ROCKWOOL Foundation Berlin
Institute for the Economy and the Future of Work (RFBerlin)

DISCUSSION PAPER SERIES 052/26

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

Michael Amior, Shmuel San

www.rfberlin.com FEBRUARY 2026

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

Authors

Michael Amior, Shmuel San

Reference

JEL Codes: J31, J42

Keywords: Wage-setting, monopsony, internal pay equity, workplace segregation, firm size

Recommended Citation: Michael Amior, Shmuel San (2026): Internal Pay Equity and the Quantity-Quality Trade-Off in Hiring. RFBerlin Discussion Paper No. 052/26

Access

Disclaimer

Imprint

RFBerlin
ROCKWOOL Foundation Berlin –
Institute for the Economy
and the Future of Work

Gormannstrasse 22, 10119 Berlin
Tel: +49 (0) 151 143 444 67
E-mail: info@rfberlin.com
Web: www.rfberlin.com

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

Michael Amior[†] and Shmuel San[‡]

February 15, 2026

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 results in workplace segregation and pay dispersion even among ex-ante identical firms. 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 provides new insights into aggregate-level, regional and sectoral variation in earnings inequality—which we explore using Israeli administrative data.

*We are grateful to Jaime Arellano-Bover, Damien Babet, Momi Dahan, Oren Danieli, Olivier Godechot, Simon Jäger, Pat Kline, Alex Mas and Elio Nimier-David (discussant) for helpful comments, and to seminar participants at Ben Gurion, Tel Aviv, Hebrew University, Yale, NYU, Princeton, the NLS-E 2025 conference, the Israeli Economic Association conference, King's College London, LSE and Columbia. We also thank Matan Goldman and Ariel Rechnitz for excellent research assistance. This study was performed in the Israel Central Bureau of Statistics research room using de-identified microdata (direct identifiers removed from data) in files prepared specifically for this project. We thank 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, King's College London; Hebrew University of Jerusalem; Centre for Economic Performance. Email: **michael.amior@kcl.ac.uk**.

‡Hebrew University of Jerusalem; Centre for Economic Policy Research. 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; Machin and Manning, 2004; 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 group morale, effort and retention (Card et al., 2012; Breza et al., 2018; Dube et al., 2019; Cullen, 2024). In this paper, we show how internal equity constraints can generate a quantity-quality trade-off in hiring in equilibrium—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 pay dispersion, even among ex-ante identical firms.

Our framework provides a novel interpretation of the AKM model (Abowd et al., 1999), which specifies log wages in terms of additive firm and worker effects. The worker effects represent internal pay differentials, which are fixed by the equity constraint. When the equity constraint binds, firms are compelled 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: i.e., the firm effects. These firm effects reflect not only variation in firm productivity but also in their hiring strategies—as they trade off quantity against quality. 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 simple conceptual basis.

Crucially, our interpretation of the AKM model allows us to reconcile log additive wages with the heavy sorting of high-skilled workers to high-paying firms (see e.g., Card et al. 2018). As Bonhomme et al. (2019) and Kline (2025) emphasize, it is difficult to rationalize log additivity if this sorting is driven by worker-firm complementarities in production. But in our model, firms use higher pay specifically to improve hiring *quality*, even at the cost of lower *quantity* (in contrast to conventional monopsony 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) heterogeneity in firm productivity, our model implies a concave or even hump-shaped relationship between firm size and 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; Bloesch and Larsen, 2023).

The equity constraint also offers a new rationale for equilibrium dispersion in firm pay. The AKM literature suggests that firms can recruit otherwise identical workers at very different wages—but why should workers ever accept offers from low-paying firms? One answer is idiosyncratic job preferences: i.e., workers are *uninterested* in alternative jobs which pay more. But as Card (2022) notes, this makes it difficult to explain evidence on turnover, job ladders and sluggish recovery from job displacement. A second answer is search frictions: search models embed a role for *luck* in wage determination, which can explain these phenomena better, but presume that workers are *unaware* of more lucrative outside options (an assumption which has provoked some debate: see Jäger et al. 2024; Caldwell et al. 2025). Our insight is 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, our model can synthesize competing explanations for 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 of workers with different education (Dillon et al., 2025), occupational rank (Babet et al., 2025) and AKM fixed effects (Song et al., 2019).[1] We argue that these trends

[1]Beyond its contribution to inequality, this growing workplace segregation has long-term implications for economic mobility which are not captured by our model (Hellerstein and Neumark, 2008). Firms' adoption

may themselves be partly attributable to technical change: in the presence of a binding equity constraint, skill-biased productivity growth makes the selective strategy more tempting for firms. And as selective 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) and Bergeaud et al. (2025) explicitly connect it with technical change; but again, we argue that outsourcing is one manifestation of a broader phenomenon.

As we argue above, technical change is one potential driving force behind these changes. But our model also shows that a growing *supply* of skills will encourage more firms to adopt the selective strategy—and thereby contribute to larger workplace segregation and earnings inequality. 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.

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, driven mostly by STEM graduates—coinciding with a rapid increase in the wage returns to STEM 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 profitability 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

of selective hiring strategies may deny low-skilled workers access to human capital externalities (Barza et al., 2024; Dillon et al., 2025) and job networks (San, 2023) associated with high-skilled peers.

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 the Veneto Worker History file (in Northern Italy), as used by Kline (2024).

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) with skill-neutral firm heterogeneity. Despite its simplicity, our model can match the key results surprisingly well: (i) the hump-shaped size-pay relationship, (ii) positive skill sorting and (iii) log additive wages. According to our estimates, the equity constraint compresses the STEM degree return by 0.52 log points (within firms) relative to the productivity differential, and the non-STEM return by 0.18 log points. This compression compels firms to adopt diverse pay and hiring strategies in equilibrium, and this in turn explains why the variance of AKM firm effects (0.032) greatly exceeds the variance of firm productivity (0.023).

We then compare our model's performance against three alternatives. First, a model with skill-neutral firm heterogeneity but no equity constraint can generate log additive wages, but fails to produce positive sorting or the hump-shaped size-pay relationship. Second, firm-worker complementarities in production can generate positive sorting; but if there is no equity constraint, this violates log additivity by introducing match effects in wages. Third, skill-varying labor supply elasticities can rationalize positive sorting (while preserving log additivity), but cannot explain the non-monotonic relationship between firm size and pay. Only our equity constraint framework can deliver all three empirical regularities.

We also use our model to assess the distributional implications of removing the equity constraint (and implicitly, the equity concerns which underpin it). This brings improvements in amenity match quality, as the low-skilled can now access the full set of firms.[2] But it also exacerbates inequality: STEM graduate welfare grows by 21%, while non-graduate welfare falls by 10%. However, a policy which prohibits the selective hiring strategy (akin to mandating uniform pay policies across firms) brings both greater equity *and* match quality.

We then explore market-level variation in firms' pay and recruitment strategies, over time, regions and sectors. First, at the aggregate level: given the growth in the relative

[2] In an alternative framework with job search frictions (instead of idiosyncratic amenity matches), this would manifest in a reduction in low-skilled unemployment: see Appendix F.

supply and productivity of STEM workers, 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, exploiting spatial variation, we show that regions with larger graduate employment shares (and larger skill expansions over time) exhibit greater firm pay dispersion and workplace 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 explore similar variation in Germany and the US; but we offer a new conceptual interpretation. As an out-of-sample validation, we use our nationally calibrated model to predict the impact of regional skill shares on these outcomes—and compare these effects to the data. Despite its parsimony, the model performs surprisingly well in both the fit and magnitude of the effects.

Our baseline sample is restricted to the private sector. But in our final exercise, we compare outcomes to the public sector. In many countries, the 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 within *individual* private sector firms are similar to the public sector—suggesting that equity constraints bind similarly tightly within private firms. Instead, what distinguishes the private sector is its fragmentation into many independent firms: this facilitates larger returns at the *aggregate* level, as firms adopt differential pay strategies, and high-skilled workers sort into high-paying firms. In this way, the public sector offers a "control" environment where administrative units cannot adopt independent pay strategies: this removes the key mechanism of our model.

Related literature. Our model's ability to explain numerous labor market phenomena provides *indirect* evidence for a binding equity constraint. But it is also consistent with how managers describe wage-setting in practice. In interviews conducted by Bewley (1999) and Galuscak et al. (2012), in both the US and Europe, managers report considerable flexibility in choosing the general pay level of a firm (as captured by the AKM firm effect in our model), but much less latitude in the choice of internal differentials (i.e., the relative worker effects). They typically attribute this to employees interacting significantly more with their co-workers (across the organizational hierarchy) than with their peers outside. Questions of fairness are therefore more salient for internal than external comparisons: external pay differentials do not generate the same emotional impact, and matter 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. Our argument is that this quality motive is a direct consequence of internal pay compression.

Crucially, as Weil (2014) emphasizes, workers care not only about pay inequity *within* job/skill categories, but also about pay differentials *between* them. Moreover, they tend to be more sensitive to wage gaps relative to *higher*-paid jobs than to those below them: Akerlof and Yellen (1990), Romer (1992) and Weil (2014) cite numerous experimental and field 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 rigid internal pay differentials.

A vertical equity constraint (bridging different skill levels) also offers a natural interpretation of domestic outsourcing, whose prevalence has grown in recent years. As Weil (2014) argues, outsourcing allows firms to escape the constraint by institutionally separating high and low-skilled employees: Drenik et al. (2023) estimate that high-paying firms only share half their pay premia with outsourced labor. Goldschmidt and Schmieder (2017), Daruich et al. (2024) and Bergeaud et al. (2025) show that low-skilled outsourced workers suffer large wage losses; and revealingly, Deibler (2022) finds large wage *gains* for workers who remain.[3] 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 particular tasks) or through the adoption of alternative production processes.

We also contribute to a growing body of work which documents constraints on internal pay differentiation. The evidence suggests that firms cannot perfectly discriminate on workers' outside options (Caldwell and Harmon, 2019; Lachowska et al., 2022; Di Addario et al., 2023; Jäger et al., 2023), on their region of employment (Hazell et al., 2022), on their age (Saez et al., 2019; Giupponi and Machin, 2022), nor between natives and migrants (Amior and Manning, 2020; Amior and Stuhler, 2023; Arellano-Bover and San, 2023).[4] In this paper,

[3] Analogously, Bergeaud et al. (2025) show that *high-skilled* workers (specifically, IT specialists and consultants) benefit from moving from large inclusive employers to specialized outsourcing firms.

[4] Our model is closely related to Amior and Stuhler (2023): that paper shows how constrained pay differ-

we apply this same logic to internal pay compression between skill groups: we show that this induces a novel quantity-quality trade-off in hiring, which can help account for numerous labor market phenomena.

We are not the first to explore the equilibrium implications of this kind of equity constraint: Romer (1984) and Akerlof and Yellen (1990) show how it can generate workplace segregation and unemployment of low-skilled workers. Our key 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 low pay—which is crucial to sustaining a quantity-quality trade-off in equilibrium. Our model is closer to Manning (1994), who introduces heterogeneous workers to an equilibrium search model where firms can only pay a single wage. We partially relax this constraint, to allow for a limited degree of internal pay differentiation: this generates a log additive (AKM) wage structure, with distinct firm and worker effects. We then show how this framework delivers a quantity-quality trade-off in hiring, which can help explain several empirical regularities in the literature. Saez et al. (2019) have also combined an internal equity constraint with wage-setting power, to explore pay differentials by age; but we allow for equilibrium wage competition between firms, which drives many of our results.[5]

Finally, our model builds on an older literature on dual labor markets: see e.g., Doeringer and Piore (1971), Gordon et al. (1982) and Bulow and Summers (1986), 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 primary sector offers superior conditions to productively similar workers (in contrast to the human capital paradigm) but rations jobs—just like our "selective" firms. The literature offers numerous hypotheses to explain this segmentation, such as divide-and-rule management, efficiency wages, and poor information on outside options; whereas 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, spatial and sectoral variation; and we conclude in Section 6.

entiation between natives and migrants (who differ in reservation wages) can generate workplace segregation and pay dispersion between ex-ante identical firms. We apply this same idea to skill groups.

[5]In earlier work, Romer (1992) 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. Frank (1984) offers 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.

2 Equilibrium wage-setting model

The economy consists of a continuum of firms (of measure k) and workers (measure n), who are either high or low-skilled. We initially assume that 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. As we will show, this constraint generates a trade-off between workforce quantity and quality, which can help shed new light on numerous labor market phenomena.

Though the model is purposefully simple, it is straightforward to extend. Appendix C incorporates heterogeneous firm productivity, Appendix D applies CES technology over skill inputs, Appendix E extends to N skill types, and Appendix F explores an alternative environment with job search frictions (instead of idiosyncratic preferences), building on Burdett and Mortensen (1998) and Manning (1994). These extensions provide valuable additional insights (which we flag where relevant), but the key results and intuitions are preserved.

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:

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

Though Ω_s is an equilibrium object, firms take it as given if there are many firms.

We now turn to production. Like Card et al. (2018), we assume that h- and l-types are perfect substitutes, but differ in productivity: h-types produce p_h and l-types 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:

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

subject to labor supply constraints:

$$l_h \leq l_h(w_h), \qquad l_l \leq l_l(w_l) \tag{5}$$

and a pay equity constraint:

$$w_l \geq \phi w_h \tag{6}$$

The supply constraints ensure that employment is bounded above by the labor supply curves: i.e., firms can only hire willing workers. The equity constraint (6) is our point of departure from standard monopsony models: firms cannot pay l-types less than a fraction $\phi \leq 1$ of the h-type wage. The $\phi = 1$ case may be interpreted as a "horizontal" constraint, where h and l-types perform similar tasks to different abilities but firms cannot pay discriminate between them (as in Manning, 1994); whereas the $\phi < 1$ case speaks more to a "vertical" constraint, which limits the extent of pay differentiation across the firm's hierarchy. The equity constraint can be microfounded using the efficiency wage model of Akerlof and Yellen (1990).[6] Note that a strictly proportional constraint, as expressed in (6), is not crucial to our story: any binding limit on internal pay differentiation can deliver the quantity-quality trade-off we describe. But strict proportionality simplifies the exposition, and it happens to fit the wage data well: as we explain below, it is consistent with log additivity.

The nature of labor market equilibrium depends on whether the equity constraint (6) 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 equity constraint does not bind, the labor supply constraints in (5) must bind instead: i.e., $l_h^* = l_h(w_h)$ and $l_l^* = l_l(w_l)$. Intuitively, since firms set wages below marginal products, they hire all workers who are willing to join them. For skill type s, the optimal wage is then:

$$w_s^* = \frac{\varepsilon}{1+\varepsilon} p_s \tag{7}$$

[6] Suppose l-type workers' 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 $e_l p_l$. I.e., they 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 will not justify the productivity losses. An alternative rationale for equation (6) is administrative costs, if a narrow wage structure is cheaper for firms to manage; but survey evidence casts doubt on this interpretation (Giupponi and Machin, 2022).

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:

$$\frac{w_l^*}{w_h^*} = \frac{p_l}{p_h} \tag{8}$$

For a non-binding equity constraint, we therefore require $\phi \leq \frac{p_l}{p_h}$.

2.2 Equilibrium if equity constraint binds

Let β denote the bite of the equity constraint:

$$\beta \equiv \phi \frac{p_h}{p_l} \tag{9}$$

i.e., the ratio of ϕ to the productivity differential. The constraint binds if $\beta > 1$, and this fixes the log wage differential between l- and h-types to $\log \phi$ within all firms. Based on equation (6), wages will then take log additive form:

$$\log w_{sf} = \eta_f + \lambda_s \tag{10}$$

where the skill effect $\lambda_s = I[s = l] \cdot \log \phi$ captures the internal wage differential between skill groups (fixed by the equity constraint), and η_f captures a firm-level wage premium (equal to $\log w_{hf}$ in the model). This simple structure provides a novel interpretation of the log additive AKM wage model: faced by a binding equity constraint (which fixes the worker effects), firms must pay a uniform wage premium to all employees (the firm effect). Unlike the worker effects, firms do have control over this uniform premium: as we now describe, variation in the firm effects reflects heterogeneity in hiring strategies, even among ex ante identical firms.

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

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

$$w_h^I = \frac{1 + \frac{1}{\beta} \cdot \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}} \cdot w_h^* < w_h^*, \qquad w_l^I = \frac{\beta + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}}{1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}} \cdot w_l^* > w_l^* \tag{11}$$

2. **Selective strategy (S).** Selective firms hire all willing h-types, so the h-type supply

constraint binds: $l_h^S = l_h\left(w_h^S\right)$. But they fully ration l-types, such that $l_l^S = 0$. This rationing only makes sense if the h-type offer w_h^S is so high that the required l-type offer (which equals ϕw_h^S if the equity constraint binds) exceeds the l-type productivity p_l.[7] Since selective firms hire only h-types, they will offer them the unconstrained optimal wage, i.e., $w_h^S = w_h^*$, which exceeds the inclusive offer w_h^L. See Appendix B.2.

Though firms are ex ante identical, they may adopt different strategies in equilibrium—with selective firms offering higher pay premia. Let σ denote the 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 ex ante identical, we cannot predict the behavior of any individual firm; but the selective share σ is uniquely determined and lies between 0 and 1. Note that equal profits 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.

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

$$\sigma = \begin{cases} 0 & \text{if } \beta < \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha} \\ \tilde{\sigma}\left(\alpha, \beta, \varepsilon\right) & \text{if } \beta \geq \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha} \end{cases} \tag{12}$$

where

$$\alpha \equiv \frac{p_h n_h}{p_h n_h + p_l n_l} \tag{13}$$

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

$$\left(1 + \frac{1-\alpha}{\alpha - \tilde{\sigma}}\right)^{1+\varepsilon} = \left(1 + \beta \frac{1-\alpha}{\alpha - \tilde{\sigma}}\right)^{\varepsilon} \tag{14}$$

[7] Job rationing takes an extreme form (with $l_l^S = 0$) in the baseline model because marginal products are fixed. But if the marginal product p_l is decreasing in l_l (e.g., if skill types are imperfect substitutes or if firms face diminishing returns), partial rationing is also possible, with l_l^S exceeding zero but still lying below the labor supply curve $l_l\left(\phi w_h^S\right)$. See Appendix D for an exposition with CES technology: selective firms do hire some l-types, but proportionally fewer than inclusive firms—due to partial job rationing.

Equation (12) 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 $\beta < \frac{(1/\alpha)^{1/\varepsilon}-\alpha}{1-\alpha}$, 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 (14).

To see how the partially segregated equilibrium works, note that increases in α, β and ε would all make the selective strategy more profitable—if σ remained unchanged. But this cannot be an equilibrium, as the selective strategy would then strictly dominate ($\pi^S > \pi^I$). It is σ that plays the equilibrating role: increases in σ reduce the relative profitability of the selective strategy, as competition for h-types intensifies—until we once again settle at equal profit, with $\pi^S = \pi^I$.

Note that wage-setting power (with finite labor supply elasticity ε) is crucial to sustaining a partially segregated equilibrium, where an inclusive hiring strategy (offering lower pay but employing *both* skill types) is viable. Under perfect competition (with $\varepsilon \to \infty$), all workers are paid their marginal product but are perfectly segregated between firms (with the σ share converging to its upper limit of α); and since there are no job rents, any quantity-quality trade-off becomes redundant. This is the case explored by Akerlof and Yellen (1990).

Finally, it is worth briefly drawing the connection with outsourcing. Selective firms are not technically outsourcing, since they maintain no relationship with the l-types workers they exclude; but the intuition is very similar. For example, suppose firms can indirectly hire l-type workers through an intermediary. Like our selective firms, they would have to sacrifice some of the rents accruing to l-type employment (to the intermediary), but this would allow them to set h-type wages freely (as the equity constraint would become redundant). Consistent with this interpretation, Deibler (2022) finds large wage gains for workers who remain in outsourcing firms.

2.3 Comparative statics: Impact of equity constraint

If the equity constraint binds, and if the labor market is not perfectly competitive, firms will face a trade-off between quantity and quality in hiring. As we will now show, this trade-off has important implications for pay dispersion, workplace segregation and inequality.

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.

If the equity constraint binds, and if its bite β exceeds $\frac{(1/\alpha)^{1/\varepsilon}-\alpha}{1-\alpha}$, at least some firms will adopt the selective strategy ($\sigma > 0$). 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 the quantity-quality trade-off: by paying more, selective firms can attract more h-types, but higher pay also compels them to ration l-types; and the equilibrium σ ensures that firms are indifferent between strategies.[8] This job rationing by selective firms delivers the workplace segregation result in part (b).

Next, consider the implications for skill wage differentials. If there is no skill segregation ($\sigma = 0$), a binding equity constraint unambiguously compresses wage gaps: inclusive firms simply redistribute earnings between skill types, in line with (11). If the bite β is sufficiently large however (such that $\sigma > 0$), this effect is offset by selective hiring—as some firms adopt the high-pay strategy and exclude l-types. But as long as skill segregation is partial (with σ below its α limit), the compression effect dominates: as Appendix B.4 shows, 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: i.e., these equity effects involve redistribution of earnings between workers alone. This is part (c) of the proposition.

Finally, we explore the implications for non-wage amenities and welfare. If β is sufficiently large (such that $\sigma > 0$), Appendix B.5 shows that the equity constraint reduces the expected value of amenity matches, for both skill types: this is part (d). For l-types, this is because they are denied access to selective firms, so they have fewer firms to choose from.[9] For h-types, the amenity loss is a consequence of firm pay dispersion: h-types are willing to sacrifice amenity match quality to secure jobs at high-paying selective firms. This presents a stark contrast to alternative models with firm-worker complementarities in production, where positive sorting is associated with *improvements* in match quality (here, in productivity).

Clearly, removing the equity constraint is not a policy-relevant counterfactual, if the constraint is underpinned by equity norms (as in Akerlof and Yellen 1990). The purpose of Proposition 1 is purely to describe the mechanics of the model. But if we do interpret the equity constraint (and the norms which underpin it) as a manipulable policy lever[10], part (d)

[8]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: higher 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 equity 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.

[9]In an alternative job search framework, selective firms' rationing of l-types would manifest in higher unemployment rather than lower-quality amenity matches: see Appendix F.

[10]Suppose we microfound the constraint in the manner of Akerlof and Yellen (1990), as described in footnote 6. Since firms always optimally conform to workers' equity norm (if it binds), we can ignore any

implies that a binding constraint is "inefficient": since aggregate earnings, profit and output are unchanged[11], the amenity losses imply aggregate efficiency losses.

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 we allow for 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 more. This can be attributed to firm heterogeneity, which we have ignored until now. In Appendix C, we incorporate skill-neutral heterogeneity in firm productivity: the h- and l-type marginal products in firm f are $p_{hf} = x_f p_h$ and $p_{lf} = x_f p_l$, where x_f is distributed log normally across firms. This heterogeneity introduces a second source of variation, which generates a countervailing *positive* correlation between firm size and pay. This positive correlation arises from the standard quantity motive in monopsony models: productive firms benefit more on the margin from larger employment, so they offer higher pay. However, firm productivity x_f does not affect the relative value of the selective and inclusive strategies; so the selective share σ is independent of x_f.

Therefore, firm wage premia may now vary for two 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 (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

implications of violations of the equity norm for workers' utility and effort in this exercise.

[11] Output in this model is fixed by assumption, as workers are equally productive in all firms. And since aggregate earnings are unchanged, the same must then be true of profit.

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).

We have focused here on the case of skill-neutral heterogeneity (i.e., no productive complementarities), as this will guide our quantitative analysis below. But it is worth emphasizing that the concavity result does not hinge on this assumption.[12]

2.5 Market-level determinants of skill wage differentials

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

To guide our conceptual discussion and the empirical analysis below, we will rely on a simple decomposition of skill wage differentials. 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:

$$E\left[\log w_h\right] - E\left[\log w_l\right] = \underbrace{\log\frac{1}{\phi}}_{\text{Within-firm}} + \underbrace{\frac{\sigma}{\alpha}\log\left(\frac{1-\sigma}{\alpha-\sigma}\right)^{\frac{1}{\varepsilon}}}_{\text{Between-firm}} \tag{15}$$

The first component on the right 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 equity constraint ϕ, which we take

[12] In particular, suppose there are productive complementarities between firm quality and worker skill. High-quality firms will then have a comparative advantage in adopting the selective strategy—and there will be some productivity cut-off above which firms are selective, and below which they are not. Below the cut-off, firm employment will be increasing in firm productivity (and hence in firm pay), as there is no strategy heterogeneity. But once we hit the cut-off (moving up the pay distribution), firm strategies switch from inclusive to selective, and firm employment decreases—a consequence of Proposition 2a. This implies a non-monotonic relationship between firm employment and pay, just as in Proposition 2b.

as given. In practice though, internal pay differentials are likely to be responsive to the other parameters—for two reasons. First, changes in relative productivity $\frac{p_h}{p_l}$ may affect workers' concept of "fair" wages; in particular, perfect pass-through to ϕ would imply that the constraint bite β in equation (9) maintains its value. Second, internal pay differentials may also be affected by the relative supply of skilled labor $\frac{n_h}{n_l}$: if larger supply encourages firms to adopt the selective strategy (see below), 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 supply $\frac{n_h}{n_l}$ of h-types. 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 (13). 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 small, such that $\beta < \frac{(1/\alpha)^{1/\varepsilon}-\alpha}{1-\alpha}$, 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, as defined by (11). Since the selective share σ is zero, there will be no workplace segregation and no between-firm component in the skill wage differential (15). 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 is sufficiently large, such that $\beta \geq \frac{(1/\alpha)^{1/\varepsilon}-\alpha}{1-\alpha}$, 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 (15), 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 different wages to otherwise identical workers, an apparent violation of the "law of one price". But why should workers ever accept offers from low-paying firms? One reason might be 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 empirical 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 has provoked some debate (Jäger et al., 2024; 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[13]: 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 in our CES extension that rationed jobs are allocated randomly.[14] 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 from the Central Bureau of Statistics. The core dataset contains employment records linking workers to firms, for the period 1990-2019.

[13]Saez et al. (2019) also explore a model with an internal equity constraint and CES technology. But our framework differs in incorporating equilibrium wage competition between firms: this means that workers who are denied access by selective firms find employment in inclusive firms.

[14]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.

We leave a detailed description to Appendix L, but summarize the main points here.

For each worker-firm match, we observe average monthly earnings, industry classification, and an indicator for public sector employment. We restrict our main analysis to the private 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 demographics and education, including field of study. For our empirical analysis, we divide workers into three education groups: (i) no college degree, (ii) non-STEM degree, and (iii) STEM degree.[15] The STEM/non-STEM distinction has become increasingly salient in recent decades (Altonji et al., 2016; Kirkebøen et al., 2016), and especially in the context of the Israeli tech boom. Though our baseline model distinguishes between just two skill types, we extend it 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".[16] 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 employment and earnings by education

Our sample period saw a large expansion of the employment shares and wage returns of college graduates. Much of the action happened in the early part of the sample, before the mid-2000s—and was specific to STEM graduates, a reflection of Israel's tech boom.

We illustrate these trends in Figure 1. Panel A shows that the STEM employment share tripled from 3% to 9% by the early 2000s, and changed little thereafter. This growth was partly driven by immigration from the former Soviet Union (FSU), who were disproportionately STEM-educated, as well as college enrollment among native Israelis: the period saw a large 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 32%.

In Panel B, we turn to wages. Conditional on basic demographics (age, gender, minority effects), the return to non-STEM degrees grew moderately from 0.28 to 0.33 (filled blue line), but the STEM return surged from 0.30 to 0.75 (filled red line); and again, much of the change occurred early in the sample.[17] Remarkably, most of the increase is explained away by firm

[15]We define science, engineering and mathematics degrees as STEM.

[16]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.

[17]The FSU immigration wave provides important context here. Many of the new arrivals had STEM degrees but were employed at unusually low wages, and this explains the initial dip in STEM returns in the

fixed effects (dashed red line): i.e., it is mostly attributable to a growing concentration of STEM graduates in high-paying firms.[18] Motivated by Proposition 3, we will argue this growing segregation was partly a response to the expanding supply of STEM graduates.

These trends were an important feature of changes in the Israeli earnings distribution; see also Cornfeld and Danieli (2015) and Dahan (2021). As Appendix Figure A2 shows, real earnings grew markedly across the board—but this growth was concentrated among the highest earners until the mid-2000s, at least when minority groups are excluded from the sample.[19] As in the US (Autor et al., 2008), returns to education made an important contribution to growing inequality—and account for about half the increase in the standard deviation of log earnings (see Appendix Figure A3).[20] But the key insight here, in the spirit of Card et al. (2013), is the role of firms' pay and hiring policies in driving these returns.

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:

$$\log w_{it} = \eta_{f(i,t)} + \lambda_i + \gamma X_{it} + e_{it} \tag{16}$$

where $\eta_{f(i,t)}$ are firm effects (for a firm f employing worker i at time t), λ_i are worker effects, and X_{it} includes time-varying controls.[21] The firm effects are identified by worker mobility

early 1990s. Their wage assimilation also contributed significantly to the growth in aggregate STEM returns after 2000: once FSU immigrants are excluded from the sample, the increase in STEM returns is almost entirely confined to the 1990s. We show this explicitly in Appendix Figure A1: compare panels A and B.

[18] Interestingly, this trend is associated with the entry of new cohorts of firms, as opposed to incumbent firms changing their pay/hiring strategies: as we show in Appendix Figure A1, fixing the value of the firm effects over time (in the dotted lines) makes little difference to the results. The importance of new firm cohorts echoes Sorkin and Wallskog (2023), who find that growing dispersion in firm pay effects was driven by the entry of more unequal firm cohorts—rather than changes in incumbents firms' pay policies. And see also Lachowska et al. (2023), who document the persistence of pay policies within firms over time.

[19] In the full sample, the 90-10 percentile gap grew significantly among women but little among men (see Panels A and B of Appendix Figure A2). But this imbalance masks important trends affecting key minority groups: the arrival of FSU immigrants 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.

[20] Appendix Figure A3 plots changes in the standard deviation of residualized log earnings, following Card et al. (2013). Excluding minority groups, this grew about 0.07 for both genders by the mid-2000s (Panels C and D). 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 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).

[21] Following Card et al. (2018), we control for quadratic and cubic polynomials of age (centered around 40), in addition to year effects. Given our focus on education, we also interact both the age and year effects with

between firms, and the worker effects by pay differentiation within them. Under our model's assumptions, the firm effects η_f have a causal interpretation: 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 selective firms ration l-types).

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. Either way, 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, for the years 2010-2019. To address measurement error in the estimated firm effects, we implement a split-sample correction (see Appendix K). Panel A shows the AKM model fits the data well, explaining 91.1% of the overall variance in log wages—similar to estimates from other countries (e.g., Card et al. 2013). The worker effects account for the largest share of wage variance (65.1%), firm effects contribute 7.6%, and the covariance between worker and firm effects explains 18.4%: this indicates 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 4.5pp. As in previous work[22], this small improvement suggests that match effects offer little additional explanatory power—consistent with our model.

The next three columns show that the AKM model exhibits similarly high explanatory power for each education group, with an R-squared of about 90% in each case. Notice also that there is significant firm-worker sorting even within education groups: this is to be expected if education is an imperfect indicator of skill.

Panel B also reports education differentials in wages; and in line with equation (15), we decompose these into within-firm and between-firm components, using our AKM estimates. 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, and $\frac{0.18-(-0.03)}{9.71-8.96} = 28\%$ of the differential between STEM graduates and non-graduates.[23] As in Figure 1, this illustrates the importance of workplace sorting in

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 they are important when moving to the data. To ease notation, we incorporate them within the worker effects in all the empirical analysis below.

[22]Using German data, Card et al. (2013) estimate an R-squared 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-squared from 93%–94% to 95% (Card et al., 2016).

[23]Notice the education differentials in worker and firm effects sum almost exactly to the raw earnings gap: this reflects the strong fit of the AKM model, for all education groups.

driving the return to education. But though substantial, the contribution of firm effects in Table 1 is smaller than in the figure: 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 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 equity constraint can explain all three empirical regularities, while competing models cannot. Finally, we illustrate the distributional implications of equity constraints using counterfactual analyses.

4.1 Relationship between firm employment and pay premia

We begin in Figure 2 by plotting the relationship between log employment and firm AKM premia, i.e., η_f from equation (16), across firms. We group firms into 20 bins according to their AKM premia, with each bin containing an equal number of firms. The y-axis shows mean log firm employment in each bin, and the x-axis shows the mean firm premia, adjusted for measurement error using a split-sample correction.[24]

Panel A reveals a striking inverse-U shape. This pattern offers strong support for a quantity-quality trade-off in hiring, and is 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. In Appendix Figure A5, we show that the inverse-U shape is preserved without the split-sample correction, and even when binning firms by their raw average wage (instead of their AKM premia).

We also find similar patterns within industries. In Panel B, we remove industry effects (87 categories) from both the y-variable (log employment) and x-variable (firm premia); and

[24] 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 (see Appendix K for details). We group firms into the 20 bins according to their sample A premia. For each bin, we then report the mean of the sample B premia on the x-axis, and the mean of log employment (using the full sample) on the y-axis. In Appendix Figure A4, we confirm that the sample B firm premia are monotonically increasing in the sample A premia, across the 20 firm bins.

the basic shape is preserved. This suggests it reflects fundamental trade-offs in firms' pay and hiring strategies, rather than simply sectoral differences in technology or optimal scale.

In Panels C and D, we exclude 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 3 plots the relationship between firms' log education-specific employment (non-graduate, non-STEM graduate and STEM graduate) and their AKM premia. For all three groups, we see a clear positive slope among the lowest-paying firms—consistent with a dominant quantity motive, with little skill rationing. But these slopes diverge markedly as we move up the distribution. Non-graduate employment follows a robust hump shape, with a sharp decline 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 predicted by our model: firms never ration high-skilled employment, so the green line simply traces out the isoelastic labor supply curve.

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 Italy. These effects are plausibly a consequence of the non-monotonic relationship we document in Figure 2, 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.

The hump shape is not particular to our data. In Appendix M, we replicate our Figure 2 analysis using the Veneto Worker History dataset from Italy. As Figure A6 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.

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 in an alternative environment with search frictions, job rationing effects may also be visible in patterns of worker *flows*. We set out such a model in Appendix F, building on Burdett and Mortensen (1998) and Manning (1994). Workers gradually ascend a job ladder to ever higher-paying firms, but this process takes time. In the presence of a binding equity constraint however, the job ladder will be "shorter" for low-skilled workers—if high-paying firms adopt selective hiring strategies and deny them access.

We test this claim in Table A1. We first divide 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. Looking first at STEM graduates (in Panel C), we see clear evidence of a job ladder: movers from the bottom quartile regularly find work across the full pay distribution, from the bottom to the top; but workers at the top rarely move to the bottom. In contrast, the ladder is significantly "shorter" for non-graduates (Panel A): very few job movers from the bottom quartile find employment at the top. This is consistent with top-quartile firms denying them jobs.

4.3 Log additivity of wages

If there is a binding equity constraint, all workers should receive higher wages in high-paying firms. And if the constraint is strictly proportional (as we assume in equation (6)), all workers should benefit equally: i.e., wages should take log additive form, with uniform firm premia. To test this claim, we estimate the AKM model (16) separately by education group and compare the group-specific firm premia. Figure 4 plots the group premia against the aggregate (i.e., full sample) premia, across 20 bins. The bins are ordered by the aggregate premia, but are defined separately by education group (and contain equal numbers of group-specific workers); since STEM workers sort into higher-paying firms, their bins (in green) are located more to the right.[25] We normalize both the group and aggregate premia to zero for firms with mean aggregate premia. If wages are log additive, the firm premia should then be identical across groups: i.e., the group premia should increase one-for-one with the aggregate premia, and should line up perfectly on the 45-degree (dashed) line. Looking at Figure 4, the data are very 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

[25] As before, we correct for measurement error using a split-sample method. We randomly divide workers into two samples ("A" and "B"), as described in Appendix K. 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-graduates, according to their sample A aggregate premia. The markers report the mean sample B aggregate premia (on the x-axis) and non-graduate premia (on the y-axis). The red and green dots repeat this exercise for non-STEM and STEM graduates, respectively.

in high and low value-added firms in Portugal.

Of course, there are many ways to account for log additivity—in isolation. But it is difficult to reconcile log additivity with the evidence above, on selective job rationing and hump-shaped employment. If the highest-paying firms demand many high-skilled workers, it makes sense that they offer them larger firm premia. But why offer equally large premia to the low-skilled, if they intend to hire so few? We argue that an equity constraint can resolve this tension, by compelling high-paying firms to trade off quantity against quality.

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: despite its simplicity, it performs surprisingly well. 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 skill types l, m and h, respectively.

Table 2 summarizes the target moments and parameter estimates. The intuition for identification is as follows. The labor supply elasticity ($\varepsilon = 5.52$) is pinned down primarily (but not exclusively of course) by the average elasticity of log firm size to AKM firm effects, and the productivity variance ($\nu = 0.02$) by the variance of AKM firm effects. We identify the (binding) equity constraint parameters (i.e., the ϕs) using skill differentials in the mean AKM worker effects (from Table 1). The aggregate wage differentials between skill groups, i.e., $E[\log w_m] - E[\log w_l]$ and $E[\log w_h] - E[\log w_l]$, then determine the extent of workplace segregation; and the model delivers the skill productivity differentials which rationalize this segregation. See Appendix G for details on our solution method.

In this model with three skill types (a special case of the N-type model in Appendix E), firms pursue one of three hierarchical strategies in equilibrium: (i) a fully inclusive "L-strategy", where firms hire all willing workers; (ii) an intermediate "M-strategy", hiring only m- and h-types; and (iii) a highly selective "H-strategy", hiring only h-types.[26] These strategies differ in the optimal h-type wage: we denote these w_h^L, w_h^M and w_h^H respectively. Holding firm productivity fixed, H-strategy firms pay 0.42 log points more than L-strategy firms, and M-strategy firms pay 0.16 more. Wages of other skill types are then fixed by the equity constraints: l-types are paid a fraction ϕ_l of their h-type co-workers' wage, and m-types are paid ϕ_m. In our calibration, $\sigma^H = 8\%$ of firms pursue the H-strategy, $\sigma^M = 17\%$ adopt the M-strategy, and the remaining 75% adopt the fully inclusive L-strategy.[27]

[26] This hierarchical structure applies if the equity constraint binds more for l-types than for m-types: i.e., if $\phi_l \cdot \frac{p_h}{p_l} > \phi_m \cdot \frac{p_h}{p_m}$. This assumption can be validated ex post, using the parameter estimates in Table 2.

[27] In principle, not all strategies need be active in equilibrium—though this happens to be the case in the

This diversity of hiring strategies contributes significantly to dispersion in firm pay. In a conventional monopsony model (where all firms hire inclusively), the variance of AKM firm effects is equal to the productivity variance, $\nu = 0.02$. But here, firm effects reflect not just *firm quality* but also *strategy*—and this explains why the firm effect variance reaches 0.032.

The results point to substantial internal pay compression. Within firms, STEM graduates earn just 0.54 log points more than non-graduates, much less than the productivity differential of 1.06.[28] This encourages large adoption of selective hiring, which partially undoes the equalizing effects of the equity constraint: on aggregate, STEM workers earn 0.74 log points more than non-graduates—significantly more than the 0.54 within-firm differential.

In contrast, 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 0.42. As a result, job rationing is less severe on this margin; and the aggregate pay differential between these skill groups (0.28) is not much larger than the within-firm differential (0.24).

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 models in Appendix I and report the estimated parameters in Table A2. Only Model 1 can match all three empirical patterns documented in Figures 2–4.

Firm size-pay relationship. Figure 5 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.[29]

national calibration. By contrast, in the regional calibration in Section 5.2, some strategies are inactive in particular regions and years.

[28] Following equation (E4) in Appendix E, the constraint bite in the three-type extension can be expressed as $\beta_l = \phi_l \cdot \frac{p_h}{p_l}$. Using our estimates, the bite is equal to $\exp(\log 1.06 - \log 0.54) = 1.96$.

[29] Of course, one can also introduce non-linearities into the firm size-pay relationship by relaxing the assumption of isoelastic labor supply (Kline, 2025). For example, if outside options were distributed according to a shifted power function (as in Card et al., 2018), the labor supply elasticity would be decreasing in firm pay. But while this can deliver *concavity* in the firm size-pay relationship, it cannot rationalize a *hump shape*; and it cannot explain why it is specifically lower-skilled workers who drive this pattern. In contrast, an equity constraint can explain both these features, as high-paying selective firms ration their demand for low-skilled labor in a quantity-quality trade-off—irrespective of the shape of the labor supply curve.

Workplace sorting. It is well known that high-skilled workers sort into high-paying firms, but Figure 3 shows this sorting takes a very particular form—with low-skilled employment hump-shaped in firm pay, and high-skilled employment increasing monotonically. Figure 6 shows that only Model 1 can deliver these patterns, as low-skilled workers are rationed by high-paying firms.[30] Though Models 3 and 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.

Log additive wages. Figure 4 shows that firms share wage premia close to equally with all skill types, consistent with the log additive AKM wage model. Model 1 delivers log additivity through the binding equity constraint, and Models 2 and 4 by assuming that firm heterogeneity is skill-neutral: see Figure 7. Model 3, however, severely violates log additivity, as the productive complementarities generate substantial worker-firm match effects.

In summary, while alternative models can match some of these empirical regularities, only an equity constraint can deliver all three. It also has intuitive appeal: as we argue in the introduction, it has a strong basis in both the theoretical and empirical literature. Of course, this should not be interpreted as evidence *against* productive complementarities or heterogeneous supply elasticities.[31] Rather, our point is that these features are insufficient on their own to deliver the key empirical results.

4.6 Resolution of empirical puzzles

The results above offer a resolution to two empirical puzzles in the wage-setting literature. First, our framework can reconcile the heavy sorting of high-skilled 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 7 shows, this fails to generate log additive wages in a vanilla monopsony framework. Intuitively, if high-paying firms wish to recruit differentially more high-skilled workers, they must reward them differentially more—if the labor supply constraints bind.

[30]Note we show the log of mean bin employment on the y-axis of Figure 6, instead of bin-level means of log employment (as in Figure 3). This is because selective firms do not employ *any* low-skilled workers (given our assumption of linear technology), and we cannot take logs of zeros. In the data however, the basic patterns are qualitatively similar whether we use log mean or mean log employment.

[31]In the presence of an equity constraint, productive complementarities or skill heterogeneity would amplify the sorting effects, while still delivering log-additive wages—as long as the constraint binds for all firms.

We are not the first to address this puzzle. Borovičková and Shimer (2024) argue that random match-specific productivity shocks can make wages appear log additive despite the presence of productive complementarities, if wages are bargained at the match level. Lamadon et al. (2024) focus instead on non-wage amenity valuations: 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 if wages are log additive). Finally, Kline (2025) proposes that wages may function as a screening device. If high-skilled workers have better outside options, and firms cannot condition wages on skill, higher offers may differentially attract high-skilled workers (see also Weiss, 1980): this can yield comparable sorting patterns to Model 4. However, these alternative stories cannot reproduce the very particular form of sorting we find in Figure 3, driven by a hump shape in low-skilled employment. To account for this, we must explain why high-paying firms choose to *ration* low-skilled employment: an equity constraint does exactly this, through the quantity-quality trade-off.

Second, our model can help explain the surprisingly weak cross-sectional relationship between firm size and wages (Sokolova and Sorensen, 2021; Bloesch and Larsen, 2023). Conventional monopsony models predict a tight relationship between the two, as firms use pay exclusively to manipulate employment. But in our model, firms also use pay to improve hiring quality, even at the cost of lower quantity. This is immediately visible in Table 2: though we estimate a labor supply elasticity ε of 5.52, the elasticity of firm size itself (denoted ε_{size}) is significantly smaller—at just 2.55. Flipping the variables, a simulation of our parsimonious model yields a small firm size premium of 0.13, which is not far from our empirical estimate of 0.05 in Israel (Sokolova and Sorensen 2021 find similar numbers elsewhere).[32]

It is notable that our estimate of the labor supply elasticity ($\varepsilon = 5.52$) 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 and inclusive), 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 2 and 3), 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.

[32]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.

4.7 Counterfactual analysis

Having established the empirical appeal of our model, we now quantify the distributional implications of internal equity constraints. First, we examine the consequences of removing the constraint (and implicitly, the equity concerns which underpin it), i.e., the reverse experiment of Proposition 1. Second, we consider a policy which prohibits firms from adopting the selective strategy. Table 3 reports impacts on expected log wages and amenity match quality, by education. Amenity effects are weighted by $\frac{1}{\varepsilon}$, to ensure they are in comparable log wage units (see equation (1)); and the utility effects sum the wage and amenity components. We leave technical derivations to Appendix H.

Panel A shows the effect of removing the equity constraint—and implicitly, the equity concerns which underpin it. This exercise does not have policy relevance, but it illustrates the mechanisms underpinning the model. Consistent with Proposition 1c, wage inequality grows: firms now set wages independently for each group, and no longer redistribute rents between them. STEM and non-STEM graduates enjoy wage gains of 21% and 4% respectively, while non-graduates suffer a 10% loss. At the same time, all three groups benefit from improved amenity matches, as in Proposition 1d. Intuitively, in the counterfactual, high-skilled workers no longer need to sacrifice amenity match quality to secure jobs 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 is insufficient to offset their wage losses, so their expected utility falls. In summary, eliminating the constraint brings improvements in amenity match quality, but exacerbates wage inequality.

However, a 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 effects are therefore reversed: non-graduates wages grow 5%, non-STEM graduate wages are little affected, and STEM wages contract by 15%. 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.[33] We therefore have efficiency gains, alongside the improvement in equity.

The second counterfactual provides a useful theoretical benchmark for interpreting a centralized collective bargaining system (which mandates uniform pay policies) or alternatively the public sector labor market—which we explore empirically in Section 5.3 below. Given

[33] The amenity effects are also *quantitatively* identical in each counterfactual. In each case, strategy differences and labor rationing are eliminated; and as a result, both counterfactuals yield the same worker allocations across firms—and hence the same improvement in expected amenity match quality.

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 Market-level outcomes

5.1 Aggregate trends in strategy shares

Until now, we have focused on empirical variation across firms. But the model also has testable implications for *market-level* outcomes. If 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—leading to greater firm pay dispersion and workplace segregation. The Israeli tech boom provides a natural setting to test these predictions. Figure 1 shows a large expansion of STEM graduate employment since 1990 (Panel A), contemporaneous with large increases in the return to STEM degrees (Panel B), much of which is driven by sorting of STEM graduates 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 (16) 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 inclusive L-strategy share fell from 91% in the 1990s to 75% in the 2010s. Conversely, the M-strategy share (hiring only m and h-types, i.e., non-STEM and STEM graduates) grew from 8% to 17%, and the H-strategy share (only STEM graduates) from 1% to 8%. This growing prevalence of selective hiring is reflected in the variance of firm pay premia, which grew from 0.026 to 0.032. The model attributes these changes to a combination of (i) growing aggregate skill shares, as documented in Figure 1, and (ii) growing skill productivity differentials—both of which make selective hiring more attractive. In particular, Table A3 shows that STEM productivity grew by 0.11 ($= 1.06 - 0.95$) log points, relative to non-graduates.

At the same time, our estimates point to a relaxation of 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 0.11 increase in productivity differentials. Our model does not provide an explanation for these changes, since the ps and ϕs are 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 explanations for the growth in earnings inequality. While Autor et al. (2008) and Dustmann et al. (2009) have reasserted the role of skill-biased technical change, Card et al. (2013) and Song et al. (2019) have highlighted the contribution of growing dispersion in firm pay—and the sorting of high-skilled workers to high-paying firms. 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; Bergeaud et al., 2025) can be interpreted as manifestations of this broader phenomenon.

5.2 Spatial variation in strategy shares

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 occurred 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 0.044 to 0.065. 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:

$$y_{rt} = \beta_0 + \beta_1 \text{GradShare}_{rt} + d_t + \varepsilon_{rt} \tag{17}$$

$$y_{rt} = \beta_0 + \beta_1 \text{GradShare}_{rt} + d_r + d_t + \varepsilon_{rt} \tag{18}$$

where y_{rt} is some outcome in region r at time t, GradShare_{rt} is the local graduate employment share (among individuals employed in region r), 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. In this analysis, 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: i.e., we do not seek to isolate skill share variation which is orthogonal to high-skilled productivity (unlike e.g., Moretti 2004). Proposition 3 makes predictions on how α affects the quantity-quality trade-off, and we seek to test these predictions 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 similar coefficients when using between-region and within-region variation: 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 are not particular to Israel: Card et al. (2025) also find a larger correlation between AKM firm and worker effects in locations with high graduate share (with a similar slope coefficient to ours).[34] Our contribution here is to demonstrate robustness to within-region variation, and to offer a new interpretation of the phenomenon. In influential work, Dauth et al. (2022) show that firm-worker sorting is also increasing in city size—and attribute this to increasing returns in the local matching technology. We find a similar effect in our Israeli data: in column 5 of Panel B (Table 4), we estimate a positive effect of log regional employment on sorting.[35] But as column 5 of Panel C shows, local graduate share captures the entire effect in a horse-race between the two variables. Additionally, only graduate share

[34]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.

[35]Our coefficient estimate is 0.020, which compares to 0.061 in Germany (Dauth et al., 2022) and 0.039 in the US (Card et al., 2025).

remains statistically significant in the presence of region fixed effects (see column 6). Still, it is worth emphasizing that our regions are smaller than those used by Dauth et al. (2022) and Card et al. (2025), and this may influence the results.[36]

The sorting effects also contribute significantly to regional variation in skill returns: see Appendix Table A6. As columns 1-2 show, wage returns to non-STEM degrees (Panel A) and STEM degrees (Panel B) are increasing in local graduate share—and much more steeply for STEM. To isolate the contribution of sorting, we rely on the decomposition of equation (15): i.e., we replace the dependent variable with skill differentials in AKM worker effects and firm effects respectively. Columns 3-4 show that much of the graduate share slope 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 (associated with the AKM firm effects) also plays an important role, and especially for STEM graduates: see columns 5-6. In the fixed effect specification, this sorting effect accounts for $\frac{0.425}{0.918} = 46\%$ of the overall graduate share slope: compare columns 2 and 6.[37]

Quantitative validation

The estimates above are *qualitatively* consistent with Proposition 3. But quantitatively, can the model rationalize effects 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 labor supply elasticity) using the national calibration from Section 4.4 (but now implemented for the 1993-1997 and 2006-2010 intervals, respectively).

Figure 8 shows how the equilibrium strategy shares vary by region and census year. In line with Proposition 3, higher-skilled regions have fewer inclusive L-strategy firms (which hire all skill types) and more selective firms: either M-strategy (hiring both m and h-types, i.e., non-STEM and STEM graduates) or H-strategy (h-types only). This pattern becomes

[36] 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.

[37] 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.

much more pronounced in 2008, following the expansion of the STEM workforce.[38]

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.[39] We present our results in Panel D of the same table, separately for each outcome. Comparing Panels A and D, the model not only captures the sign of the skill share slopes—but even performs reasonably on magnitude, especially for the mean and variance outcomes. For workplace sorting, the model overpredicts the graduate share effect by a factor of about two.[40] In Figure 9, 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 well—with R-squared between 0.44 and 0.73, and slope coefficients between 0.42 and 1.04.

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 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 Section 4.7.

To test this interpretation, we estimate the AKM model of equation (16) 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. Using the estimated

[38] 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 only, L+M, L+H, and L+M+H). See Table A7 for the distribution of configurations and Appendix J for further details.

[39] 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.

[40] Beyond our model's parsimony, and the neglect of other sources of regional variation, there are good conceptual reasons to expect an overprediction. When extrapolating from the national calibration (using regional variation in 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 share relative to the model's predictions.

AKM effects, we then decompose the wage returns to education (separately for each sector) into between-firm and within-firm components, in line with equation (15).

We present our estimates in Table 5. In Israel, the return to non-STEM degrees is slightly larger in the public sector—a consequence mostly of health and education professionals. But the STEM return is much larger in the private sector: 0.747 versus 0.472. Only a small fraction of this gap can be attributed to *within-firm* wage differentials, which are similar across sectors (0.523 versus 0.480, as identified by the worker effects): this suggests that differences in the equity constraint do not play a major role. Rather, consistent with our hypothesis, the bulk of the gap (0.225 versus -0.002) is driven by the *between-firm* component: i.e., sorting of STEM workers into high-paying firms. This mechanism is absent in the public sector, whose administrative units cannot easily compete on pay. As the final row shows, the variance of AKM firm effects is much larger in the private sector: 0.035 versus 0.022.

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 an internal equity constraint generates a quantity-quality trade-off in hiring. When the constraint binds, firms must choose between two hiring strategies: (i) a selective strategy, paying high wages to attract high-quality workers, while rationing lower-skilled employment, and (ii) an inclusive strategy, maintaining lower wages to employ a larger but lower-skilled workforce. Unlike in conventional monopsony models, firms use higher pay to improve hiring quality, even at the cost of lower quantity.

This simple insight can shed light on three key empirical regularities. First, it provides a novel interpretation of the (empirically successful) log additive AKM wage model: faced by a binding equity constraint (which fixes the worker effects), firms must pay a uniform wage premium to all employees (the firm effect), which reflects their chosen hiring strategy and not just their productivity. Second, our model shows how log additivity can be reconciled with sorting of high-skilled workers to high-paying firms, via a differential job rationing effect. And third, it can account for a striking hump-shaped relationship between firm size and pay premia, which we document in this paper. Consistent with our model, this extreme concavity is driven by high-paying firms rationing low-skilled workers. Using detailed administrative data from Israel, we find strong empirical support for our model's predictions. Our very parsimonious model can successfully reproduce the three key empirical regularities, in contrast to alternative monopsony models.

Finally, our hypothesis provides a new interpretation of temporal, regional and sectoral variation in earnings inequality. At the aggregate level, many countries have experienced growing dispersion in firm pay, alongside increasing workplace segregation of high and low-skilled workers. We attribute this phenomenon to a proliferation of the selective hiring strategy: as the relative supply and productivity of high-skilled workers have grown, this strategy has become increasingly attractive to firms. Similarly, we argue that regional differences in these same outcomes (firm pay dispersion and workplace segregation) can be attributed to regional variation in workforce skill composition; and we show how this variation can provide an out-of-sample validation of our national model. Finally, we interpret the public sector as a "control" environment, where administrative units cannot adopt independent pay strategies. By shutting down the key mechanism of our model, this can explain the relatively compressed wage structure observed in the public sector—even with no significant difference in the bite of the equity constraint.

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.

Bergeaud, A., C. Malgouyres, C. Mazet-Sonilhac, and S. Signorelli (2025). Technological Change and Domestic Outsourcing. *Journal of Labor Economics 43*(4), 1135–1168.

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.

Carry, P., B. Kleinman, and E. Nimier-David (2025). Location Effects or Sorting? Evidence from Firm Relocation. NBER Working Paper No. 33779.

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.

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

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.

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.

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.

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.

Machin, S. and A. Manning (2004). A Test of Competitive Labor Market Theory: The Wage Structure among Care Assistants in the South of England. *ILR Review*, 371–385.

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.

Moretti, E. (2004). Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data. *Journal of econometrics 121*(1-2), 175–212.

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

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 earnings	0.42	0.32	0.43	0.48
AKM model (%):				
Var. worker effect	65.1	66.2	68.4	62.3
Var. firm effect	7.6	7.9	6.9	8.6
2 × Cov(worker, firm)	18.4	15.0	15.0	20.2
R-squared	91.1	89.2	90.3	91.1
Comparison match model (%)				
R-squared	95.6	94.7	95.3	95.8
Panel B: Sample means and size				
Worker-years				
N.	7,884,004	4,645,994	2,527,748	710,262
Share N.	1.00	0.59	0.32	0.09
Av. log earnings	9.12	8.96	9.24	9.71
Av. worker effect	9.11	8.99	9.22	9.53
Av. firm effect	0.01	-0.03	0.03	0.18
Workers				
N.	1,435,461	843,292	475,336	116,833
Share N.	1.00	0.59	0.33	0.08
Firms				
N.	1,039,243			
Av. firm size	29.9			

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}	2.55	Elasticity of firm size w.r.t firm effect	ε	5.52	Elasticity of labor supply w.r.t pay
V_{AKMf}	0.032	Variance of firm effect	ν	0.023	Variance of firm productivity
$\log \frac{\phi_m}{\phi_l}$	0.24	Relative m v l-type worker effect	σ^M	0.17	Share firms with M-strategy
$\log \frac{1}{\phi_l}$	0.54	Relative h v l-type worker effect	σ^H	0.08	Share firms with H-strategy
$E[\log w_m] - E[\log w_l]$	0.28	Relative expected m v l-type log wage	$\log \frac{w_h^M}{w_h^L}$	0.16	Relative M v L-strategy log wage
$E[\log w_h] - E[\log w_l]$	0.74	Relative expected h v l-type log wage	$\log \frac{w_h^H}{w_h^L}$	0.42	Relative H v L-strategy log wage
$\log \frac{n_m}{n_l}$	-0.61	Relative m v l-type employment	$\log \frac{\phi_m^\varepsilon \Omega_m}{\phi_l^\varepsilon \Omega_l}$	-1.04	Relative m v l-type labor-supply intercept
$\log \frac{n_h}{n_l}$	-1.88	Relative h v l-type employment	$\log \frac{\Omega_h}{\phi_l^\varepsilon \Omega_l}$	-2.85	Relative h v l-type labor-supply intercept
			Implied parameters		
			$\log \frac{p_m}{p_l}$	0.42	Relative m v l-type log productivity
			$\log \frac{p_h}{p_l}$	1.06	Relative h v l-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

	Exp log wage	Exp amenity	Exp utility
Panel A: Removing equity constraint			
Non-graduates	-0.099	0.052	-0.047
Non-STEM graduates	0.035	0.030	0.065
STEM graduates	0.213	0.084	0.297
Panel B: Prohibiting selective strategy			
Non-graduates	0.054	0.052	0.106
Non-STEM graduates	-0.002	0.030	0.027
STEM graduates	-0.154	0.084	-0.070

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	0.308	0.362	0.048	0.050	0.601	0.730
	(0.056)	(0.071)	(0.005)	(0.019)	(0.104)	(0.258)
Within-R^2	0.380	0.511	0.538	0.259	0.325	0.272
Panel B: Effect of regional employment						
Log employment	0.011	0.009	0.001	0.000	0.020	0.025
	(0.003)	(0.015)	(0.000)	(0.003)	(0.005)	(0.038)
Within-R^2	0.199	0.014	0.221	0.001	0.154	0.013
Panel C: Controlling for both						
Graduate share	0.297	0.367	0.051	0.051	0.608	0.741
	(0.064)	(0.064)	(0.009)	(0.019)	(0.133)	(0.247)
Log employment	0.001	0.012	-0.000	0.001	-0.001	0.031
	(0.003)	(0.009)	(0.000)	(0.002)	(0.007)	(0.030)
Within-R^2	0.381	0.536	0.541	0.262	0.325	0.293
Panel D: Model-predicted outcomes						
Graduate share	0.292	0.297	0.029	0.048	1.172	1.845
	(0.027)	(0.031)	(0.008)	(0.009)	(0.241)	(0.261)
Within-R^2	0.844	0.822	0.292	0.580	0.404	0.596
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Sample	98	98	98	98	98	98

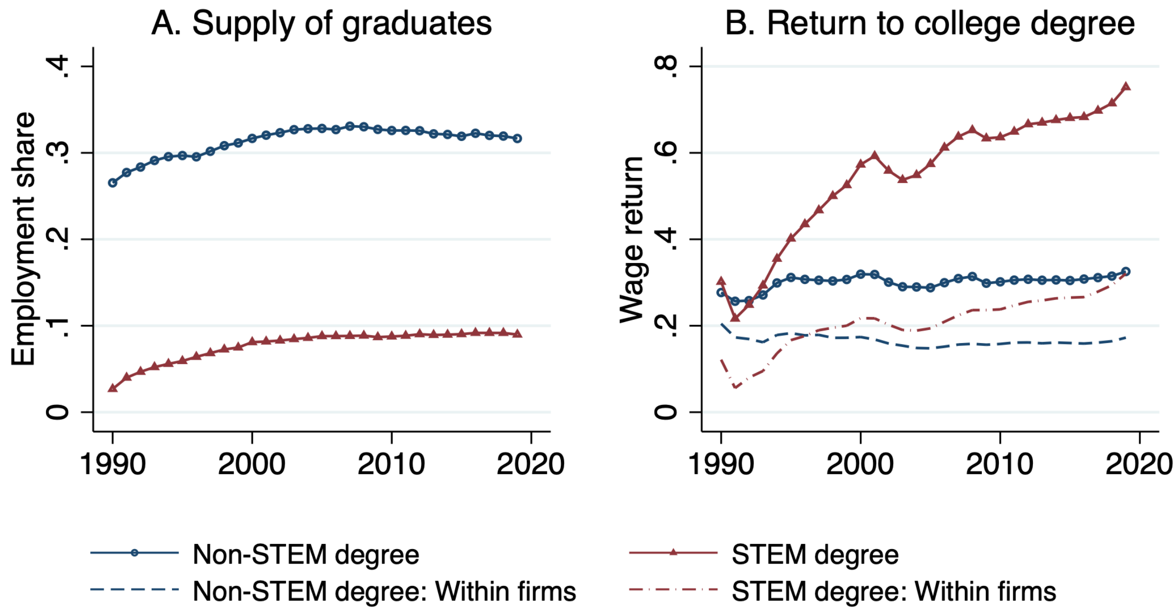
Notes: Panel A shows the relationship between regional graduate share and key local outcomes. Panel B reproduces these estimates, but with log regional employment instead of graduate share on the right-hand side. Panel C controls for both variables simultaneously. Panel D estimates effects on model-predicted outcomes, extrapolated from the national calibration. Odd-numbered columns exploit cross-sectional variation across regions, using equation (17). Even-numbered columns control for region fixed effects, as in equation (18), 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
Non-STEM graduates v non-graduates		
Log earnings	0.283	0.297
Worker effect	0.222	0.309
Firm effect	0.061	-0.014
STEM graduates v non-graduates		
Log earnings	0.747	0.472
Worker effect	0.523	0.480
Firm effect	0.223	-0.002
Var. firm effect	0.035	0.022

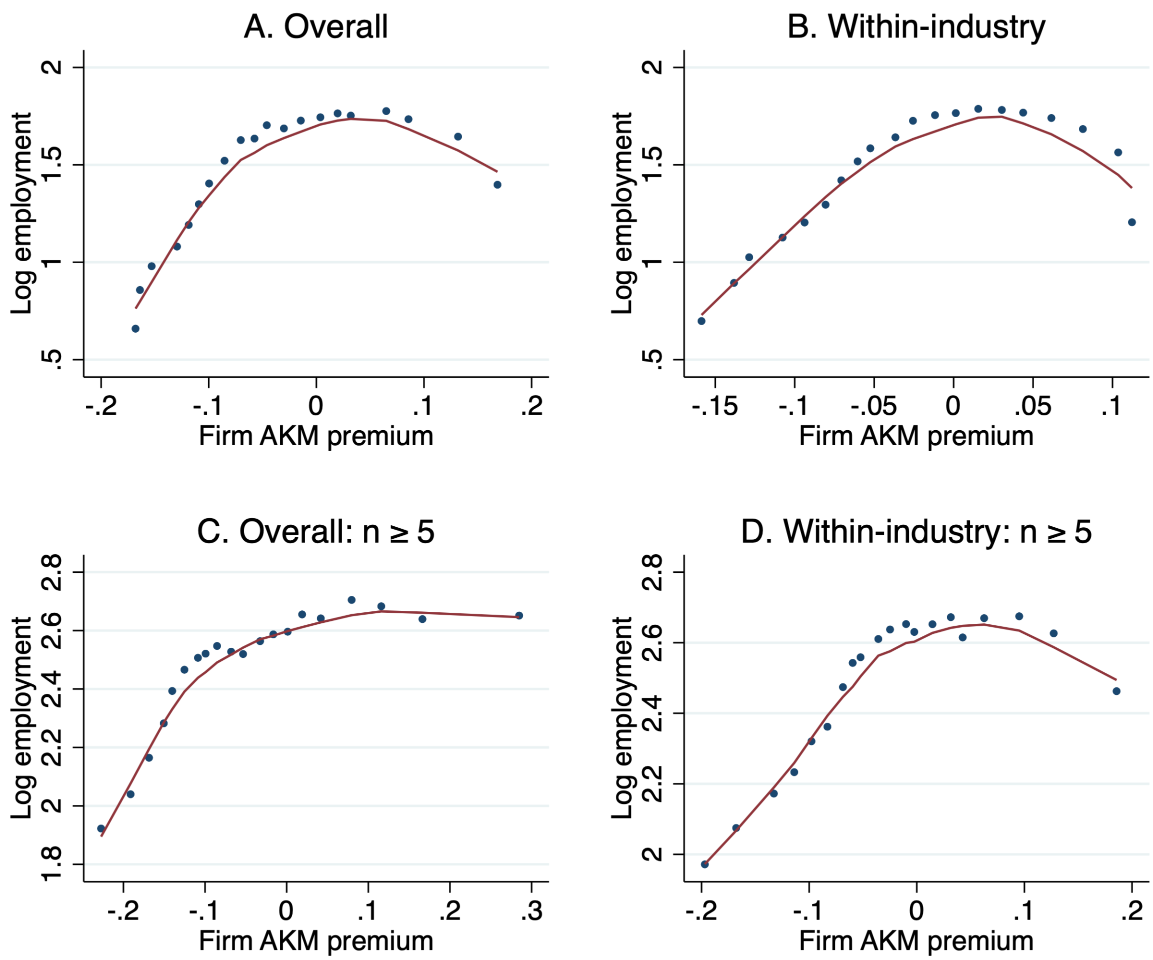
Notes: This table reports mean wage differentials between (i) STEM graduates and non-graduates and (ii) non-STEM graduates and non-graduates, separately for the private and public sectors, for the period 2010-2019. We disaggregate these wage differentials into contributions from AKM worker and firm effects, in line with equation (15). For this decomposition exercise, we rely on a common AKM model estimated using data from both sectors. In the final row, 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: Education employment shares and wage returns



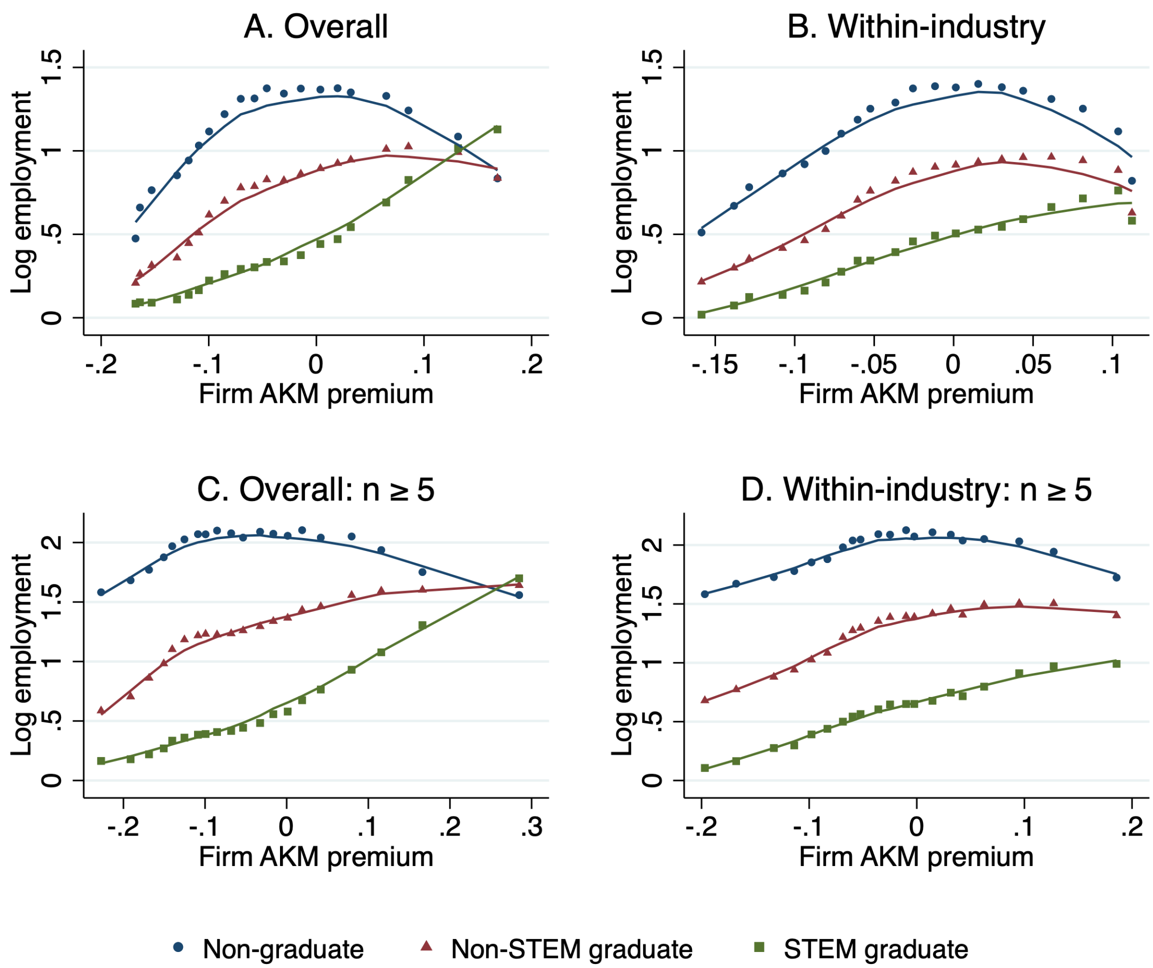
Notes: Panel A shows trends in the aggregate employment shares of STEM and non-STEM graduates. Panel B shows wage returns to STEM and non-STEM degrees. The filled lines condition on basic demographics only: interactions between gender, minority effects (for Arabs, ultra-orthodox Jews and FSU immigrants) and an age cubic, separately by year. The dashed lines are within-firm returns, which condition additionally on firm fixed effects (again, separately by year).

Figure 2: Employment by firm pay premium



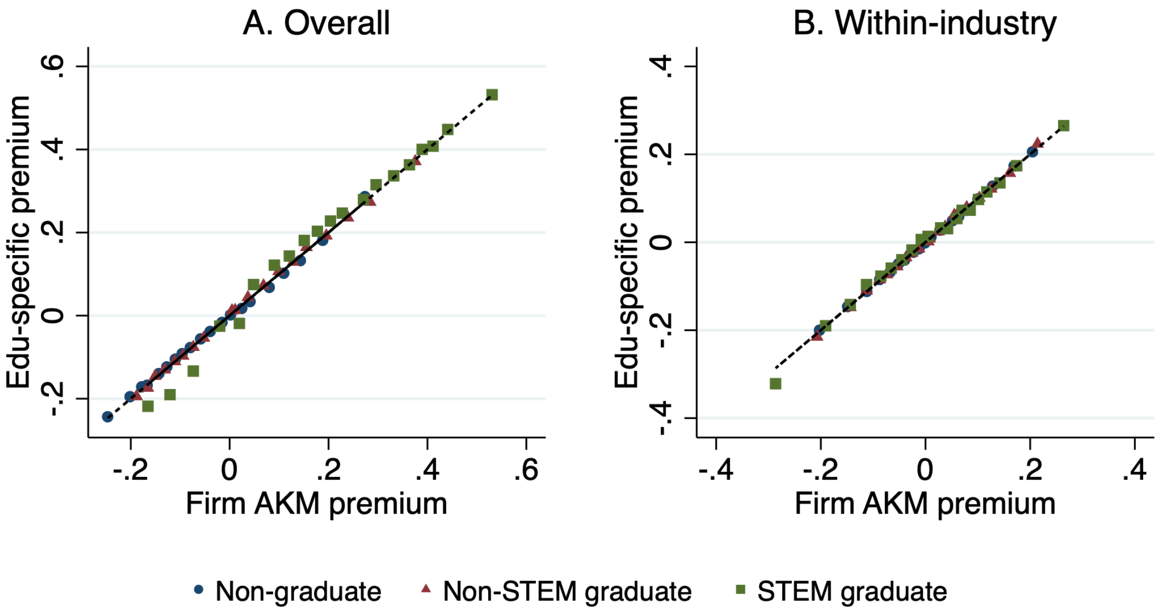
Notes: Panel A shows mean log firm employment across 20 bins (with equal numbers of firms), arranged by AKM firm premia. Firm premia are normalized to their worker-weighted mean. We implement a split-sample procedure to correct for measurement error in the firm premia, as described in Section 4.1. In 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 3: Education-specific employment by firm pay premium



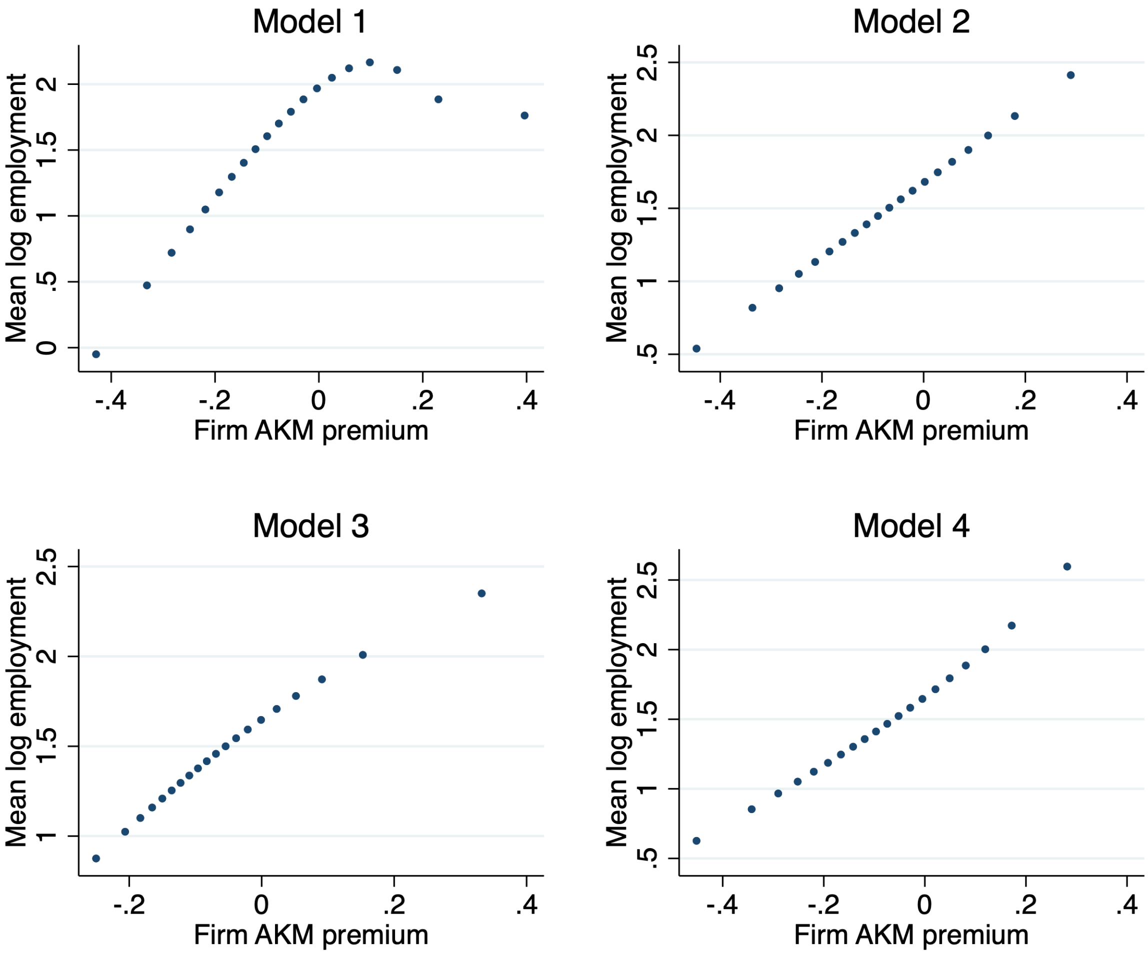
Notes: These plots repeat the exercise of Figure 2, but now showing mean log firm employment separately for three education groups: non-graduates, non-STEM graduates and STEM graduates. Sample consists of private sector firms in 2010-2019.

Figure 4: Education-specific firm pay premia



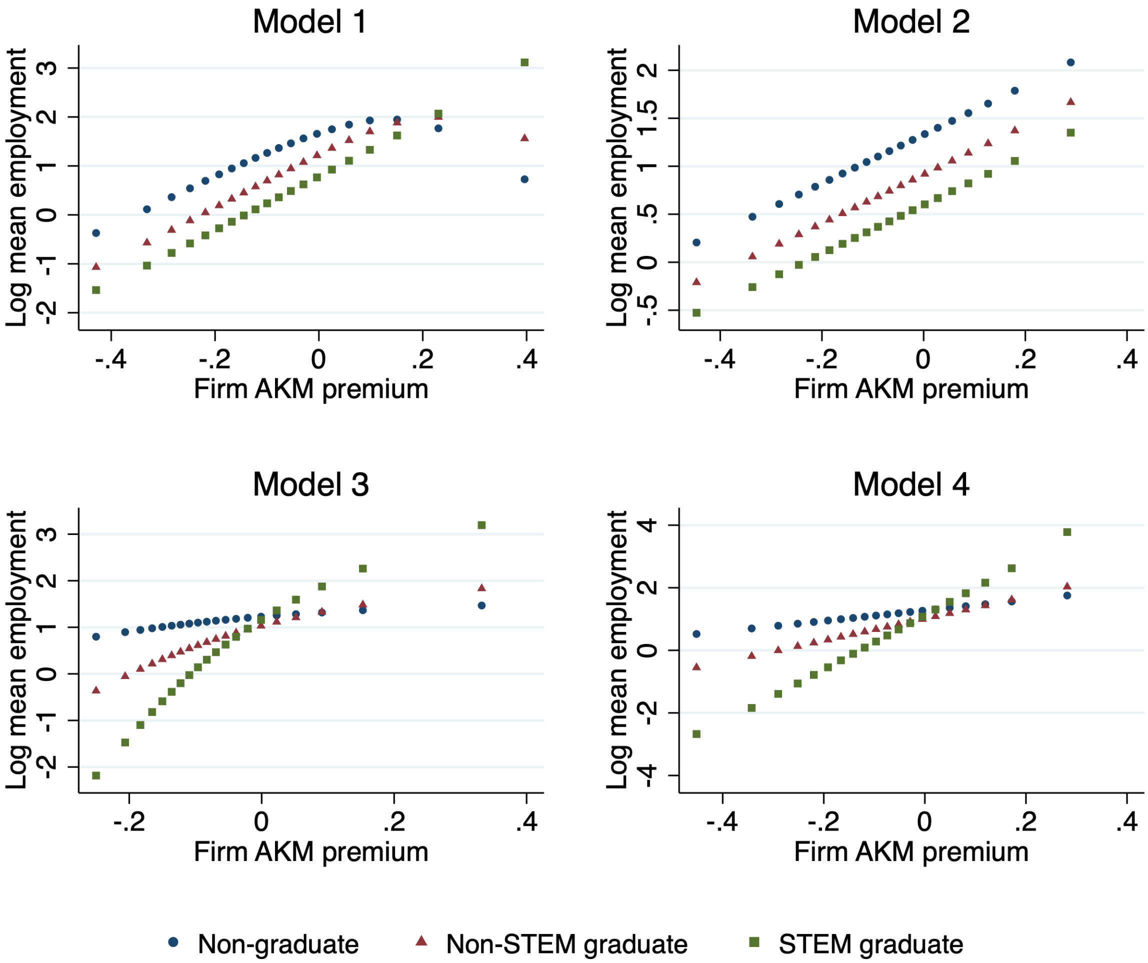
Notes: This figure estimates AKM firm premia separately by education group, and plots these group-specific premia against the aggregate (i.e., full sample) firm premia, across 20 bins (ordered by the aggregate premia). The bins are defined separately by education group, and contain equal numbers of group-specific 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 5: Calibrations of employment by firm pay premium



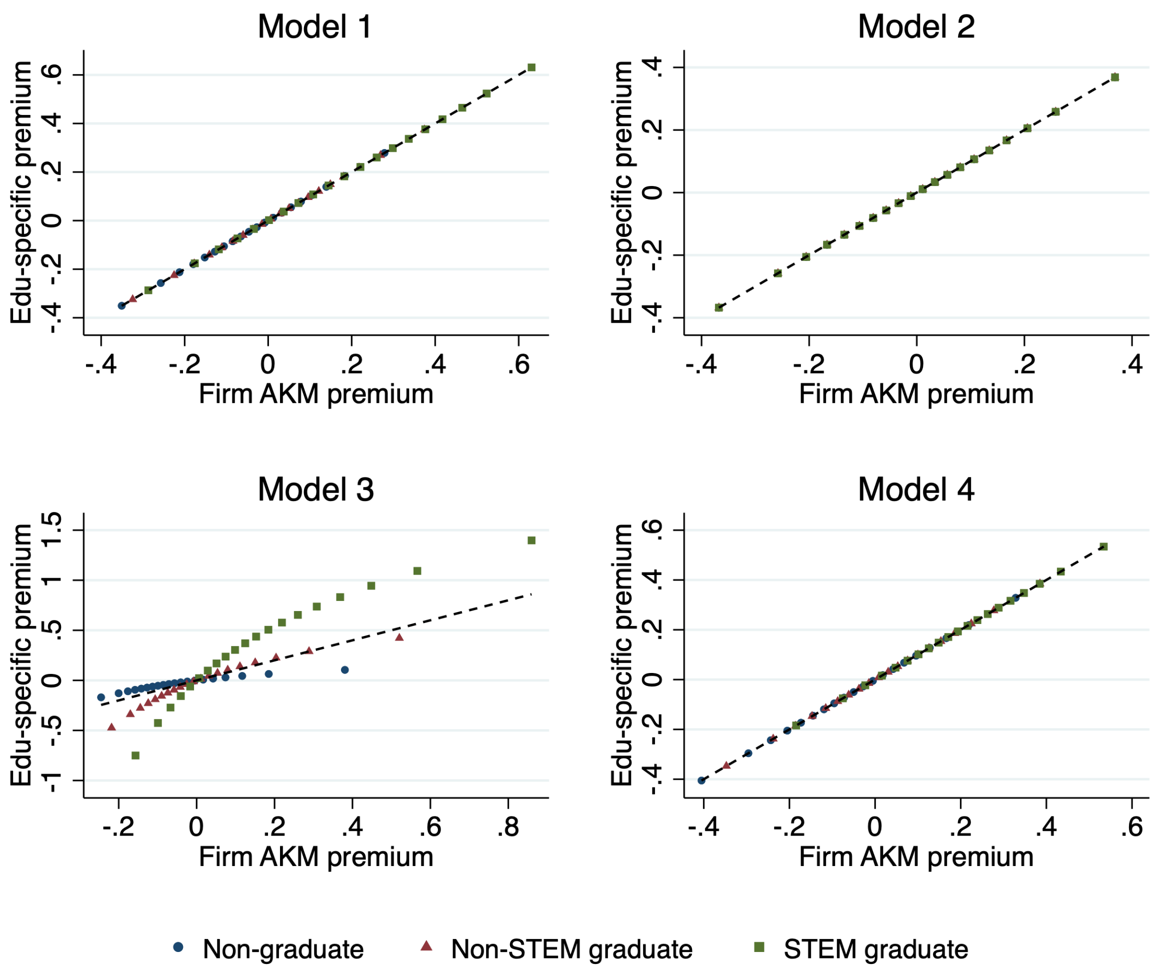
Notes: Figure shows mean log firm employment across 20 bins (each with equal numbers of firms), arranged by AKM firm premia, separately for the four models described in the text. Firm premia are normalized to their worker-weighted mean.

Figure 6: Calibrations of education-specific employment by firm pay premium



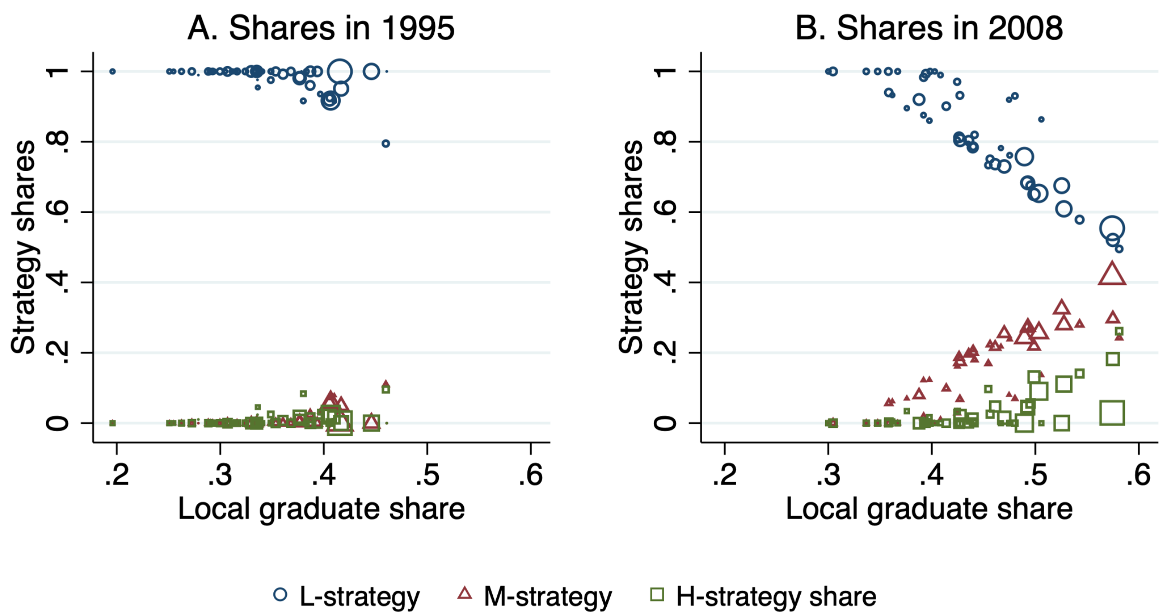
Notes: Figure shows log of mean firm employment by education group across 20 bins (each with equal numbers of firms), arranged by AKM firm premia, separately for the four models described in the text. Firm premia are normalized to their worker-weighted mean.

Figure 7: Calibrations of education-specific firm pay premia



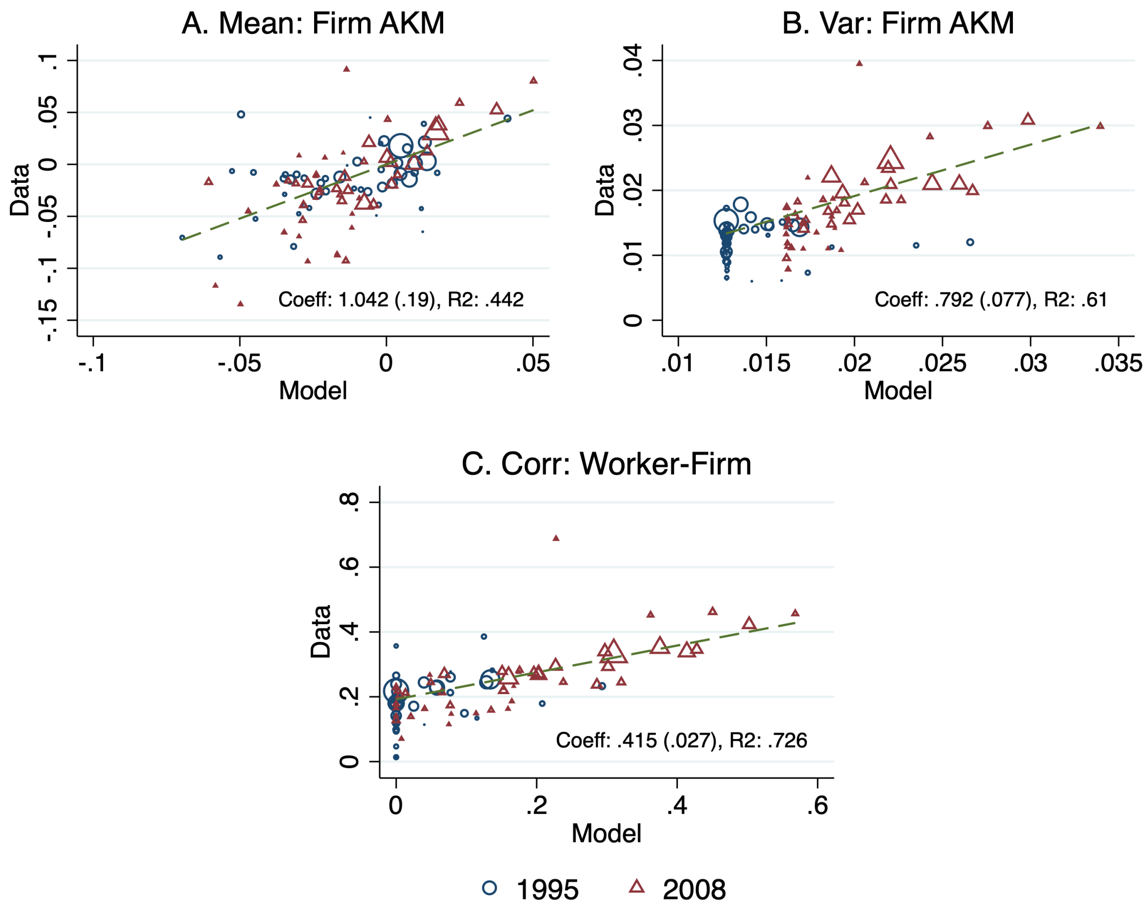
Notes: Figure shows education-specific firm wage premia across 20 bins (each with equal numbers of firms), arranged by aggregate AKM firm premia, separately for the four models described in the text. Firm premia are normalized to their worker-weighted mean.

Figure 8: 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 9: 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)	5.516	2.546	2.447	-
STEM-degree (ε_h)	-	-	-	8.762
Non-STEM (ε_m)	-	-	-	3.492
No degree (ε_l)	-	-	-	1.555
Productivity parameters				
Firm productivity variance (ν)	0.023	0.032	0.278	0.028
Log productivity (relative to STEM)				
Non-STEM intercept ($\log \frac{p_m}{p_h}$)	-0.641	-0.463	0.102	-0.170
No-degree intercept ($\log \frac{p_l}{p_h}$)	-1.057	-0.744	-0.075	-0.151
Non-STEM slope ($\theta_m - 1$)	-	-	-0.588	-
Non-degree slope ($\theta_l - 1$)	-	-	-0.874	-
Equity constraint parameters				
Non-STEM pay ratio ($\log \phi_m$)	-0.312	-	-	-
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}	2.56	2.57	2.55	ε	3.91	4.63	5.52
V_{AKMf}	0.026	0.035	0.032	ν	0.021	0.028	0.023
$\log \frac{\phi_m}{\phi_l}$	0.21	0.20	0.23	σ^M	0.08	0.14	0.17
$\log \frac{1}{\phi_l}$	0.22	0.40	0.54	σ^H	0.01	0.05	0.08
$E[\log w_m] - E[\log w_l]$	0.24	0.24	0.28	$\log \frac{w_h^M}{w_h^L}$	0.21	0.17	0.16
$E[\log w_h] - E[\log w_l]$	0.32	0.57	0.74	$\log \frac{w_h^H}{w_h^L}$	0.58	0.47	0.42
$\log \frac{n_m}{n_l}$	-0.76	-0.58	-0.61	$\log \frac{\phi_m^\varepsilon \Omega_m}{\phi_l^\varepsilon \Omega_l}$	-0.94	-0.90	-1.04
$\log \frac{n_h}{n_l}$	-2.35	-1.90	-1.88	$\log \frac{\Omega_h}{\phi_l^\varepsilon \Omega_l}$	-2.64	-2.57	-2.85
				Implied parameters			
				$\log \frac{p_m}{p_l}$	0.46	0.41	0.42
				$\log \frac{p_h}{p_l}$	0.95	0.99	1.06

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	0.387	0.044	0.490	0.065
Non-STEM graduate share	0.324	0.038	0.397	0.047
STEM graduate share	0.063	0.016	0.094	0.034
Mean: Log wage	8.981	0.113	9.086	0.137
Mean: Firm AKM	0.000	0.019	0.000	0.034
Mean: Worker AKM	0.000	0.098	0.000	0.107
Var: Firm AKM	0.014	0.002	0.020	0.005
Corr: Worker, firm	0.210	0.041	0.297	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	0.152	0.116	0.034	-0.010	0.112	-0.239
	(0.109)	(0.087)	(0.005)	(0.023)	(0.086)	(0.192)
STEM grad share	0.643	0.662	0.077	0.123	1.652	1.910
	(0.113)	(0.112)	(0.014)	(0.017)	(0.119)	(0.273)
Within-R^2	0.504	0.651	0.593	0.477	0.597	0.557
Panel B: Model						
Non-STEM grad share	0.195	0.199	-0.009	0.003	0.011	0.358
	(0.011)	(0.034)	(0.002)	(0.006)	(0.065)	(0.277)
STEM grad share	0.501	0.415	0.110	0.104	3.670	3.654
	(0.015)	(0.031)	(0.004)	(0.007)	(0.226)	(0.229)
Within-R^2	0.961	0.875	0.935	0.886	0.906	0.825
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 (17). Even-numbered columns control for region fixed effects, as in equation (18), 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						
Graduate share	0.689	0.303	0.488	0.156	0.176	0.170
	(0.098)	(0.234)	(0.079)	(0.206)	(0.028)	(0.063)
Within-R^2	0.330	0.067	0.250	0.021	0.362	0.261
Panel B: STEM graduates v non-graduates						
Graduate share	1.687	0.918	1.146	0.481	0.523	0.425
	(0.244)	(0.593)	(0.215)	(0.465)	(0.054)	(0.136)
Within-R^2	0.434	0.108	0.338	0.046	0.536	0.252
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Region FE	No	Yes	No	Yes	No	Yes
Sample	98	98	98	98	98	98

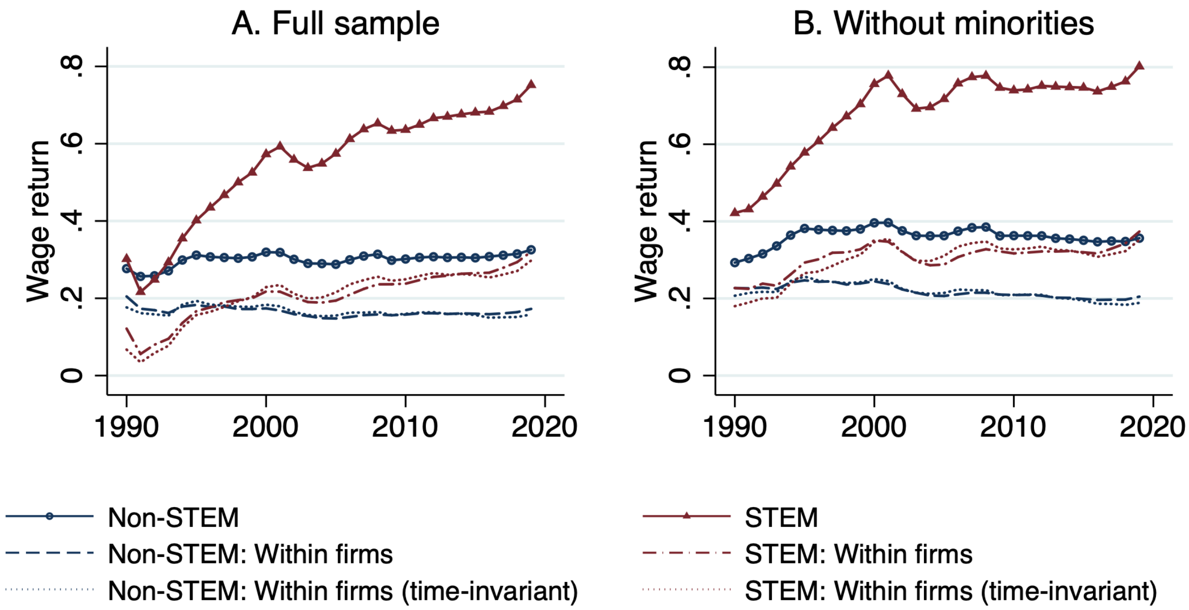
Notes: Table shows the relationship between regional graduate share and local skill returns. Odd-numbered columns exploit cross-sectional variation across regions, using equation (17). Even-numbered columns control for region fixed effects, as in equation (18), relying on within-region changes for identification. 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	0.673	0.163
$L + M$ strategies	0.020	0.347
$L + H$ strategies	0.122	0.000
$L + M + H$ strategies	0.184	0.490

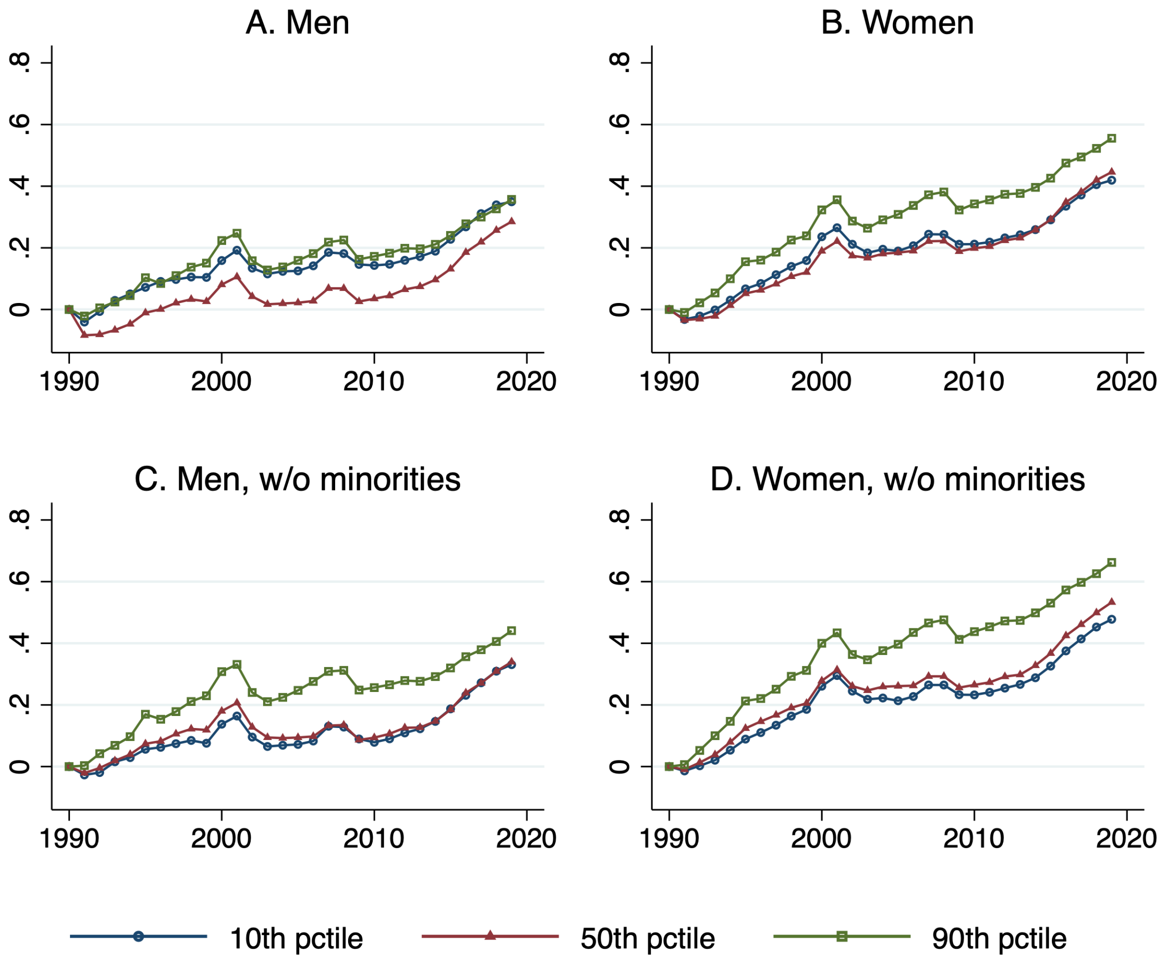
Notes: This table reports shares of regions with each equilibrium configuration, by census year. The first row shows the share of regions populated exclusively by inclusive L-strategy firms (i.e., hiring all skill types). Regions in the second row contain a mix of L- and M-strategy firms, and those in 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: Additional detail on education wage returns



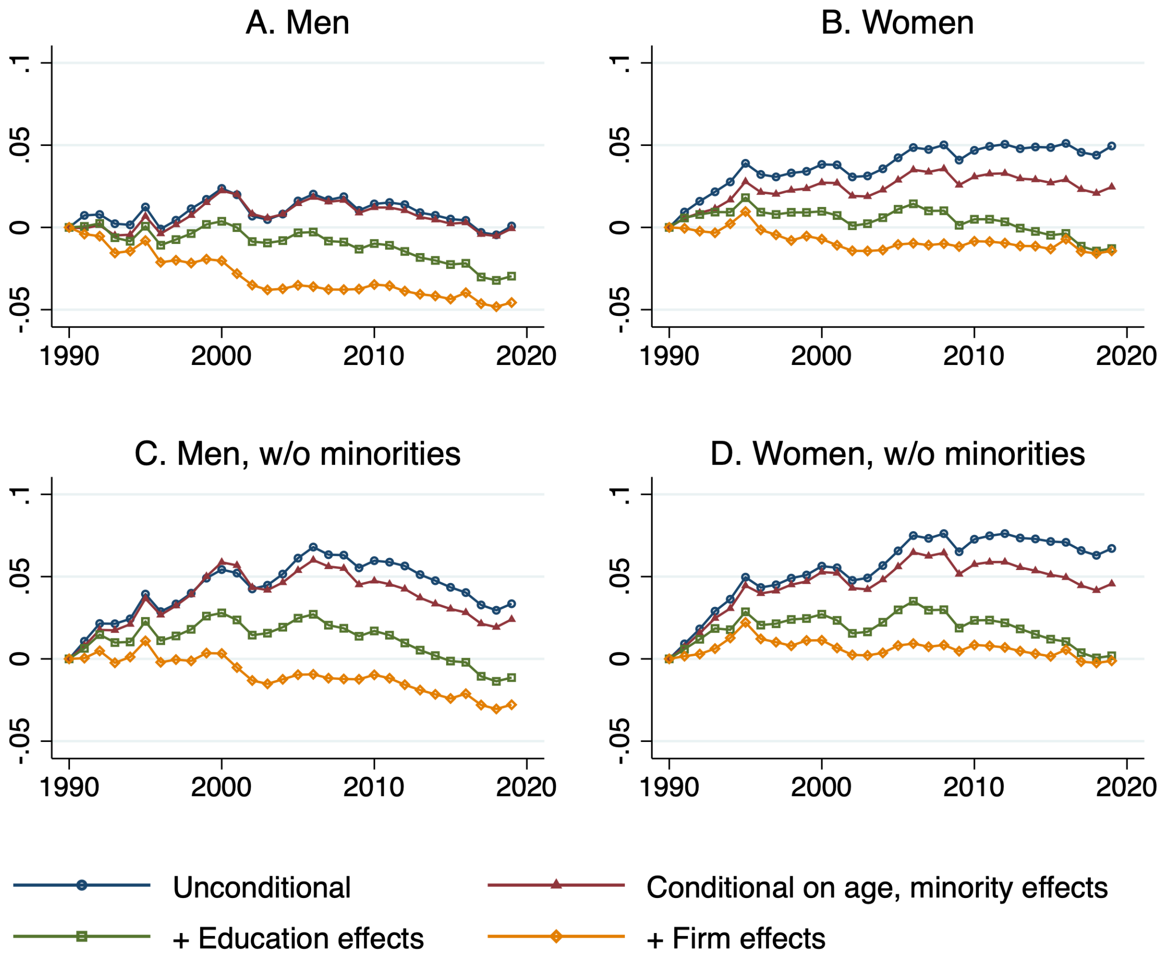
Notes: This figure offers additional detail on the evolution of education wage returns, to complement Figure 1 in the main text. As before, the filled lines show wage returns which condition on basic demographics only: interactions between gender, minority effects (for Arabs, ultra-orthodox Jews and FSU immigrants) and an age cubic, separately by year. The dashed lines are within-firm returns, which condition additionally on firm fixed effects (again, separately by year). The dotted lines are new to this figure: these condition on time-invariant firm fixed effects (i.e., with values with fixed across all years). To implement this, 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). Panel B replicates this exercise, but for a restricted sample which excludes key minority groups: Arabs, ultra-orthodox Jews, and FSU immigrants.

Figure A2: Percentiles of real earnings over time



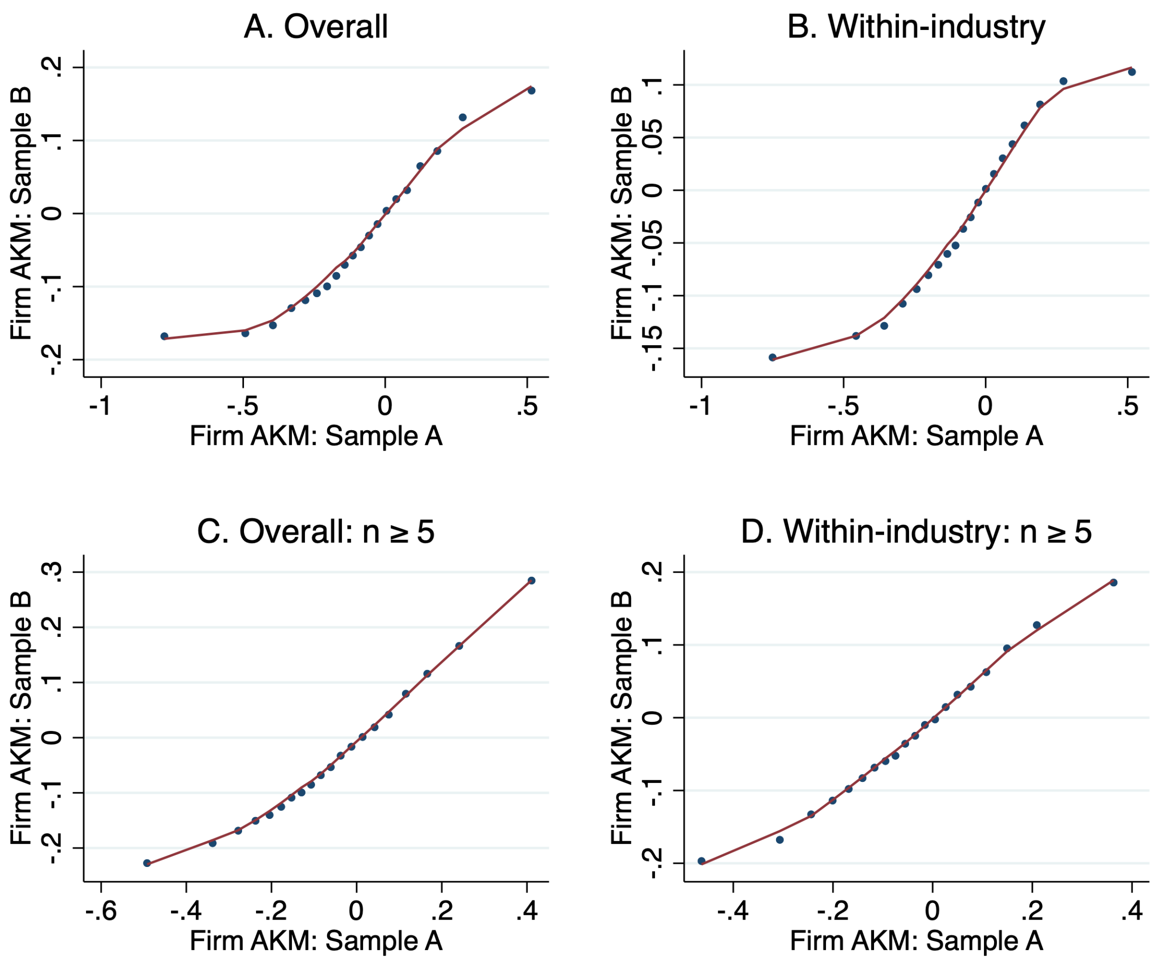
Notes: This figure shows the evolution of percentiles of real earnings from 1990 to 2019, separately for different sub-samples of the workforce. Panels A and B show trends for men and women respectively. Panels C and D exclude Arabs, ultra-orthodox Jews and FSU immigrants from the sample.

Figure A3: Standard deviations of residualized log earnings



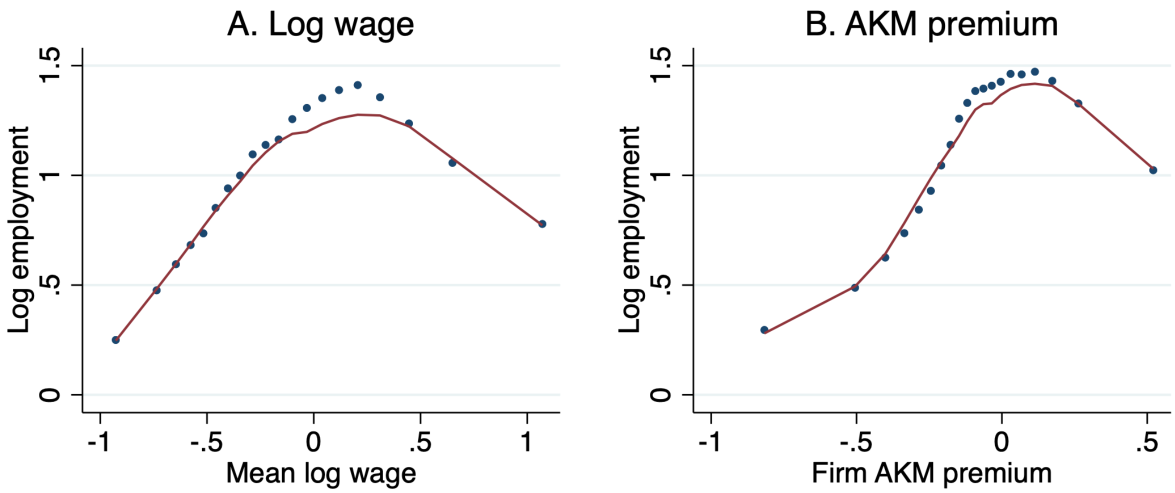
Notes: This figure shows the evolution of standard deviations of residualized log earnings from 1990 to 2019. Panel A and B show trends for men and women respectively, and Panels C and D exclude Arabs, ultra-orthodox Jews, and FSU immigrants from the sample. Each line shows standard deviations after 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 A4: First stage for split-sample correction



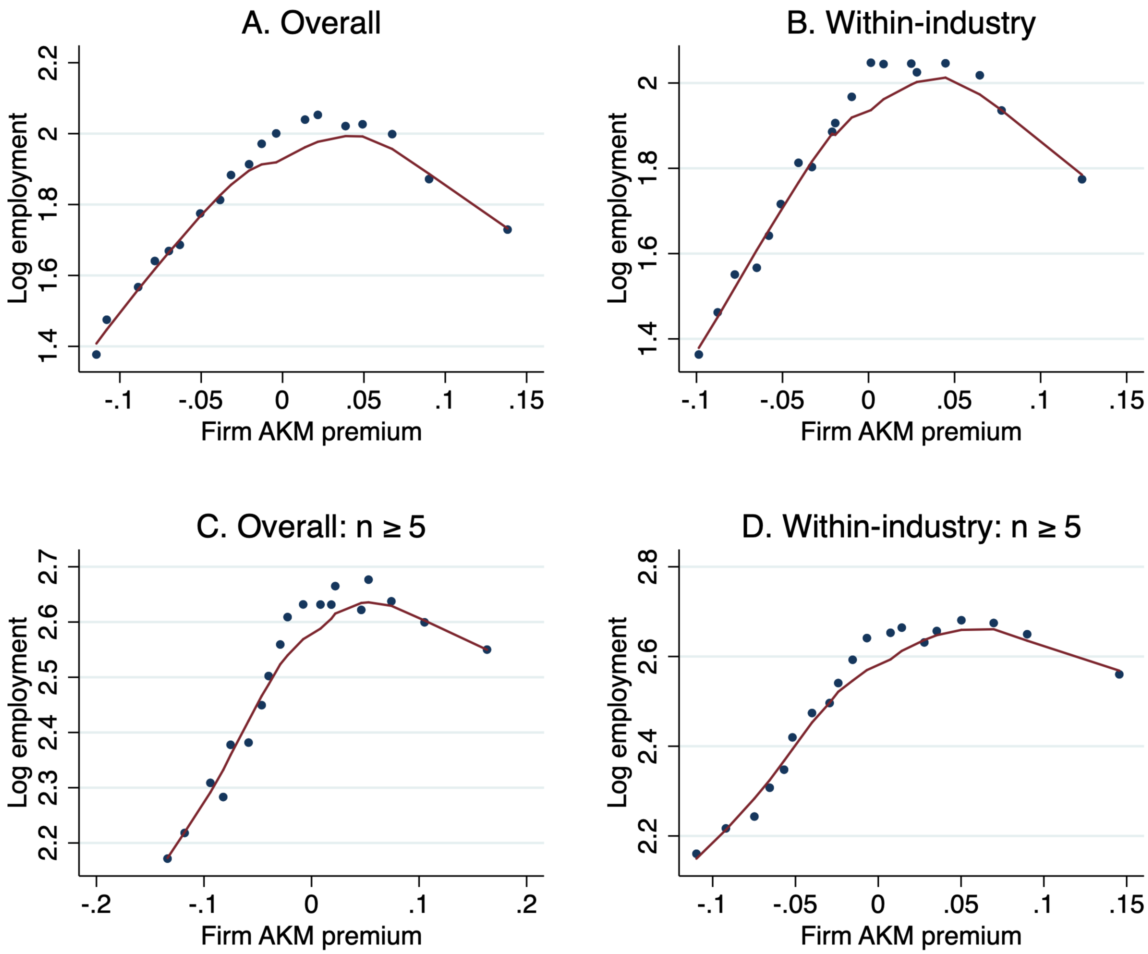
Notes: This figure shows the relationship between the AKM firm effects in the two random worker samples. This can be interpreted as a "first stage" relationship for our split-sample correction in the main text. In Panel B, we remove industry fixed effects from both the y-variable and the x-variable.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

Figure A5: Employment by mean firm wage 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 A6: 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 (11)

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 (7), and $\beta = \phi \frac{p_h}{p_l}$ from (9), 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 the first equation in (11); and the second 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: $l_l < l_l(\phi w_h)$. Since marginal products are fixed, 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:

$$(p_h - w_h)\, l_h'(w_h) = l_h(w_h) \tag{B5}$$

Using (2), this implies:

$$w_h^S = \frac{\varepsilon}{1+\varepsilon} p_h = w_h^* \tag{B6}$$

where w_h^* is the optimal unconstrained wage.

B.3 Derivation of equilibrium equations (12) and (14)

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:

$$\Omega_h = \frac{n_h}{k}\left[\sigma\left(w_h^S\right)^{\varepsilon} + (1-\sigma)\left(w_h^I\right)^{\varepsilon}\right]^{-1} \tag{B7}$$

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:

$$\Omega_l = \frac{n_l}{k}\left[(1-\sigma)\left(\phi w_h^I\right)^{\varepsilon}\right]^{-1} \tag{B8}$$

where n_l is the measure of l-type workers. Since l-type workers cannot access selective firms, the outside option in (B8) only accounts for inclusive firms.

Using the definitions of β and α in equations (9) and (13), the ratio of the two intercepts can be written as:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha}\cdot\frac{\beta}{\phi^{1+\varepsilon}}\left[1 + \frac{\sigma}{1-\sigma}\left(\frac{w_h^S}{w_h^I}\right)^{\varepsilon}\right] \tag{B9}$$

Finally, replacing w_h^I and w_h^S with (B3) and (B6), we have:

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

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:

$$\pi^I = \frac{\varepsilon^{\varepsilon}}{(1+\varepsilon)^{1+\varepsilon}}\cdot\frac{\left(1+\frac{\phi^{1+\varepsilon}}{\beta}\cdot\frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon}}{\left(1+\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}\right)^{\varepsilon}}\cdot\Omega_h p_h^{1+\varepsilon} \tag{B11}$$

Similarly, inserting the optimal selective wage (B6) into equation (B4), and replacing $l_h(w_h)$ with (2), the selective profit can be written as:

$$\pi^S = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \Omega_h p_h^{1+\varepsilon} \tag{B12}$$

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 (B11) and (B12), this implies:

$$\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon} > \left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^\varepsilon \tag{B13}$$

But imposing $\sigma = 0$ on (B10), we have:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha} \cdot \frac{\beta}{\phi^{1+\varepsilon}} \tag{B14}$$

Applying this to (B13) yields:

$$\beta < \frac{\left(\frac{1}{\alpha}\right)^{\frac{1}{\varepsilon}} - \alpha}{1-\alpha} \tag{B15}$$

which is the threshold condition for a $\sigma = 0$ equilibrium in equation (12).

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 (B11) and (B12), this implies:

$$\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon} = \left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^\varepsilon \tag{B16}$$

Imposing this on (B10) yields:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha-\sigma} \cdot \frac{\beta}{\phi^{1+\varepsilon}} \tag{B17}$$

And replacing $\frac{\Omega_l}{\Omega_h}$ in equation (B16) with (B17):

$$\left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{1+\varepsilon} = \left(1 + \beta \frac{1-\alpha}{\alpha-\sigma}\right)^\varepsilon \tag{B18}$$

This is an implicit equation which solves for $\tilde{\sigma}$ in equation (14), 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

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:

$$E[\log w_h|\beta > 1] = \frac{(1-\sigma)\, l_h\left(w_h^I\right)\log w_h^I + \sigma l_h\left(w_h^S\right)\log w_h^S}{(1-\sigma)\, l_h\left(w_h^I\right) + \sigma l_h\left(w_h^S\right)} \tag{B19}$$

In a counterfactual unconstrained economy, all firms offer h-types the unconstrained optimum w_h^*, as defined by equation (7). As Appendix B.2 shows, the optimal selective wage w_h^S is equal to w_h^*. Using (B19), the impact of the equity constraint can then be written as:

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

As (11) shows, the inclusive wage w_h^I is less than the unconstrained optimum w_h^*. And thus, $E[\log w_h] - \log w_h^* < 0$: i.e., the equity constraint reduces the expected log h-type wage.

We next turn to l-type wages. In the presence of the equity constraint, l-types always earn the inclusive wage w_l^I, since selective firms deny them access. Using the notation above, the impact of the pay constraint (relative to the unconstrained optimum) can then be written as:

$$E[\log w_l|\beta > 1] - \log w_l^* = \log\frac{w_l^I}{w_l^*} \tag{B21}$$

From equation (11), the inclusive wage w_l^I must exceed the unconstrained optimum w_l^*. So (B21) 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) and there is no unemployment, 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 (7) to (3), the labor supply intercepts for skill type s will equal:

$$\Omega_s^* = \left(\frac{1+\varepsilon}{\varepsilon}\right)^{\varepsilon}\frac{n_s}{k}\cdot p_s^{-\varepsilon} \tag{B22}$$

Using this expression, the binding labor supply curve (2) and optimal wage (7), the profit function (4) can be written as:

$$\pi^* = \frac{1}{\alpha} \cdot \frac{n_h}{k} \cdot \frac{p_h}{1+\varepsilon} \tag{B23}$$

where the h-type output share α is defined by (13).

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 (B11):

$$\pi|\beta > 1 = \frac{\varepsilon^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}} \cdot \frac{\left(1 + \frac{\phi^{1+\varepsilon}}{\beta} \cdot \frac{\Omega_l}{\Omega_h}\right)^{1+\varepsilon}}{\left(1 + \phi^{1+\varepsilon} \frac{\Omega_l}{\Omega_h}\right)^{\varepsilon}} \cdot \Omega_h p_h^{1+\varepsilon} \tag{B24}$$

Inserting expressions for the h-type labor supply intercept Ω_h from (B7), the equilibrium intercept ratio $\frac{\Omega_l}{\Omega_h}$ from (B17), the optimal unconstrained wage w_h^* from (7), and the optimal inclusive wage w_h^I from (B3), this can be written as:

$$\pi|\beta > 1 = \frac{\left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{1+\varepsilon}}{\sigma\left(1 + \beta\frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon} + (1-\sigma)\left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon}} \cdot \frac{n_h}{k} \cdot \frac{p_h}{1+\varepsilon} \tag{B25}$$

In a non-segregated equilibrium, the selective share σ will equal zero; and equation (B25) will collapse to the unconstrained profit π^* in (B23). In a partially segregated equilibrium, the equal profit condition in equation (B18) ensures that $\left(1 + \beta\frac{1-\alpha}{\alpha-\sigma}\right)^{\varepsilon} = \left(1 + \frac{1-\alpha}{\alpha-\sigma}\right)^{1+\varepsilon}$; so again, equation (B25) 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: Expected amenity effects

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

$$\bar{u}_s = \log \int_f w_{sf}^{\varepsilon} df + \gamma \tag{B26}$$

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:

$$\bar{a}_s = \log \int_f w_{sf}^{\varepsilon} df - \varepsilon E\left[\log w_s\right] + \gamma \tag{B27}$$

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. Since all firms pay the unconstrained optimum w_h^*, the h-type $\bar{a}_h^*$ 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 (B27) and (B20), this can be written as:

$$\bar{a}_h|\beta > 1 = \log\left[(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma\right] - \frac{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}}{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \sigma} \log\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon} + \gamma \tag{B28}$$

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

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

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 (B29) 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.

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 (B27) and (B21), this can be written as:

$$\bar{a}_l|\beta > 1 = \log(1-\sigma)\, w_l^I - \log w_l^I + \gamma \tag{B30}$$

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

$$(\bar{a}_l|\beta > 1) - \bar{a}_l^* = \log(1-\sigma) \tag{B31}$$

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.

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

In the baseline model with productively identical firms, Proposition 2a states that high-pay (i.e., selective) firms will have lower employment overall. To prove this result, we derive expressions for firm size for the selective and inclusive strategies.

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

$$l_h\left(w_h^S\right) = \Omega_h\left(w_h^*\right)^{\varepsilon} \tag{B32}$$

where Ω_h is the h-type labor supply intercept.

Inclusive firms employ both h- and l-types, at wages w_h^I and $w_l^I = \phi w_h^I$ respectively. Using equation (2) and the inclusive wage w_h^I in (B3), their firm size can be written as:

$$l_h\left(w_h^I\right)+l_l\left(w_l^I\right) = \Omega_h\left(w_h^I\right)^{\varepsilon}+\Omega_l\left(\phi w_h^I\right)^{\varepsilon} = \left(1+\frac{\Omega_l}{\Omega_h}\phi^{\varepsilon}\right)\left(\frac{1+\frac{1}{\beta}\cdot\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}{1+\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}\right)^{\varepsilon}\Omega_h\left(w_h^*\right)^{\varepsilon} \tag{B33}$$

Since we are comparing selective and non-selective firms, we must be in an equilibrium with positive selective share ($\sigma > 0$); so the equal profits condition in (B16) will apply. Imposing this on (B33):

$$l_h\left(w_h^{NS}\right)+l_l\left(w_l^{NS}\right) = \frac{1+\frac{\phi^{\varepsilon}\Omega_l}{\Omega_h}}{1+\frac{\phi}{\beta}\cdot\frac{\phi^{\varepsilon}\Omega_l}{\Omega_h}}\Omega_h\left(w_h^*\right)^{\varepsilon} \tag{B34}$$

Since $\frac{\phi}{\beta} = \frac{p_l}{p_h} < 1$, this expression must exceed the selective firm size (B32). This confirms that selective firms have lower employment overall.

B.7 Derivation of equation (15): Expected skill differential

The expected log h-type wage is given by equation (B19). And since l-types are denied access to selective firms, they all receive the inclusive wage $w_l^I = \phi w_h^I$. Subtracting one from the other, the expected skill differential is:

$$E\left[\log w_h\right]-E\left[\log w_l\right] = -\log\phi w_h^I+\frac{(1-\sigma)\,l_h\left(w_h^I\right)\log w_h^I+\sigma l_h\left(w_h^S\right)\log w_h^S}{(1-\sigma)\,l_h\left(w_h^I\right)+\sigma l_h\left(w_h^S\right)} \tag{B35}$$

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

$$E\left[\log w_h\right]-E\left[\log w_l\right] = -\log\phi+\frac{\sigma}{(1-\sigma)\left(\frac{w_h^I}{w_h^*}\right)^{\varepsilon}+\sigma}\log\frac{w_h^*}{w_h^I} \tag{B36}$$

To derive an expression for the $\frac{w_h^I}{w_h^*}$ term, we use (B17) to eliminate $\frac{\Omega_l}{\Omega_h}$ from equation (B3). This yields:

$$\frac{w_h^I}{w_h^*} = \frac{1 + \frac{1-\alpha}{\alpha-\sigma}}{1 + \beta \frac{1-\alpha}{\alpha-\sigma}} = \left(\frac{\alpha - \sigma}{1 - \sigma} \right)^{\frac{1}{\varepsilon}} \tag{B37}$$

where the second equality follows from (14). And finally, after using (B37) to eliminate $\frac{w_h^I}{w_h^*}$ from equation (B36), we reach equation (15) in the main text.

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

Suppose first that $\beta < \frac{(1/\alpha)^{1/\varepsilon} - \alpha}{1-\alpha}$. The selective share σ is then fixed at zero (see (12)), and all firms adopt the inclusive strategy. Replacing $\frac{\Omega_l}{\Omega_h}$ with (B14) in equation (B3), the h-type inclusive wage will equal:

$$w_h^I = \frac{1}{\alpha + (1-\alpha)\beta} \cdot w_h^* \tag{B38}$$

which is increasing in the h-type productivity p_h, both via α and via the optimal unconstrained wage term w_h^*. But since the equity constraint binds, the l-type wage is a fixed share ϕ of w_h^I. Therefore, any productive benefits of larger p_h are shared equally with l-types.

Now suppose instead that $\beta \geq \frac{(1/\alpha)^{1/\varepsilon} - \alpha}{1-\alpha}$; so the selective share σ exceeds zero: see (12). For the purposes of this proof, it is useful to define the function $\Lambda(\beta, \varepsilon)$ as the solution of the implicit equation:

$$(1 + \Lambda)^{1+\varepsilon} = (1 + \beta\Lambda)^{\varepsilon} \tag{B39}$$

This is identical to the equilibrium equation (B18), except with $\frac{1-\alpha}{\alpha-\sigma}$ replaced by Λ, which exceeds zero if $\sigma > 0$. Using this definition, we can summarize equilibrium by:

$$\Lambda(\beta, \varepsilon) = \frac{1-\alpha}{\alpha - \sigma} \tag{B40}$$

But since Λ is fixed by the exogenous parameters β and ε (and invariant to α), equation (B40) implies that σ must be increasing in α. This proves the first part of the proposition.

Next, consider the between-firm component in equation (15), which is equal to $\frac{\sigma}{\alpha} \log \left(\frac{1-\sigma}{\alpha-\sigma} \right)^{\frac{1}{\varepsilon}}$. Using (B40), this can be re-written as:

$$\text{Between-firm} = \left[1 - \frac{1-\alpha}{\alpha} \cdot \frac{1}{\Lambda(\beta, \varepsilon)} \right] \log \left(1 + \Lambda(\beta, \varepsilon) \right)^{\frac{1}{\varepsilon}} \tag{B41}$$

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

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 $\frac{(1/\alpha)^{1/\varepsilon}-\alpha}{1-\alpha}$; so the equilibrium selective share σ exceeds 0.

C.1 Characterization of equilibrium with heterogeneous firms

We begin by characterizing equilibrium in this extended model. Building from equation (7), 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 (B3), 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 (B11) and (B12), 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 (B11) and (B12), 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.

Using this result, we now 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. $\tilde{w}_h^I \sim F^I$. Given the orthogonality between firm productivity x and strategy choice, it 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 $\tilde{w}_h^I(1) = \log \frac{1+\frac{1}{\beta}\cdot\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}}{1+\phi^{1+\varepsilon}\frac{\Omega_l}{\Omega_h}} \frac{\varepsilon}{1+\varepsilon} p_h$. Hence, both the F^S and F^I distributions have identical variance (equal to ν, the same as for firm productivity x), but inclusive firms offer lower pay on average.

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 ε exponent) of selective firms (weighted by the selective firm share σ) and inclusive firms (weighted $1-\sigma$). The second line follows from (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-types can be written as:

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

Putting these together, the intercept ratio is identical to equation (B17) in the baseline model:

$$\frac{\Omega_l}{\Omega_h} = \frac{1-\alpha}{\alpha}\cdot\frac{\beta}{\phi^{1+\varepsilon}}\left[1 + \frac{\sigma}{1-\sigma}\left(\frac{w_h^I(1)}{w_h^I(1)}\right)^{\varepsilon}\right] = \frac{1-\alpha}{\alpha-\sigma}\cdot\frac{\beta}{\phi^{1+\varepsilon}} \tag{C9}$$

And hence, the equilibrium selective share σ will take an identical form to the baseline model, as specified by equation (B18).

Finally, let $\kappa = \tilde{w}_h^S(x) - \tilde{w}_h^I(x)$ denote the pay differential between equally productive inclusive and selective firms. For the reasons explained above, this differential is independent of x. Inserting (C9) into (C3), the κ differential can be written as:

$$\kappa = \tilde{w}_h^S(x) - \tilde{w}_h^I(x) = \log\left[1 + \frac{1-\alpha}{1-\sigma}(\beta - 1)\right] \tag{C10}$$

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 here is that firm pay may vary for two reasons: (i) heterogeneity in productivity x and (ii) choice of selective or inclusive strategy. It is the mixture of (i) and (ii) that produces the shape described by Proposition 2b.

It is first useful to define the selective share $\sigma(\tilde{w}_h)$, among firms which pay h-types $\tilde{w}_h$:

$$\begin{aligned}\sigma(\tilde{w}_h) &= \frac{\sigma f^S(\tilde{w}_h)}{(1-\sigma) f^I(\tilde{w}_h) + \sigma f^S(\tilde{w}_h)} = \frac{\sigma f^S(\tilde{w}_h)}{(1-\sigma) f^S(\tilde{w}_h+\kappa) + \sigma f^S(\tilde{w}_h)} \\ &= \left[\frac{1-\sigma}{\sigma}\exp\left(-\frac{\kappa}{\nu^2}\left(\tilde{w}_h - \tilde{w}_h^S(1) + \frac{1}{2}\kappa\right)\right) + 1\right]^{-1}\end{aligned} \tag{C11}$$

where σ is the unconditional selective share, and $\tilde{w}_h^S(1) = \log\frac{\varepsilon}{1+\varepsilon}p_h$. The second equality follows from the definitions of F^S and F^I in (C6) and (C7), and the definition of κ in (C10): 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 $\tilde{w}_h^S(1)$ and $\tilde{w}_h^S(1) - \kappa$ respectively, and variance ν^2. Equation (C11) shows that the selective share $\sigma(\tilde{w}_h)$ 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 greater the representation of selective firms.

Next, we consider how log firm employment varies over the firm pay distribution. Let $E[\log l|\tilde{w}_h]$ 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 $\tilde{w}_h$:

$$E[\log l|\tilde{w}_h] = \sigma(\tilde{w}_h) E\left[\log l^S|\tilde{w}_h\right] + [1-\sigma(\tilde{w}_h)] \log\left[\log l^I|\tilde{w}_h\right] \tag{C12}$$

where $\sigma(\tilde{w}_h)$ is defined by (C11). Since selective firms recruit only h-types, their expected employment is given by the h-type labor supply curve. For inclusive firms, expected employment is given by the sum of the h- and l-type labor supply curves. So we have:

$$E[\log l|\tilde{w}_h] = \sigma(\tilde{w}_h)\log l_h\left(e^{\varepsilon\tilde{w}_h}\right) + [1-\sigma(\tilde{w}_h)]\log\left[l_h\left(e^{\varepsilon\tilde{w}_h}\right) + \log l_l\left(\phi^\varepsilon e^{\varepsilon\tilde{w}_h}\right)\right] \tag{C13}$$

Inserting the labor supply curve (2) and rearranging:

$$E[\log l|\tilde{w}_h] = \log(\Omega_h + \Omega_l\phi^\varepsilon) + \varepsilon\tilde{w}_h - \log\left(1 + \frac{\Omega_l\phi^\varepsilon}{\Omega_h}\right)\sigma(\tilde{w}_h) \tag{C14}$$

The first term on the right-hand side is a constant. The second term is increasing linearly in $\tilde{w}_h$, with slope ε: this is the contribution of the upward-sloping 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:

$$\frac{d}{d\tilde{w}_h}E[\log l|\tilde{w}_h] = \varepsilon - \frac{\kappa}{\nu^2}\log\left(1+\frac{\Omega_l\phi^{\varepsilon}}{\Omega_h}\right)\sigma(\tilde{w}_h)[1-\sigma(\tilde{w}_h)] \tag{C15}$$

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:

$$\frac{d^2}{d\tilde{w}_h^2}E[\log l|\tilde{w}_h] = -\frac{\kappa}{\nu^2}\log\left(1+\frac{\Omega_l\phi^{\varepsilon}}{\Omega_h}\right)\sigma(\tilde{w}_h)[1-\sigma(\tilde{w}_h)][1-2\sigma(\tilde{w}_h)] \tag{C16}$$

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.

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. Intuitively, since κ is the pay differential between selective and inclusive firms, it determines the relative dominance of the "quality motive". Conversely, ν^2 is the firm productivity variance, and this determines the salience of the "quantity motive". So as the $\frac{\kappa}{\nu^2}$ ratio grows, firms become more willing to trade-off quantity for quality; and hence the greater curvature of expected l-type employment, $E[\log l|\tilde{w}_h]$.

D Extension with CES technology

In the baseline model, h- and l-types are perfect substitutes. As Section 2.2 shows, this means that selective firms do not hire any l-types: i.e. perfect rationing. More generally though, if h- and l-types are imperfect substitutes, selective firms may 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 with no search frictions: some l-types will be fortunate to find work in selective firms, but others not. To explore this more formally, we now study the case of CES technology. This delivers some new insights, but the key intuitions from the baseline model are otherwise preserved.

Suppose firms produce output y (with price normalized to 1) according to the following CES technology:

$$y(l_h, l_l) = p\left(\theta l_h^{\gamma} + (1-\theta) l_l^{\gamma}\right)^{\frac{1}{\gamma}} \tag{D1}$$

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. This specification delivers the baseline model, if we set $\gamma = 1$, $p_h = \theta p$ and $p_l = (1-\theta)\,p$. For simplicity, we assume that firms are ex ante identical, as in the baseline model.

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

$$\max_{w_h, w_l, l_h, l_l} \pi\,(w_h, w_l, l_h, l_l) = y\,(l_h, l_l) - w_h l_h - w_l l_l \tag{D2}$$

subject to the labor supply constraints:

$$l_h \leq l_h\,(w_h)\,, \qquad l_l \leq l_l\,(w_l) \tag{D3}$$

and the pay equity constraint:

$$w_l \geq \phi w_h \tag{D4}$$

D.1 Equilibrium if equity constraint does not bind

If the equity constraint does not bind, the labor supply constraints in (D3) must bind, i.e. $l_h^* = l_h\,(w_h^*)$ and $l_l^* = l_l\,(w_l^*)$. The optimal wages are then fixed mark-downs $\frac{\varepsilon}{1+\varepsilon}$ on the marginal products:

$$w_h^* = \frac{\varepsilon}{1+\varepsilon} y_h\,(l_h^*, l_l^*) = \frac{\varepsilon}{1+\varepsilon}\left[\left(\frac{1-\theta}{\theta}\right)^{\frac{1+\varepsilon}{1+\varepsilon(1-\gamma)}}\left(\frac{\Omega_l^*}{\Omega_h^*}\right)^{\frac{\gamma}{1+\varepsilon(1-\gamma)}} + 1\right]^{\frac{1-\gamma}{\gamma}} \theta^{\frac{1}{\gamma}} p \tag{D5}$$

$$w_l^* = \frac{\varepsilon}{1+\varepsilon} y_l\,(l_h^*, l_l^*) = \frac{\varepsilon}{1+\varepsilon}\left[\left(\frac{\theta}{1-\theta}\right)^{\frac{1+\varepsilon}{1+\varepsilon(1-\gamma)}}\left(\frac{\Omega_h^*}{\Omega_l^*}\right)^{\frac{\gamma}{1+\varepsilon(1-\gamma)}} + 1\right]^{\frac{1-\gamma}{\gamma}} (1-\theta)^{\frac{1}{\gamma}} p \tag{D6}$$

where $y_h\,(l_h, l_l) \equiv \frac{\partial y(l_h, l_l)}{\partial l_h}$ and $y_l\,(l_h, l_l) \equiv \frac{\partial y(l_h, l_l)}{\partial l_l}$ are 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 that these Ω_s^* intercepts are equal to $(w_s^*)^{-\varepsilon} n_s$ for skill $s \in \{h, l\}$. The optimal wage differential will then equal:

$$\frac{w_l^*}{w_h^*} = \frac{y_l\,(l_h^*, l_l^*)}{y_h\,(l_h^*, l_l^*)} = \left[\frac{1-\theta}{\theta}\left(\frac{\Omega_h^*}{\Omega_l^*}\right)^{1-\gamma}\right]^{\frac{1}{1+\varepsilon(1-\gamma)}} = \frac{1-\theta}{\theta}\left(\frac{n_h}{n_l}\right)^{1-\gamma} \tag{D7}$$

where n_s is the aggregate measure of type-s workers. From equation (D7), the equity constraint will therefore not bind if $\phi \leq \frac{1-\theta}{\theta}\left(\frac{n_h}{n_l}\right)^{1-\gamma}$.

D.2 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 (10). In equilibrium, firms will adopt one of two pay strategies, just as in the baseline model: inclusive or selective.

The intuition is the following. Just as in the baseline model, the equity constraint and the h-type supply constraint always bind. Firms must then pick between two options. The first is to set wages low enough such that the l-type supply constraint also binds, in which case firms only have direct control over wages (as both l-type and h-type employment are *supply*-rationed): this is the inclusive strategy. Alternatively, firms can set wages high enough such that l-type employment is *demand*-rationed, in which case they must exert direct control over l-type employment also: this is the selective strategy. We discuss each strategy in turn and then solve for the equilibrium share of firms that 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:

$$\lambda^I \equiv \frac{l_l^I}{l_h^I} = \frac{l_l\left(\phi w_h^I\right)}{l_h\left(w_h^I\right)} = \frac{\phi^\varepsilon \Omega_l}{\Omega_h} \tag{D8}$$

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 optimal wages, 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 these bind also). This simplifies the problem to:

$$\max_{w_h} \pi\left(w_h\right) = y\left(l_h\left(w_h\right), l_l\left(\phi w_h\right)\right) - w_h l_h\left(w_h\right) - \phi w_h l_l\left(\phi w_h\right) \tag{D9}$$

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

$$w_h^I = \frac{\left[(1-\theta)\left(\lambda^I\right)^\gamma + \theta\right]^{\frac{1}{\gamma}}}{1+\phi\lambda^I} \cdot \frac{\varepsilon}{1+\varepsilon} p \tag{D10}$$

The associated profit is:

$$\pi^I = \frac{1}{\varepsilon}\left(1+\phi\lambda^I\right)\Omega_h\left(w_h^I\right)^{1+\varepsilon} \tag{D11}$$

Selective strategy (S)

Selective firms hire all willing h-type workers (so the h-type supply curve binds), but freely choose l-type employment to maximize profit; so the l-type supply constraint need 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.

If selective firms only partially ration their employment of l-types, we require some tie-break rule to determine which l-types are "fortunate" (and are admitted to 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 supply constraints binds); but we allow firms to freely choose l-type employment l_l:

$$\max_{w_h, l_l} \pi\left(w_h, l_l\right) = y\left(l_h\left(w_h\right), l_l\right) - w_h l_h\left(w_h\right) - \phi w_h l_l \tag{D12}$$

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

$$w_h^S = \left[(1-\theta) + \theta\left(\lambda^S\right)^{-\gamma}\right]^{\frac{1-\gamma}{\gamma}} \frac{1-\theta}{\phi} \cdot p \tag{D13}$$

where

$$\lambda^S \equiv \frac{l_l^S}{l_h^S} = \frac{l_l^S}{l_h\left(w_h^S\right)} < \frac{l_l\left(\phi w_h^S\right)}{l_h\left(w_h^S\right)} = \frac{l_l\left(\phi w_h^I\right)}{l_h\left(w_h^I\right)} \equiv \lambda^I \tag{D14}$$

is the optimal ratio of l-type to h-type employment in selective firms. Since selective firms ration l-type employment, we have $l_l^S < l_l\left(w_l^S\right)$; and therefore, equation (D14) shows that λ^S is smaller than λ^I (note the penultimate equality follows from the log linearity of the labor supply function). The λ^S ratio is pinned down by the first order condition for l_l:

$$\frac{\theta}{1-\theta}\varepsilon\phi\left(\lambda^S\right)^{1-\gamma} = (1+\varepsilon) + \phi\lambda^S \tag{D15}$$

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

$$\pi^S = \frac{1}{\varepsilon}\left(1+\phi\lambda^S\right)\Omega_h\left(w_h^S\right)^{1+\varepsilon} \tag{D16}$$

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:

$$\Omega_h = \frac{n_h}{k}\left[(1-\sigma)\left(w_h^I\right)^{\varepsilon} + \sigma\left(w_h^S\right)^{\varepsilon}\right]^{-1} \tag{D17}$$

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:

$$\Omega_l = \frac{n_l}{k}\left[(1-\sigma)\left(\phi w_h^I\right)^{\varepsilon} + \frac{\lambda^S}{\lambda^I}\sigma\left(\phi w_h^S\right)^{\varepsilon}\right]^{-1} \tag{D18}$$

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 pins down the extent of rationing. Putting equations (D17) and (D18) together, we have:

$$\frac{\sigma}{1-\sigma} = \frac{\lambda^I - \frac{n_l}{n_h}}{\frac{n_l}{n_h} - \lambda^S}\left(\frac{w_h^I}{w_h^S}\right)^{\varepsilon} \tag{D19}$$

which pins down the equilibrium selective share σ. Since inclusive firms disproportionately hire l-types, we must have $\lambda^I > \frac{n_l}{n_h}$ and $\frac{n_l}{n_h} > \lambda^S$: 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 (D11) and (D16), this implies:

$$\left(\frac{w_h^I}{w_h^S}\right)^{1+\varepsilon} > \frac{1+\phi\lambda^S}{1+\phi\lambda^I} \tag{D20}$$

But imposing $\sigma = 0$ on (D19), we have:

$$\lambda^I = \frac{n_l}{n_h} \tag{D21}$$

Intuitively, since all firms are inclusive (and offer the same wages), their skill ratio λ^I must equal the aggregate ratio. Using equations (D10), (D13) and (D21), we can then re-write (D20) as:

$$\left(\frac{\left[(1-\theta)\left(\frac{n_l}{n_h}\right)^\gamma + \theta\right]^{\frac{1}{\gamma}}}{\left[(1-\theta)+\theta\left(\lambda^S\right)^{-\gamma}\right]^{\frac{1-\gamma}{\gamma}}} \cdot \frac{\phi}{1-\theta} \cdot \frac{\varepsilon}{1+\varepsilon}\right)^{1+\varepsilon} > \left(1+\phi\lambda^S\right)\left(1+\phi\frac{n_l}{n_h}\right)^{\varepsilon} \tag{D22}$$

where the selective firms' skill ratio λ^S is pinned down by the exogenous parameters in equation (D15). 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 (D10) and (D13). 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 (D11) and (D16), this implies:

$$\left(\frac{w_h^I}{w_h^S}\right)^{1+\varepsilon} = \frac{1+\phi\lambda^S}{1+\phi\lambda^I} \tag{D23}$$

Equilibrium can then be characterized by the following five equations: (D10), (D13), (D15), (D19) and (D23). 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.

Generalization of Proposition 2a: Negative firm size premium

In the baseline model, Proposition 2a shows that selective firms have lower employment. This result is preserved under CES technology, where $\sigma > 0$. To see why, notice that total employment in selective and inclusive firms can be written as:

$$l_h\left(w_h^S\right) + l_l\left(\phi w_h^S\right) = \Omega_h\left(w_h^S\right)^{\varepsilon}\left(1+\lambda^S\right) \tag{D24}$$

$$l_h\left(w_h^I\right) + l_l\left(\phi w_h^I\right) = \Omega_h\left(w_h^I\right)^{\varepsilon}\left(1+\lambda^I\right) \tag{D25}$$

respectively. Using equation (D23), the ratio of the two is equal to:

$$\frac{l_h\left(w_h^S\right) + l_l\left(\phi w_h^S\right)}{l_h\left(w_h^I\right) + l_l\left(\phi w_h^I\right)} = \left(\frac{1+\phi\lambda^S}{1+\phi\lambda^I}\right)^{\frac{1}{1+\varepsilon}} < 1 \tag{D26}$$

Since $\lambda^S < \lambda^I$, selective firms must have lower employment overall.

E Extension with N skill types

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:

$$\max_{\{w_s;l_s\}_{s=1}^N} \pi\left(w_1, ..., w_N; l_1, ..., l_N\right) = \sum_{s=1}^{N}\left(p_s - w_s\right) l_s \tag{E1}$$

where skill types are perfect substitutes, and they are ordered such that skill-specific productivity p_s is increasing in s. Firms are subject to labor supply constraints:

$$l_s \leq l_s\left(w_s\right) \tag{E2}$$

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

$$w_s \geq \phi_s w_N \tag{E3}$$

for every skill type s. We normalize ϕ_N to 1, so the Nth equation of (E3) is redundant. Analogously to the baseline model, we can also define the "bite" β_s of each equity constraint as:

$$\beta_s \equiv \phi_s \frac{p_N}{p_s} \tag{E4}$$

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.1 Equilibrium strategies

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

$$\log w_{sf} = \eta_f + \lambda_s \tag{E5}$$

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:

$$p_s - \phi_s w_N = \left(\frac{1}{\beta_s} - \frac{w_N}{p_N} \right) \phi_s p_N \tag{E6}$$

using equation (E4). 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 (E6) 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 s-type employment of strategy-z firms. 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.

Though firms are identical, they may choose different pay strategies in equilibrium—just as in the baseline model. 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:

$$\sum_z \sigma^z = 1 \tag{E7}$$

E.2 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 (E1) as:

$$\max_{w_N} \pi^z \left(w_N\right) = \sum_{s \geq z}^{N} \left(p_s - w_s\right) l_s \tag{E8}$$

The first-order condition is then:

$$\sum_{s \geq z} \phi_s \left(p_s - \phi_s w_N\right) l_s' \left(\phi_s w_N\right) = \sum_{s \geq z} \phi_s l_s \left(\phi_s w_N\right) \tag{E9}$$

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

$$w_N^z = \frac{\sum_{s \geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}}{\sum_{s \geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}} \cdot \frac{\varepsilon}{1+\varepsilon} p_N \tag{E10}$$

Finally, using (E8), optimal profit of strategy-k firms is:

$$\pi^z = \frac{\varepsilon^\varepsilon}{\left(1+\varepsilon\right)^{1+\varepsilon}} \cdot \frac{\left(\sum_{s \geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{1+\varepsilon}}{\left(\sum_{s \geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{\varepsilon}} \cdot \Omega_N p_N^{1+\varepsilon} \tag{E11}$$

E.3 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:

$$\Omega_s = \frac{n_s}{k} \left[\sum_{z \leq s} \sigma^z \left(\phi_s w_N^z\right)^\varepsilon\right]^{-1} \tag{E12}$$

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

$$\frac{\phi_s^\varepsilon \Omega_s}{\Omega_N} = \frac{\alpha_s}{\alpha_N} \cdot \frac{\beta_s}{\phi_s} \cdot \frac{\sum_z \sigma^z \left(\frac{w_N^z}{w_N^N}\right)^\varepsilon}{\sum_{z \leq s} \sigma^z \left(\frac{w_N^z}{w_N^N}\right)^\varepsilon} \tag{E13}$$

where

$$\alpha_s \equiv \frac{n_s p_s}{\sum_x n_x p_x} \tag{E14}$$

is the output share of s-type workers.

Also, from (E10), notice the optimal wage of strategy-N firms is:

$$w_N^N = \frac{\varepsilon}{1+\varepsilon} p_N \tag{E15}$$

So the relative wage $\frac{w_N^z}{w_N^N}$ in equation (E13) is equal to:

$$\frac{w_N^z}{w_N^N} = \frac{\sum_{s \geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}}{\sum_{s \geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}} \tag{E16}$$

for strategy $z < N$.

E.4 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). For all other strategies z, there are two possibilities. Either no firms adopt strategy z, so we have:

$$\sigma^z = 0 \tag{E17}$$

which 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 (E11), equal profits implies:

$$\frac{\left(\sum_{s \geq z} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{1+\varepsilon}}{\left(\sum_{s \geq z} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{\varepsilon}} = \frac{\left(\sum_{s} \frac{\phi_s}{\beta_s} \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{1+\varepsilon}}{\left(\sum_{s} \phi_s \cdot \frac{\phi_s^\varepsilon \Omega_s}{\Omega_N}\right)^{\varepsilon}} \tag{E18}$$

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 (E13) for strategies $z = 1, ..., N-1$; (ii) the relative wage equations (E16) for strategies $z = 1, ..., N-1$; (iii) one equilibrium condition, either (E17) or (E18), for every strategy $z = 2, ..., N$; and (iv) equation (E7), which ensures the strategy shares sum to 1.

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 equity constraint. The set-up 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 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)-(6). 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 $l_s(w_s)$, the supply of type-s workers to firms paying wage w_s. Suppose that all workers (whether employed or unemployed) draw job offers 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, suppose that workers receive a zero utility flow when unemployed. Since the offer rate λ does not vary by employment status, unemployed workers will then accept any positive wage offer.

Let $F_s(w_s)$ be the equilibrium distribution of offers for type-s workers, and let $G_s(w_s)$ be their distribution of realized wages. In equilibrium, $G_s(w_s)$ will of course depend on $F_s(w_s)$. 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 steady-state:

$$u_s \lambda F_s(w_s) n_s = \delta(1-u_s) G_s(w_s) n_s + \lambda(1-F_s(w_s))(1-u_s) G_s(w_s) n_s \tag{F1}$$

where n_s is the measure of type-s workers, and u_s is their unemployment rate. The type-s inflow to this group of firms, on the left-hand side of (F1), is composed exclusively of the unemployed. And the outflow, on the right, consists of two components: (i) separations to unemployment (at rate δ), and (ii) quits to firms which pay above w_s. Note that (F1) is only defined for wages w_s below productivity p_s, as firms will never employ workers at a loss.

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

$$u_s = \frac{\delta}{\delta + \lambda F_s(p_s)} \tag{F2}$$

The job finding rate (out of unemployment) is equal to $\lambda F_s(p_s)$, because only those firms which offer wages below p_s will employ type-s workers. Substituting this into (F1) and rearranging gives:

$$G_s(w_s) = \frac{\delta}{\delta + \lambda [1 - F(w_s)]} \cdot \frac{F_s(w_s)}{F_s(p_s)} \tag{F3}$$

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

$$\frac{\lambda}{k} u_s n_s + \frac{\lambda}{k} (1 - u_s) G_s(w_s) n_s = [\delta + \lambda (1 - F_s(w_s))] l_s(w_s) \tag{F4}$$

The left-hand side shows the total inflow of type-s workers 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 less than w_s. The right-hand side of (F4) shows the total outflow from this firm, which consists of separations to unemployment (at rate δ) and quits to firms which pay more than w. Using (F2) and (F3), this steady-state condition implies:

$$l_s(w_s) = \frac{\delta + \lambda}{\delta + \lambda F_s(p_s)} \cdot \frac{\delta \lambda}{[\delta + \lambda (1 - F_s(w_s))]^2} \cdot \frac{n_s}{k} \tag{F5}$$

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, just as in Burdett and Mortensen (1998): i.e. $l_s = l_s(w_s)$ for each skill type s. The firm's problem can then be simplified to:

$$\max_{w_h, w_l} \pi(w_h, w_l) = \max_{w_h} \pi_h(w_h) + \max_{w_l} \pi_l(w_l) \tag{F6}$$

where $\pi_s(w_s)$ is the profit earned from skill type s:

$$\pi_s(w_s) = (p_s - w_s) l_s(w_s) \tag{F7}$$

We therefore have a distinct Burdett-Mortensen model for each skill type, which can be solved in the usual way.

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 allow them to increase employment discretely, as they poach workers from the mass point). Additionally, the lowest offer in $F_s(w_s)$ must equal the reservation wage of unemployed workers (which we have assumed to be zero); otherwise, the lowest-paying firm could profit by reducing its offer to the reservation wage (at no cost to employment). Finally, 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 $\pi_s(w_s) = \pi_s(0)$ for all w_s on the support. Using equations (F5), (F6) and (F7), this implies:

$$\frac{n_s}{k} \cdot \frac{\delta\lambda(p_s - w_s)}{[\delta + \lambda(1 - F_s(w_s))]^2} = \frac{n_s}{k} \cdot \frac{\delta\lambda p_s}{(\delta + \lambda)^2} \tag{F8}$$

Rearranging then yields the unconstrained type-s offer distribution:

$$F_s^*(w_s) = \frac{\delta + \lambda}{\lambda}\left[1 - \left(\frac{p_s - w_s}{p_s}\right)^{\frac{1}{2}}\right] \tag{F9}$$

Implications for skill wage premium and skill sorting

For convenience, suppose that firms are ranked identically in their 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 (F9), the type-s wage of the percentile F firm will then be:

$$w_s^*(F) = \left[1 - \left(1 - \frac{\lambda}{\delta + \lambda}F\right)^2\right] p_s \tag{F10}$$

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

$$\frac{w_l^*(F)}{w_h^*(F)} = \frac{p_l}{p_h} \tag{F11}$$

Just as in the baseline model, it follows that log wages will be additively separable in firm and worker effects. Finally, notice that equations (F5) and (F10) imply:

$$\frac{l_l^*(F)}{l_h^*(F)} = \frac{n_l}{n_h} \tag{F12}$$

[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 pay l-types more also).

i.e. the relative employment of skill types is independent of firm percentile F, so there is no assortative matching. Intuitively, since workers of each skill type care equally about firm rank F, they sort identically across the firm rank distribution; and firms hire all willing workers, because the labor supply constraints bind.

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 $\phi > \frac{p_l}{p_h}$, so the equilibrium in (F11) is not feasible. The equity constraint will then bind, with $w_l = \phi w_h$ in all firms. 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 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). Firms adopt this strategy if their wage offers are sufficiently low, specifically if their l-type offer $w_l = \phi w_h$ is smaller than the l-type productivity p_l.

2. **Selective strategy (S).** Selective firms hire all willing h-type workers, so the h-type supply constraint binds: $l_h^S = l_h(w_h)$. But they fully ration l-types, i.e. $l_l^S = 0$. Firms adopt this strategy if their wage offers are sufficiently high, specifically if their l-type offer $w_l = \phi w_h$ exceeds the l-type productivity p_l.

Unlike in the baseline model, there is now a distribution of wage offers corresponding to each strategy. Firms which pay l-types less than their productivity p_l (or equivalently, pay h-types less than $\frac{1}{\phi}p_l$) adopt the inclusive strategy, and those which pay above this cut-off adopt the selective strategy.

Since the equity constraint binds (and $w_l = \phi w_h$ in all firms), it is sufficient to derive the equilibrium distribution for h-type offers, which we denote F, i.e.:

$$F(w_h) \equiv F_h(w_h) = F_l(\phi w_h) \tag{F13}$$

Since the w_h cut-off for adopting the selective strategy is $\frac{1}{\phi}p_l$, the selective share of firms is equal to:

$$\sigma = 1 - F\left(\frac{1}{\phi}p_l\right) \tag{F14}$$

Since firms are ex ante identical, equilibrium requires that all wage offers on the support of F yield equal profit. And since the lowest offer must be zero (for the reasons above), it follows that $\pi(w_h) = \pi(0)$ for all offers w_h on the support. This condition can be written as:

$$(p_h - w_h)\, l_h(w_h) + \max\{p_l - \phi w_h, 0\}\, l_l(w_l) = p_h l_h(0) + p_l l_l(0) \tag{F15}$$

Looking at the left-hand side, all firms recruit h-types; and the h-type supply constraint always binds. But firms only recruit l-types (i.e., adopt the inclusive strategy) if the l-type offer $w_l = \phi w_h$ is below their productivity p_l; hence the second term in (F15). On the right-hand side, firms offering $w_h = 0$ are necessarily inclusive, so the labor supply constraint binds for both skill types. Using (F5) and (F13) and rearranging, equation (F15) implies:

$$F(w_h) = \frac{\delta+\lambda}{\lambda}\left\{1 - \left[\frac{(p_h - w_h)\, n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)} \cdot \max\{p_l - \phi w_h, 0\}\, n_l}{p_h n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)} \cdot p_l n_l}\right]^{\frac{1}{2}}\right\} \tag{F16}$$

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

As in the baseline model, equilibrium can take one of two forms: zero workplace segregation ($\sigma = 0$) or partial segregation ($\sigma > 0$). In the latter case[2], we can solve for σ by applying the expression in (F16) to equation (F14):

$$\sigma = 1 - \frac{\delta+\lambda}{\lambda}\left\{1 - \left[\frac{\left(p_h - \frac{p_l}{\phi}\right) n_h}{p_h n_h + \frac{\delta+\lambda}{\delta+\lambda(1-\sigma)} \cdot p_l n_l}\right]^{\frac{1}{2}}\right\} \tag{F17}$$

which pins down 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 (F16) shows that the offer distribution $F(w_h)$ is continuous, but kinked around the firm strategy cut-off $w_h = \frac{p_l}{\phi}$. The key intuitions from the baseline model can be gleaned from variation around this cut-off.

[2] To determine which case materializes, we can impose $\sigma = 0$ in equation (F16), and equate F to 1, to solve for the implied maximum wage offer. If this maximum offer is less than the selective strategy cut-off $\frac{p_l}{\phi}$, this confirms that no firm will adopt the selective strategy in equilibrium; so σ is indeed zero. Otherwise, σ must exceed zero.

First, h-type workers sort disproportionately into higher-paying firms. This is because all firms above the cut-off adopt the selective strategy—and exclusively hire h-type labor.

Second, firms just above the cut-off have lower employment than firms just below. This is because selective firms (above the cut-off) ration l-type labor, whereas h-type employment is increasing continuously in firm pay (even around the cut-off). 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, workers gradually work their way up a job ladder to ever higher-paying firms. However, in the presence of a binding equity constraint, this job ladder will be "shorter" for l-type workers—since high-paying firms (above the cut-off $w_h = \frac{p_l}{\phi}$) adopt selective hiring strategies and deny them access. 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. There are two equity constraints: firms must pay l- and m-types a fraction ϕ_l and ϕ_m (respectively) of the h-type wage. Assuming the equity constraint has stronger bite for l-types, i.e., $\beta_l > \beta_m$ where $\beta_s \equiv \phi_s \frac{p_h}{p_s}$ (we will validate this assumption ex post), Appendix E shows that firms may pursue one of three hierarchical strategies in equilibrium: (i) hire all willing workers; (ii) hire only m- and h-type workers; and (iii) hire only h-type workers. We call these the L-, M- and H-strategies respectively. Let σ^L, σ^M and σ^H denote the equilibrium shares of L-, M- and H-strategy firms, where $\sigma^L + \sigma^M + \sigma^H = 1$; and let w_h^L, w_h^M and w_h^H denote the wage paid in each strategy to type-h workers.

As in Appendix C, we assume 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 firm productivity variance ν and 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 (for each 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 (E18) in the N-type model. Equal profits for the L- and H-strategies implies:

$$\frac{w_h^L}{w_h^H} = \left(1 + \phi_m \frac{\phi_m^\varepsilon \Omega_m}{\Omega_h} + \phi_l \frac{\phi_l^\varepsilon \Omega_l}{\Omega_h}\right)^{-\frac{1}{1+\varepsilon}} \tag{G1}$$

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

$$\frac{w_h^M}{w_h^H} = \left(1 + \phi_m \cdot \frac{\phi_m^\varepsilon \Omega_m}{\Omega_h}\right)^{-\frac{1}{1+\varepsilon}} \tag{G2}$$

Next, we have two equations for equilibrium ratios of the labor supply intercepts. From equation (E13) in the N-type model, these are:

$$\frac{\phi_l^\varepsilon \Omega_l}{\Omega_h} = \frac{n_l}{n_h} \cdot \frac{1 + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon} \tag{G3}$$

$$\frac{\phi_m^\varepsilon \Omega_m}{\Omega_h} = \frac{n_m}{n_h} \cdot \frac{1 + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon}{\frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + \frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon} \tag{G4}$$

Finally, we have two expressions for the expected log wages of l-types and m-types, expressed relative to h-types. Integrating over the firm strategy distribution, these are:

$$E[\log w_l] - E[\log w_h] = \log \phi_l + \frac{\frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon \log \frac{w_h^L}{w_h^M} + \log \frac{w_h^L}{w_h^H}}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^\varepsilon + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^\varepsilon + 1} \tag{G5}$$

$$E\left[\log w_m\right] - E\left[\log w_h\right] = \log \phi_m + \frac{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}} - \frac{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\frac{\sigma_l}{\sigma_h}\left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \frac{\sigma_m}{\sigma_h}\left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} + 1} \tag{G6}$$

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_{size}) and (ii) the variance of AKM firm effects (V_{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 skill type, organize this data in "long" form (with each row corresponding to a firm × skill type), and then estimate an AKM model by regressing log wages on firm and skill fixed effects, and save the firm premia as η_f. We then regress log employment on the firm effects η_f:

$$\log l_f = \mu_0 + \mu_1 \eta_f + \epsilon_f \tag{G7}$$

The estimated coefficient μ_1 provides our model-based moment for ε_{size}. The variance of the estimated firm effects η_f across all firms provides our model-based moment for V_{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 $m^* = (\varepsilon_{size}, V_{AKM})$ 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., $\sum_j |m_{ij} - m_j^*| < \tau$, where $j \in \{1, 2\}$ indexes the two moments. We set the learning rate $\eta = 0.1$ and tolerance $\tau = 10^{-3}$. At each iteration, we re-solve the equilibrium equations from Step 1 given the updated ε and ν.

The final estimated parameters imply a labor supply elasticity of $\varepsilon = 5.52$ and productivity variance of $\nu = 0.023$. With these values, the model successfully replicates both the average firm size-wage premium relationship ($\varepsilon_{size} = 2.55$ in both model and data) and the overall dispersion in firm wage premia ($V_{AKM} = 0.032$ 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 (i) the removal of the equity constraint in Appendix H.1 and (ii) 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:

$$w_s^{CF1} = w_h^* = \frac{\varepsilon}{1+\varepsilon} p_s \tag{H1}$$

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:

$$\log w_h^{CF1} - E\left[\log w_h\right] = -\frac{\sigma^L \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \sigma^M \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\sigma^L \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma^M \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} + \sigma^H} \tag{H2}$$

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). To reach (H2), we have integrated over the firm strategy distribution, applied the labor supply function in (2), and used the fact that 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 the expected m-type wage is:

$$\begin{aligned}\log w_m^{CF1} - E\left[\log w_m\right] &= \log \frac{w_m^{CF1}}{w_h^{CF1}} + \left(\log w_h^{CF1} - E\left[\log w_h\right]\right) + \left(E\left[\log w_h\right] - E\left[\log w_m\right]\right) \\ &= -\log \beta_m - \frac{\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^L}{w_h^H} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} \log \frac{w_h^M}{w_h^H}}{\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}} \end{aligned} \tag{H3}$$

where the second line uses the definition of β_s in (E4), the expected skill differential in (G6), and the counterfactual impact in (H2). Finally, the impact on the expected l-type wage is:

$$\begin{aligned}\log w_l^{CF1} - E\left[\log w_l\right] &= \log \frac{w_l^{CF1}}{w_h^{CF1}} + \left(\log w_h^{CF1} - E\left[\log w_h\right]\right) + \left(E\left[\log w_h\right] - E\left[\log w_l\right]\right) \\ &= -\log \beta_l - \log \frac{w_h^L}{w_h^H} \end{aligned} \tag{H4}$$

where the second line uses the definition of β_s in (E4), the expected skill differential in (G5) and the counterfactual impact in (H2).

Impact on expected utility

We now turn to expected utility. Note we weight utility by $\frac{1}{\varepsilon}$ for this exercise, to express it in log wage units: see equation (1). Using equation (B26), the impact on h-type utility is:

$$\begin{aligned}\frac{1}{\varepsilon}\left(\bar{u}_h^{CF1} - \bar{u}_h\right) &= \frac{1}{\varepsilon} \log \frac{\left(w_h^{CF1}\right)^{\varepsilon}}{\sigma_l \left(w_h^L\right)^{\varepsilon} + \sigma_m \left(w_h^M\right)^{\varepsilon} + \sigma_h \left(w_h^H\right)^{\varepsilon}} \\ &= -\frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon} + \sigma_h\right] \end{aligned} \tag{H5}$$

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

$$\begin{aligned}\frac{1}{\varepsilon}\left(\bar{u}_m^{CF1} - \bar{u}_m\right) &= \frac{1}{\varepsilon} \log \frac{\left(w_m^{CF1}\right)^{\varepsilon}}{\sigma_l \phi_m^{\varepsilon} \left(w_h^L\right)^{\varepsilon} + \sigma_m \phi_m^{\varepsilon} \left(w_h^M\right)^{\varepsilon}} \\ &= -\log \beta_m - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon} + \sigma_m \left(\frac{w_h^M}{w_h^H}\right)^{\varepsilon}\right] \end{aligned} \tag{H6}$$

And for l-types:

$$\frac{1}{\varepsilon}\left(\bar{u}_l^{CF1} - \bar{u}_l\right) = -\log \beta_l - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^{\varepsilon}\right] \tag{H7}$$

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-type case in equation (E16), the optimal L-strategy wage (for h-type workers) can be written as:

$$w_h^{CF2} = w_h^L = \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} \cdot \frac{\varepsilon}{1+\varepsilon} p_h \tag{H8}$$

To derive equation (H8), we have replaced the intercept ratios with the aggregate employment ratios, i.e. $\frac{\phi_s^\varepsilon \Omega_s}{\Omega_h} = \frac{n_s}{n_h}$ for $s = \{l, m\}$. This follows from the fact that all firms adopt the same strategy (and pay the same wage): see equation (E12) in the N-type model.

Since the equity constraints bind, m-types receive $\phi_m w_h^{CF2}$ and l-types $\phi_l w_h^{CF2}$. Building from (H2), the impact on the expected log wages of h-, m- and l-types are therefore:

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

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

$$\log w_l^{CF2} - E\left[\log w_l\right] = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \log \frac{w_h^L}{w_h^H} \tag{H9}$$

Impact on expected utility

We now turn to expected utility. As before, we weight utility by $\frac{1}{\varepsilon}$, to express it in log wage units. Building from (H5), the impact on expected h, m- and l-type utility are:

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

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

$$\frac{1}{\varepsilon}\left(\bar{u}_l^{CF2} - \bar{u}_l\right) = \log \frac{1 + \frac{\phi_m}{\beta_m} \cdot \frac{n_m}{n_h} + \frac{\phi_l}{\beta_l} \cdot \frac{n_l}{n_h}}{1 + \phi_m \cdot \frac{n_m}{n_h} + \phi_l \cdot \frac{n_l}{n_h}} - \frac{1}{\varepsilon} \log \left[\sigma_l \left(\frac{w_h^L}{w_h^H}\right)^\varepsilon\right] \tag{H10}$$

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 the three alternative models (Models 2, 3 and 4) we compare to our baseline equity constraint framework (Model 1). In each case, we explain how we calibrate the parameters; and at the end of the appendix, we report our parameter estimates.

I.1 Model 2: Skill-neutral firm heterogeneity

In Model 2, we impose skill-neutral heterogeneity in firm productivity (as in the baseline model), but remove the equity constraint. Firms f therefore pay the unconstrained optimum to each skill type $s = \{h, m, l\}$, in line with equation (7). For skill s, the optimal wage is:

$$w_{sf} = \frac{\varepsilon}{1+\varepsilon} p_{sf} \tag{I1}$$

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

$$p_{sf} = x_f p_s \tag{I2}$$

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. The labor supply functions $l_s(w)$ are given by equation (2), and the labor supply intercepts Ω_s by (3).

Expressing outcomes relative to h-types, the model can be summarized by six parameters: the labor supply elasticity ε, the variance of firm productivity ν, the base productivity differentials, $\log \frac{p_m}{p_h}$ and $\log \frac{p_l}{p_h}$, and the relative labor supply intercepts, $\frac{\Omega_m}{\Omega_h}$ and $\frac{\Omega_l}{\Omega_h}$.

We calibrate these parameters to match six empirical moments from Table 2: the average elasticity of firm size with respect to AKM firm premia, ε_{size}; the variance of AKM firm effects, V_{AKM}; the mean log wage differentials between skill groups, $E[\log w_m] - E[\log w_h]$ and $E[\log w_l] - E[\log w_h]$; and the aggregate skill employment ratios, $\frac{n_m}{n_h}$ and $\frac{n_l}{n_h}$.

I.2 Model 3: Skill-biased firm heterogeneity

Model 3 allows for skill-biased productivity differences across firms, but again removes the equity constraint. The marginal product of s-type workers in firm f is:

$$p_{sf} = x_f^{\theta_s} p_s \tag{I3}$$

where the θ_s are skill-specific productivity elasticities with respect to firm heterogeneity x_f. We normalize θ_h to 1, and estimate θ_m and θ_l. The model is otherwise identical to Model 2.

The model can be characterized by eight parameters: the labor supply elasticity ε, the firm productivity variance ν, the skill-specific productivity elasticities, θ_m and θ_l, the base productivity differentials, $\log \frac{p_m}{p_h}$ and $\log \frac{p_l}{p_h}$, and relative labor supply intercepts, $\frac{\Omega_m}{\Omega_h}$ and $\frac{\Omega_l}{\Omega_h}$.

We calibrate these parameters to match the six moments we used in Model 2, plus skill differentials (l- v h-type and m- v h-type) in the mean AKM firm effects of workers' employers (see the "Av. firm effect" row in Table 1). This gives us eight moments in total, the same number we use in the baseline model.

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:

$$u_{isf} = \varepsilon_s \log w_{sf} + a_{if} \tag{I4}$$

where ε_s is the skill-specific labor supply elasticity. Like Model 3, 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, $\log \frac{p_m}{p_h}$ and $\log \frac{p_l}{p_h}$, and the relative labor supply intercepts, $\frac{\Omega_m}{\Omega_h}$ and $\frac{\Omega_l}{\Omega_h}$. We calibrate these parameters to match the same eight moments as in Model 3.

I.4 Estimation procedure and results

To estimate each model, we follow a similar two-step iterative procedure to the baseline model (see Appendix G). In Step 1, for a given labor supply elasticity ε and firm productivity variance ν, 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 ε_{size} and V_{AKM} moments.

We report our parameter estimates in Table A2. Model 2 yields a labor supply elasticity of 2.55 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 the local 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 skill 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; or 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 intercepts (E13), (ii) the relative wages across active strategies (E16), and (iii) the equal profit conditions (E18) 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 z (those with $\sigma^z = 0$) 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 8.

K Split-sample method and variance decomposition

Like Babet et al. (2025) and Carry et al. (2025), we employ a split-sample method to correct for measurement error, which we apply throughout our analysis—including for the AKM variance decomposition (in Table 1) and bin plots. In this appendix, we describe our approach in greater detail.

K.1 Split-sample procedure

Following Babet et al. (2025), we implement a firm-splitting algorithm that creates two balanced samples while maintaining maximal connectivity. The procedure works as follows:

1. Identify workers as movers if they appear in at least two firms during the sample period; otherwise, classify them as stayers.
2. Draw a shock from a uniform distribution (ϵ_{it}) for each worker-year observation, and draw another (ζ_{ft}) for each firm-year combination.
3. Randomly select a year t from the panel.
4. For each firm f operating in year t:
   - Let m_{fts} be the median of ϵ_{it} in firm f for group s (stayers or movers). The median takes the middle point in case of an even number of elements.
   - If $\zeta_{ft} < 0.5$:
     - For stayers: assign worker i to sample A if $\epsilon_{it} < m_{fts}$; else assign to sample B
     - For movers: assign worker i to sample A if $\epsilon_{it} \leq m_{fts}$; else assign to B
   - If $\zeta_{ft} \geq 0.5$:
     - For stayers: assign worker i to sample A if $\epsilon_{it} \leq m_{fts}$; else assign to sample B
     - For movers: assign worker i to sample A if $\epsilon_{it} < m_{fts}$; else assign to B
5. Choose another year randomly and repeat the same process for workers not yet assigned to a sample.
6. Continue for all years of the panel.

This procedure ensures that: (i) each worker appears in only one sample, (ii) each firm with more than one worker has workers in both samples, and (iii) movers contribute to connectivity in both samples.

K.2 AKM variance decomposition

To implement the variance decomposition, we first estimate the AKM model (16) separately for samples A and B. For this analysis, we rely on the estimated worker effects from sample A, denoted $\hat{\lambda}_i^A$ (which includes also the estimated time and age effects, slightly abusing notation), and our two estimates of the firm effect (from each sample), denoted $\hat{\eta}_f^A$ and $\hat{\eta}_f^B$.

The variance of firm fixed effects is calculated as the covariance between estimates from the two samples:

$$Var(\eta_f) = Cov(\hat{\eta}_f^A, \hat{\eta}_f^B) \tag{K1}$$

To compute the covariance between firm and worker effects, we use estimated effects from the two samples:

$$Cov(\lambda_i, \eta_{f(i,t)}) = Cov(\hat{\lambda}_i^A, \hat{\eta}_{f(i,t)}^B) \tag{K2}$$

And building on Carry et al. (2025), we compute the variance of worker fixed effects as:

$$Var(\lambda_i) = Cov(\hat{\lambda}_i^A, \hat{\lambda}_{it}^{A(B)}) \tag{K3}$$

where $\hat{\lambda}_{it}^{A(B)} = \log w_{it} - \hat{\eta}_{f(i,t)}^B$ are the worker fixed effects from sample A (plus the error term) implied by the sample B firm effects. Finally, the R^2 of the model is calculated as the sum of the explained variance components relative to the total variance of log wages:

$$R^2 = \frac{Var(\lambda_i) + Var(\eta_{f(i,t)}) + 2 \cdot Cov(\lambda_i, \eta_{f(i,t)})}{Var(\log w_{it})} \tag{K4}$$

L Preparation of Israeli administrative data

This appendix provides additional details on data preparation and variable definitions.

Earnings: Our raw earnings data consist of observations at worker × firm × year level, with monthly employment indicators and total annual compensation for each employment spell. Before proceeding, we implement several data cleaning procedures: (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, and (iv) 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 monthly earnings 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 earnings.

To focus on workers with substantial labor market attachment, we exclude worker-year observations with monthly earnings below 25% of the national average that year.[3] And to

[3] 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.

reduce the influence of possible spurious outliers, we also exclude the top 1% of earnings observations within each year and sector (private or public) from our sample. Our final sample spans 1990-2019 and includes workers aged 25-64 in each year.

Education: We use the Central Bureau of Statistics' 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 degree), (ii) non-STEM graduate (BA-equivalent or higher degree in non-STEM field), and (iii) STEM graduate (BA-equivalent or higher degree in STEM field). We define science, engineering and mathematics degrees as STEM.

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 sample size for all analyses, we incorporated the three smallest regions into neighboring regions.

Industry: We use a consistent two-digit industry classification with 87 codes, based on the ISIC Rev. 4 scheme.

M 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 matched employer-employee administrative records for Italy's Veneto region over 1975-2001. The data cover the universe of private sector employment. We estimate AKM firm effects using log daily earnings (the ratio of annual earnings to days worked) for years between 1992 and 2001, and implement the same split-sample correction for measurement error as in our main analysis.

In Figure A6, we plot the relationship between log firm employment and AKM firm premia, across 20 firm bins. As in the Israeli data, we again see a hump-shaped relationship, with employment initially increasing and then decreasing in firm wage premia—both for the aggregate data and after residualizing by two-digit industry (55 codes). 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.