since the equity constraint does not bind, firms will earn a positive profit on all marginal recruits; so they optimally hire all willing workers (the labor supply constraints bind). And since workers only care about firm rank F, both skill types sort identically across the firm rank distribution.

In summary, the unconstrained equilibrium shares the same key features as the baseline model. Though there is now a non-degenerate distribution of wage offer (a consequence of on-the-job search and direct wage competition between firms), all firms offer the same skill premium, and recruit skill types in equal proportions.

F.3 Equilibrium if equity constraint binds

Suppose now that the equity constraint does bind, i.e. $\phi > \frac{p_l}{p_h}$. Firms are then compelled to compress their internal pay differential to ϕ; so $\frac{w_{lf}}{w_{hf}} = \phi$ in all firms f. Just as in the baseline model, firms will adopt one of two pay strategies:

1. **Inclusive strategy (I).** Inclusive firms hire all willing workers, so the labor supply constraints bind for both skill types: i.e., $l_h^I = l_h(w_h)$ and $l_l^I = l_l(w_l)$. To accommodate both types, firms compress pay internally to satisfy the equity constraint, redistributing wages between h- and l-types (relative to the unconstrained optimum). This strategy

[1] Intuitively, firms which pay higher wages to h-types recruit more of them, and therefore benefit disproportionately from hiring more l-types (and so will optimally compensate them more also).

is optimal if and only if the l-type offer does not exceed their productivity: i.e. as long as $w_l = \phi w_h \leq p_l$.

2. **Selective strategy (S).** Selective firms hire all willing h-type workers, so the h-type labor supply constraint binds: i.e., $l_h^S = l_h(w_h)$. But they ration l-type employment. This strategy is optimal if and only if the l-type offer exceeds their productivity: i.e. if $w_l = \phi w_h > p_l$. In this scenario, since technology is linear, firms will optimally reject all l-types: i.e. $l_l^S = 0$.

Since the l-type wage w_l is equal to ϕw_h in all firms (assuming the equity constraint binds), it is sufficient to derive the equilibrium distribution for h-type offers, which we denote F, i.e.:

$$F(w_h) \equiv F_h(w_h) = F_l(\phi w_h) \tag{C77}$$

Firms with h-type offers above the cutoff $\frac{1}{\phi}p_l$ will adopt the selective strategy (and only employ h-types); and firms offering below $\frac{1}{\phi}p_l$ will adopt the inclusive strategy (and hire both skill types). The selective share of firms is therefore equal to:

$$\sigma = 1 - F\left(\frac{1}{\phi}p_l\right) \tag{C78}$$

Just as in the unconstrained case, equilibrium requires that all wage offers in the support of F yield equal profit. And since the lowest offer must be zero (for the reasons above), it follows that:

$$\pi(w_h) = \pi(0) \tag{C79}$$

for all offers w_h on the equilibrium support. Given the cutoff for the selective strategy, this can be written as:

$$(p_h - w_h)\, l_h(w_h) + \max\{p_l - \phi w_h, 0\}\, l_l(w_l) = p_h l_h(0) + p_l l_l(0) \tag{C80}$$

Looking at the left-hand side, all firms recruit h-types; and their labor supply constraint always binds. But firms only recruit l-types (i.e., adopt the inclusive strategy) if the the l-type offer $w_l = \phi w_h$ is below their productivity p_l; hence the second term in (C80). On the right-hand side, firms offering a zero wage are necessarily inclusive, so the labor supply constraint binds for both skill types. Using (C68) and (C77) and rearranging, equation (C80) implies:

$$F(w_h) = \frac{\delta+\lambda}{\lambda}\left\{1 - \left[\frac{(p_h - w_h)\, n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)} \cdot \max\{p_l - \phi w_h, 0\}\, n_l}{p_h n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)} \cdot p_l n_l}\right]^{\frac{1}{2}}\right\} \tag{C81}$$

This expresses the equilibrium offer distribution F in terms of (i) the model's exogenous parameters and (ii) the selective share σ of firms. F is generally smooth, except for a kink at $\frac{p_l}{\phi}$.

To solve for the maximum wage offer w_h^{max}, we can replace $F(w_h)$ with 1 in equation (C81):

$$1 = \frac{\delta+\lambda}{\lambda}\left\{1 - \left[\frac{(p_h - w_h^{max})\,n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)}\cdot \max\{p_l - \phi w_h^{max}, 0\}\,n_l}{p_h n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)}\cdot p_l n_l}\right]^{\frac{1}{2}}\right\} \tag{C82}$$

As in the baseline model, equilibrium can take one of two forms: zero workplace segregation ($\sigma = 0$) or partial segregation ($\sigma > 0$). If the equilibrium σ is equal to zero, the maximum offer w_h^{max} implied by equation (C82) must be less than the cutoff $\frac{1}{\phi}p_l$. This would confirm that no firms benefit from adopting the selective strategy. If however the implied maximum offer w_h^{max} exceeds the cutoff $\frac{1}{\phi}p_l$ (after imposing $\sigma = 0$), this would indicate that firms would in fact profit from the selective strategy; so an equilibrium with $\sigma = 0$ is infeasible. To solve for σ in this case, we can then apply the expression in (C81) to equation (C78):

$$\sigma = 1 - \frac{\delta+\lambda}{\lambda}\left\{1 - \left[\frac{\left(p_h - \frac{p_l}{\phi}\right)n_h}{p_h n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)}\cdot p_l n_l}\right]^{\frac{1}{2}}\right\} \tag{C83}$$

which determines σ in terms of the exogenous parameters alone.

Implications for skill sorting and firm size premium

In equilibria with partial workplace segregation ($\sigma > 0$), equation (C81) shows that the offer distribution $F(w_h)$ is continuous—but kinked around the cutoff $w_h = \frac{p_l}{\phi}$. The kink is a consequence of the distribution of workers around the cutoff. Since discontinuously more workers are located just below (l-types are exclusively recruited below the cutoff), discontinuously more firms also offer wages just below to compete over these workers.

The key intuitions from the baseline model, in the presence of a binding equity constraint, can be gleaned from variation around this cutoff. First, h-type workers sort disproportionately into higher-paying firms. This is because (selective) firms above the cutoff exclusively hire h-type labor.

Second, (selective) firms just above the cutoff have lower employment than (inclusive) firms just below. Again, this is because selective firms ration l-type labor. The inclusive and selective strategies nevertheless yield equal profits, since selective firms compete more effectively for h-type workers (who deliver larger profit margins in equilibrium). This trade-off generates a locally negative firm size premium—at this point in the offer distribution.

Implications for job ladder

Above, we have shown that the central predictions of the baseline model are unaffected by the introduction of search frictions. But a search framework delivers additional testable predictions on job mobility. Just as in the standard Burdett-Mortensen model (Burdett and Mortensen, 1998), workers gradually work their way up a job ladder to ever higher-paying firms---but given search frictions, this process takes time and is subject to luck. However, in the presence of a binding equity constraint, this job ladder will be "taller" for h-type workers: i.e., h-types have exclusive access to higher-paying firms. This is because high-paying firms (above the cutoff $w_h = \frac{p_l}{\phi}$) adopt selective hiring strategies and deny employment to l-type workers. This implies heterogeneous patterns of job mobility (across the firm pay distribution) by skill type, and we test this claim empirically in Section 4.2 in the main text.

G Quantification of model's parameters

This appendix provides the technical details for quantifying the model parameters in Section 4.4. We implement this exercise in an extension with heterogeneous firms (as in Appendix C) and three skill types (a special case of Appendix E).

The three skill types correspond to non-graduates, non-STEM graduates and STEM graduates, and we denote them l, m and h, respectively. Assuming the equity constraint has stronger bite for l-types, i.e., $\beta_l > \beta_m$ (we will validate this assumption ex post), Appendix E shows that firms may pursue one of three strategies in equilibrium: (i) L-strategy: hire all willing workers, and pay wages w_s^L to s-type workers; (ii) M-strategy: hire only m- and h-type workers, and pay wages w_s^M to s-type workers; and (iii) H-strategy: hire only h-type workers, and pay them w_h^H. Let σ^L, σ^M and σ^H denote the equilibrium shares of L, M and H-strategy firms, where $\sigma^L + \sigma^M + \sigma^H = 1$.

As in Appendix C, the marginal product of s-type workers in firm f is equal to $p_{sf} = x_f p_s$, where $\log x_f$ is distributed normally across firms with mean 0 and variance ν.

G.1 Solution method: Step 1

To solve for the parameter values, we iterate over two steps. In the first step, for given labor supply elasticity ε, we solve for six parameters, using six moments and six equations. The six parameters are: $\frac{w_h^L}{w_h^H}$, $\frac{w_h^M}{w_h^H}$, $\frac{\phi_l^\varepsilon \Omega_l}{\Omega_h}$, $\frac{\phi_m^\varepsilon \Omega_m}{\Omega_h}$, σ^L, σ^M; and the six moments are: ϕ_m, ϕ_l, $\frac{n_m}{n_h}$, $\frac{n_l}{n_h}$, $E[\log w_m] - E[\log w_h]$, $E[\log w_l] - E[\log w_h]$.

We now set out the six equations. Recall from Appendix C that optimal wages (by strategy) and profits are log additive in firm productivity. It follows that the intercept ratios, i.e. $\frac{\Omega_l}{\Omega_h}$ and $\frac{\Omega_m}{\Omega_h}$, are independent of the firm productivity distribution; and the equilibrium strategy shares (i.e. σ^L, σ^M and σ^H) are orthogonal to firm productivity. We can therefore solve for $\frac{w_h^L}{w_h^H}$, $\frac{w_h^M}{w_h^H}$, $\frac{\phi_l^\varepsilon \Omega_l}{\Omega_h}$, $\frac{\phi_m^\varepsilon \Omega_m}{\Omega_h}$, σ^L and σ^M independently of the firm productivity distribution.

We have two equilibrium conditions for equal profits, which follow from equation (C63) in the N-type model. Equal profits for the L- and H-strategies implies:

$$\frac{w_h^L}{w_h^H} = \left(1 + \phi_m \frac{\phi_m^\varepsilon \Omega_m}{\Omega_h} + \phi_l \frac{\phi_l^\varepsilon \Omega_l}{\Omega_h}\right)^{-\frac{1}{1+\varepsilon}} \tag{D1}$$

and equal profits for the M- and H-strategies implies:

$$\frac{w_h^M}{w_h^H} = \left(1 + \phi_m \cdot \frac{\phi_m^\varepsilon \Omega_m}{\Omega_h}\right)^{-\frac{1}{1+\varepsilon}} \tag{D2}$$

Next, we have two equations for equilibrium ratios of the labor supply intercepts. From equation (C58) in the N-type model, these are:

$$\frac{\phi_l^\varepsilon \Omega_l}{\Omega_h} = \frac{n_l}{n_h} \cdot \frac{1 + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon} \tag{D3}$$

and

$$\frac{\phi_m^\varepsilon \Omega_m}{\Omega_h} = \frac{n_m}{n_h} \cdot \frac{1 + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon}{\frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon} \tag{D4}$$

Finally, we use two expressions for the expected log wages of l-types and m-types, expressed relative to h-types. These are:

$$E\left[\log w_l\right] - E\left[\log w_h\right] = \log \phi_l + \frac{\frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon \log \frac{w_h^L}{w_h^M} + \log \frac{w_h^L}{w_h^H}}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + 1} \tag{D5}$$

and

$$E\left[\log w_m\right] - E\left[\log w_h\right] = \log \phi_m + \frac{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}} - \frac{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} + 1} \tag{D6}$$

G.2 Solution method: Step 2

In the second step, we pick the firm productivity variance ν and the labor supply elasticity ε to match two additional moments: (i) the average elasticity of firm size with respect to AKM firm effects (denoted ε_{data}) and (ii) the variance of AKM firm effects (var_{AKM}). To estimate these moments in the model, we first simulate a panel of 1 million firms, drawing log firm productivity $\tilde{x}_f$ from a normal distribution with mean zero and variance ν.

For each simulated firm f, we compute employment and wages by education group, organize this data in "long" form (with each row corresponding to a firm $\times$ education group), and then estimate an AKM model by regressing log wages on firm and education fixed effects, and save the firm premia as η_f. We then regress log employment on the firm effects η_f:

$$\log l_f = \alpha + \gamma \eta_f + \epsilon_f \tag{D7}$$

The estimated coefficient γ provides our model-based moment for ε_{data}. The variance of the estimated firm effects η_f across all firms provides our model-based moment for var_{AKM}.

Following San (2023), we implement an iterative gradient descent procedure to find values of the firm productivity variance ν and labor supply elasticity ε that equate the model-based and empirical moments. The procedure updates parameters in each iteration according to the moments most affected by those parameters, based on the model's structure. Specifically, at each iteration i, we:

1. Compute model moments $m_i = (m_{i1}, m_{i2})$ for current parameter values $\theta_i = (\varepsilon_i, \nu_i)$.
2. Update parameters according to $\theta_{i+1} = \theta_i + \eta(m^* - m_i)$, where $m^* = (\varepsilon_{data}, var_{AKM})$ are the empirical target moments and η is the learning rate

The algorithm continues until the distance between model and empirical moments falls below a tolerance level τ: i.e., $\sum_j |m_{ij} - m_j^*| < \tau$. We set the learning rate $\eta = 0.1$ and tolerance $\tau = 10^{-3}$. At each iteration, we resolve the equilibrium equations from Step 1 given the updated ε.

The final estimated parameters imply a labor supply elasticity of $\varepsilon = 3.78$ and productivity variance of $\nu = 0.02$. With these values, the model successfully replicates both the average firm size-wage premium relationship ($\varepsilon_{data} = 3.61$ in both model and data) and the overall dispersion in firm wage premia ($var_{AKM} = 0.0355$ in both model and data).

H Derivation of counterfactual outcomes

In this appendix, we derive expressions for the impact of two counterfactuals in a model with three skill types $s = \{l, m, h\}$. We consider the removal of the equity constraint in Appendix H.1 and the prohibition of selective hiring strategies in Appendix H.2.

For convenience, we focus throughout on log outcomes. This allows us to abstract from heterogeneous firm productivity in the analysis: firm productivity enters through a log-linear intercept (as Section C shows, strategy choices are orthogonal to firm productivity), which is eliminated when computing differences between counterfactual and baseline outcomes.

H.1 Counterfactual with no equity constraint

Impact on expected log wages

In the counterfactual, all workers earn the unconstrained optimum wage, for $s = \{l, m, h\}$, in all firms. Denoting counterfactual outcomes with a $CF1$ superscript, wages for skill type s are therefore:

$$w_s^{CF1} = w_h^* = \frac{\varepsilon}{1+\varepsilon} p_s \tag{D8}$$

We now derive the impact on expected log wages for each skill type. Since h-types are employed by all firms in the baseline model, the counterfactual impact can be written as:

$$\begin{aligned} \log w_h^{CF1} - E\left[\log w_h\right] = \log w_h^{CF} & \\ & - \frac{\sigma^L l_h\left(w_h^L\right)\log w_h^L + \sigma^M l_h\left(w_h^M\right)\log w_h^M + \sigma^H l_h\left(w_h^H\right)\log w_h^H}{\sigma^L l_h\left(w_h^L\right) + \sigma^M l_h\left(w_h^M\right) + \sigma^H l_h\left(w_h^H\right)} \\ = & - \frac{\sigma^L\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon \log\frac{w_h^L}{w_h^H} + \sigma^M\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon \log\frac{w_h^M}{w_h^H}}{\sigma^L\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon + \sigma^M\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \sigma^H} \end{aligned} \tag{D9}$$

where σ^L, σ^M and σ^H are the shares of L, M and H-strategy firms respectively (using the notation of Appendix E). The second line uses the labor supply function in (2), and also the fact the H-strategy wage w_h^H is equal to the unconstrained optimum w_h^* (since these firms hire only h-types).

The impact on m-types is:

$$\begin{aligned}\log w_m^{CF1} - E[\log w_m] &= \left(\log w_m^{CF1} - \log w_h^{CF1}\right) + \left(\log w_h^{CF1} - E[\log w_h]\right) \\ &\quad + \left(E[\log w_h] - E[\log w_m]\right) \\ &= -\log \beta_m - \frac{\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}}\end{aligned} \tag{D10}$$

where the second line uses the definition of β_s in (C49), the expected skill differential in (D6), and the counterfactual impact in (D9).

Finally, the impact on l-types is:

$$\begin{aligned}\log w_l^{CF1} - E[\log w_l] &= \left(\log w_l^{CF1} - \log w_h^{CF1}\right) + \left(\log w_h^{CF1} - E[\log w_h]\right) \\ &\quad + \left(E[\log w_h] - E[\log w_l]\right) \\ &= -\log \beta_l - \log \frac{w_h^L}{w_h^H}\end{aligned} \tag{D11}$$

where the second line uses the expected skill differential in (D5) and the counterfactual impact in (D9).

Impact on expected utility

We now turn to the effect on expected utility. Note we weight utility by $\frac{1}{\varepsilon}$ for this exercise, to ensure it is in log wage units: see equation (1). Using equation (B27), the impact on h-type utility can be written as:

$$\begin{aligned}\frac{1}{\varepsilon}\left(\bar{u}_h^{CF1} - \bar{u}_h\right) &= \frac{1}{\varepsilon} \log \frac{\left(w_h^{CF1}\right)^{\varepsilon}}{\sigma_l \left(w_h^L\right)^{\varepsilon} + \sigma_m \left(w_h^M\right)^{\varepsilon} + \sigma_h \left(w_h^H\right)^{\varepsilon}} \\ &= -\frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} + \sigma_h\right]\end{aligned} \tag{D12}$$

which again uses the equality between w_h^{CF1} and w_h^H. For m-types, the impact is:

$$\begin{aligned}\frac{1}{\varepsilon}\left(\bar{u}_m^{CF1} - \bar{u}_m\right) &= \frac{1}{\varepsilon} \log \frac{\left(w_m^{CF1}\right)^{\varepsilon}}{\sigma_l \phi_m^{\varepsilon} \left(w_h^L\right)^{\varepsilon} + \sigma_m \phi_m^{\varepsilon} \left(w_h^M\right)^{\varepsilon}} \\ &= -\log \beta_m - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}\right]\end{aligned} \tag{D13}$$

And for l-types:

$$\frac{1}{\varepsilon}\left(\bar{u}_l^{CF1}-\bar{u}_l\right)=-\log\beta_l-\frac{1}{\varepsilon}\log\left[\sigma_l\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon}\right] \tag{D14}$$

The impact on expected amenities (weighted by $\frac{1}{\varepsilon}$) is simply the difference between the expected utility and log wage effects.

H.2 Counterfactual with no selective strategy

Effects on expected log wages

In this counterfactual, all firms adopt the inclusive L-strategy, and employ all workers who are willing to work: i.e., the labor supply constraints always bind. Building from the N-types case in equation (C61), the optimal L-strategy wage (for h-type workers) can be written as:

$$w_h^{CF2}=w_h^L=\frac{1+\frac{\phi_m}{\beta_m}\cdot\frac{\phi_m^{\varepsilon}\Omega_m}{\Omega_h}+\frac{\phi_l}{\beta_l}\cdot\frac{\phi_l^{\varepsilon}\Omega_l}{\Omega_h}}{1+\phi_m\frac{\phi_m^{\varepsilon}\Omega_m}{\Omega_h}+\phi_l\frac{\phi_l^{\varepsilon}\Omega_l}{\Omega_h}}\cdot\frac{\varepsilon}{1+\varepsilon}p_h \tag{D15}$$

Since all firms adopt the same strategy (and pay the same wage), the labor supply intercept ratios collapse to the aggregate employment ratios:

$$\frac{\phi_s^{\varepsilon}\Omega_s}{\Omega_h}=\frac{n_s}{n_h} \tag{D16}$$

for $s=\{l,m\}$. More formally, this can be seen from equation (C57) in the N-type model. Imposing this on the equation above, we have:

$$w_h^{CF2}=\frac{1+\frac{\phi_m}{\beta_m}\cdot\frac{n_m}{n_h}+\frac{\phi_l}{\beta_l}\cdot\frac{n_l}{n_h}}{1+\phi_m\cdot\frac{n_m}{n_h}+\phi_l\cdot\frac{n_l}{n_h}}\cdot\frac{\varepsilon}{1+\varepsilon}p_h \tag{D17}$$

Imposing the equity constraints, m-types receive $\phi_m w_h^{CF2}$ and l-types receive $\phi_l w_h^{CF2}$.

Building from (D9), the impact on the expected log h-type wage is:

$$\log w_h^{CF2}-E\left[\log w_h\right]=\log\frac{1+\frac{\phi_m}{\beta_m}\cdot\frac{n_m}{n_h}+\frac{\phi_l}{\beta_l}\cdot\frac{n_l}{n_h}}{1+\phi_m\cdot\frac{n_m}{n_h}+\phi_l\cdot\frac{n_l}{n_h}}-\frac{\sigma_l\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon}\log\frac{w_h^L}{w_h^H}+\sigma_m\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}\log\frac{w_h^M}{w_h^H}}{\sigma_l\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon}+\sigma_m\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}+\sigma_h} \tag{D18}$$

For m-types, we have:

$$\log w_m^{CF2} - E\left[\log w_m\right] = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \frac{\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}} \tag{D19}$$

And for l-types:

$$\log w_l^{CF2} - E\left[\log w_l\right] = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \log \frac{w_h^L}{w_h^H} \tag{D20}$$

Impact on expected utility

We now turn to the effect on expected utility. Note we weight utility by $\frac{1}{\varepsilon}$ for this exercise, to ensure it is in log wage units: see equation (1). Building from (D12), the impact on expected h-type utility is:

$$\frac{1}{\varepsilon}\left(\bar{u}_h^{CF2} - \bar{u}_h\right) = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} + \sigma_h\right] \tag{D21}$$

For m-types, we have:

$$\frac{1}{\varepsilon}\left(\bar{u}_m^{CF2} - \bar{u}_m\right) = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}\right] \tag{D22}$$

And for l-types:

$$\frac{1}{\varepsilon}\left(\bar{u}_l^{CF2} - \bar{u}_l\right) = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon}\right] \tag{D23}$$

As before, the impact on expected amenities (weighted by $\frac{1}{\varepsilon}$) is the difference between the expected utility and log wage effects.

I Alternative models

In this appendix, we describe three alternative models that we compare to our baseline equity constraint model. For each model, we explain the key differences and how we calibrate the parameters.

I.1 Model 2: Skill-neutral firm heterogeneity

Our first alternative specification removes the equity constraint and assumes firms differ only in skill-neutral productivity.

Since there is no equity constraint, firms pay the unconstrained optimum to each skill type $s = \{h, m, l\}$, in line with equation (10). For skill s, the optimal wage is:

$$w_{sf} = \frac{\varepsilon}{1+\varepsilon} p_{sf} \tag{D24}$$

where p_{sf} is the marginal product of s-type workers in firm f:

$$p_{sf} = x_f p_s \tag{D25}$$

where x_f is distributed log-normally across firms, with mean 0 and variance ν; and p_s represents base productivity for skill type s.

Labor supply of skill type s to a firm paying wage w is:

$$l_s(w) = \Omega_s w^{\varepsilon} \tag{D26}$$

and the labor supply intercept is:

$$\Omega_s = \left(\sum_f w_{sf}^{\varepsilon} \right)^{-1} n_s \tag{D27}$$

Expressing outcomes relative to h-types, the model can be summarized by six parameters. First, the labor supply elasticity ε determines workers' responsiveness to wage differences across firms. Second, the variance of firm productivity ν governs pay dispersion across firms. The next two are the base productivity differentials, $\log(p_m/p_h)$ and $\log(p_l/p_h)$, which capture skill-specific differences in worker productivity. And finally, the relative labor supply intercepts, Ω_m/Ω_h and Ω_l/Ω_h, reflect the relative availability of each worker type adjusted for their outside options.

We calibrate these parameters to match six empirical moments. The first two are the average elasticity of firm size with respect to AKM firm premia ($\varepsilon_{data} = 3.61$) and the variance of AKM firm effects ($var_{AKM} = 0.0355$): these help identify the labor supply elasticity ε and productivity variance ν. The next two moments are the mean log wage differentials between education groups: between non-STEM and STEM graduates, i.e. $E[\log w_m] - E[\log w_h] = -0.67$, and between non-graduates and STEM graduates, i.e. $E[\log w_l] - E[\log w_h] = -0.47$. The final two moments are the aggregate employment ratios: the ratio of non-STEM to

STEM graduate employment, i.e., $n_m/n_h = 3.81$, and the ratio of non-graduate to STEM graduate employment, i.e., $n_l/n_h = 6.85$.

In all three alternative models, we follow a similar two-step estimation procedure to the baseline model. In Step 1, for a given labor supply elasticity ε, we solve for the other parameters to match the wage differentials and aggregate employment ratios. In Step 2, we update ε and ν based on the firm size-wage premia elasticity and AKM variance moments. The calibrated parameters are reported in Table A2.

I.2 Model 3: Skill-biased firm heterogeneity

Our second alternative model allows for skill-biased productivity differences across firms. The marginal product of s-type workers in firm f is:

$$p_{sf} = x_f^{\theta_s} p_s \tag{D28}$$

where the θ_s are skill-specific productivity elasticities with respect to firm heterogeneity x_f. We normalize $\theta_h = 1$ and estimate θ_m and θ_l. The rest of the model is identical to the previous model.

The model can be characterized by eight parameters: the labor supply elasticity ε, the variance of firm productivity ν, the skill-specific productivity elasticities θ_m and θ_l, the base productivity differentials $\log(p_m/p_h)$ and $\log(p_l/p_h)$, and the relative labor supply intercepts, Ω_m/Ω_h and Ω_l/Ω_h.

We calibrate these parameters to match the same eight moments as in the baseline model: the average elasticity of firm size with respect to AKM firm effects (ε_{data}), the variance of AKM firm effects (var_{AKM}), the mean firm AKM effects by education group relative to STEM graduates (AKM_{fm} and AKM_{fl}), the log wage differentials between education groups, i.e. $E[\log w_m] - E[\log w_h]$ and $E[\log w_l] - E[\log w_h]$, and the aggregate employment ratios, i.e. n_m/n_h and n_l/n_h.

I.3 Model 4: Skill-varying labor supply elasticities

The final model imposes skill-neutral heterogeneity in firm productivity, but permits the labor supply elasticity to vary by skill group. The utility of worker i of skill type s in firm f now takes the form:

$$u_{isf} = \varepsilon_s \log w_{sf} + a_{if} \tag{D29}$$

where ε_s is the skill-specific labor supply elasticity. Like the previous model, this model also has eight parameters: the base labor supply elasticity ε_h, the elasticity differentials $\varepsilon_m - \varepsilon_h$

and $\varepsilon_l - \varepsilon_h$, the variance of firm productivity ν, the base productivity differentials, $\log(p_m/p_h)$ and $\log(p_l/p_h)$, and the labor supply intercepts, Ω_m and Ω_l.

We calibrate these parameters to match the same eight moments as Model 3 (and the baseline model): the elasticity of firm size with respect to AKM effects (ε_{data}), the variance of AKM firm effects (var_{AKM}), the mean firm AKM effects by education group relative to STEM graduates (AKM_{fm} and AKM_{fl}), the log wage differentials between education groups, i.e. $E[\log w_m] - E[\log w_h]$ and $E[\log w_l] - E[\log w_h]$, and the aggregate employment ratios, i.e. n_m/n_h and n_l/n_h.

I.4 Estimation results

For each model, we implement an iterative procedure similar to our baseline model, using gradient descent with step size $\eta = 0.1$ and convergence tolerance of 10^{-3}. The calibrated parameters for all model variants are presented in Table A2. Model 2 yields a labor supply elasticity of 3.61 and sizable productivity gaps across education groups. Model 3 generates substantial skill-biased productivity differences across firms, with larger productivity heterogeneity for high-skilled workers. Model 4 produces considerable heterogeneity in labor supply elasticities across skill groups, with high-skilled workers being the most responsive to wage differences.

J Quantitative validation of regional outcomes

In Section 5.2, we use our nationally calibrated model to predict the impact of observable regional variation in skill shares. For each of the 49 regions, and for both the 1995 and 2008 census years, we use observed employment ratios of non-graduates to STEM graduates ($\frac{n_l}{n_h}$) and non-STEM graduates to STEM graduates ($\frac{n_m}{n_h}$) as model inputs. And for all remaining parameter values, we rely on our national-level calibration for the 2000-2009 interval: see Table A3. Separately for each of the 98 region-year pairs, we then solve for the equilibrium shares of firms adopting each strategy (σ^L, σ^M, σ^H) and the corresponding skill differentials in firm pay premia—as a measure of workplace segregation.

As before, we rely on the three-type extension to the model. In equilibrium, there are three possible equilibrium pay strategies: L (hire all skill types), M (hire only m- and h-types), and H (hire only h-types). But not all strategies are necessarily active in equilibrium, and there are four possible equilibrium configurations: only the L-strategy is active; only L and M are active; only L and H; all three strategies are active.

The solution algorithm involves systematically checking all four configurations (separately for each region-year pair), using the equilibrium equations of Appendix E. For any given set of active strategies, we solve the system of equations consisting of: (i) the relative labor supply intercept equations (C58) for each skill type; (ii) the relative wage equations (C61) derived from first-order conditions for each active strategy; and (iii) the equal profit conditions (C63) for strategies with positive shares. A particular configuration of strategies constitutes a valid equilibrium if: strategy shares are non-negative and sum to one for active strategies; the equal profit condition holds for all active strategies; and inactive strategies (those with $\sigma^S = 0$) yield profits below those of active strategies. For all 98 region-year pairs, there is exactly one equilibrium configuration that satisfies all conditions.

In Table A7, we report the shares of regions with each equilibrium configuration, separately by census year. Between 1995 and 2008, we see a large reduction in the number of regions populated exclusively by L-strategy firms, consistent with Figure 12.

K Preparation of Israeli administrative data

This appendix provides additional details on the data preparation and definitions of key variables.

Wages: Our raw data contains observations at the worker $\times$ firm $\times$ year level, with monthly employment indicators and total annual compensation for each employment spell. We implement several data cleaning procedures to ensure accurate wage measurements: (i) removing observations with missing worker or firm identifiers, (ii) standardizing the treatment of monthly indicators by replacing missing values with zeros, (iii) eliminating exact duplicates across all variables, and (iv) for cases where worker-firm combinations appear multiple times within a year, consolidating by taking the maximum value of monthly indicators and summing the annual earnings.

From this cleaned dataset, we construct an annual panel by assigning individuals to the firm where they worked during November. For each worker-firm match, we impute a monthly salary by dividing total annual earnings by the number of months employed at that firm. In cases where workers had multiple employers in November, we assign the worker to the firm paying the higher monthly salary.

To focus on workers with substantial labor market attachment, we exclude worker-year observations with monthly earnings below 25% of the national average wage that year.[2] And to reduce the influence of possible spurious outliers, we also exclude the top 1% of wage

[2]For context, the statutory minimum wage in Israel ranged between 40-50% of the average wage during our sample period, reaching 48.8% in 2015. Our threshold therefore excludes workers earning approximately half the minimum wage or less, likely representing part-time or marginal employment.

observations within each year and education group (non-graduate, non-STEM graduate, STEM graduate) from our earnings sample. When exploiting regional variation, we exclude the top 1% within region-education-year cells. Our final sample spans 1990-2019 and includes workers aged 25-64 in each year.

Education: We use the Central Bureau of Statistics (CBS) education registry to classify workers into three mutually exclusive and time-invariant education categories, based on the highest degree they obtained during our sample period. These categories are: (i) non-graduate (no BA-equivalent or higher qualification), (ii) non-STEM graduate (BA-equivalent or higher degree in a non-STEM field), and (iii) STEM graduate (BA-equivalent or higher degree in a STEM field).

Workplace location: We rely on workplace geographical identifiers from 20% samples of the Israeli census conducted in 1995 and 2008, which we merge into the main employment records. We aggregate these identifiers into 49 regional units based on Israel's "natural regions", as defined by the Central Bureau of Statistics. These natural regions are constructed to ensure demographic, economic, and social homogeneity of the constituent populations. To deliver sufficient statistical power for all analyses, we incorporated the three smallest regions into their neighboring regions.

Industry: We use a consistent 2-digit industry classification for each firm across the entire sample period.

L Replication using Veneto Worker History dataset

This appendix reproduces the relationship between log firm size and AKM wage premia using the Veneto Worker History (VWH) dataset, which contains detailed employer-employee linked administrative records for Italy's Veneto region over 1975-2001. The data cover the universe of private sector employment in the region, with particularly good coverage of small establishments that characterize the region's manufacturing sector. We estimate AKM firm effects using log annual earnings for years between 1992 and 2001, and implement the same split-sample approach as in our main analysis to address measurement error.

We plot the relationship in Figure A2, across 20 firm bins ranked by their AKM wage premia. As in the Israeli data, we again see a hump-shaped relationship between firm size and firm wage premia, with employment initially increasing and then decreasing with the wage premium—both for the aggregate data and after residualizing by industry. These results build on previous work by Kline (2024), who highlights non-monotonicities in the reverse relationship (from firm size to pay) in this same data.

This evidence suggests that the quantity-quality trade-off is a more general phenomenon, arising from fundamental constraints on firms' wage-setting, rather than from country-specific institutions or policies.