Which Workers Earn More at Productive Firms? Position Specific Skills and Individual Worker Hold-up Power*

Justin Bloesch — Cornell University

Birthe Larsen — Copenhagen Business School

Bledi Taska — TalentNeuron

July 18, 2026

Abstract

We propose and test a theory of individual worker hold-up power to explain why workers in high-wage occupations earn a larger wage premium at productive firms than do workers in low-wage occupations. We derive a model where firms set wages and produce using a novel generalization of Kremer (1993)'s O-Ring production function. Workers have greater hold-up power at productive firms if (i) labor inputs are organized into indivisible positions, (ii) separations by experienced incumbent workers generate multiplicative output losses, and (iii) workers learn position-specific skills while on the job. This creates an efficiency wage motivation, where productive firms are incentivized to share rents to improve retention, and differentially so in occupations where experienced workers are more able to disrupt production. Using Danish administrative data, we report a range of empirical evidence supporting the model's predictions. We discuss implications for wage inequality, the gender wage gap, and the effects of minimum wages.

Note: Due to a pending update in a data use agreement, some summary statistics and empirical evidence reflect an out-of-date sample and will be noted in the draft. Results will be updated upon renewed data access.

*Contact: Justin Bloesch (jb2722@cornell.edu), Birthe Larsen (bl.eco@cbs.dk), and Bledi Taska (bt540@nyu.edu).

We thank Antoine Bertheau, Adrien Bilal, Gabriel Chodorow-Reich, John Coglianese, Harris Eppsteiner, Jason Furman, Xavier Gabaix, Andrew Garin, Ed Glaeser, Fane Groes, Lawrence Katz, Omeed Maghzian, Christian Moser, Namrata Narain, Anna Stansbury, Ludwig Straub, James Stock, and Jacob Weber and seminar participants at Harvard University, the 2021 Search and Matching Conference, the University of Copenhagen, the Copenhagen Business School, the 2025 Barcelona Summer Forum, and 2025 Society of Economic Dynamics the for thoughtful discussions and comments. This research was supported by the James M. and Kathleen D. Stone PhD Scholarship in Inequality and Wealth Concentration.

1 Introduction

There is a long-established empirical finding that some firms and industries systematically pay higher wages, and that at least part of this premium reflects rents (Krueger and Summers, 1988; Sorkin, 2018). Historically, firm wage premia were strongly correlated with firm size (Brown and Medoff, 1989; Song et al., 2019), and both high- and low-educated workers benefited similarly from employment at large firms (Song et al., 2019), consistent with a common firm wage premium (Abowd et al., 1999).

More recent evidence points to a different pattern. The large-firm wage premium has declined in the US (Bloom et al., 2018) and is very small in Denmark (Bloesch et al., 2025). At the same time, changes in firm productivity have a stronger effect on the wages of workers in high-wage occupations and those more attached to the firm.[1] Figure 1a reports a similar pattern across occupations in the cross-section in Denmark from 2008-2019, plotting individual worker hourly wages by occupation against firm value added per worker. The figure shows the steepest slope of wages to firm productivity for managers, an intermediate slope for middle-wage occupations, and the flattest slope for low-wage occupations. Figure 1b shows that with the exception of very small firms, the relationship between firm size and wages is weak for all occupations. These patterns lead us to ask: why would productive firms, but not necessarily large firms, share rents with workers in high-wage occupations, but not with workers in low-wage occupations?

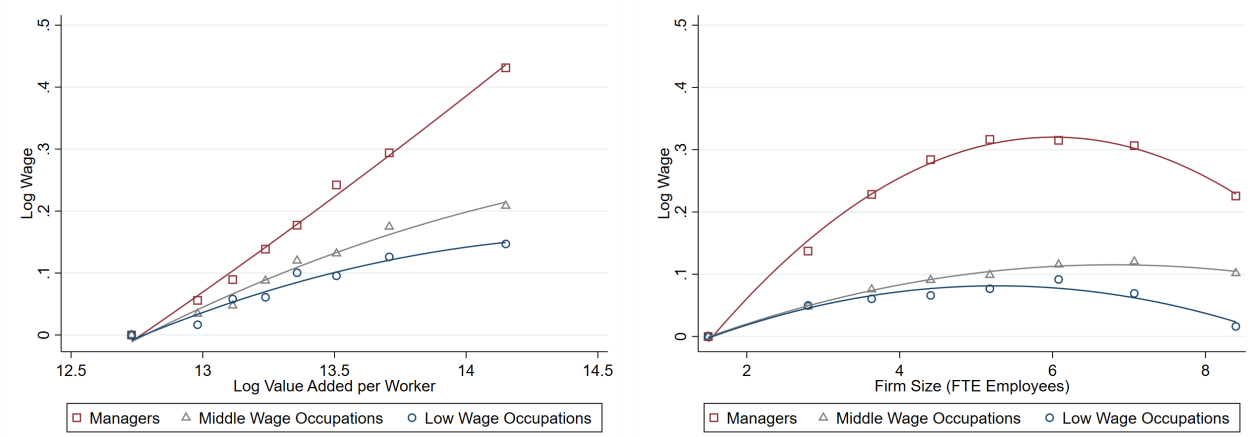
To answer this question, we derive a novel model where productive firms share rents only with workers in occupations where workers have individual hold-up power. At a high level, a worker has individual hold-up power if their job is important to the firm *and* if the worker is hard to replace. More precisely, workers have individual hold-up power when (i) labor in production is organized into distinct, differentiated positions that are decided in advance of hiring and training workers, (ii) firm output is disrupted in a multiplicative way when a worker's job is not performed or is not performed well, and (iii) workers' skills are position-specific, i.e., skills are acquired on the job and are not transferable across positions or firms. Together, these properties mean that if a worker unexpectedly separates, the firm's production is disrupted, and no other worker can step in and fully perform the separated worker's job. If output disruptions from losing an experienced worker are larger at higher productivity firms, then a worker in a given occupation can hold up more output at a high productivity firm than at a low productivity firm. The threat of worker turnover and incomplete contracts incentivize productive firms to pay higher wages to decrease this

[1] See Juhn et al. (2018), Kline et al. (2019), Friedrich et al. (2019), Chan et al. (2020), Maibom and Vejlin (2021), Garin and Silvério (2024).

differentially costly turnover, but only in occupations where workers have this hold-up power.

Figure 1: Wages, Firm Productivity, and Firm Size by Occupation

(a) Wages and Firm Productivity by Occupation (b) Wages and Firm Size by Occupation



This figure reports the average wage of workers in three broad occupation groups, ordered by either their firm's log value added per worker or log firm size in terms of full-time equivalent workers. Averages are taken from 2008-2019, and each dot represents approximately equal numbers of full-time equivalent workers. Middle-Wage Occupations include professionals and technicians, and low-wage occupations include crafts, assembly, administrative, service, and manual occupations.

For intuition, consider a simple example of a restaurant with two positions, using only a cook and a waiter (with a non-working owner). The production function is Leontief: if the tasks in either position are left undone, then the restaurant sells nothing. Suppose also that training a new cook takes longer than training a waiter, as even a cook with industry experience must take time to learn the new recipes. In this setting, the cook has hold-up power, but the waiter does not. Both positions are essential in production, but only the cook acquires position-specific skills, while the waiter is replaceable. If the restaurant becomes more popular and can charge higher prices, the amount of profit that the cook can hold up increases, and the owner will be incentivized to raise the cook's pay. The waiter's pay, on the other hand, will be set to the going market rate.

While the previous example is limited to a firm with only two positions, our more general model formalizes these intuitions in firms with a continuum of positions, diminishing returns to labor, and multiple occupations. We derive a novel production function with three key features: (i) labor is divided into "positions" that can be filled with at most one worker at a time, (ii) a firm's output depends both on the number of workers and on the share of positions filled with experienced workers, as each new position introduces a potential new possible failure point in production, and (iii) losses of experienced workers decrease output in

a way that is proportionate to the firm's output per worker. The cost of losing an experienced worker differs across occupations, generating variation in hold-up power across occupations.

We then embed firms with this production function into a static equilibrium labor market, where each period has multiple stages. First firms decide the number of positions in each occupation they will produce with, and then firms set wages and hire workers. After being hired, workers learn skills that are specific to their position, and only one worker can learn their position's skills at a time. After these workers become experienced, they draw idiosyncratic disutility of remaining with the firm, which is private information and is not contractable. Workers can quit their jobs and rematch with firms who are re-filling their positions, but these matches are less productive, and so firms are incentivized to pay high wages to avoid this costly turnover. This incentive to pay high wages is particularly strong at high productivity firms for workers in high-hold up occupations, as the multiplicative nature of our production function makes these separations differentially costly.

Our model generates variation in firm productivity in a way that is non-standard for the rent-sharing literature, revealing an important relationship between rent-sharing, firm productivity, and firms' labor share. Unlike in models with linear production functions or models with inelastic labor supply curves, in our model a firm's total factor productivity has no effect on wages and affects only the firm's size. Instead, variation in output per worker is largely driven by firms' output elasticity of labor: firms with a low output elasticity of labor have high measured output per worker and low labor share of value added. Due to the multiplicative nature of our production function, the output costs of losing an experienced worker with hold-up power is greater at firms with a low output elasticity of labor (and therefore low labor share and high productivity). Because these firms have greater costs of turnover, they are differentially incentivized to pay high wages, but only to workers with this hold-up power. Thus, rents are the greatest for workers with hold-up power in firms with low labor shares.

At its core, our story is one of efficiency wages (Katz, 1986): a firm pays high wages to decrease worker turnover, which increases output given the number of workers. This efficiency wage motivation is stronger in occupations where incumbents are harder to replace, and differentially so in hard-to-replace occupations at more productive firms. Our two main contributions therefore are (i) to develop a tractable model of individual hold-up power in multi-worker firms where this efficiency wage motivation differs systematically across occupations and firm productivity but not firm size, and (ii) to empirically test the model's predictions. The key testable prediction of our model is that the occupations where workers earn more at productive firms, both in steady state and in response to long-run changes in firm productivity, are the same occupations where the cost of worker separations is higher

at more productive firms. Additionally, our model makes three predictions about levels: the level of turnover costs is higher in higher-hold up occupations, workers in higher hold-up occupations earn a higher level of wages conditional on education, and large firms have no differential incentive to pay high wages. We find support for all of these predictions.

We test the model's predictions using Danish administrative data and in US job postings data. First, we estimate the relationship between wages and firm productivity across occupations in the cross-section and find that managers exhibit the steepest wage-productivity gradients, followed by middle-wage occupations, while low-wage occupations show the weakest relationship. Consistent with the model, firm size has little explanatory power for wages once firm productivity is taken into account. We then estimate the pass-through of persistent productivity shocks to wages using an instrumental variables approach (Card et al., 2018; Juhn et al., 2018; Maibom and Vejlin, 2021). We deviate from the previous literature by instrumenting changes in the labor share with longer changes in firm value added per worker, which we show isolates the kind of productivity innovations that generate rent sharing in our model. We find a similar pattern as in the cross-section: pass-through of persistent productivity changes to wages is highest for managers and lowest for low-wage occupations.

Next, we estimate the effect of worker separations on firm profits using worker deaths (Isen, 2015; Jäger, 2016; Bertheau et al., 2021). We find that profit losses are largest for managers, moderate for middle-wage occupations, and smallest for low-wage occupations. Moreover, these losses are more sensitive to firm productivity in high wage occupations, providing direct evidence of individual hold-up power.

Finally, using US job posting data, we construct a measure of task differentiation at the occupation level, following the intuition of Lazear (2009) that specific skills are unique combinations of general skills. We show that occupations with more differentiated task requirements from other jobs in the firm, our proxy for position-specificity, exhibit both higher wage levels and greater pass-through of productivity to wages. We also find that conditional on our task differentiation score, more educated occupations do not show higher pass-through of productivity to wages. This provides evidence that position specific skills, rather than general skills, is more predictive of rent-sharing.

While our model shares with models of wage bargaining and multi-worker firms the feature that wages can depend on firms' inframarginal productivity, our model differs in key ways. Our model differs from Stole and Zwiebel (1996) because rent sharing survives even when firms choose their quantity of labor.[2] Relative to Wong (2025) where wages depend on

[2]In Stole and Zwiebel (1996), when firms can choose their quantity of labor, they over-hire to equalize wages with workers' outside options, and so there is no difference in wages across high and low productivity firms in steady state.

profits per worker due to collective bargaining, in our model workers can individually extract higher wages at firms with high inframarginal productivity due to individual hold-up power.

Our model differs from monopsony-based models of rent sharing. In such models, workers earn rents at productive firms because firms with high total factor productivity are constrained by an upward sloping labor supply curve, endogenously generating both high value added per worker and high wages. These monopsony-based models make predictions about rent sharing that are at odds with empirical evidence. Models with isolelastic labor supply curves (Lamadon et al., 2019; Chan et al., 2020) imply a pass-through of value added per worker to wages of 1, which is far above empirical estimates (Card et al., 2018). Models with firms' labor supply elasticity declining with firm size can achieve realistic rent sharing but predict that the firm size-wage premium is steeper for large firms and that large firms have higher pass-through of TFP shocks to wages, which contradicts empirical evidence from Bloom et al. (2018) and Chan et al. (2024), respectively. In contrast, our model of individual worker hold-up can rationalize empirical estimates of pass-through of value added per worker to wages between 0-0.2 and is consistent with evidence from Bloesch et al. (2025) that firms do not need to pay higher wages to become larger in the long run.

The rest of the paper is organized as follows. Section 2 presents the model and derives its main predictions. Section 3 describes the data. Section 4 presents our empirical evidence: cross-sectional evidence, pass-through regressions, task differentiation scores, and the effect of worker deaths on firms. In Section 5 we estimate our hold-up power parameters for three broad occupation groups. We discuss implications for inequality, the minimum wage, and the gender wage gap in Section 6. Section 7 concludes.

2 Model

In this section, we derive a repeated static model where workers learn position-specific skills while on the job, and multi-worker firms trade off wage costs against the retention of experienced workers.

We begin by describing the firm's production function, which is the most novel feature of the model. We then describe the environment and timing, and then we characterize the firm's problem and equilibrium. We show that firms with a low output elasticity of labor have high measured output per worker. Variation within a firm in the output elasticity of labor generates co-movement between firms' output per worker and wages, with a steeper slope for occupations with greater hold-up power.

2.1 Production Function

We now introduce the production function. Firms are denoted j and occupations are denoted k. Each firm is endowed with a technology that determines its total factor productivity $A_{j,t}$, the elasticity of output with respect to capital and labor inputs, respectively $\beta_{j,t} \in [0, 1)$ and $\alpha_{j,t} \in (0, 1)$ with $\alpha_{j,t} + \beta_{j,t} < 1$, and the within-labor weights on each occupation $\sigma_{jk,t}$, with $\sum_k \sigma_{jk,t} = 1$. We assume that $A_{j,t}$, $\beta_{j,t}$, $\alpha_{j,t}$, and $\sigma_{jk,t}$ may change over time.

Firms have four inputs into production: the capital stock $K_{j,t}$, the number of positions for each occupation $P_{jk,t}$, the number of experienced "senior" workers $S_{jk,t}$, and the number of inexperienced "junior" workers $J_{jk,t}$. The firm's output will depend on both the number of workers $J_{jk,t} + S_{jk,t}$ and the *composition* of workers, defined here as the share of positions that are filled with experienced S workers. Firm j's value added function is

$$Y_{j,t} = A_{j,t} K_{j,t}^{\beta_{j,t}} \prod_{k \in \mathcal{K}(j)} \left(\underbrace{(S_{jk,t} + J_{jk,t})}_{\substack{\# \text{ of} \\ \text{workers}}} \times \underbrace{\left(\frac{S_{jk,t}}{P_{jk,t}} \right)^{\delta_k + \frac{h_k}{\alpha_{j,t}}}}_{\substack{\text{share of positions with} \\ \text{experienced workers}}} \right)^{\sigma_{jk,t}\alpha_{j,t}} . \tag{1}$$

where $Y_{j,t}$, is firm j's value added in period t. We impose $\delta_k \geq 0$, $h_k \geq 0$, and $\delta_k + h_k > 0$. Firms are limited to have no more workers than positions: $S_{jk,t} + J_{jk,t} \leq P_{jk,t}$. The number of workers must be non-negative, so $0 \leq S_{jk,t} \leq P_{jk,t}$ and $0 \leq J_{jk,t} \leq P_{jk,t} - S_{jk,t}$.

The overall structure builds on standard Cobb Douglas production: firms have total productivity $A_{j,t}$, and output increases with capital inputs according to elasticity $\beta_{j,t}$. Total output increases with respect to labor inputs according to elasticity $\alpha_{j,t}$, within which occupations have weights $\sigma_{jk,t}$.

Within each occupation, labor inputs are increasing in the total number of workers $S_{jk,t} + J_{jk,t}$, where senior and junior workers are treated similarly. The second term is the composition term $\frac{S_{jk,t}}{P_{jk,t}}$, which is the fraction of positions that are filled with experienced senior workers at the time of production. As the firm loses senior workers given the number of positions, the composition term falls with elasticity $\delta_k + h_k/\alpha_j$, where δ_k and h_k are occupation-specific constants, and α_j is the same parameter as the output elasticity of labor for firm j. While we have not introduced wages yet, it will later become clear that δ_k represents turnover costs that are proportionate to wages, while h_k represents turnover costs that are proportionate to the firm's total output per worker.

The constraint that there can be at most one worker per position $S_{jk,t} + J_{jk,t} \leq P_{jk,t}$ gives the composition term in the production function bite. Intuitively, to increase production, firms needs to open more positions. However, every new position opens a new failure point,

and with this functional form, failure to perform any given job can affect output by more than just "at the margin."

To further illustrate the role of the composition term $\frac{S_{jk,t}}{P_{jk,t}}$ and its related parameters δ_k and h_k, we will define and characterize the marginal products of junior and senior workers. Define the marginal product of a junior worker as $\frac{\partial Y_{j,t}}{\partial J_{jk,t}}$ and let $N_{jk,t} \equiv J_{jk,t} + S_{jk,t}$ denote the total number of workers. The marginal product of a junior worker is

$$MRP^J_{jk,t} \equiv \frac{\partial Y_{j,t}}{\partial J_{jk,t}} = \sigma_{jk,t}\alpha_{j,t}\frac{Y_{j,t}}{N_{jk,t}}.$$

This term resembles the marginal product of labor under standard Cobb-Douglas production. We will later show that when firms are optimizing, firms equalize the marginal product of a junior worker with the wage: $MRP^J_{jk,t} = w_{jk,t}$.

Next consider how output would change if the firm lost a senior S worker but did not replace them with a junior worker, which we define as the "marginal product of a senior worker" $\frac{\partial Y_{j,t}}{\partial S_{jk,t}}$:

$$\begin{aligned} MRP^S_{jk,t} \equiv \frac{\partial Y_{j,t}}{\partial S_{jk,t}} =& \sigma_{jk,t}\alpha_j \left(1 + \left(\delta_k + \frac{h_k}{\alpha_j}\right)\left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{-1}\right)\frac{Y_{j,t}}{N_{jk,t}} \\ =& MRP^J_{jk,t} + \underbrace{\left(\delta_k MRP^J_{jk,t} + h_k\frac{\sigma_{jk,t}Y_{j,t}}{N_{jk,t}}\right)\left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{-1}}_{\text{Additional Output Losses}}. \end{aligned}$$

This term shows that losing a senior worker means losing both "regular" marginal product $MRP^J_{jk,t}$ as well the term $h_k\frac{\sigma_{jk,t}Y_j}{N_{j,k}}$, which is output per worker scaled by occupation k's weight in labor $\sigma_{jk,t}$. We will explain the intuition of this term momentarily. Lastly, there is an additional term $\left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{-1}$, which shows that losing senior workers gets more costly as the proportion of senior workers to positions falls further.

To see why h_k affects output per worker multiplicatively, consider the case where the firm j only produces using one occupation k, and evaluate $\frac{\partial Y_{j,t}}{\partial S_{jk,t}}$ at the point where every position is filled with a senior worker $S_{jk,t} = P_{jk,t}$:

$$MRP^S_{jk,t}\Big|_{S_{jk,t}=P_{jk,t},|\mathcal{K}(j)|=1} = MRP^J_{jk,t}(1+\delta_k) + \underbrace{h_k\frac{Y_{j,t}}{N_{j,t}}}_{\substack{\text{Multiplicative losses} \\ \text{in output per worker}}}$$

Therefore, when an occupation has $h_k > 0$, the disruption to output from losing a senior worker is multiplicative in the firm's average product or output per worker. As we alluded to previously that we will have the result, $MRP^J_{jk,t} = w_{jk,t}$, δ_k reflects turnover costs in

units of wages. Thus our two parameters δ_k and h_k respectively capture "marginal" versus "inframarginal" turnover costs, of which only the latter depends on the total productivity of the firm.

2.2 Environment and Timing

We now describe the model environment and timing. There are three key features that help give workers hold-up power. First, the number of positions is decided in advance and is fixed for the rest of the period. Second, workers become experienced S workers by learning position-specific skills while on the job, and each position can be occupied by at most one worker at a time. Third, limited commitment by workers means that workers do not internalize the costs that firms bear when workers leave the firm.

Let there be a mass 1 of firms and mass 1 of workers. Workers denoted i are differentiated by their skill level $s \in \mathcal{S}$, with population shares ϕ_s. Firms employ workers in occupations $k \in \mathcal{K}$. Each occupation k can use only one type of worker $s(k)$, but workers in skill group s may work in potentially multiple different occupations.[3]

To employ workers, firms must first create "positions", which can be thought of as a set of tasks that, once bundled together into a position, must be performed by one person. Define $P_{jk,t}$ as the number of positions for firm j in occupation k in period t. The number of positions is created at the beginning of the period and cannot be changed for the rest of the period. Additionally, only one worker at a time can be productive in a position.

At the beginning of the period, all workers start unemployed. Firms begin by choosing their level of capital inputs $K_{j,t}$ and the number of positions $\{P_{jk,t}\}$. The firm pays a rental rate r^K per unit of capital. After capital and positions are chosen, unemployed workers and positions match within each skill market. On the firm side, the total number of positions available to workers of skill group s is $P_{s,t} = \sum_j \sum_{k \in s} P_{jk,t}$. The number of matches in a skill sub-market is the minimum of the measure of positions and measure of unemployed searchers. We will focus on "slack" equilibria where the measure of positions in each skill sub-market s is weakly less than the measure of unemployed searchers, so each position will be matched to a worker with probability 1.[4]

After firms have matched their positions with workers, firms pay a small onboarding cost ϵ_k for each hired worker. Now that workers are in their respective positions, they learn how

[3]For example, this could mean that managerial and professional occupations draw from college educated workers, while production and services jobs employ non-college workers. Our model assumes that all college workers can equally work in managerial and professional jobs. Thus we abstract from occupation-specific skills, and workers are differentiated in their overall skill level and whether they have position-specific skills.

[4]We verify that this is in fact an equilibrium outcome.

to do their particular job: i.e., they learn skills that are position-specific while on the job, so firms have $P_{jk,t}$ trained workers. After workers have been hired, firms choose a wage for each occupation $w_{jk,t}$. Firms are subject to an equal-pay constraint for all workers within an occupation, so senior and junior workers within an occupation are paid the same wage.[5] Firms commit to pay $w_{jk,t}$ for the period.

After workers have learned their job-specific skills, workers have an opportunity to quit their job, knowing that they will have an opportunity to re-match before production occurs. Workers draw job-specific disultility of labor $\chi_{ijk,t}$, which is private information and is not known by the firm. Let the value of quitting be $V_{s(k),t}$, which is common for all workers of the same skill group s. Note that workers employed in different occupations k but have the same skill s have the same value of their outside option. Trained workers will quit if the value of quitting exceeds the value of utility from remaining in their job:

$$\mathcal{U}(w_{jk,t}, \chi_{ijk,t}) < V_{s(k),t}.$$

From the firm's perspective, any individual workers' quit probability is random, but the total quit rate $Q(w_{jk,t}, V_{s(k),t})$ is a function of the firm's wage for that occupation k and the outside value, with $\frac{\partial Q}{\partial w} < 0$. Then the number of senior workers in each occupation is who stay in firm j is

$$S_{jk,t} = (1 - Q(w_{jk,t}, V_{s(k),t}))P_{jk,t}.$$

As the last step before production, firms have an opportunity to hire "junior" replacement workers $J_{jk,t}$ for the positions that were just vacated. The replacement workers come from a pool of both workers who had a job earlier in the period and quit as well as workers who were never successfully matched in the first round of hiring. Matching is again random. As with hiring workers earlier in the period, firms have to pay the ϵ_k onboarding cost for each replacement worker. Firms are also limited to one worker per position, so $J_{jk,t} + S_{jk,t} \leq P_{jk,t}$.

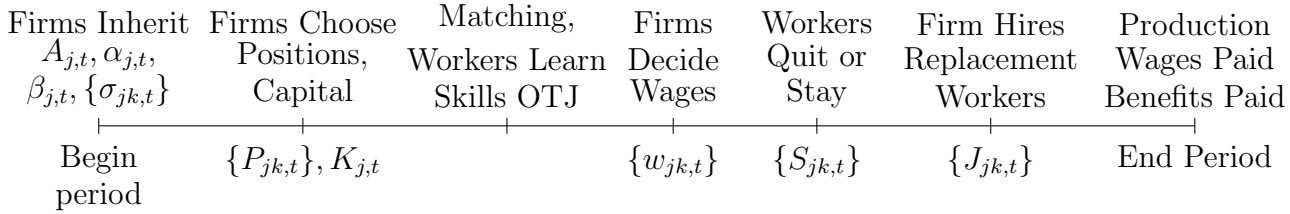
Once replacement "junior" workers are hired, then the firm produces and pays wages. If workers are re-hired as replacement workers, their utility is just $w_{jk,t}$. If workers are unemployed after the second round of matching, they consume their skill-specific unemployment benefits b_s. Workers' value V_s of being unemployed before the second round of matching is just their expected consumption at the end of the period. Thus we have

$$V_s = b_s + \frac{\sum_{k \in s} \int_j J_{jk,t}(w_{jk,t} - b_s)}{\phi_s - \sum_{k \in s} \int_j S_{jk,t}},$$

[5] In Appendix A.6 we relax this assumption in a partial equilibrium exercise, allowing firms to pay replacement workers their outside option.

where $\phi_s - \sum_k \int_j S_{jk,t}$ is the measure of unemployed workers of skill level s during the second round of matching. The timing of the model can be summarized by the following figure

Figure 2: Model Timing



Parameterization To close the model, we parameterize the workers' utility function and labor disutility draw. Let the utility of a trained worker i staying in firm j in occupation k be

$$\mathcal{U}(w_{jk,t}, \chi_{ijk,t}) = \frac{w_{jk,t}}{\chi_{ijk,t}}.$$

Let $\chi_{ijk,t}$ be distributed Pareto $(1, \gamma_s(k))$. This means that for the value of the outside option $V_{s(k)}$, the worker will stay if

$$V_{s(k)} \leq \frac{w_{jk,t}}{\chi_{ijk,t}} \Leftrightarrow \chi_{ijk,t} \leq \frac{w_{jk,t}}{V_{s(k)}}.$$

Therefore the worker will stay according to the following probability:

$$P[\text{stay}] = 1 - \left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}.$$

Using the law of large numbers, that means that the firm's retention function $1 - Q(w_{jk,t})$ will be

$$S_{jk,t} = \left(1 - \left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}\right) P_{jk,t}.$$

2.3 Firm's Problem and Equilibrium

We now state the firm's problem. We derive the firm's optimal wage, the pass-through elasticity from productivity to wages, and the surplus of the match to the firm. All of these results hold in partial equilibrium. Then we characterize the equilibrium.

Firm's Problem The firm's problem is

$$\max_{\substack{K_{j,t},\{P_{jk,t}\},\\ \{w_{jk,t}\},\{S_{jk,t}\},\{J_{jk,t}\}}} A_{j,t}K_{j,t}^{\beta_{j,t}} \prod_{k\in\mathcal{K}(j)} \left((S_{jk,t}+J_{jk,t}) \left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{\delta_k+\frac{h_k}{\alpha_{j,t}}} \right)^{\sigma_{jk,t}\alpha_{j,t}} - r^K K_{j,t} - \sum_k w_{jk,t}(J_{jk,t}+S_{jk,t}) - \sum_k \epsilon_k(J_{jk,t}+P_{jk,t})$$

$$\begin{aligned} S_{jk,t} &= \left(1-\left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}\right) P_{jk,t} \text{ if } w_{jk,t} > V_{s(k)} \\ &= 0 \qquad\qquad \text{otherwise} \\ J_{jk,t}+S_{jk,t} &\leq P_{jk,t}. \end{aligned} \tag{2}$$

Firms maximize value added minus capital costs, wage costs, and onboarding costs, subject to the constraints that the number of senior workers equals the share of trained workers that are retained, and that firms can have at most one worker per position.

Next, we derive the key equations from the model and map them into equations that can be estimated using our data.

Optimal Wages To get closed form solutions, we take the limit as $\epsilon_k \to 0\ \forall k$. [6] In this limit, optimal wages are

$$w^*_{jk,t} = \left(1+\gamma_{s(k)}\left(\delta_k+\frac{h_k}{\alpha_{j,t}}\right)\right)^{\frac{1}{\gamma_{s(k)}}} V_{s(k)}. \tag{3}$$

This equation captures the key force in the model: firms trade off wage costs with turnover costs. As both δ_k and h_k increase, turnover costs rise, incentivizing firms to pay higher wages to avoid this costly turnover. A higher value of δ_k increases only the level of wages. A higher value of h_k raises the level of wages but also interacts with the output elasticity of labor $\alpha_{j,t}$: when the output elasticity of labor $\alpha_{j,t}$ is lower, firms will generally have a higher output per worker and there are higher turnover costs for occupations with $h_k > 0$. Therefore to avoid the higher turnover costs, firms that employ occupations with $h_k > 0$ with lower values of $\alpha_{j,t}$ will pay higher wages. Taking elasticities, we have

$$\varepsilon_{w_{jk},\alpha_j} = \frac{-h_k}{\alpha_j + \gamma_{s(k)}(\alpha_j\delta_k + h_k)} \tag{4}$$

Thus, changes in firms' output elasticity of labor affects wages *only* if workers have hold-up power $h_k > 0$.

[6] For $\epsilon_k = 0$ the number of positions would not be pinned down, since training workers would be completely free. However, the expression for optimal wages is not closed-form when $\epsilon_k > 0$. Therefore, we take the limit of $\epsilon_k \to 0$, which gives us a closed form expression for wages while preserving the unique solution.

Note in equation (3) that total factor productivity $A_{j,t}$ has no effect on wages. Instead, total factor productivity only affects firm size. Define the total number of workers $N_{jk,t} \equiv J_{jk,t} + S_{jk,t}$. Optimal employment for occupation k is

$$N^*_{jk,t} = \frac{\sigma_{jk,t}\alpha_{j,t}}{w^*_{jk,t}} \left[A_{j,t} \left(\frac{\beta_{j,t}}{r^K} \right)^{\beta_{j,t}} \prod_k \left(\frac{\sigma_{jk,t}\alpha_{j,t}}{w^*_{jk,t}} \left(\frac{S^*_{jk,t}}{P^*_{jk,t}} \right)^{\delta_k + \frac{h_k}{\alpha_{j,t}}} \right)^{\sigma_{jk,t}\alpha_{j,t}} \right]^{\frac{1}{1-\alpha_{j,t}-\beta_{j,t}}}.$$

As this expression shows, total factor productivity $A_{j,t}$ affects the optimal number of workers but has no effect on wages.

In practice, what would be the source of variation in $\alpha_{j,t}$ over time within a firm? One explanation could be that the capital and labor intensity of production changes over time, i.e., $\beta_{j,t}$ rises while $\alpha_{j,t}$ falls. Another possibility is that a firm's market power changes over time, changing the $\beta_{j,t} + \alpha_{j,t}$. Recall that $Y_j()$ is a firm's value added function, not necessarily its revenue function. In Appendix A.7, we provide a microfoundation of finite product demand where low $\alpha_{j,t}$ can arise from product market power, which lowers the elasticity of value added with respect to labor.

Pass-through of Productivity to Wages When trying to estimate the relationship between wages and productivity, we will not be able to directly observe α_j. However, we can observe output per worker (i.e., firm's average product of labor APL) and the firm's labor share of value added. To map the model's results to observables, we first highlight a few relationships that arise when firms are optimizing.

First, we have the intermediate results that junior workers are paid their marginal product of labor and firms replace every worker who quits:

$$MRP^J_{jk,t} = w_{jk,t}$$
$$J_{jk,t} = P_{jk,t} - S_{jk,t}.$$

This also means that $\alpha_{j,t}$ will be equal to the labor share, and that the share of value added that is spent on the wages of workers in occupation k will be $\sigma_{jk,t}\alpha_{j,t}$:

$$\text{Labor Share}_{j,t} \equiv \frac{\sum_k w_{jk,t}(S_{jk,t} + J_{jk,t})}{Y_{j,t}} = \sum_k \sigma_{jk,t}\alpha_{j,t} = \alpha_{j,t}. \tag{5}$$

Define output per worker as $APL_{j,t} \equiv \frac{Y_{j,t}}{\sum_k N_{jk,t}}$. Output per worker also rises as the output elasticity of labor falls $\alpha_{j,t}$. The elasticity of measured output per worker $APL_{j,t}$ with respect to $\alpha_{j,t}$ is

$$\varepsilon_{APL_j,\alpha_j} = -1 - \sum_k \frac{N_{jk}}{N_j} \left(\frac{h_k}{\alpha_j + \gamma_{s(k)}(\delta_k \alpha_j + h_k)} \right),$$

which is strictly negative and weakly less than -1. Thus, optimizing firms with low output elasticities of labor have high measured output per worker.

Combining expressions, we can derive the elasticity of wages with respect to observed output per worker due to variation in α_j: $\varepsilon^{(\alpha_j)}_{w_{jk},APL_j} = \varepsilon_{w_{jk},\alpha_j}/\varepsilon_{APL_j,\alpha_j}$.

$$\varepsilon^{(\alpha_j)}_{w_{jk},APL_j} = \frac{\dfrac{h_k}{\alpha_j + \gamma_{s(k)}(\delta_k \alpha_j + h_k)}}{\underbrace{1 + \sum_{\ell} \frac{N_\ell}{N_j}\left(\frac{h_\ell}{\alpha_j + \gamma_{s(\ell)}(\delta_\ell \alpha_j + h_\ell)}\right)}_{\text{Firm-level denominator: employment-weighted pass-through}}}. \tag{6}$$

The numerator is simply the (minus) elasticity of wages with respect to the firm's output elasticity of labor $\alpha_{j,t}$. The denominator takes into account that as $\alpha_{j,t}$ changes, the optimal composition of employment shifts, since changes in $\alpha_{j,t}$ affect the relative optimal wages of different occupations within the firm, and therefore the optimal employment in each occupation. Importantly, the elasticity of wages to output per worker $\varepsilon^{(\alpha_j)}_{w_{jk},APL_j}$ for each occupation k has the same denominator, so within a firm, occupations with higher values of h_k will have higher pass-through of (measured) productivity to wages.

In total, variation in the firm's returns to scale $\alpha_{j,t}$ moves wages, the labor share, and output per worker. In Section 4.2, we will use an instrumental variables strategy to estimate $\varepsilon_{w_{jk},\alpha_j}$ for each occupation using these pass-through formulas.

Firm Surplus from Senior Workers When we turn to empirics, we will use worker separations in different occupations at high- and low- labor share firms to build evidence for which occupations give workers hold-up power.

In the model, firms find it optimal to replace the workers who quit: $J_{jk,t} = P_{jk,t} - S_{jk,t}$. However, the firm is indifferent between replacing the final (or marginal) worker at the point of $J_{jk,t} = P_{jk,t} - S_{jk,t}$, since the marginal product of the last junior worker $MRP^J_{jk,t}$ is equal to the wage $w_{jk,t}$. Thus, regardless of whether the firm replaces a worker, the surplus in the match to the firm from retaining a senior worker is just the senior worker's marginal product minus their wages. This will be useful when we later estimate the profit losses of firms losing a single worker.

As a share of the worker's wage, the surplus of the match to the firm is:

$$\frac{MRP^S_{jk} - w_{jk}}{w_{jk}} = \left(\delta_k + \frac{h_k}{\alpha_j}\right)\left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{-1}.$$

When firms are optimizing, the share of experienced workers who stay in the firm will be $S_{jk,t}/P_{jk,t} = \left(\gamma\left(\delta_k + \frac{h_k}{\alpha_j}\right)\right) / \left(1 + \gamma\left(\delta_k + \frac{h_k}{\alpha_j}\right)\right)$. Thus when firms are optimizing, the

surplus of retaining an experienced worker as a fraction of their wage will be

$$\frac{MRP^S_{jk,t} - w_{jk,t}}{w_{jk,t}} = \frac{1}{\gamma_{s(k)}} + \delta_k + \frac{h_k}{\alpha_{j,t}}. \tag{7}$$

As mentioned earlier in Section 2.1, δ_k is the turnover cost in units of wages, and so a higher value of δ_k increases the surplus to the firm as a fraction of the wage. A higher value of h_k raises the surplus relative to the wage but is scaled by the inverse of the firm's labor elasticity α_j. Lastly, the match surplus to the firm as a share of the worker's wage depends on the workers' quit elasticity $\gamma_{s(k)}$. This result stems from the fact that losing senior workers gets more costly as the senior share $\frac{S_{jk,t}}{P_{jk,t}}$ falls. With a lower value of $\gamma_{s(k)}$, workers are less wage sensitive, so firms know they can offer lower wages without losing as many workers. This means that firms end up exploiting this monopsony power more and choose a lower value of $\frac{S_{jk,t}}{P_{jk,t}}$. At these lower values of $\frac{S_{jk,t}}{P_{jk,t}}$, losing an additional worker is more costly. Thus, the presence of $\gamma_{s(k)}$ in this equation shows that when firms have more monopsony power, firms exploit this power more, resulting in greater turnover by incumbent workers, which makes the marginal incumbent worker more valuable.

Equilibrium An equilibrium is a set of capital $\{K_j\}_j$, positions $\{P_{jk}\}_{j,k}$, wages $\{w_{jk}\}_{j,k}$, senior workers $\{S_{jk}\}_{j,k}$, junior workers $\{J_{jk}\}_{j,k}$, and outside options $\{V_s\}_s$ such that (i) firms maximize profits and (ii) workers maximize utility.

Note that the labor market does not need to clear. Indeed, as δ_k and h_k increase, firms choose higher wages and fewer positions, generating unemployment in equilibrium. This is characteristic of efficiency wage models, which we will explain more in Section 2.6.

2.4 Summary of Predictions

This model yields six main predictions. (i) Workers in high hold-up occupations should earn higher wages compared to workers with similar outside options but in lower hold-up occupations. (ii) Firms with higher productivity as measured by value added per worker should pay higher wages in the cross-section, but only to workers in occupations with hold-up power. (iii) Conditional on firm productivity, wages should not be related to firm size. (iv) Persistent changes in firm productivity should pass through more strongly to workers' wages in high hold-up occupations. (v) Profit losses associated with involuntary worker separations from the firm's perspective should be larger for workers in high hold-up occupations. (vi) Profit losses from involuntary separations should be larger at lower labor share firms in high hold-up occupations.

2.5 Discussion of Key Assumptions

In this section, we show that relaxing any of the three key assumptions - (i) position-specific rather than firm-specific skills, (ii) only one worker can learn position-specific skills at a time, (iii) multiplicative output losses following the separation of an experienced incumbent — eliminates the effect of h_k on wages. Intuitively, these assumptions ensure that experienced incumbents have an exclusive ability to perform the tasks in their position in the current period, and no other worker can replace them in the short run. If any of the assumptions are relaxed, turnover becomes purely a cost at the margin rather than a disruption to the firm's inframarginal productivity, eliminating hold-up power.

Position-Specific Skills vs. Firm-Specific Skills In our model, workers must learn the skills for a particular position. This limits the number of workers who can be trained to the number of positions, exposing the firm to the large costs of separations after workers have learned their position's skills. Suppose instead that skills were firm-specific so firms could train more workers than positions. In this environment, firms would train enough employees to ensure that the number of experienced workers who remain equals the number of positions. Firms would then trade off wage costs against expected training costs rather than against disruptions to production. Since firms could ensure that $S_{jk,t}/P_{jk,t} = 1$, then firms would never face inframarginal losses, and firms with different average products would pay the same wages.[7]

Indivisible Positions/One Worker per Position Our model relies on the assumption that only one worker can learn a position's specific skills at a time, which we see as the static-problem analogue to learning-by-doing. Suppose instead that the firm could train multiple workers in the same position. In this case, for very important positions, firms could 'double up' by training two workers and paying to have a back-up worker ready in case the first worker leaves. In this case, firms would not be exposed to inframarginal losses due to losing the one worker who can do the tasks of a position. Furthermore, positions are indivisible in the sense that if a position is vacated, the firm cannot divide up the tasks associated with a position or re-assign them to other workers.

Multiplicative Production. If workers have position-specific skills but output is not individually complementary (i.e., $h_k = 0$), then worker separations do not disrupt the firm's inframarginal output. In this case, higher productivity firms do not pay higher wages, and

[7] In this extension, wages would satisfy $(1 - w^{-\gamma})^2 = \epsilon\gamma w^{-\gamma-1}$ and not depend on δ_k, h_k, or α_k.

the elasticity of wages with respect to output per worker satisfies $\varepsilon_{w_{jk},APL_j} = 0$. Without complementarities, there is no economically meaningful distinction between position-specific and firm-specific skills.

Incomplete Contracts One assumption of this model is that workers cannot post a bond (Shapiro and Stiglitz, 1984; Dickens et al., 1989), which employers could revoke if workers were trained and then quit before production. If this were available, firms could require a sufficiently high bond that workers would never quit, resulting in firms never facing extra output losses from losing senior workers. If there is no threat of inframarginal losses, then productive firms would have no incentive to pay higher wages.

2.6 Relation to the Literature

Efficiency Wage Interpretation At its core, this model is an efficiency wage model: when firms pay a higher wage, firms get more output per worker. Define $X_{jk,t} \equiv \left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{\delta_k + \frac{h_k}{\alpha_{j,t}}}$, so value added is $Y_{j,t} = A_{j,t} K_{j,t}^{\beta_{j,t}} \left(\prod_{k \in \mathcal{K}(j)} N_{jk,t} X_{jk,t}\right)^{\sigma_{jk,t}\alpha_{j,t}}$. As an efficiency wage model, the Solow efficiency wage conditions holds when firms are optimizing:

$$\varepsilon_{X_{jk},w_{jk}} = 1,$$

which is equivalent to the optimal wage condition in equation (3). This condition states that when firms are optimizing, a one-percent increase in wages at the margin raises output per worker by exactly one percent, reflecting the trade-off between higher wage costs and improved retention of experienced workers.

Our model differs from the famous Shapiro and Stiglitz (1984) shirking model in a key way: in our model, retention of a given worker is probabilistic, while in the Shapiro and Stiglitz (1984), firms need only to satisfy a known no-shirking condition. This probabilistic feature of retention in our model helps generates rent sharing: since productive firms have more to lose from losing an experienced worker in a high h_k occupation, productive firms have a stronger incentive to pay high wages to improve retention. In a model like Shapiro and Stiglitz (1984), firms would need only to satisfy the worker's no-shirking condition, which does not vary with firm productivity, and so wages would not vary with firm productivity.

Comparison with Existing Rent Sharing Models The two most common families of models in which firms share rents with workers are monopsony models and bargaining models.

In monopsony models, workers earn rents at productive firms because firms with high total factor productivity are constrained by an upward sloping labor supply curve, endogenously generating both high value added per worker and high wages. Models with isolelastic labor supply curves (Lamadon et al., 2019; Chan et al., 2020) imply a pass-through of value added per worker to wages of 1, which is far above empirical estimates (Card et al., 2018). Models with firms' labor supply elasticity declining with firm size can achieve realistic rent sharing Kline (2025) but predict that the firm size-wage premium is steeper for large firms and that large firms have higher pass-through of TFP shocks to wages, which contradicts empirical evidence from Bloom et al. (2018) and Chan et al. (2024), respectively. Monopsony models as an explanation for rent sharing also are at odds with evidence from Bloesch et al. (2025) who find that firms do not need to pay higher wages to become larger in the long run.

The main bargaining model with multi-worker firms is Stole and Zwiebel (1996). In this model, wages can depend on average firm productivity, as firms and workers take into account what workers' marginal product would be if rounds of negotiations broke down. This relationship between average productivity and wages is stronger for workers with higher bargaining power. However, this dependence of wages on average productivity (in steady state) only arises in a special case of the model where employment is fixed. When firms can choose employment, firms intentionally over-hire to lower the worker's marginal product, driving the bargained wage down to the worker's outside option. Thus, the version of Stole and Zwiebel (1996) that is most comparable to our model, when firms can choose employment, does not generate a relationship between firm average productivity and wages.

Relation to Incentive Pay Our model is distinct from standard models of incentive pay, in which firms use explicit performance-contingent compensation schemes to align workers' efforts with firm objectives. In incentive pay models, higher pay can raise output by inducing greater effort along a measurable performance margin, and compensation is explicitly tied to individual or team performance outcomes.

In contrast, the wage premia in our model do not arise from effort provision or moral hazard. Instead, higher wages increase output by improving the retention of experienced workers whose separation would disrupt production. Because skills are position-specific and production features complementarities, the firm benefits from paying higher wages even when worker effort is fixed. Wages, therefore, serve as a retention device rather than as an incentive for higher effort.

2.7 Mapping to Estimation

In this subsection, we describe how we use our model to motivate empirical exercises that will be useful for estimating model parameters.

Persistent vs Transitory Shocks and Pass-through Estimation Later in Section 4.2 we will use pass-through regressions similar to Juhn et al. (2018) and Card et al. (2018) that distinguish between persistent and transitory shocks to productivity. In the model presented in Section 2, variation in $\alpha_{j,t}$ generates persistent changes in output per worker, but there are no shocks that generate transitory shocks to output per worker.

Therefore in Appendix A.3, we extend the model to have unanticipated changes in TFP that generate transitory productivity innovations that also affect wages. Specifically, we decompose $A_{j,t} = A^0_{j,t} z_{j,t}$ into a component $A^0_{j,t}$ that is known prior to firms choosing their capital and positions and a surprise component $z_{j,t}$ that is learned after capital and the number of positions is chosen, but before wages are set. Fluctuations in $z_{j,t}$ therefore affect output and optimal wages (as higher z increases the value of retaining workers), but firms are not able to bring average productivity back down by hiring more workers within a period. We show that this extension motivates the instrumental-variables estimation that we perform in Section 4.2 if the variance of z_t shocks is sufficiently small.

One of the challenges with estimating pass-through elasticities is that many different shocks may affect output per worker and wages in ways that are different that variation in $\alpha_{j,t}$. For example, variation in $\beta_{j,t}$ or $\sigma_{jk,t}$ could move output per worker but does not change optimal wages. The following proposition shows that instrumenting changes in the labor share with longer differences in output per worker isolates the effect of persistent variation in the output elasticity α_j on wages. In the following proposition, we outline an estimation procedure that isolates the effect of variation on the output elasticity α_j on wages.

Proposition 1 *Let $W(z, \alpha)$ and $\theta(z, \alpha)$ be vectors of first difference in observed log wages and firms' log labor share, respectively, as generated by the model. Also let Z be a vector of long differences in measured log output per worker, where the time period of Z covers the first differences of wages and labor but does not share end points. Assume that h_k, δ_k, and $\gamma_{s(k)}$ remain fixed.*

For small shocks to $z_{j,t}$ and small variation in $\alpha_{j,t}$, a regression of changes in wages on changes in the labor share, instrumented with longer periods of changes in output per worker, approximates the elasticity $\varepsilon_{w_{jk},\alpha_j}$

$$(Z'\theta)^{-1}(Z'W) \approx \frac{-h_k}{\alpha_j + \gamma_{s(k)}(\alpha_j \delta_k + h_k)}. \tag{8}$$

This result holds even with shocks to $A_{j,t}$ or changes in $\sigma_{jk,t}$, since neither has a persistent effect on the firm's labor share. We also show in Appendix A.3 that the instrumental variable regression sweeps out transitory measurement error in output, similar to Juhn et al. (2018), as well as persistent measurement error in wages.

Estimation of Firm Surplus Using Exogenous Separations In equation (7) we derived the surplus in the match to the firm as a function of workers' wages and other parameters. Thinking ahead to estimation, we now show that the size of this surplus can be measured by estimating the effect of exogenous worker separations on the firm's value added less the wage bill.

Proposition 2 *Define firm j's wage bill as $WB_{j,t} \equiv \sum_k \sum_{i \in k} w_{ijk,t}$. Suppose a senior worker in occupation k at firm j separates exogenously. Regardless of whether or not the firm replaces the worker, the per-period profit loss to the firm from losing worker i is given by*

$$\Delta \left(Y_{j,t} - WB_{j,t} \right) \approx -w_{ijk,t} \left(\frac{1}{\gamma_{s(k)}} + \delta_k + \frac{h_k}{\alpha_{j,t}} \right). \tag{9}$$

Proof: see Appendix A.4. This proposition states that in response to an exogenous worker separation, the change in value added net of the wage bill relative to the deceased worker's prior wage equals the match surplus share of the worker's prior wage. This result holds in the model regardless of whether the firm replaces the exogenously separated worker, since at the margin, replacement workers are paid their marginal product $w_{jk,t} = MRP^J_{jk,t}$.

Indirect Inference Equations (8) and (9) show how to map moments based on observables (value added, value added per worker, labor share, wages, and wage bill) back into parameters. In practice, we will estimate three moments for each occupation: the effect of worker deaths on profits at low labor share firms, the effect of worker deaths on profits at high labor share firms, and pass-through elasticities. Taking as given external evidence on separation elasticities to pin down γ, the model has three moments and two parameters, so our model will be over-identified.

3 Data

In the empirical part of this paper, we use two different sources of data: administrative data on workers and firms from Denmark from 2008 to 2019, and job posting data in the US from Lightcast from 2010 to 2018.

3.1 Danish Administrative Data

We use administrative data from Denmark on workers and firms from 2008-2019 to generate a merged employer-employee panel data set. For the worker data, we use the IDAN registry, which reports workers' earnings, hours, occupation, firm, and establishment.[8] Beginning in 2008, wage information is drawn from the e-indkomst monthly online reporting database, which is regarded as highly reliable. We also use the BFL database, which reports similar information to IDAN but on a monthly level. The demographic registries provide information on age and education. The timing of worker deaths is obtained in the DOD registry, which records the date of all deaths in Denmark.

For firm data, we use the FIRM registry, which is an annual dataset containing data on firm sales, value added, employment (measured in full-time equivalents), gross salaries, and gross profits. The unit of observation is the firm, which is identified by the variable *cvrnr* rather than by establishments, which are identified with *arbnr*, as firm financial variables such as value added are reported only at the firm level. Financial data is reported from May of year t to April of year $t+1$.

In our empirical exercises, we use overlapping samples of Danish firm and worker data. In the section estimating the effect of worker deaths, we use data from 1,331 firms that experienced one worker death during the sample period, all of which have between 5 and 100 full-time equivalent employees, as well as up to five matched control firms. The summary statistics for this sample can be found in the first three rows of Table 1. On average, the firms in the sample employ [45.9] full-time equivalent workers and produce approximately [$4.3] million in value added.

In the pass-through exercises, we use a much larger sample of workers and firms, including all full-time workers employed at a private sector firm that report value added and industry and employ, on average, at least five full-time equivalent workers. Summary statistics on the firms in this sample can be found in the bottom three rows of Table 1. There are [100,993] unique firms, with [8,081,081 worker]×year observations. There are approximately [1 million worker]× firm observations per year, out of a labor force of about [2.5] million.[9]

[8]Many workers' occupations change in 2009 but are the same in years 2008 and 2010. For workers whose employer is the same between 2008-2010 and the occupation is the same in 2008 and 2010, we impute the occupation in 2009.

[9]Public-sector employment, part-time employment, and self-employment account for the remainder of the workforce.

Table 1: Firm Summary Statistics

Note: The following table reflects data from an out-of-date sample. Results will be updated upon renewed data access.

	Deaths Sample - Firms			
N =1218	mean	sd		
Full Time Equivalent (FTE)	45.9	23.7		
Value Added ($, million)	4.327	3.894		
VA-salaries ($, million)	.851	2.936		
	Pass-through Sample - Firms			
Average taken across:	Firms		Workers	
N= 8,081,081	mean	sd	mean	sd
Full Time Equivalent (FTE)	75.9	346	1,794	3,966
Value Added ($, million)	9.14	63.1	221.6	661
VA/Worker ($, thous)	106.7	389	116.3	178

[This table reports summary statistics for the firm samples used in the deaths and pass-through analyses. In the deaths analysis, we restrict our attention to firms with an average number of full-time equivalent employees between 5 and 100 in the three years before the death. The average firm has 45.9 workers and produces approximately $4.3 million worth of value added each year (deflated to 2005 dollars). In the pass-through sample, we use larger firms, restricting attention to firms with an average of at least 5 full-time equivalents over the sample. The average firm in the sample has 76 full-time equivalent employees, but the mean of firm size across workers is 1,794 employees. Value added per worker is similar depending on if we weight by firms or workers, at about $100,000 per worker.]

Grouping Occupations Danish occupations are categorized according to the Danish International Standard Classification of Occupations, or DISCO codes, which consists of ten one digit occupation groups. We exclude agricultural and armed forces occupations from the analysis. Summary statistics on workers' wages, average tenure, differentiation score (which we will explain in Section 4.3), average years of schooling, and share of employment by DISCO group in the 2008-2019 pass-through sample are shown in Table 2.

Table 2: Mean Characteristics of 1-Digit Occupation Groups

	(1)	(2)	(3)	(4)	(5)	
DISCO Group	Log Wage	Tenure	Diff. Score	Educ	Emp Share	Broad Group
Managers	5.77	7.6	1.89	14.8	5.9	Manager
	(.54)	(7.1)	(.45)	(2.2)		
Professionals	5.59	5.7	1.41	16.2	13.6	Mid-wage
	(.36)	(5.6)	(.84)	(2.2)		
Technicians	5.38	6.3	0.62	14.5	15.7	Mid-wage
	(.35)	(6.3)	(.77)	(1.9)		
Crafts	5.26	6.6	-0.01	13.5	13.2	Low-wage
	(.34)	(6.8)	(.84)	(1.9)		
Assembly	5.23	7.2	-0.87	12.4	11.5	Low-wage
	(.23)	(6.9)	(1.27)	(2.3)		
Administrative	5.15	6.4	-0.02	13.5	10.6	Low-wage
	(.30)	(6.6)	(.74)	(2.0)		
Service	4.91	3.6	-2.01	12.2	15.7	Low-wage
	(.37)	(4.3)	(.88)	(2.1)		
Manual	5.03	4.5	-1.51	11.9	13.8	Low-wage
	(.37)	(5.3)	(.89)	(2.5)		

This table reports summary statistics for 1-digit DISCO occupation groups in Danish matched employer–employee data, using the sample described in Section 3. The data cover the period 2008–2019 and include workers employed at firms with an average of at least 5 employees that report value added. All statistics are computed at the worker level and then averaged within occupation groups. Log wages are annual log earnings deflated to 2005 prices. Tenure measures years of tenure with the current employer. Diff. Score is a standardized measure of task differentiation, with higher values indicating occupation where a job's requires skills differ from the required skills of other jobs in the firm, which we will explain in Section 4.3. Educ reports average years of completed schooling. Emp Share reports the average share of total employment accounted for by each occupation group over the sample period. The final column maps each DISCO group into the broad occupation categories.

Throughout the paper, we divide workers into three broad occupation groups: managers (DISCO group 1), professionals and technicians are grouped into "middle-wage" occupations (1-digit DISCO codes 2 and 3), and crafts, assembly, administrative, service, and manual occupations are grouped into "low-wage occupations" (1-digit DISCO codes 7, 8, 4, 5, and 9, respectively).

At the top of the labor market, both managers and professionals exhibit high wages, high levels of education, and high levels of within-firm task differentiation. We nevertheless distinguish managers from professionals and technicians because we conjecture that managerial inputs may be more complementary in production than non-managerial inputs. We combine other occupation groups based on occupational characteristics and worker flows across

occupation groups.

Institutional Setting Denmark is known for its "flexicurity" model of labor market institutions, characterized by a generous welfare state and high union density, but with limited employment protections, high labor mobility, and industry-level collective bargaining. There is no statutory national minimum wage. Instead, minimum wages are negotiated at the industry level through collective bargaining between labor unions and employer organizations, while firms retain considerable scope to set individual wages above these negotiated minima. Labor mobility, measured by hiring and separation rates, is high by OECD standards and is more similar to the US than to most continental European economies (Caldwell and Harmon (2019)). In recent decades, firm-level bargaining has become increasingly important in determining wages (Dahl et al. (2013)).

3.2 Lightcast Job Posting Data

We use vacancy data from Lightcast, which covers nearly the entire universe of online job postings in the United States from 2010-2018. The job postings include information such as location, employer, and 6-digit SOC occupation codes. Further, the job postings include detailed information about the skill requirements posted with every vacancy, which Lightcast cleans and categorizes into broad skill categories. We describe merging the 6-digit SOC codes used by Lightcast to Danish ISCO in Section 4.3.

The Lightcast data have several advantages for our analysis. Most importantly, the standardized skill information is derived directly from the text of job postings, allowing us to measure skill requirements at a highly detailed level. This level of detail is essential because we construct an index of how unique a job's skill requirements are relative to other positions within the same firm. Our measure therefore captures differences in skill requirements across positions, even within the same occupation and firm.[10]

4 Empirical Evidence

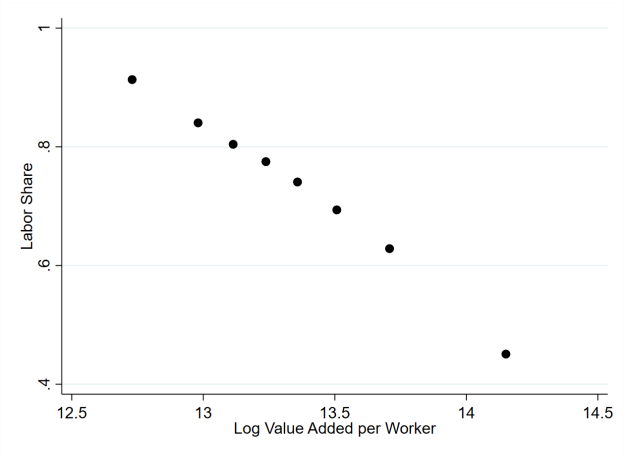
This section presents the main empirical evidence. We first estimate the pass-through of persistent productivity changes to wages and then examine the effects of worker deaths on firm profits.

[10] Jensen (2024) use Danish job vacancy data provided by HBS Economics and classify the text of job postings into nine broad skill categories, which is too coarse for our purposes. Bagger et al. (2022) use vacancy data from Jobindex A/S, which are useful for measuring the number of vacancies but do not provide the detailed skill classifications available in Lightcast.

4.1 Cross-Sectional Evidence

We begin by documenting cross-sectional empirical relationships predicted by the model. First, Figure 3 reports the cross-sectional relationship between firms' log value added per worker and the firms' labor share of value added. Both of these statistics are calculated as the average from 2008-2019. The two variables are negatively related: firms with a high value added per worker have low labor shares, with an elasticity of around -0.5.

Figure 3: Firm Value Added per Worker and Labor Share



This figure reports the relationship between log firm value added per worker and firm labor share. Variables are constructed as a firm-level average between 2008 and 2019.

Next, we report the elasticity of wages with respect to firm size and firm productivity in the cross-section, both for all occupations and separately for our three occupation groups. We restrict our sample of workers to workers who have been in their firm for at least two full years. Because we are starting with descriptive evidence and are not using our instrumental variables strategy yet, we want a measure of output per worker that approximates $\alpha_{j,t}$. To do so, we construct a measure of output $\hat{Y}_{j,t}$ that is residualized to remove variation due to the composition of its occupations. We explain how we residualize value added per worker in Appendix B.2. Our full cross-sectional specification takes the form:

$$\log(w_{ijk,t}) = \omega_N \log N_{j,t} + \omega_Y \hat{Y}_{j,t} + \sum_{k \in \mathcal{K}} \omega_k \hat{Y}_{j,t} \times \mathbb{1}\{i \in k\} + X_i\beta + Z_j\tau + F_k\theta + e_{ijk,t}, \quad (10)$$

where X_i is a vector of person controls, Z_j is a vector of firm controls, F_k denotes detailed occupation fixed effects, and $e_{ijk,t}$ is a person-firm-occupation-year error.

Table 3: Cross-Sectional and Pass-Through Estimates by Occupation Group

Note: The following table reflects data from an out-of-date sample. Results will be updated upon renewed data access.

	$\log w$ OLS			
	(1)	(2)	(3)	(4)
Productivity $\hat{Y}$	.075**	.	.	.
	(.005)			
$\log FTE$	.	.015**	.	-.023**
		(.003)		(.003)
$\hat{Y} \times \mathbb{1}\{\text{Managers}\}$	.	.	.249**	.292**
			(.015)	(.012)
$\hat{Y} \times \mathbb{1}\{\text{Middle Wage}\}$	.	.	.072**	.116**
			(.007)	(.007)
$\hat{Y} \times \mathbb{1}\{\text{Low Wage}\}$	.	.	.042**	.093**
			(.007)	(.006)
Occ FE Digit	3	3	3	3
Observations (Thous.)	5,108	5,108	5,108	5,108
R-squared	.472	.458	.483	.486
Clustered SE's	Firm	Firm	Firm	Firm

Standard errors in parentheses
** $p < 0.01$, * $p < 0.05$

This table reports estimates for equation (10). The first column reports including results including only firm productivity, and the second column reports results including only firm size. The third column includes firm productivity interacted with dummies for each broad occupation group, and the fourth column again includes firm size as a control. The standard deviations for average $\hat{Y}$ and the log of FTE are 0.63 and 1.96, respectively. All regressions include individual worker controls, firm age, workers' lagged tenure, years of education, years of education squared, potential experience, interactions of education and potential experience, as well as dummies for sex, year, and 1-digit industry code.

We report results for equation (10) in Table 3. In the first column, we set $\omega_N = \omega_k = 0 \; \forall k$ and include only our measure of output per worker, showing that the elasticity of wages with respect to output per worker is about .075. The second column reports the regression that includes only log number of full-time equivalent employees ($\omega_Y = \omega_k = 0 \; \forall k$), which reports an elasticity of .015. Column 3 includes as regressors firm productivity interacted with a dummy for our three large occupation groups, excluding the coefficient on firm size ($\omega_N = 0$). Similar to what we saw in Figure 1a, managers have the steepest slope of wages with respect to firm productivity, followed by middle-wage occupations, and low-wage occupations have the flattest slope. Column 4 includes the full specification, which steepens the slope of wages with respect to output per worker for each occupation. Confirming the patterns in Figure 1, managers have the steepest slope of wages with respect to productivity, middle-wage

occupations have a substantially lower slope, and low-wage occupations have the smallest slope.

4.2 Pass-through of Persistent Productivity Changes to Wages

From the model, we are interested in changes in wages due to shifts in the firm's output elasticity of labor, as these are the shifts that generate persistent changes in output per worker persistent rent-sharing for workers with hold-up power. We focus on the elasticity of wages of each occupation k with respect to the labor elasticity α_j because it is strictly a function of parameters, whereas the elasticity of wages with respect to output per worker depends on the composition of occupations and the source of variation. As described in Proposition 1, we can recover the elasticity of wages for each occupation k with respect to the labor elasticity α_j by instrumenting changes in the labor share with longer changes in output per worker. Define $\Delta_3 x_t \equiv x_{t+2} - x_{t-1}$ and $\Delta_7 x_t \equiv x_{t+4} - x_{t-3}$. We implement various versions of the pass-through regression using:

$$\Delta_3 \log w_{ijk,t} = \sum_{k \in \mathcal{K}} \tau_k \, \Delta_3 \mathcal{Y}_{j,t} \times \mathbb{1}\{i \in k\} + X_i \beta + Z_j \tau + F_k \theta + e_{ijk,t}, \tag{11}$$

where $\Delta_3 \mathcal{Y}_{j,t}$ is the difference in firm-level regressor $\mathcal{Y}$ from period $t-1$ to $t+2$, X_i is a vector of person controls, $Z_{j,t}$ is a vector of firm controls, F_k denotes detailed occupation fixed effects. We run specifications where $\mathcal{Y}$ may be log value added per worker, residualized value added, and the firm labor share. The sample of workers include workers who were in the firm continuously from period $t-3$ to $t+2$, so that the worker had at least two full years of tenure with the firm and stayed in the firm throughout the measured duration. Standard errors are clustered at the firm level.

We choose a horizon of $s = 2$, so we take differences of wages and the labor share over three years. Taking longer horizons of changes in wages and the labor share should reduce the share of the variance due to temporary shocks and measurement error, making our estimates more precise. For our instrument, we choose seven year changes in value added per worker. Building on the argument in Juhn et al. (2018), if temporary shocks or measurement error last for more than one year, using a wider bracket of time periods for the instrumental variable filters out the effects of the temporary shocks or measurement error. Using a seven year window for the instrument for three year differences in the dependent and independent variable will sweep out variation from lagged effects of temporary shocks.

Columns (1) and (2) of Table 4 report OLS regression of three year changes in log wages on three year changes in value added per worker and residualized value added. Managers have higher pass-through of productivity to wages than middle-wage and low-wage occupations.

The pass-through elasticities are fairly small, between one-fifth and one-third of the cross-sectional elasticities. Column (3) reports the regression of three year changes in log wages on three year changes in the firm's labor share. These point estimates are all negative as would be implied by the pass-through elasticity in equation 4. However, all the point estimates are close to 0, consistent with positive bias that would mechanically arise from wages being both on the left and right hand sides of the equation.

Table 4: Within-Firm Pass-through of Productivity to Wages

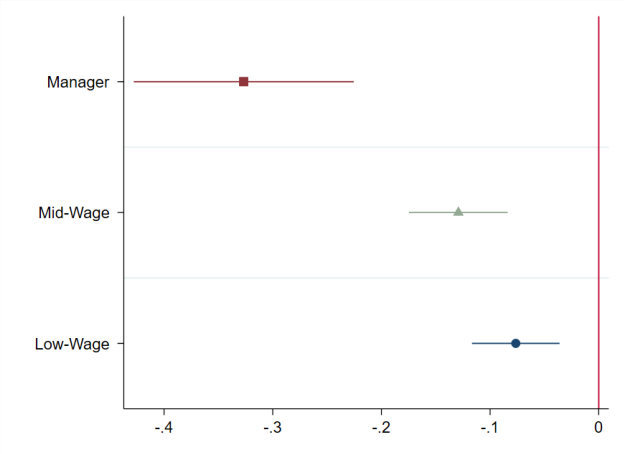
	OLS			IV	
	$\Delta_3 \log w$	$\Delta_3 \log w$	$\Delta_3 \log w$	$\Delta_3 \log w$	$\Delta_3 \log w$
	(1)	(2)	(3)	(4)	(5)
Manager	.051***	.049***	-.009**	-.277***	-.327***
	(.004)	(.003)	(.004)	(.055)	(.052)
Middle Wage	.018***	.019***	-.006***	-.107***	-.129***
	(.002)	(.002)	(.002)	(.035)	(.023)
Low Wage	.018***	.018***	-.001	-.053***	-.076***
	(.002)	(.002)	(.002)	(.021)	(.021)
Constant	.044***	.045***	.043***	.	.
	(.011)	(.011)	(.011)		
Independent Variable	$\Delta_3 \log(Y/N)$	$\Delta_3 \hat{Y}$	Δ_3 Lab Sh	Δ_3 Lab Sh	Δ_3 Lab Sh
Instrument	.	.	.	$\Delta_7 \log(Y/N)$	$\Delta_7 \hat{Y}$
Observations	1,854,716	1,821,646	1,869,005	892,550	804,977
R-squared	.049	.050	.046	-.010	-.018

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

This table reports estimates of equation (11). All regressions include worker controls, firm controls, detailed occupation fixed effects, and year fixed effects. Columns (1)–(3) report OLS estimates using three-year changes. Columns (4)–(5) report IV estimates of the semi-elasticity of wages with respect to the labor share, instrumenting three-year changes in the labor share with seven-year changes in log value added per worker or residualized value added. Robust standard errors are clustered at the firm level.

Columns (4) and (5) of Table 4 report the IV estimates where the independent variable is the three year change in the labor share. In column (4), the instrument is seven year changes in log value added per worker. In Column (5), the instrument is seven year changes in the residualized measure of value added. Managers have a semi-elasticity of wages with respect to the labor share of -.327, middle-wage occupations have a semi-elasticity of -.129, and low-wage occupations have a semi-elasticity of -.076. The results from column (5) are reported graphically in Figure 4.

Figure 4: Pass-through Semi-Elasticity of Labor Share to Wages, Three Occupation Groups



This figure reports the results of the IV pass-through estimation using equation (11) for three broad occupation groups. Managers exhibit the steepest coefficient, with a point estimate of -.327. Middle-wage occupations (professionals and technicians) have an intermediate coefficient of -.129, while low-wage occupations (crafts, assembly, administrative, service, and manual jobs) have the smallest coefficient at -.076.

4.3 Supporting Evidence: Task Differentiation in US Job Postings

In the previous section, we showed that managers have the highest pass-through of productivity to wages, followed by middle-wage occupations, with low-wage occupations having the lowest pass-through. Through the lens of our model, this would suggest that managers have the greatest hold-up power h, while low-wage occupations have the lowest hold-up power.

In this section, we use more granular information about occupations to provide evidence that supports our hypothesis that heterogeneity in hold-up power is driving pass-through of productivity changes to wages. As we explained in Section 2, one of the key features that give workers hold-up power is that skills are position-specific, and so it is hard for other workers even with the same firm to perform each others' tasks. Therefore we want to measure, for each occupation, on average how different are tasks in a given job are from the tasks of other jobs within the firm.

To achieve this, we construct a"differentiation score" using online job postings from Lightcast for each 6-digit SOC occupation. To construct the differentiation score, we apply a clustering algorithm that allocates job postings within an establishment to clusters based on the set of skill subcategories associated with each posting. The algorithm is called ROCK (Guha et al. (2000)), or "RObust Clustering using linKs," and is designed for clustering categorical data.[11] Lightcast, formerly Burning Glass Technologies, extracts and standardizes

[11]A common application is to cluster transaction data in which sets of goods are frequently purchased

the skills listed in job postings and organizes them into a hierarchical taxonomy consisting of individual skills, skill subcategories, and broader skill categories. The intermediate level, skill subcategories, is the level of aggregation that we use in the ROCK clustering algorithm.[12] Specifically, each posting is represented by a vector of indicators for whether a skill from each subcategory appears in the posting. Examples of the relationship between individual skills, skill subcategories, and skill categories are provided in Appendix Figure **??**. There are 442 distinct skill subcategories in the processed data. An example allocation of job postings to ROCK clusters is shown in Appendix Figure B.2.

To compute the differentiation score by detailed (6-digit SOC) occupation, we aggregate three statistics across establishments: (i) the share of job postings assigned to a nonmissing ROCK cluster (ii) among clustered postings, the log of the posting-weighted average cluster size and (iii) the share of clustered postings assigned to relatively small clusters, defined as clusters containing no more than 15 postings. We then take the first principal component of these three measures and orient its sign so that higher values indicate greater differentiation. The three components capture whether postings are difficult to group with other postings at the same establishment and, conditional on being grouped, whether they belong to relatively small or large groups of jobs with similar skill requirements. Summary statistics for the differentiation score by occupation, aggregated to the 1-digit DISCO level, are reported in column (3) of Table 2. We construct these statistics by merging U.S. SOC occupations to DISCO codes in the Danish administrative data,[13] and then computing weighted averages of the differentiation scores within each DISCO group.

Consistent with our hypothesis about different levels of hold-up power, managers exhibit the highest levels of task differentiation, followed by middle-wage occupations (professionals and technicians). Low-wage occupations, including crafts, assembly, administrative, service, and manual jobs, have substantially lower differentiation scores on average. This pattern mirrors differences in estimated hold-up power across occupation groups while also highlighting meaningful heterogeneity in task differentiation among occupations with similar average levels of education. At the detailed occupation level, examples of differentiation scores include "Computer and Information Systems Managers" (1.67), "Pharmacists" (-0.27), "Construction Equipment Operators" (0.49), and "Cooks, fast food" (-2.21).

Next, we will estimate similar pass-through regressions as in the previous section, but

together.

[12]The earlier Burning Glass taxonomy referred to this intermediate level as "skill clusters." In the current Lightcast data, the corresponding fields are labeled "skill subcategories." We use the latter terminology to distinguish these taxonomy categories from the clusters produced by the ROCK algorithm.

[13]A special thanks to Eskil Heinesen for providing the crosswalk of U.S. SOC occupations to Danish ISCO codes.

now including our task differentiation measure as an interaction term. Additionally, we will compare our task differentiation measure with the average level of education in an occupation, showing that task differentiation does a better job of predicting pass-through than education.

4.3.1 Levels and Pass-through: Task Differentiation vs. Education

We now estimate regressions of the level of wages on firm productivity, average occupational education levels, average occupational task differentiation, and interactions of our occupation-specific measures and firm productivity according to the following specifications:

$$\log(w_{ijk,t}) = \eta_1 \hat{Y}_{j,t} + \eta_2 \overline{\text{Educ}}_{k(i)} + \eta_3 \text{Diff}_{k(i)} + X_i\beta + Z_j\tau + e_{ijk,t} \tag{12}$$

$$\log(w_{ijk,t}) = \eta_1 \hat{Y}_{j,t} + \eta_2 \overline{\text{Educ}}_{k(i)} \times \hat{Y}_{j,t} + \eta_3 \text{Diff}_{k(i)} \times \hat{Y}_{j,t} + X_i\beta + Z_j\tau + F_k\theta + e_{ijk,t}. \tag{13}$$

The specification in equation (12) estimates the effect of workers being more educated on average in an occupation or the level of task differentiation on the level of wages. Estimates are reported in the first three columns of Table 5. After controlling for both individual and average occupational education, an occupation that is one standard deviation above average in task differentiation will pay around 5 percent higher wages than an occupation with an average level of task differentiation.

The specification in equation (13) includes occupation fixed effects and interaction terms of firm productivity and average occupational education and task differentiation. Column 4 of Table 5 reports the interaction regress including only the interaction on regression, finding that the elasticity of wages with respect to value added per worker residuals is .014 percentage points higher for every year of average educational attainment. Column 5 includes only the interaction term on task differentiation, showing that a one-standard deviation in task differentiation increases the pass-through elasticity by .030. When including both average education and task differentiation in Column 6, we have that the coefficient on the interaction between task differentiation and firm productivity is positive and statistically significant, while the coefficient on the interaction term with education is effectively zero. This suggests that higher task differentiation, not higher overall education levels, predicts that wages rises with productivity more in the cross-section.

Table 5: Task Differentiation and Education: Wage Levels

Note: The following table reflects data from an out-of-date sample. Results will be updated upon renewed data access.

	$\log w_t$ Levels - OLS					
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{Y}_t$	.121***	.130***	.118***	.112***	.107***	.106***
	(.009)	(.010)	(.009)	(.006)	(.006)	(.009)
$\overline{\text{Educ Years}}_j$	.094***		.074***			
	(.002)		(.003)			
Diff. Score_j		.109***	.047***			
		(.005)	(.005)			
$\hat{Y}_t \times \overline{\text{Educ Years}}_j$				.014***		.002
				(.003)		(.002)
$\hat{Y}_t \times \text{Diff. Score}_j$					.030***	.028***
					(.003)	(.002)
Occupation FE	No	No	No	Yes	Yes	Yes
Observations (Thous.)	4,091	4,091	4,091	4,091	4,091	4,091
R-squared	.397	.373	.493	.490	.491	.491
Clustered SE's	Firm	Firm	Firm	Firm	Firm	Firm

This table reports OLS results for equation (12) and (13). Sample sizes are smaller than in Table 3 because this table only includes occupations with successfully merged differentiation scores. Controls in each regression include the firm age, workers' lagged tenure, individual years of education, years of education squared, potential experience, interactions of education and potential experience, as well as dummies for sex, year, and 1-digit industry code.

Next, we estimate pass-through of productivity to wages, including interaction terms on average education and task differentiation. The regressors are three year changes in the firm residualized value added $\Delta_3 \hat{Y}_{j,t}$, as well as interactions between these productivity changes, the differentiation score, and average occupational education attainment. The instruments are the seven year changes in residualized productivity and corresponding interactions, and we include detailed occupational fixed effects to absorb time-invariant differences in wage levels across occupations. We estimate:

$$\Delta_3 \log(w_{ijk,t}) = \chi_1 \Delta_3 \hat{Y}_{j,t} + \chi_2 \Delta_3 \hat{Y}_{j,t} \times \overline{\text{Educ}}_j + \chi_3 \Delta_3 \hat{Y}_{j,t} \times \text{Diff}_j + X_i\beta + Z_j\tau + F_k\theta + e_{ijk,t} \quad (14)$$

Table 6 reports IV estimates of the pass-through of firm productivity shocks to wages, interacted with occupation-level characteristics. Column (1) includes interactions between productivity changes and average occupational education and shows that higher-education occupations do not exhibit significantly higher pass-through. Column (2) instead interacts productivity changes with the task differentiation score and shows a strong, positive, and sta-

tistically significant relationship between task differentiation and pass-through. Column (3) includes both interaction terms simultaneously. After conditioning on average occupational education, the interaction between productivity changes and task differentiation remains positive and statistically significant.

Table 6: Task Differentiation and Education: Pass-Through

Note: The following table reflects data from an out-of-date sample. Results will be updated upon renewed data access.

	$\Delta_3 \log w_t$		
	Pass-through - IV		
	(1)	(2)	(3)
$\Delta_3 \hat{Y}_t$	.063***	.058***	.058***
	(.008)	(.008)	(.008)
$\Delta_3 \hat{Y}_t \times \overline{\text{Educ Years}}_j$	.005		-.003
	(.004)		(.004)
$\Delta_3 \hat{Y}_t \times \text{Diff. Score}_j$		.016***	.020***
		(.006)	(.006)
Observations (Thous.)	935	935	935
R-squared	.023	.022	.022

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports estimates for equation (14). Controls in each regression include the firm age, workers' lagged tenure, individual years of education, years of education squared, potential experience, interactions of education and potential experience, as well as dummies for sex, year, and 1-digit industry code.

Taken together, these results indicate that task differentiation is a much stronger predictor of heterogeneity in wage pass-through than average occupational education. We interpret this pattern as evidence that individual hold-up power is not determined solely by traditional measures of skill or education: workers in highly educated occupations need not exhibit high hold-up power if the production process lacks complementarities or if skills are not position specific. This is particularly relevant for our highest earning occupations: managers, professionals, and technicians. Managers have no higher education than the other two occupation groups but exhibit substantially higher task differentiation and higher pass-through elasticities.

In the next section, we will provide further evidence that separations by managers and middle-wage occupations generate greater profit losses than for low-wage occupations, and consistent with our theory of differences in hold-up power, that the profit losses are higher

at low-labor share firms.

4.4 Firm Profit Losses from Worker Deaths

In this section, we estimate the effect of worker deaths on value added net of the wage bill relative to the deceased worker's prior wage using a matched event study specification with a continuous treatment intensity. We estimate the effect for our three broad occupation groups, and we split firms into high- and low-labor share firms. We find that losses in value added net of the wage bill are the greatest following manager deaths and the least following deaths of low-wage workers. We further find that losses are higher at low labor share firms than at high labor share firms for managers and middle-wage occupations, consistent with higher levels of hold-up power for these two occupation groups.

Sample and Variable Construction Our sample includes deaths between 2008 and 2016 and firm outcomes from 2005 to 2019. We restrict the sample to deaths of non-owner workers[14] who were employed at the firm in the month prior to the death and who worked at least 100 hours in that month.[15] We restrict attention to firms with 5 to 100 full-time equivalent employees to ensure that the effect of a worker death is large relative to the typical variance of output and profit. Given the length of our sample, we use an event window of seven periods, denoted $s = -3, ..., 3$, with the death occurring in period $s = 0$. Firms k are matched to five control firms based on exact year and 1-digit industry in period s = -3. The selected control firms minimize the sum of squared differences in log firm size (measured in full-time equivalent workers), log wage bill, and log value added in periods s = -3. We construct the control group by averaging outcomes across the matched firms using equal weights. Treated firms are excluded from the potential controls.

We match treated to control firms in period $s = -3$ because matching on the period immediately preceding the death would induce mechanical mean reversion in our setting. Our empirical analysis compares worker deaths across firms with high and low labor shares, but measured labor shares are affected by transitory productivity shocks and measurement error. Firms classified as low labor share based on period $s = -1$ are therefore likely to have unusually high value added net of the wage bill in that period relative to otherwise similar firms. Matching in the same period would consequently pair treated firms with

[14]Individuals in this sample are very unlikely to be employers: our deaths sample is derived from the BFL database, where very few workers have a value of "A" for the variable *type*, indicating that the individual is the owner of a firm.

[15]Other studies such as Bertheau et al. (2021) and Jäger (2016) restrict the sample to exclude deaths where the worker had a long-term illness or had prior visits to the hospital.

controls that have systematically different labor shares, leading to spurious convergence after the matching period. To avoid this issue, we construct the matched sample using firm characteristics measured in period $s = -3$ and then evaluate pre-trends over periods $s = -2$ and $s = -1$. The absence of differential pre-trends provides evidence that the matching procedure successfully removes these mechanical differences before the worker death occurs.

We will estimate the effect of worker deaths on two different outcomes: value added and the wage bill. Let $\mathcal{O}_{j,t}$ be a given outcome in year t, where t is calendar time. We construct a within-match j, within-period t difference $\tilde{\mathcal{O}}_{j,t} = (\mathcal{O}_{j,t}^{treated} - \bar{\mathcal{O}}_{j,t}^{control})/N_{j,t-1}$, where $\bar{\mathcal{O}}_{j,t}^{control}$ is the average outcome of the matched control firms that are matched to firm j, and $N_{j,t-1}$ is the number of full-time equivalent workers in the year prior to the shock. $w_{j,t-1}$ is the prior annual wage of the deceased worker at firm j, and $\tilde{w}_{j,t-1} = w_{j,t-1}/N_{j,t-1}$ scales the size of the shock by the firm's prior size. We scale both sides by firm size $N_{j,t-1}$ to reduce heteroskedasticity, as the variance of $(\mathcal{O}_{j,s}^{treated} - \bar{\mathcal{O}}_{j,s}^{control})$ increases with firm size.

Monetary values are deflated into 2005 Danish kroner and converted to dollars using 6 DKK/USD. We winsorize the differences in value added net of the wage bill $\tilde{\mathcal{O}}_{j,s}$ at the 2.5th and 97.5th percentiles separately for pre- and post-treatment observations. Note that deaths occur within the $s = 0$ period, so the coefficient β_0 will reflect a partial treatment effect.

Specification and Results To first build descriptive evidence about worker deaths, we estimate an event study specification. Let t be calendar time and s be event time, where a worker death occurs in period $s = 0$. We estimate

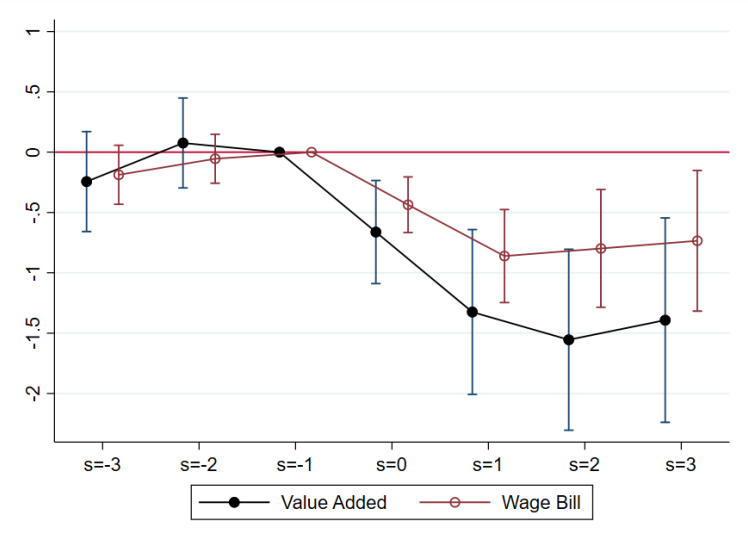
$$\tilde{\mathcal{O}}_{j,t} = \psi_j + \sum_{\substack{s=-3 \\ s\neq -1}}^{3} \beta_s \cdot D_{j,t-s} \cdot \tilde{w}_{j,t-1} + \varepsilon_{j,t}, \tag{15}$$

pooling across deaths by workers in all occupations. The regression in equation (15) recovers a dynamic sequence of coefficients β_s that trace out the effect of a worker separation over time.

Figure 5 reports the results of equation (15), pooling across occupation groups and using both value added and the wage bill as outcome variables. In the year of the death, value added falls by about 70% of the workers' salary and the wage bill falls by around 45% of the deceased worker's salary. Effects are smaller in period 0 because deaths occur during the fiscal year of the firms, so the effects of the death are not felt for the entire year. By period 1, value added falls by over 130% of the deceased worker's salary, and the wage bill falls by around 80% of the deceased worker's salary. Value added and the wage bill remain persistently depressed three years after the death. Visually inspecting the pre-trends suggests

that firms that experience worker deaths are on similar trajectories as firms without worker deaths in terms of value added and their wage bill.

Figure 5: Effect of Worker Death on Firm Value Added and Wage Bill



This figure reports event-study estimates of the effect of worker deaths on firm outcomes using equation (15). The figure plots the estimated coefficients β_s from a regression of within-match differences in firm outcomes on event-time indicators interacted with the size of the shock, pooling across occupation groups. Outcomes are measured as value added and the wage bill, expressed in units of the deceased worker's prior annual wage. Each point represents the estimated effect in period s relative to the pre-treatment period, with $s = 0$ denoting the year of death. Standard errors are clustered at the match level and 95% confidence intervals are shown.

Next, we estimate the effect of worker deaths on value added net of the wage bill separately for our three occupation groups. We estimate the following regression

$$\tilde{\mathcal{O}}_{j,t} = \psi_k^j + \sum_k \sum_{s=-3}^{-2} \beta_{k,s} \cdot D_{j,t-s}^k \cdot \tilde{w}_{j,t-1}^k + \sum_k \beta_{k,post} \cdot D_{j,t}^{k,post} \cdot \tilde{w}_{j,t-1}^k + \varepsilon_{j,t}, \tag{16}$$

where $\mathcal{O}_{j,t} = Y_{j,t}^{treated} - WB_{j,t}^{treated}$, $\bar{\mathcal{O}}_{j,t}^{control}$ is the average value added net of the wage bill of the matched control firms that are matched to firm k, $\tilde{\mathcal{O}}_{j,t} = (\mathcal{O}_{j,t}^{treated} - \bar{\mathcal{O}}_{j,t}^{control})/N_{j,t-1}$, $D_{j,t-s}^k$ is an indicator for event time s at firm j where the deceased worker's occupation is occupation k, $D_{j,t}^{k,post}$ is an indicator for event time $s \in \{0,1,2,3\}$ at firm j for occupation k, and $\tilde{w}_{j,t-1}^k$ is the wage of the deceased worker in firm j in occupation k in the year prior to their death. The coefficient $\beta_{k,post}$ captures the average effect of a death of a worker in occupation k on value added net of the wage bill over the treatment periods from $s = 0$ to $s = 3$.

Figure 6a reports the estimates of the effect of a worker death on firms' value added net of the wage bill for each occupation group, measured as annual losses expressed in years of prior wages. Pooling all occupations, the average annual loss of value added net of the wage bill is around .5, indicating that profits fall by 50% of workers' prior salary in response to a death. For managers the effect is the highest, with profit losses around 90% of workers' prior salary. For middle-wage occupations, losses are close to 60% of workers' prior salary. For low-wage occupations, the point estimate for profit losses is around 30% of workers' prior salary, but this is not statistically significant.

Finally, we estimate the effect of worker deaths on value added net of the wage bill by occupation and by firm labor share, splitting the sample between firms that have a labor share above or below 0.8. This division roughly splits the sample in half between high and low labor share firms. We estimate:

$$\begin{aligned}\tilde{\mathcal{O}}_{j,t} = \psi_k^j + &\sum_k \sum_{s=-3}^{-2} \beta_{k,s}^{low} \cdot D_{j,t-s}^k \cdot \{\tilde{\alpha}_{j,t-3} > .8\} \cdot \tilde{w}_{j,t-1}^k + \sum_k \beta_{k,post}^{low} \cdot D_{j,t}^{k,post} \cdot \{\tilde{\alpha}_{j,t-3} > .8\} \cdot \tilde{w}_{j,t-1}^k \\ &+ \sum_k \sum_{s=-3}^{-2} \beta_{k,s}^{high} \cdot D_{j,t-s}^k \cdot \{\tilde{\alpha}_{j,t-3} < .8\} \cdot \tilde{w}_{j,t-1}^k + \sum_k \beta_{k,post}^{high} \cdot D_{j,t}^{k,post} \cdot \{\tilde{\alpha}_{j,t-3} < .8\} \cdot \tilde{w}_{j,t-1}^k \\ &+ \varepsilon_{j,t}, \end{aligned} \tag{17}$$

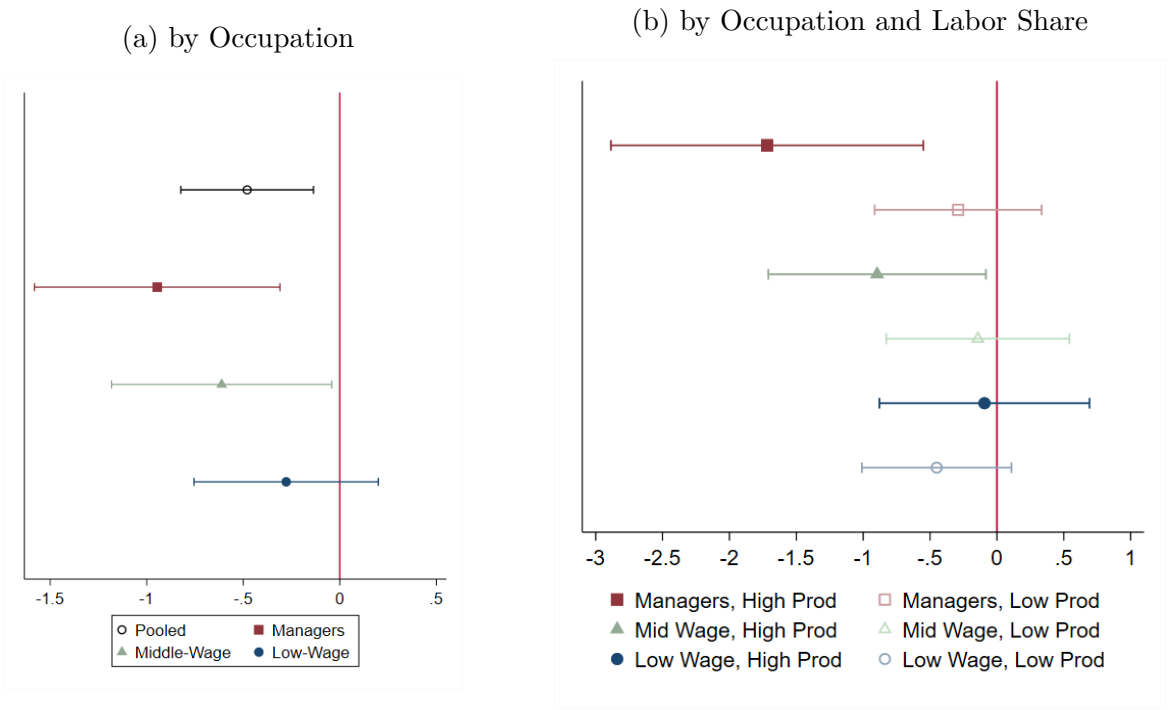
where the notation is the same as before, except that we separately estimate the effects of workers deaths for firms with labor shares in period $t-3$ that are above and below 0.8.

Figure 6b plots estimates for $\{\beta_{k,post}^{high}\}$ and $\{\beta_{k,post}^{low}\}$. Recall that firms with low labor shares have high productivity, and firms with high labor shares have low productivity, so the superscript *high* refers to productive, low labor share firms. We find that deaths by managers and workers in middle-wage occupations at high-productivity (low labor share) firms generate large and statistically significant profit losses, but the effects are not statistically significant in low-productivity firms. For workers in low-wage occupations, worker deaths do not have a significant effect on firm profits at either high or low productivity firms. We report the point estimates for this figure in a table in Appendix A.5.

Bertheau et al. (2021) also estimate the effects of worker deaths on firm profits in Denmark. Our analysis differs in a few ways. Because we do not have access to the SPELLS dataset, it is difficult to identify prior to 2008 which workers are still employed at the time of their death. To deal with this smaller sample in terms of years of data, we have two other differences in our methodology. First, we include worker deaths at firms with up 100 full-time equivalent workers, while Bertheau et al. (2021) limit their analysis to firms with 30 workers or fewer. The challenge of using larger firms is that the variance of profit changes becomes

large relative to the size of the shock due to a worker death. We address this by scaling our treatment intensity by in the inverse of firm size, which reduces heteroskedasticity and increases precision. We also scale the treatment intensity of the shock with workers' prior wage, as advised by our theory. Additionally, our sample of worker deaths uses all deaths, similar to Smith et al. (2019), rather than including only unexpected deaths as Bertheau et al. (2021) do, as using unexpected deaths would reduce our already small sample by over 60%. Finally, our theory motivates us to look at different effects of worker deaths by firm labor share, which is not the focus of Bertheau et al. (2021).

Figure 6: Effect of Death on Value Added Net of the Wage Bill, in Units of Prior Wages of the Deceased Worker



This figure reports event-study estimates of the effect of worker deaths on firm value added net of the wage bill using equation (16) and (17). Estimates are expressed in units of the deceased worker's prior annual wage and reported as average per-year losses over the treatment periods $s = 0$ to $s = 3$. Results are shown separately for the three occupation groups used throughout the paper: managers, middle-wage occupations, and low-wage occupations. Panel (b) further splits the sample by labor share, comparing low- and high-labor share firms. Error bars denote 95% confidence intervals.

Overall, our estimates support the hypothesis that workers in higher-wage occupations also have greater hold-up power, which makes exogenous separations of workers more costly at more productive, lower labor share firms. As previously discussed, we will use these estimates of the effect of worker deaths as well as our pass-through elasticities to estimate

h_k and δ_k for our three occupation groups.

5 Parameter Estimation

We now estimate our turnover costs parameters h_k and δ_k for our three broad occupation groups. We estimate these parameters with a hybrid of indirect inference and moment matching.

We begin by externally calibrating various parameters. We set $\gamma = 2$ to match the general finding in that the elasticity of separation rate with respect to firm-specific wages are around 2 Bassier et al. (2022). We parameterize the model so there are two types of firms H and L that have common values of $\{\sigma_{jk,t}\}$ and A_j but differ in their average value of α_j. Let the population of H firms be ζ_H and the population of L firms be $1-\zeta_H$. Firms are subject to TFP shocks and follow a process of shocks to $\alpha_{j,t}$. We normalize the population of each set of firms to 0.5. Because firms have a Cobb-Douglas production structure, we can calibrate the occupation shares σ_{MG}, σ_{MW}, and σ_{LW} to match empirical wage bill shares. We normalize the value of unemployment $b = 0$. Lastly, after we estimate h_k and δ_k for each occupation, we can directly back out the populations of C and NC workers needed to generate the relative wage differences we see in the data between workers in low-wage occupations versus in managerial and middle-wage occupations.

Table 7: Calibrated Parameters

Parameter	Value	Target moment	Target Value	Source	Model Value
A	1	Total Factor Productivity	-	Normalization	-
γ	2.00	Separation Elasticity	-2.1	Bassier et al. (2022)	-2
α_H	0.63	Labor Share, High Wage Firms	0.63	Figure 3	0.63
α_L	0.83	Labor Share, Low Wage Firms	0.83	Figure 3	0.83
σ_{MG}	0.096	Wage Bill Share, Managers	0.096	Table 2	0.096
σ_{MW}	0.358	Wage Bill Share, Mid-wage	0.358	Table 2	0.358
σ_{LW}	0.546	Wage Bill Share, Low-wage	0.546	Table 2	0.546
ϕ_C	0.36	Relative wage C vs NC	1.53	Table 2	1.55
ζ_H	2/3	Emp Share with $WB_j/VA_j < .8$	0.57	Figure 3	0.60
b	0.00	Normalization			

To estimate our parameters of interest, we simulate the model and calculate pass-through elasticities and costs of worker deaths in the model, comparing model-generated moments with the pass-through estimates in Figure 4 and the profit losses from worker deaths by occupation and at high- and low-labor share firms in Figure 6(b).

The simulated panel is then used to run the auxiliary IV regressions. For each occupation $j \in \{G, M, S\}$, the model pass-through moment is the IV coefficient in

$$\Delta_3 \log w_{jk,t} = \tau_j \Delta_3 LS_{j,t} + u_{jk,t},$$

where $LS_{j,t}$ is the firm-level labor share in levels. The regressor $\Delta LS_{j,t}$ is instrumented with the seven-period change in log value added per worker. The resulting simulated IV coefficients (τ_G, τ_M, τ_S) are the model counterparts to the empirical pass-through semi-elasticities.

The worker-death moments are computed analytically. For occupation k, the model moment at firm type $j \in \{H, L\}$ is

$$\beta_k^j(\theta) = -\left(\frac{1}{\gamma} + \delta_k + \frac{h_k}{\alpha_j}\right).$$

The estimator minimizes a diagonal, standard-error-weighted distance between the empirical and model moments:

$$\widehat{\theta} = \arg\min_{\theta} \sum_k \left(\frac{\tau_k(\theta) - \widehat{\tau}_k}{\widehat{\sigma}_{\tau_k}}\right)^2 + \sum_k \left(\frac{\beta_k^L(\theta) - \widehat{\beta}_{k,post}^{low}}{\widehat{\sigma}_{kL}}\right)^2 + \sum_k \left(\frac{\beta_k^H(\theta) - \widehat{\beta}_{k,post}^{high}}{\widehat{\sigma}_{kH}}\right)^2.$$

Our estimates for h_k and δ_k can be found in Table 8a. We find $h_{\text{managers}} = .282$, $h_{\text{middle-wage}} = .078$, and $h_{\text{low-wage}} = .043$. Our estimation procedure consistently finds $\delta_k = 0$ for each occupation.

Because we have three empirical moments and only two parameters for each occupation, the model is over-identified, and there are benefits and downsides to using either the pass-through or profit losses estimates to inform our parameters. With worker deaths, we have quasi-experimental evidence, but the estimates are very imprecise and the point estimates lie outside of the range of values that can be generated by the model. With pass-through estimation, the moment is coming from an equilibrium correlation, but the standard errors are smaller. Our estimator naturally puts more weight on empirical estimates that are more precise, and so the estimated values of $\{h_k\}$ are largely determined by the pass-through elasticities. Estimation details can be found in Appendix A.5.

6 Implications

6.1 The Skill Premium and Wage Inequality

A distinct feature of efficiency wage models is that wages are pushed above market clearing levels. As we saw in Table 2, occupations with more educated workers also endow workers with more hold-up power, potentially widening the gap in wages between more- and less-educated workers. How much of the skill premium can be accounted for by higher skill workers being employed in occupations with greater hold-up power?

To answer this question, we solve for a counterfactual equilibrium where managers and middle-wage occupations have the same hold-up power as low-wage occupations, setting

Table 8: Equilibrium Wages and Employment

(a) Baseline

Occupation	$\hat{h}$	$\hat{\delta}$	$\bar{w}_{\text{High } \alpha}$	$\bar{w}_{\text{Low } \alpha}$	Employment Level	Employment Share
Manager	0.282	0.000	1.219	1.288	0.059	0.063
Mid-Wage	0.078	0.000	1.028	1.051	0.267	0.286
Low-Wage	0.043	0.000	0.692	0.702	0.609	0.651

(b) Alternative Equilibrium: Less Hold-up Power

Occupation	h^{alt}	δ^{alt}	$\bar{w}_{\text{High } \alpha}$	$\bar{w}_{\text{Low } \alpha}$	Employment Level	Employment Share
Manager	0.043	0.000	1.070	1.084	0.072	0.076
Mid-Wage	0.043	0.000	1.070	1.084	0.270	0.284
Low-Wage	0.043	0.000	0.724	0.733	0.609	0.640

(c) Binding Minimum Wage for the Low-Wage Occupation

Occupation	h	δ	$\bar{w}_{\text{High } \alpha}$	$\bar{w}_{\text{Low } \alpha}$	Employment Level	Employment Share
Manager	0.282	0.000	1.214	1.283	0.059	0.066
Mid-Wage	0.078	0.000	1.024	1.047	0.267	0.296
Low-Wage	0.043	0.000	0.733	0.733	0.577	0.639

$h_G = h_M = h_S = .043$. Table 8b reports average wages and employment for the three occupations. Compared to our baseline estimated values, lowering the hold-up power of managers and middle-wage occupations compresses the wage distribution: average manager wages fall from around 1.10 to .96, average wages for middle-wage occupations rise from .92 to .96, and wages in low-wage occupations rise from .745 to .785.

This compression happens for two reasons. First, for managers who start with a lot of hold-up power, decreasing hold-up power allows firms to pay lower wages. This increases the quantity of labor demanded for managers, which increases the marginal product of labor of the other occupations. Both employing more managers and increasing the marginal product of middle-wage occupations increases demand for middle-wage occupations, raising wages in equilibrium. Similarly, greater employment of managers and middle-wage occupations increases demand for low-wage occupations, driving up equilibrium wages. In our baseline model, the average college worker earns wages that are 55% higher than non-college workers. When hold-up power is reduced such that managers and middle-wage occupations have the same hold-up power as low-wage occupations, then the average wage in equilibrium of college workers is only 48% higher than non-college workers. Thus greater hold-up power in occupations where workers are more highly educated increases the skill premium.

While accounting for changes in wage inequality over time is beyond the scope of this paper, we speculate that "de-specification" of non-college workers (Stansbury and Summers, 2020; Edmond and Mongey, 2019) may have contributed to rising college wage premiums over time.

6.2 Gender Wage Gap

While our focus on three broad occupation groups helped with estimating model parameters, there is potentially a great deal of heterogeneity in hold-up power within these broad occupation groups. If men and women are allocated differently across occupations with different hold-up power, this could contribute to the gender wage gap.

Across all individuals, the average occupational differentiation score for men is 0.40 and -0.41 for women. In Figure B.3 of Appendix B.4, we show that even within 1-digit occupations, women are much more concentrated in 4-digit occupations that have low task differentiation. Using our coefficient from Table 5 that a one standard deviation increase in task differentiation is associated with a 4.7 percent increase in wages, the 0.81 standard deviation difference between the task differentiation score of men's and women's occupation would alone generate a 3.8 percent gender wage gap. This can account for the majority of the between-occupation component of the gender wage gap.

Different levels of hold-up power, if well-proxied by task differentiation, could help explain why men earn higher wage premia at high paying firms than women (Card et al., 2016). Plugging in an average task differentiation score of 0.40 for men and -0.41 into our pass-through results from Table 6 would imply that men would have a pass-through elasticity of productivity to wages of .066 and women would have a pass-through elasticity of only .050. Thus, men may earn differentially more at high-paying, high-productivity firms because men are in the occupations where productive firms are incentivized to share rents.

6.3 Minimum Wage

In our model, firms are able to elastically increase the number of workers without having to increase wages, as wages affect the average tenure of workers but not the number of workers that a firm employs. This differs from monopsonistic models of rent sharing, where high wages help attract a greater number of workers. One of the key predictions of monopsony models is that marginal increases in the minimum wage above the decentralized wage should increase employment. This is no longer true in our model, as a higher minimum wage increases the cost of labor, which decreases labor demand. Table 8c reports the equilibrium wages and employment when a minimum wage increases the average wage of workers in the

low-wage occupation by 10%. As a result of the higher minimum wage, employment falls by 10%. Thus, the presence of rent-sharing does not necessarily imply positive employment effects of small minimum wage increases on employment.

7 Conclusion

In this paper, we show that the combination of individual production complementarities and position-specific skills generates hold-up power for individual workers, and firms are incentivized to share rents with workers in high hold-up occupations. Our model predicts that occupations in which pass-through of productivity to wages is the highest are also occupations where worker turnover is more costly and differentially so at high productivity firms. We find that managers exhibit both the highest hold-up power and pass-through, while low-wage service, manual, and administrative jobs have the smallest measure on both dimensions. We show that a model where heterogeneous hold-up power is the source of rent sharing has implications for the skill premium, the employment effect of minimum wages, and the gender wage gap.

Establishing that workers have different degrees of individual hold-up power opens multiple avenues for future research. One direction to explore is if position specificity and hold-up power are changing over time. Among less educated workers, the composition of jobs has been shifting away from moderate hold-up jobs, like in crafts and assembly occupations, to low hold-up service occupations. A further question is if automation has not only decreased the share of work in moderate hold-up, blue collar occupations, like in manufacturing, but also decreased the specificity of workers in these occupations as well.

Individual worker hold-up power may also be a useful framework for exploring the borders of the firm, workplace fissuring, and internal pay equity constraints. The observation that industry pay premia in the US were once uniform across occupations (Dickens and Katz (1987)) but now firm premia in the US only benefit college educated workers (Bloom et al. (2018)), combined with the pay losses from outsourcing documented in Goldschmidt and Schmieder (2017), suggest that changes in internal pay equity constraints may be important for explaining changes in wage inequality. In particular, knowing which occupations have individual hold-up power may predict which workers benefit from erosion of internal pay constraints. Individual hold-up power may also be useful for understanding the boundaries of the firm, as employers may use the boundary of the firm to mitigate hold-up power while outsourcing the lower hold-up jobs.

References

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

Bagger, Jesper, Francois Fontaine, Manolis Galenianos, and Ija Trapeznikova. 2022. "Vacancies, Employment Outcomes and Firm Growth: Evidence from Denmark." *Labour Economics* 75 102103.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu. 2022. "Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies." *Journal of Human Resources* 57 (S): S50–s86.

Bertheau, Antoine, Pierre Cahuc, Simon Jager, and Rune Vejlin. 2021. "Turnover Costs: Evidence from Unexpected Worker Separations." Working Paper.

Bloesch, Justin, Birthe Larsen, and Anders Yding. 2025. "Monopsony with Recruiting." *Available at SSRN*.

Bloom, Nicholas, Fatih Guvenen, Benjamin S Smith, Jae Song, and Till von Wachter. 2018. "The Disappearing Large-Firm Wage Premium." In *AEA Papers and Proceedings*, Volume 108. 317–22.

Brown, Charles, and James Medoff. 1989. "The Employer Size-wage Effect." *Journal of Political Economy* 97 (5): 1027–1059.

Caldwell, Sydnee, and Nikolaj Harmon. 2019. "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks." Job Market Paper, The University of California, Berkeley.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *Journal of Labor Economics* 36 (S1): S13–S70.

Card, David, Ana Rute Cardoso, and Patrick Kline. 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *The Quarterly Journal of Economics* 131 (2): 633–686.

Chan, M, E Mattana, S Salgado, and M Xu. 2024. "Dynamic Wage Setting: The Role of Monopsony Power and Adjustment Costs."Technical report, Working paper. 12, 4.1.

Chan, Mons, Sergio Salgado, and Ming Xu. 2020. "Heterogeneous Passthrough from TFP to Wages." Working Paper.

Dahl, Christian M, Daniel Le Maire, and Jakob R Munch. 2013. "Wage Dispersion and Decentralization of Wage Bargaining." *Journal of Labor Economics* 31 (3): 501–533.

Dickens, William, and Lawrence F Katz. 1987. "Inter-industry Wage Differences and Theories of Wage Determination."

Dickens, William T, Lawrence F Katz, Kevin Lang, and Lawrence H Summers. 1989. "Employee Crime and the Monitoring Puzzle." *Journal of labor economics* 7 (3): 331–347.

Edmond, Chris, and Simon Mongey. 2019. "Unbundling Labor." Working Paper.

Friedrich, Benjamin, Lisa Laun, Costas Meghir, and Luigi Pistaferri. 2019. "Earnings Dynamics and Firm-Level Shocks." National Bureau of Economic Research Working Paper no. 25786.

Garin, Andrew, and Filipe Silvério. 2024. "How Responsive are Wages to Firm-Specific Changes in Labour Demand? Evidence from Idiosyncratic Export Demand Shocks." *Review of Economic Studies* 91 (3): 1671–1710.

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

Guha, Sudipto, Rajeev Rastogi, and Kyuseok Shim. 2000. "ROCK: A Robust Clustering Algorithm for Categorical Attributes." *Information Systems* 25 (5): 345–366.

Isen, Adam. 2015. "Dying to Know: Are Workers Paid Their Marginal Product."Technical report, Working Paper.

Jäger, Simon. 2016. "How Substitutable are Workers? Evidence from Worker Deaths." Working Paper.

Jensen, Mathias Fjællegaard. 2024. "Gender Differences in Returns to Skills: Evidence from Matched Vacancy-Employer-Employee Data." *Mimeo.*

Juhn, Chinhui, Kristin McCue, Holly Monti, and Brooks Pierce. 2018. "Firm Performance and the Volatility of Worker Earnings." *Journal of Labor Economics* 36 (S1): S99–S131.

Katz, Lawrence F. 1986. "Efficiency Wage Theories: A Partial Evaluation." *NBER Macroeconomics Annual* 1 235–276.

Kline, Patrick M. 2025. "Labor Market Monopsony: Fundamentals and Frontiers."

Kline, Patrick, Neviana Petcova, Heidi Williams, and Owen Zidar. 2019. "Who Profits from Patents? Rent-Sharing at Innovative Firms." *The Quarterly Journal of Economics* 1343 1404.

Kremer, Michael. 1993. "The O-ring Theory of Economic Development." *The Quarterly Journal of Economics* 108 (3): 551–575.

Krueger, Alan B, and Lawrence H Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica: Journal of the Econometric Society* 259–293.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 2019. "Imperfect Competition, Compensating Differentials and Rent Sharing in the US Labor Market." *American Economic Review* 112 (1): 169–212.

Lazear, Edward P. 2009. "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Political Economy* 117 (5): 914–940.

Maibom, Jonas, and Rune Vejlin. 2021. "Passthrough of Firm Performance to Income and Employment Stability." IZA Discussion Paper no. 14131.

Shapiro, Carl, and Joseph E Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *The American economic review* 74 (3): 433–444.

Smith, Matthew, Danny Yagan, Owen Zidar, and Eric Zwick. 2019. "Capitalists in the Twenty-First Century." *The Quarterly Journal of Economics* 134 (4): 1675–1745.

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

Sorkin, Isaac. 2018. "Ranking Firms using Revealed Preference." *The Quarterly Journal of Economics* 133 (3): 1331–1393.

Stansbury, Anna, and Lawrence H Summers. 2020. "The Declining Worker Power Hypothesis: An Explanation for the Recent Evolution of the American Economy."Technical report, National Bureau of Economic Research.

Stole, Lars A, and Jeffrey Zwiebel. 1996. "Intra-firm Bargaining under Non-binding Contracts." *The Review of Economic Studies* 63 (3): 375–410.

Wong, Horng Chern. 2025. "Understanding high-wage firms." *Unpublished. Available at https://horngchernwong. github. io/horngchernwong. com/HWFs. pdf.*

A For Online Publication: Theory Appendix

A.1 Deriving Large Firms with Complementarities from Generalized O-ring

Consider the O-Ring production function described in Kremer (1993), in which production is composed of N tasks. Let $q_\ell \in \{0, 1\}$ be an indicator of whether task ℓ was successfully performed. Then total firm output takes the form:

$$Y = K^\beta N \prod_{\ell \in 1}^{N} q_\ell,$$

with $\beta \in [0, 1)$. The defining feature of this production function is extreme complementarities that each position has with the rest of firm output. This generates the extreme result that total output is lost if a single worker fails to do their job:

$$\frac{\Delta Y}{\Delta q_\ell}\Big|_{q_i=1 \forall i \neq \ell} = K^\beta N.$$

This is an extreme assumption that if a single task is not performed, the firm loses its entire output. Additionally, larger firms that produce with more tasks N have more to lose from a single non-performing worker. To soften this extreme form of complementarities, we introduce a generalized O-ring production function.

Generalized O-Ring Consider a generalization of the O-Ring production function such that if a single worker fails to perform a given task j, the firm loses a fraction of total output that scales inversely with the size of the firm, rather than total output. Consider a function of the form:

$$Y_j = A_j K_j^{\beta_j} \prod_k (S_{jk} + J_{jk})^{\sigma_{jk,t}\alpha_j} \prod_{k \in \mathcal{K}(j)} \prod_{p=0}^{P_{jk}-S_{jk}} \frac{P_{jk} - p - \sigma_{jk,t}(\delta_k \alpha_j + h_k)}{P_{jk} - p} \tag{18}$$

In the case of one occupation $|\mathcal{K}(j)| = 1$, this production function nests the O-ring production function when $\alpha_j = 1$, $\delta_k = 0$, and $h_k = P_{jk}$. That is, if losing one senior worker knocks out the output of all P_{jk} positions, then our production function is equivalent to the O-ring production in Kremer (1993).

Next we show that our production function from equation (1) of the main text approximates this multiplicative production in equation (18).

$$
\begin{aligned}
\log Y_j &= \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) + \sum_{k\in\mathcal{K}(j)} \sum_{p=0}^{P_{jk}-S_{jk}} \log\left(\frac{P_{jk}-p-\sigma_{jk,t}(\delta_k\alpha_j+h_k)}{P_{jk}-p}\right) \\
&= \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) + \sum_{k\in\mathcal{K}(j)} \sum_{p=0}^{P_{jk}-S_{jk}} \log\left(1-\frac{\sigma_{jk,t}(\delta_k\alpha_j+h_k)}{P_{jk}-p}\right) \\
&\approx \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) + \sum_{k\in\mathcal{K}(j)} \sum_{p=0}^{P_{jk}-S_{jk}} \frac{-\sigma_{jk,t}(\delta_k\alpha_j+h_k)}{P_{jk}-p} \\
&= \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) - \sigma_{jk,t}(\delta_k\alpha_j+h_k) \sum_{k\in\mathcal{K}(j)} \sum_{p=0}^{P_{jk}-S_{jk}} \frac{1}{P_{jk}-p} \\
&\approx \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) - \sigma_{jk,t}(\delta_k\alpha_j+h_k) \sum_{k\in\mathcal{K}(j)} \int_{p=0}^{P_{jk}-S_{jk}} \frac{1}{P_{jk}-p}dp \\
&= \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) + \sigma_{jk,t}(\delta_k\alpha_j+h_k) \sum_{k\in\mathcal{K}(j)} \log\left(P_{jk}-p\right)\Big|_{p=0}^{P_{jk}-S_{jk}} \\
&= \log A_j + \beta_j \log K_j + \sum_k (\sigma_{jk,t}\alpha_j)\log(S_{jk}+J_{jk}) + \sigma_{jk,t}(\delta_k\alpha_j+h_k) \sum_{k\in\mathcal{K}(j)} \left(\log(S_{jk}) - \log\left(P_{jk}\right)\right).
\end{aligned}
$$

Exponentiating again, we have

$$
\begin{aligned}
Y_j &\approx A_j K_j^{\beta_j} \prod_{k\in\mathcal{K}(j)} (S_{jk}+J_{jk})^{\sigma_{jk,t}\alpha_j} \left(\frac{S_{jk}}{P_{jk}}\right)^{\sigma_{jk,t}(\alpha_j\delta_k+h_k)} \\
&= A_j K_j^{\beta_j} \prod_{k\in\mathcal{K}(j)} \left((S_{jk}+J_{jk})\left(\frac{S_{jk}}{P_{jk}}\right)^{\left(\delta_k+\frac{h_k}{\alpha_j}\right)}\right)^{\sigma_{jk,t}\alpha_j}
\end{aligned}
$$

Thus our production function in equation (1) approximates the multiplicative function in equation (18).

A.2 Deriving the Pass-Through Elasticities

Firms solve equation (2) from the main text. The firm's Lagrangian is:

$$
\begin{aligned}
\mathcal{L} = A_j K_{j,t}^{\beta_{j,t}} \prod_{k\in\mathcal{K}(j)} \left((S_{jk,t}+J_{jk,t})\left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{\left(\delta_k+\frac{h_k}{\alpha_{j,t}}\right)}\right)^{\sigma_{jk,t}\alpha_{j,t}} &- r^K K_{j,t} - \sum_k w_{jk,t}(S_{jk,t}+J_{jk,t}) - \sum_k \epsilon_k(J_{jk,t} \\
+ \sum_k \lambda_{jk,t}\left[S_{jk,t} - \left(1-\left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}\right)P_{jk,t}\right] & \\
+ \sum_k \mu_{jk,t}\left[P_{jk,t} - S_{jk,t} - J_{jk,t}\right]. &
\end{aligned}
$$

The first order conditions (substituting back in for Y_j) are:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial K_{j,t}}: &\ \beta_{j,t}\frac{Y_{j,t}}{K_{j,t}}r^K = 0 \\
\frac{\partial \mathcal{L}}{\partial J_{jk,t}}: &\ \sigma_{jk,t}\alpha_{j,t}\frac{Y_{j,t}}{N_{jk,t}} - w_{jk,t} - \epsilon_k - \mu_{jk,t} = 0 \\
\frac{\partial \mathcal{L}}{\partial S_{jk,t}}: &\ \sigma_{jk,t}\alpha_{j,t}\frac{Y_{j,t}}{N_{jk,t}} + \sigma_{jk,t}\alpha_{j,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{Y_{j,t}}{S_{jk,t}} - w_{jk,t} + \lambda_{jk,t} - \mu_{jk,t} = 0 \\
\frac{\partial \mathcal{L}}{\partial P_{jk,t}}: &\ -\sigma_{jk,t}\alpha_{j,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{Y_{j,t}}{P_{jk,t}} - \epsilon_k - \lambda_{jk,t}\left(1 - \left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}\right) + \mu_{jk,t} = 0 \\
\frac{\partial \mathcal{L}}{\partial w_{jk,t}}: &\ -(S_{jk,t} + J_{jk,t}) - \lambda_{jk,t}P_{jk,t}\gamma_{s(k)}V_{s(k)}^{\gamma_{s(k)}}w_{jk,t}^{-\gamma_{s(k)}-1} = 0
\end{aligned}$$

with complementary slackness conditions

$$\begin{aligned}
\mu_{jk,t} &\geq 0, \\
P_{jk,t} - S_{jk,t} - J_{jk,t} &\geq 0, \\
\mu_{jk,t}\left(P_{jk,t} - S_{jk,t} - J_{jk,t}\right) &= 0.
\end{aligned}$$

Deriving $\varepsilon_{w_{jk,t},\alpha_{j,t}}$ Combining the first order conditions on S and J yields

$$-\lambda_{jk,t} = \sigma_{jk,t}\left(\alpha_{j,t}\delta_k + h_k\right)\frac{Y_{j,t}}{S_{jk,t}} + \epsilon_k.$$

Plugging this expression for $\lambda_{jk,t}$ into the first order condition on $P_{jk,t}$ and using $\frac{S_{jk,t}}{P_{jk,t}} = 1 - \left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}$ yields

$$\begin{aligned}
\mu_{jk,t} =& \sigma_{jk,t}\alpha_{j,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{Y_{j,t}}{P_{jk,t}} + \lambda_{jk,t}\left(\frac{S_{jk,t}}{P_{jk,t}}\right) + \epsilon_k \\
=& \sigma_{jk,t}\alpha_{j,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{Y_{j,t}}{P_{jk,t}} + \epsilon_k - \left(\sigma_{jk,t}\left(\alpha_{j,t}\delta_k + h_k\right)\frac{Y_{j,t}}{S_{jk,t}} + \epsilon_k\right)\left(\frac{S_{jk,t}}{P_{jk,t}}\right) \\
=& \epsilon_k\left(1 - \frac{S_{jk,t}}{P_{jk,t}}\right).
\end{aligned}$$

Since $1 - S_{jk,t}/P_{jk,t} = (V_{s(k)}/w_{jk,t})^{\gamma_{s(k)}} > 0$, $\epsilon_k > 0$ implies that $\mu_{jk,t} > 0$, so by complementary slackness, firms re-fill all their positions $J_{jk,t} + S_{jk,t} = P_{jk,t}$. In the limit when $\epsilon_k \to 0$, the shadow value of the vacancy constraint converges to zero, but the limiting equilibrium is the full-rehiring allocation, which we use throughout the remainder of the derivation.

Plugging the expression for $\lambda_{jk,t}$ into the first order condition on w_{jk}, and taking the limit as $\epsilon_k \to 0$, we have

$$S_{jk,t} + J_{jk,t} = \left(\sigma_{jk,t}(\delta_k\alpha_{j,t} + h_k)\frac{Y_{j,t}}{S_{jk,t}}\right)\gamma_{s(k)}\left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}} w_{jk,t}^{-1}P_{jk,t}$$

Using the limiting full-rehiring allocation $S_{jk,t} + J_{jk,t} = P_{jk,t}$, this becomes

$$w_{jk,t} = \sigma_{jk,t} \frac{Y_{j,t}}{S_{jk,t}} (\delta_k \alpha_{j,t} + h_k) \gamma_{s(k)} \left(\frac{V_{s(k)}}{w_{jk,t}} \right)^{\gamma_{s(k)}}.$$

We then make use of $\mu_{jk,t} = 0$. The first order condition on $J_{jk,t}$ can be written as

$$w_{jk,t} = \alpha_{j,t} \sigma_{jk,t} \frac{Y_{j,t}}{S_{jk,t}} \frac{S_{jk,t}}{S_{jk,t} + J_{jk,t}}.$$

Plugging this into the previous expression yields

$$\alpha_{j,t} \sigma_{jk,t} \frac{Y_{j,t}}{S_{jk,t}} \frac{S_{jk,t}}{S_{jk,t} + J_{jk,t}} = \sigma_{jk,t} \frac{Y_{j,t}}{S_{jk,t}} (\delta_k \alpha_{j,t} + h_k) \gamma_{s(k)} \left(\frac{V_{s(k)}}{w_{jk,t}} \right)^{\gamma_{s(k)}}$$

Imposing full rehiring $S_{jk,t} + J_{jk,t} = P_{jk,t}$ and $\frac{S_{jk,t}}{P_{jk,t}} = \left(1 - \left(\frac{V_{s(k)}}{w_{jk,t}} \right)^{\gamma_{s(k)}} \right)$, the above expression becomes

$$1 - \left(\frac{V_{s(k)}}{w_{jk,t}} \right)^{\gamma_{s(k)}} = \left(\delta_k + \frac{h_k}{\alpha_{j,t}} \right) \gamma_{s(k)} \left(\frac{V_{s(k)}}{w_{jk,t}} \right)^{\gamma_{s(k)}}$$

Rearranging yields out optimal wage expression

$$w^*_{jk,t} = \left(1 + \gamma_{s(k)} \left(\delta_k + \frac{h_k}{\alpha_{j,t}} \right) \right)^{\frac{1}{\gamma_{s(k)}}} V_{s(k)}.$$

Therefore the elasticity of $w_{jk,t}$ with respect to $\alpha_{j,t}$ is

$$\varepsilon_{w_{jk}, \alpha_j} = \frac{-h_k}{\alpha_{j,t} + \gamma_{s(k)} (\delta_k \alpha_{j,t} + h_k)}.$$

Deriving $\varepsilon^{(\alpha_j)}_{w_{jk}, APL_j}$ Recall that if firms fully rehire, so $S_{jk,t} + J_{jk,t} \equiv N_{jk,t} = P_{jk,t}$. In the limit as $\epsilon_k \to 0$ costs go to zero, the employment FOC implies

$$w_{jk,t} N_{jk,t} = \alpha_{j,t} \sigma_{jk,t} Y_{j,t}.$$

Therefore,

$$APL_{j,t} = \frac{Y_{j,t}}{\sum_k N_{jk,t}} = \frac{Y_{j,t}}{\sum_k \frac{\alpha_{j,t} \sigma_{jk,t} Y_{j,t}}{w_{jk,t}}} = \frac{1}{\alpha_{j,t}} \frac{1}{\sum_k \frac{\sigma_{jk,t}}{w_{jk,t}(\alpha_{j,t})}}.$$

Taking logs,

$$\log(APL_{j,t}) = -\log(\alpha_{j,t}) - \log\left(\sum_k \sigma_{jk,t} w_{jk,t}^{-1}(\alpha_{j,t})\right),$$

$$\frac{\partial \log(APL_{j,t})}{\partial \alpha_{j,t}} = -\frac{1}{\alpha_{j,t}} + \frac{\sum_k \sigma_{jk,t} w_{jk,t}^{-2}(\alpha_{j,t}) \frac{\partial w_{jk,t}}{\partial \alpha_{j,t}}}{\sum_k \sigma_{jk,t} w_{jk,t}^{-1}(\alpha_{j,t})},$$

$$\begin{aligned}\varepsilon_{APL_j,\alpha_j} &= -1 + \frac{\sum_k \sigma_{jk,t} w_{jk,t}^{-1}(\alpha_j) \varepsilon_{w_{jk,t},\alpha_j}}{\sum_k \sigma_{jk,t} w_{jk,t}^{-1}(\alpha_j)} \\ &= -1 + \sum_k \frac{N_{jk,t}}{N_{j,t}} \varepsilon_{w_{jk},\alpha_{j,t}}.\end{aligned}$$

Thus,

$$\begin{aligned}\varepsilon_{w_{jk},APL_j}^{(\alpha_j)} &= \varepsilon_{w_{jk},\alpha_j} \varepsilon_{\alpha_j,APL_j} \\ &= \frac{\varepsilon_{w_{jk},\alpha_j}}{\sum_\ell \frac{N_{j\ell}}{N_j} \varepsilon_{w_{j\ell},\alpha_j} - 1}.\end{aligned}$$

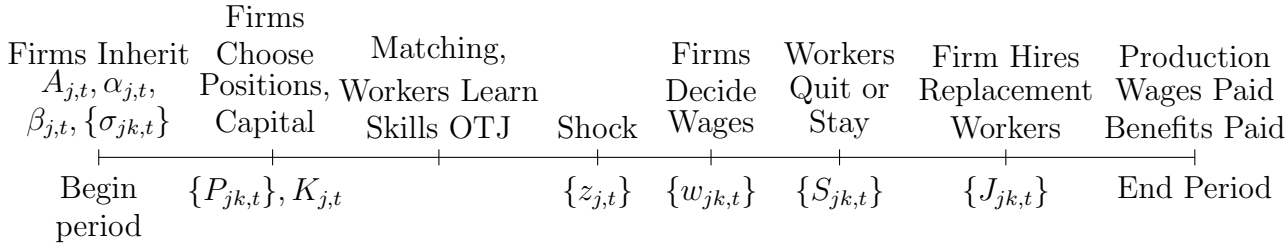
Substituting in the expression for $\varepsilon_{w_{jk},\alpha_j}$ gives

$$\varepsilon_{w_{jk},APL_j}^{(\alpha_j)} = \frac{\frac{h_k}{\alpha_j + \gamma_{s(k)}(\delta_k \alpha_j + h_k)}}{1 + \sum_\ell \frac{N_\ell}{N_j} \left(\frac{h_\ell}{\alpha_j + \gamma_{s(\ell)}(\delta_\ell \alpha_j + h_\ell)}\right)}.$$

A.3 Transitory Shocks and Instrumental Variables Derivation

Timing with Transitory Shocks The timing within a period is the same as in Section 2, except that firms draw an unanticipated productivity shock z_t after choosing the number of positions but before choosing wages. Let $z_t = \exp(\epsilon_t^z)$ where ϵ_t^z is drawn from a distribution with mean zero and standard deviation σ^z. The timing of the model becomes

Figure A.7: Model Timing



Firms now solve their problem in two steps and use backward induction. Towards the end of the period, the firm has already chosen its capital $K_{j,t}$ and number of positions $\{P_{jk,t}\}$

and needs to decide its wages $\{w_{jk,t}\}$, retention $\{S_{jk,t}\}$, and rehiring $\{J_{jk,t}\}$ conditional on the realization of $z_{j,t}$. Firms solve

$$\max_{\{w_{jk,t}\},\{S_{jk,t}\},\{J_{jk,t}\}} z_{j,t}A_{j,t}K_{j,t}^{\beta_{j,t}}\left(\prod_{k=1}^{\mathcal{K}}\left(S_{jk,t}+J_{jk,t}\right)\left(\frac{S_{jk,t}}{P_{jk,t}}\right)^{\delta_k+\frac{h_k}{\alpha_{j,t}}}\right)^{\sigma_{jk,t}\alpha_{j,t}} - \sum_{k\in\mathcal{K}(j)} w_{jk,t}\left(S_{jk,t}+J_{jk,t}\right) \tag{19}$$

subject to:

$$\text{Probability of retaining workers:} \qquad S_{jk,t} = \left(1-\left(\frac{V_{s(k)}}{w_{jk,t}}\right)^{\gamma_{s(k)}}\right)^{\alpha_{j,t}\delta_k+h_k} \quad \text{if } w_{jk,t}\geq V_{s(k)}$$
$$=0 \qquad \text{otherwise} \tag{20}$$

$$\text{One worker per position:} \quad J_{jk,t}+S_{jk,t} \leq P_{jk,t} \tag{21}$$

The solution to the above problem gives the firm a policy rule for wages $w^*_{jk,t}(z_{j,t})$, retention $S^*_{jk,t}(z_t)$, and rehiring $J^*_{jk,t}(z_{j,t})$ as a function of the shock $z_{j,t}$. Then at the beginning of the period, firms choose the capital stock $K_{j,t}$ and the number of positions for each occupation $\{P_{jk,t}\}$ taking into account the optimal policy rules and integrating over the distribution of $z_{j,t}$. Thus at the beginning of the period, firms solve

$$\max_{\{P_{jk,t}\},K_{j,t}} \int_{z_{j,t}} \left[z_{j,t}A_{j,t}K_{j,t}^{\beta_{j,t}}\left(\prod_{k=1}^{\mathcal{K}}\left(S^*_{jk,t}(z_{j,t})+J^*_{jk,t}(z_{j,t})\right)\left(\frac{S^*_{jk,t}(z_{j,t})}{P_{jk,t}}\right)^{\delta_k+\frac{h_k}{\alpha_{j,t}}}\right)^{\sigma_{jk,t}\alpha_{j,t}} - r^K K_{j,t} \right.$$
$$\left. - \sum_{k\in\mathcal{K}(j)} \epsilon_k P_{jk,t} - \sum_k w^*_{jk,t}(z_{j,t})\left(S^*_{jk,t}(z_{j,t})+J^*_{jk,t}(z_{j,t})\right) \right] d\Gamma_z(z_{j,t} \mid z_{j,t-1}), \tag{22}$$

Response of Wages To Shocks For sufficiently small shocks, we can use the envelope theorem, which tells us that locally $\frac{\partial \Pi_{j,t}}{\partial z_{j,t}} = Y_{j,t}$. Since positions are chosen before the realization of $z_{j,t}$, the firm chooses them based on expected profits. Around the deterministic optimum, mean-zero productivity shocks affect expected profits only to second order. Accordingly, for sufficiently small shocks we approximate the optimal number of positions by its deterministic value: $P^*_{jk,t}(z_{j,t-1}) \approx P^*_{jk,t}(\bar{z}_{j,t})$.

Log-linearizing around this deterministic number of positions $P_{jk,t}$, we can characterize how wages respond to realizations of $z_{j,t}$. In general, optimal wages are increasing in z: $\frac{\partial w_{jk,t}}{\partial z_{j,t}} > 0$. Using hatted variables to denote percent deviations, the percent deviation of wages with respect to shocks to z yields

$$\hat{w}_{jk,t} = \frac{\hat{z}_{j,t} + \sum_{k'\neq k}\sigma_{k'}\alpha_j\,\hat{w}_{jk',t}}{\gamma_{s(k)}+1-\sigma_{jk,t}\alpha_j+\frac{\alpha_j}{\delta_k\alpha_j+h_k}}. \tag{23}$$

where the response $\hat{w}_{jk,t}$ depends both on the shock $z_{j,t}$ and the response of the wages of other occupations $\hat{w}_{jk',t}$. Notice that the denominator includes the term $\frac{\alpha_j}{\delta_k \alpha_j + h_k}$, which can be rewritten as $\left(\delta_k + \frac{h_k}{\alpha_j}\right)^{-1}$. This is the inverse of the firm's total turnover costs, which means that the response of wages to unanticipated shocks is increasing in turnover costs. Define total turnover costs $\tau_{jk} \equiv \delta_k + \frac{h_k}{\alpha_j}$. This implies that the responsiveness of wages to z shocks is increasing in total turnover costs $\frac{\partial^2 \hat{w}_{jk,t}}{\partial z_{j,t} \partial \tau_{j,k}} > 0$.

The intuition is the following. A higher value of z increases the productivity of all workers, which the firm cannot offset by creating more positions in the short run. In the limit as turnover costs $\delta_k + \frac{h_k}{\alpha_j} \to 0$, replacement workers become as productive as the trained workers they replace. Therefore for occupations with very low turnover costs, a higher value of z has no impact on the costs of losing a senior worker. As the replacement costs parameters $\delta_k + \frac{h_k}{\alpha_j}$ get bigger, the difference in productivity between incumbents and replacement hires relative to the zero-shock wage widens with larger values of z, motivating firms to raise wages by more.

Statistical Model of Wages Let $\tilde{w}$ be the observed wage, w^* is the true wage, $e^w_{j,t}$ is a common firm-by-period measurement error, and $e^w_{i,t}$ is a worker-by-period measurement errors.

$$\begin{aligned} \tilde{w}_{ijk,t} =& w^*_{jk,t} \mathcal{E}^w_{i,t} E^w_{j,t} \\ \log \tilde{w}_{ijk,t} =& \log w^*_{jk,t} + \epsilon^w_{i,t} + e^w_{j,t}. \end{aligned}$$

For sufficiently small shocks, current variables $w^*_{jk,t}$ and $Y^*_{j,t}$ can be decomposed in approximation into a deterministic component and shock component:

$$\begin{aligned} w^*_{jk,t} =& \bar{w}^*_{jk} Z^w_{\alpha_{j,t},h}(z_{j,t}) \\ Y^*_{j,t} =& \bar{Y}^*_j Z^Y_{\alpha_{j,t},h}(z_{j,t}) \\ N^*_{j,t} =& \bar{N}^*_j \end{aligned}$$

where $Z^w_{\alpha_{j,t},h}(z_{j,t})$ and $Z^Y_{\alpha_{j,t},h}(z_{j,t})$ are functions that govern how wages and output deviate from the no-shock wage and output levels as a (α_j and h_k specific) function of $z_{j,t}$. In logs, these are

$$\begin{aligned} \log w^*_{jk,t} &= \log \bar{w}^*_{jk} + z^w(z_{j,t}) \\ \log Y^*_{j,t} &= \log \bar{Y}^*_j + z^Y(z_{j,t}) \end{aligned}$$

We estimate an IV where we instrument the log labor share with output per worker, where the instrument $\Delta_3 \frac{Y^*_{j,t}}{N^*_{j,t}}$ is the three-period difference in output per worker. Our results would hold

with longer differences, as long as the changes in Y/N are wider than and not overlapping with changes in wages. The key assumption is that changes in this variable are orthogonal to changes in measurement error and changes in $Z()$ are uncorrelated over time.

$$\beta_{IV} = \frac{cov(\Delta \log \tilde{w}_{ijk,t}, \Delta_3 \frac{Y^*_{j,t}}{N^*_{j,t}})}{cov(\Delta \log \frac{N_{j,t}\sum_{i\in j}\tilde{w}_{ijk,t}}{Y_{j,t}}, \Delta_3 \frac{Y^*_{j,t}}{N^*_{j,t}})}$$

Breaking up the log wage expression:

$$\begin{aligned}
\Delta \log \tilde{w}_{ijk,t} =& \Big(\log w^*_{jk,t} + z^w(z_{j,t}) + \epsilon^w_{i,t} + e^w_{j,t} \Big) - \Big(\log w^*_{jk,t-1} + z^w(z_{j,t-1}) + \epsilon^w_{i,t-1} + e^w_{j,t-1} \Big) \\
=& \frac{1}{\gamma_{s(k)}} \left[\log \left(1 + \gamma_{s(k)} \left(\delta_k + \frac{h_k}{\alpha_{j,t}} \right) \right) - \log \left(1 + \gamma_{s(k)} \left(\delta_k + \frac{h_k}{\alpha_{j,t-1}} \right) \right) \right] \\
&+ \Delta z^w(z_{j,t}, z_{j,t-1}) + \Delta \epsilon^w_{i,t} + \Delta e^w_{j,t} \\
\approx& \frac{-h_k}{\alpha_{j,t-1} + \gamma_{s(k)}(\delta_k \alpha_{j,t-1} + h_k)} \frac{\alpha_{j,t} - \alpha_{j,t-1}}{\alpha_{j,t-1}} + \Delta z^w(z_{j,t}, z_{j,t-1}) + \Delta \epsilon^w_{i,t} + \Delta e^w_{j,t},
\end{aligned}$$

where the approximation makes use of small changes $\alpha_{j,t}$. The log labor share is

$$\begin{aligned}
\log \frac{N_{j,t}\sum_{i\in j}\tilde{w}_{ijk,t}}{Y_{j,t}} &= \log \frac{N_{j,t}\sum_i w^*_{jk,t} Z^w_{\alpha_t,h}(z_{j,t}) \mathcal{E}^w_{i,t} E^w_{j,t}}{\bar{Y}^*_j Z^Y_{\alpha_j,h}(z_{j,t})} \\
&= \log \frac{N_{j,t} w^*_{jk,t} Z^w_{\alpha_t,h}(z_{j,t}) E^w_{j,t} \sum_i \mathcal{E}^w_{i,t}}{\bar{Y}^*_j Z^Y_{\alpha_j,h}(z_{j,t})} \\
&= \log \alpha_{j,t} + \log \frac{Z^w_{\alpha_t,h}(z_{j,t}) E^w_{j,t} \sum_i \mathcal{E}^w_{i,t}}{Z^Y_{\alpha_j,h}(z_{j,t})} \\
&\approx \log \alpha_{j,t} + z^w_{\alpha_t,h}(z_{j,t}) + e^w_{j,t} + \sum_i \epsilon^w_{i,t} - z^Y_{\alpha_j,h}(z_{j,t}),
\end{aligned}$$

where the approximation assumes small shocks to $z_{j,t}$. The first difference of the log labor share is:

$$\Delta \log \frac{\tilde{w}_{ijk,t} N_{j,t}}{Y_{j,t}} \approx \frac{\alpha_{j,t} - \alpha_{j,t-1}}{\alpha_{j,t-1}} + \Delta z^w_{\alpha_t,h}(z_{j,t}, z_{j,t-1}) + \Delta e^w_{j,t} + \sum_i \Delta \epsilon^w_{i,t} - \Delta z^Y_{\alpha_j,h}(z_{j,t}, z_{j,t-1})$$

Plugging these expressions into the original β_{IV} formula, and noting that under the model assumptions that measurement error and transitory productivity shocks are orthogonal to longer changes in output per worker $Cov(\Delta z^w, \Delta_3 \frac{Y}{N}) = 0$, $Cov(\Delta \epsilon^w_{i,t}, \Delta_3 \frac{Y}{N}) = 0$, and

$Cov(\Delta e^w_{j,t}, \Delta_3 \frac{Y}{N}) = 0$, we have

$$\begin{aligned}
\beta_{IV} =& \frac{cov\left(\frac{-h_k}{\alpha_{j,t-1}+\gamma_{s(k)}(\delta_k\alpha_{j,t-1}+h_k)}\frac{\alpha_{j,t}-\alpha_{j,t-1}}{\alpha_{j,t-1}} + \Delta z^w(z_{j,t}, z_{j,t-1}) + \Delta\epsilon^w_{i,t} + \Delta e^w_{j,t}, \Delta_3\frac{Y^*_{j,t}}{N^*_{j,t}}\right)}{cov\left(\frac{\alpha_{j,t}-\alpha_{j,t-1}}{\alpha_{j,t-1}} + \Delta z^w_{\alpha_t,h}(z_{j,t}, z_{j,t-1}) + \Delta e^w_{j,t} + \sum_i \Delta\epsilon^w_{i,t} - \Delta z^Y_{\alpha_j,h}(z_{j,t}, z_{j,t-1}), \Delta_3\frac{Y^*_{j,t}}{N^*_{j,t}}\right)} \\
=& \frac{cov\left(\frac{-h_k}{\alpha_{j,t-1}+\gamma_{s(k)}(\delta_k\alpha_{j,t-1}+h_k)}\frac{\alpha_{j,t}-\alpha_{j,t-1}}{\alpha_{j,t-1}}, \Delta_3\frac{Y^*_{j,t}}{N^*_{j,t}}\right)}{cov\left(\frac{\alpha_{j,t}-\alpha_{j,t-1}}{\alpha_{j,t-1}}, \Delta_3\frac{Y^*_{j,t}}{N^*_{j,t}}\right)} \\
=& \frac{-h_k}{\alpha_{j,t-1}+\gamma_{s(k)}(\delta_k\alpha_{j,t-1}+h_k)}.
\end{aligned}$$

Thus, under the maintained assumptions, the IV estimand identifies the structural wage pass-through elasticity derived in the previous subsection.

A.4 Proof: Estimating Surplus

Recall proposition 2:

Define firm j's wage bill as $WB_{j,t} \equiv \sum_k \sum_i w_{ij,t}$. Suppose a senior worker in occupation k at firm j separates exogenously. Regardless of whether or not the firm replaces the worker, the per-period profit loss to the firm from losing worker i is given by

$$\Delta\left(Y_j - WB_j\right) \approx -w_{ijk}\left(\frac{1}{\gamma_{s(k)}} + \delta_k + \frac{h_k}{\alpha_j}\right).$$

Proof: Let $S^*_{jk,t}$ be the number of senior workers in firm j in occupation k that the firm would have without an additional exogenous separation, and let $Y_{j,t}(S^*_{jk,t})$ denote output absent the additional exogenous separation. Define $\Delta^S_{jk,t}$ as the measure of workers who are exogenous separated after endogenous separations have occurred. Suppose that firms replace their exogenously separated workers. Then output in period t is

$$\begin{aligned}
Y_{j,t} =& Y_{j,t}(S^*_{jk,t})\left(\frac{S_{jk,t}}{S^*_{jk,t}}\right)^{\sigma_{jk,t}(\alpha_{j,t}\delta_k + h_k)} \\
=& Y_{j,t}(S^*_{jk,t})\left(\frac{S^*_{jk,t} - \Delta^S_{jk,t}}{S^*_{jk,t}}\right)^{\sigma_{jk,t}(\alpha_{j,t}\delta_k + h_k)} \\
\approx& Y_{j,t}(S^*_{jk,t})\left(1 - \sigma_{jk,t}\left(\delta_k\alpha_{j,t} + h_k\right)\frac{\Delta^S_{jk,t}}{S^*_{jk,t}}\right)
\end{aligned}$$

Using the first-order approximation above,

$$Y_{j,t}(S^*_{jk,t} - \Delta^S_{jk,t}) - Y_{j,t}(S^*_{jk,t}) \approx -Y_{j,t}\sigma_{jk,t}(\alpha_{j,t}\delta_k + h_k)\frac{\Delta^S_{jk,t}}{S^*_{jk,t}}.$$

Using the firm's optimality condition,

$$w_{jk,t}N_{jk,t} = \sigma_{jk,t}\alpha_{j,t}Y_{j,t},$$

and full rehiring, $N_{jk,t} = P_{jk,t}$, we can rewrite this as

$$\Delta Y_{j,t} \approx -w_{jk,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{P_{jk,t}}{S^*_{jk,t}}\Delta^S_{jk,t}.$$

At the firm's optimum,

$$\frac{S^*_{jk,t}}{P_{jk,t}} = \frac{\gamma_{s(k)}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)}{1 + \gamma_{s(k)}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)}.$$

Thus the output loss from an exogenous separation is

$$\Delta Y_{j,t} \approx -w_{jk,t}\left(\frac{1}{\gamma_{s(k)}} + \delta_k + \frac{h_k}{\alpha_{j,t}}\right)\Delta^S_{jk,t}.$$

Since the separated worker is replaced by a junior worker who is paid the same wage, then the wage bill is unchanged.

$$\Delta WB_{j,t} = 0.$$

Hence firm surplus changes by

$$\Delta(Y_{j,t} - WB_{j,t}) \approx -w_{jk,t}\left(\frac{1}{\gamma_{s(k)}} + \delta_k + \frac{h_k}{\alpha_{j,t}}\right)\Delta^S_{jk,t}.$$

Now suppose instead that firms do not replace their exogenously separated worker with a junior worker. Then when a worker is exogenously separated, output is

$$\begin{aligned} Y_{j,t} =& Y_{j,t}(S^*_{jk,t})\left(\frac{S_{jk,t}}{S^*_{jk,t}}\right)^{\sigma_{jk,t}(\alpha_{j,t}\delta_k + h_k)}\left(\frac{S_{jk,t}+J_{jk,t}}{S^*_{jk,t}+J^*_{jk,t}}\right)^{\sigma_{jk,t}\alpha_j} \\ \approx& Y_{j,t}(S^*_{jk,t})\left(1 - \sigma_{jk,t}\left(\delta_k\alpha_{j,t} + h_k\right)\frac{\Delta^S_{jk,t}}{S^*_{jk,t}} - \sigma_{jk,t}\alpha_{j,t}\frac{\Delta^S_{jk,t}}{N^*_{jk,t}}\right) \end{aligned}$$

Again using the firm's optimality condition, we can rewrite the two terms as

$$\Delta Y_{j,t} \approx -w_{jk,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{N^*_{jk,t}}{S^*_{jk,t}}\Delta^S_{jk,t} - w_{jk,t}\Delta^S_{jk,t}.$$

Since the firm does not replace the separated worker, the wage bill falls by

$$\Delta WB_{j,t} = -w_{jk,t}\Delta^S_{jk,t}.$$

Hence firm surplus changes by

$$\begin{aligned} \Delta(Y_{j,t} - WB_{j,t}) &= \Delta Y_{j,t} - \Delta WB_{j,t} \\ &\approx -w_{jk,t}\left(\delta_k + \frac{h_k}{\alpha_{j,t}}\right)\frac{N^*_{jk,t}}{S^*_{jk,t}}\Delta^S_{jk,t}. \end{aligned}$$

Once again at the firm's optimal choice for $S^*_{jk,t}/P_{jk,t}$ and full rehiring, then

$$\Delta(Y_{j,t} - WB_{j,t}) \approx -w_{jk,t}\left(\frac{1}{\gamma_{s(k)}} + \delta_k + \frac{h_k}{\alpha_{j,t}}\right)\Delta^S_{jk,t}.$$

Thus, whether or not firms hire a replacement worker, the change in value added net of the wage bill is the same. We attain the result in Proposition 2 when the increment of lost senior workers $\Delta^S_{jk,t} = 1$.

A.5 Estimation Details

The exercise estimates the occupation-specific turnover parameters in a version of the hold-up model with two firm types and three occupations. The two firm types differ in their production-function curvature, α_H and α_L, where H denotes the high-wage, low-α firm type and L denotes the low-wage, high-α firm type. The three occupations are managers G, middle-wage workers M, and low-wage workers S. Managers and middle-wage workers draw from the same skill market and share the value V_C, while low-wage workers draw from a separate skill market with value V_{NC}. These groups are based on managers, professionals, and technicians having on average more than two years of college, and all other occupations having fewer than two years of college, as reported in Table 2 of the main text.

Parameters The estimated parameter vector is

$$\theta = (\delta_G, \delta_M, \delta_S, h_G, h_M, h_S)\,.$$

Each occupation has its own value of both δ_k and h_k. The remaining parameters are calibrated. In particular, the code sets

$$\gamma = 2, \qquad \alpha_H = .63, \qquad \alpha_L = .83,$$

with the share of H firms $\zeta_H = 2/3$. These labor shares are based on the average of the top four and bottom four bins from Figure 3. Production shares are $\sigma_G = .096$, $\sigma_M = .358$, and $\sigma_S = 1 - \sigma_G - \sigma_M$.

Empirical Moments The empirical moments are the three occupation-specific pass-through semi-elasticities and two sets of worker-death moments. The pass-through moments are interpreted as semi-elasticities: changes in log wages with respect to changes in the level of the labor share. The empirical moments used in the code are:

	Managers	Middle-Wage	Low-wage
Pass-through semi-elasticity	−0.33	−0.13	−0.08
Death moment, low-wage firms	−0.30	−0.15	−0.45
Death moment, high-wage firms	−1.72	−0.90	−0.08

The empirical moment from the pass-through regressions is $\hat{\tau}_k$ from Section 4.2. The empirical moment for the losses from worker deaths is taken from $\hat{\beta}^{k,high}_{post}$ and $\hat{\beta}^{k,low}_{post}$ from Section 4.4.

Model Moments For a candidate parameter vector θ, the code first solves the stationary equilibrium, which pins down occupation-firm wages, employment, and the two worker values V_C and V_{NC}.

Firms are subject to two shocks. First, firms alternate between $\alpha_j + \Delta_t^\alpha$ and $\alpha_j - \Delta_t^\alpha$ with probability 0.1 each period, where $\Delta^\alpha = 0.01$. Firms also randomly draw a value of $z \in [.99, 1.01]$, each with probability 0.5. The equilibrium is solved taking into account the stationary distribution of shocks to z and α.

The pass-through moments are simulated. The code fixes random draws for the TFP shock, the alpha-switching process, and wage measurement error so that the objective function is deterministic across evaluations. For each unique value of realized α and TFP A, the code solves the nonlinear wage system after employment is chosen based on firms' expectation of TFP.

The simulated panel is then used to run the auxiliary IV regressions. For each occupation $j \in \{G, M, S\}$, the model pass-through moment is the IV coefficient in

$$\Delta \log w_{jk,t} = \tau_j \Delta LS_{j,t} + u_{jk,t},$$

where $LS_{j,t}$ is the firm-level labor share in levels. The regressor $\Delta LS_{j,t}$ is instrumented with the three-period change in log value added per worker. The resulting simulated IV coefficients (τ_G, τ_M, τ_S) are the model counterparts to the empirical pass-through semi-elasticities.

The worker-death moments are computed analytically. For occupation k, the model moment at firm type $j \in \{H, L\}$ is

$$\beta_k^j(\theta) = -\left(\frac{1}{\gamma} + \delta_k + \frac{h_k}{\alpha_j}\right).$$

Thus the high-wage-firm death moments use α_H, while the low-wage-firm death moments use α_L.

Criterion Function The estimator minimizes a diagonal, standard-error-weighted distance between the empirical and model moments:

$$\widehat{\theta} = \arg\min_\theta \sum_k \left(\frac{\tau_k(\theta) - \widehat{\tau}_k}{\widehat{\sigma}_{\tau_k}}\right)^2 + \sum_k \left(\frac{\beta_k^L(\theta) - \widehat{\beta}_{k,post}^{low}}{\widehat{\sigma}_{kL}}\right)^2 + \sum_k \left(\frac{\beta_k^H(\theta) - \widehat{\beta}_{k,post}^{high}}{\widehat{\sigma}_{kH}}\right)^2.$$

The optimization is performed over the logs of the six parameters, so the estimated values of δ_k and h_k are constrained to be positive.

A.6 Relaxing Equal Wages between Senior and Junior Workers

In this section, we relax the assumption that firms have to pay all workers the same wage. Instead, we impose that firms pay the same wage to all experienced incumbents, but the firm can pay replacement workers their outside option. Since workers' value if unmatched in the second round of matching is just the unemployment benefit b_s, and since all firms make take-it-or-leave-it offers, all firms can re-fill all their positions by paying $w^J_{jk,t} = b_{s(k)}$. This means that a worker's value of quitting is $V_{s(k)} = b_s$.

With firms being able to pay replacement workers a separate wage, the firm's problem becomes

$$\max_{\substack{K_{j,t},\{P_{jk,t}\},\\ \{w_{jk,t}\},\{S_{jk,t}\},\{J_{jk,t}\}}} A_{j,t}K_{j,t}^{\beta_{j,t}} \prod_{k\in\mathcal{K}(j)} \left((S_{jk,t} + J_{jk,t}) \left(\frac{S_{jk,t}}{P_{jk,t}} \right)^{\delta_k + \frac{h_k}{\alpha_{j,t}}} \right)^{\sigma_{jk,t}\alpha_{j,t}} - r^K K_{j,t} - \sum_k \epsilon_k (J_{jk,t} + P_{jk,t}) - \sum_k w^J_{jk,t} J_{jk,t} - \sum_k w^S_{jk,t} S_{jk,t}$$

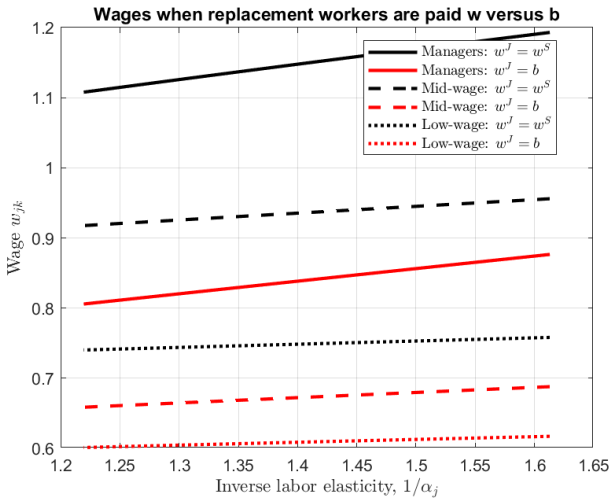
$$\text{s.t.} \qquad S_{jk,t} = \left(1 - \left(\frac{V_{s(k)}}{w_{jk,t}} \right)^{\gamma_{s(k)}} \right) P_{jk,t}$$

$$J_{jk,t} + S_{jk,t} \leq P_{jk,t}$$

$$V_{s(k)} \leq w^J_{jk,t} \qquad \text{if} J_{jk,t} > 0.$$

We then calculate the same equilibrium as in Section 5, except that we set the unemployment benefit $b_s = .5\ \forall s$. We then solve in partial equilibrium what firms would pay if they could pay $w_{jk,t} = b_{s(k)}$. Figure A.8 plots the level of wages

Figure A.8: Optimal Wage Under Equal Wages $w_{jk}^J = w_{jk}^S$ and Outside Option Matching $w_{jk}^J = b_{s(k)}$



This figure plots the optimal wage for trained workers w_{jk}^* under different values of the degree of hold-up power h_k, the underlying curvature that determines average product α_j, and wage policies for untrained workers. The black lines indicate the assumption in the main text, that untrained workers receive the same wage as trained workers. The red lines indicate that untrained workers receive their unemployment benefit b_s. For any value of h_j, firms pay a substantially lower wage level to trained workers if the untrained workers just get their outside option. However, the slope of wages with respect to α_j^{-1} is functionally identical across the assumptions on wages for untrained workers.

A.7 Finite Product Demand

In the main text, we wrote the firm's problem in terms of value added, and we did not specify whether overall curvature in the value added function came from finite product demand elasticities or sub-unity returns to scale. In this section, we derive a microfoundation for our value added function where firms can have a low labor elasticity of value added α_j in part because of low product demand elasticities. In this world, firms with high markups will have high measured value added per worker and low labor shares.

Final goods firms competitively bundle intermediate goods using a modified CES production function, where y_j is the quantity of good i, q_j is the quality of good i and D_j is an exogenous product demand shifter:

$$Y = \left(\int_j D_j y_j^{\frac{\sigma-1}{\sigma}} q_j \right)^{\frac{\sigma}{\sigma-1}}.$$

The final goods bundler takes intermediate good price p_j and quality q_j as given and chooses

quantities $\{y_j\}$:

$$\max_{\{y_j\}} = \left(\int_j D_j y_j^{\frac{\sigma-1}{\sigma}} q_j \right)^{\frac{\sigma}{\sigma-1}} - p_j y_j.$$

Taking first order conditions yields

$$D_j q_j y_j^{-\frac{1}{\sigma}} \propto p_j.$$

We can normalize the demand for a single good to be:

$$y_j^{\frac{1}{\sigma}} = D_j \frac{q_j}{p_j}.$$

Production Let each firm's production function be constant returns to scale with physical elasticity $\tilde{\alpha_j}$ with respect to labor and $1 - \tilde{\alpha_j}$ with respect to capital. The quantity of firm j's output depends only on capital inputs and the quantity labor used:

$$y_{j,t} = \tilde{A}_{j,t} K_{j,t}^{1-\tilde{\alpha}_{j,t}} \prod_{k \in \mathcal{K}(j)} (S_{jk,t} + J_{jk,t})^{\sigma_{jk,t}\tilde{\alpha}_{j,t}}$$

The quality of firms' output depends on how well-staffed the firm is with experienced workers relative to the number of positions within an occupation. Specifically,

$$q_{j,t} = \prod_{k \in \mathcal{K}(j)} \left(\frac{S_{jk,t}}{P_{jk,t}} \right)^{\sigma_{jk,t} h_k}.$$

Total revenue is therefore

$$\begin{aligned}
p_j y_j =& D_j y_j^{\frac{\sigma-1}{\sigma}} q_j \\
=& D_j \left(\tilde{A}_{j,t} K_{j,t}^{1-\tilde{\alpha}_{j,t}} \prod_{k \in \mathcal{K}(j)} (S_{jk,t} + J_{jk,t})^{\sigma_{jk,t}\tilde{\alpha}_{j,t}} \right)^{\frac{\sigma-1}{\sigma}} \prod_{k \in \mathcal{K}(j)} \left(\frac{S_{jk,t}}{P_{jk,t}} \right)^{\sigma_{jk,t} h_k} \\
=& D_j \tilde{A}_j^{\frac{\sigma-1}{\sigma}} \left(K_{j,t}^{1-\tilde{\alpha}_{j,t}} \prod_{k \in \mathcal{K}(j)} \left((S_{jk,t} + J_{jk,t}) \times \left(\frac{S_{jk,t}}{P_{jk,t}} \right)^{\frac{h_k}{\frac{\sigma-1}{\sigma}\tilde{\alpha}_{j,t}}} \right)^{\sigma_{jk,t}\tilde{\alpha}_{j,t}} \right)^{\frac{\sigma-1}{\sigma}}.
\end{aligned}$$

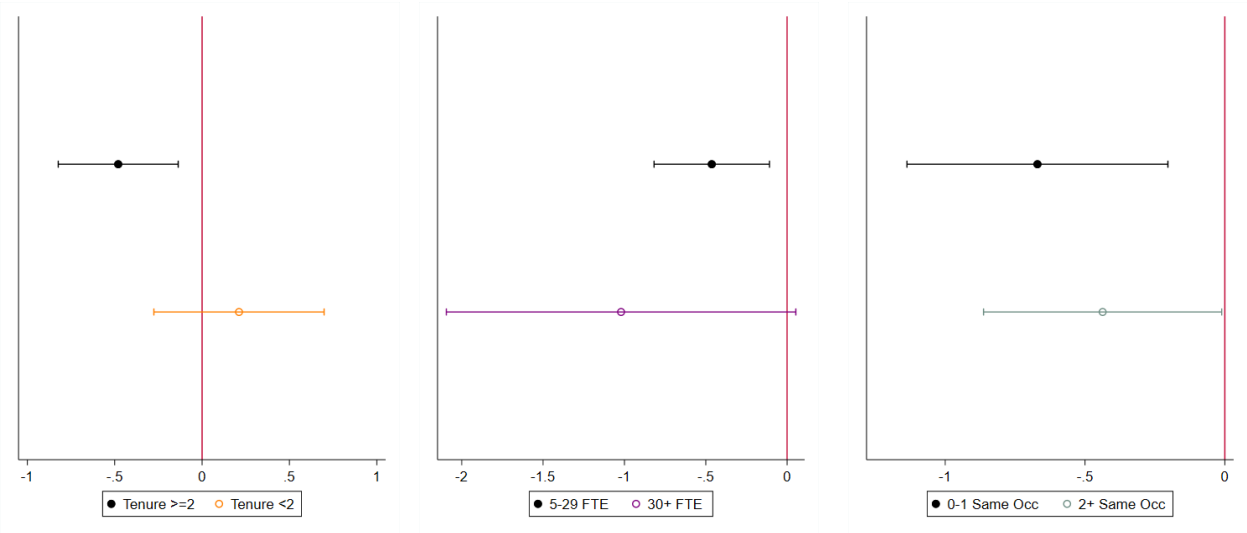
With $A_j = \tilde{A}^{\frac{\sigma-1}{\sigma}}$, $\beta_{j,t} = (1 - \tilde{\alpha_j})\frac{\sigma-1}{\sigma}$, and $\alpha_j = \tilde{\alpha_j}\frac{\sigma-1}{\sigma}$, this model is equivalent to the model in the main text when $\delta_k = 0$. Under this extension, the formulas for optimal wages and the labor share are identical to equations (3) and (5) of the main text, respectively. Thus, a low output elasticity of labor can be due to market power, high capital share, or low physical returns to scale, and our model is unchanged.

B For Online Publication: Empirical Appendix

B.1 Additional Death Results

Figure B.1: Profit Losses Relative to Prior Wage: Heterogeneity

(a) Worker Tenure (b) Firm Size (c) # of Workers in Same Occ



The figure reports the effect of worker deaths on firm profits relative to the prior wage of the deceased worker, looking at various dimensions of heterogeneity. Panel (a) splits workers between workers who have at least two full years of tenure in the firm and workers who have less than two years of tenure at the time of their death. Panel (b) splits firms between 5-30 full-time-equivalent workers (FTE) and firms with 31-100 FTE. Panel (c) splits deaths by the number of workers in the same occupation as the deceased worker. The results suggest that losing workers with greater tenure is more costly, but neither firm size nor the number of workers in the same occupation generates a significant difference in the cost of a worker death.

B.2 Residual Productivity Measure

In this appendix section, we describe our process for residualizing firm output per worker. The goal is to have a measure of firm-level productivity that we can use in cross-sectional regressions of wages on productivity. To make this measure as informative as possible, we want three features. First, we want a measure of output per worker that approximates the output elasticity of labor $\alpha_{j,t}$. Second, we want to remove variation in output per worker due to the composition of workers, e.g., an engineering firm should have a higher value added per worker than a cleaning services firms simply because the engineering firm hires workers with

higher wages. Third, our measure should not use firm-specific wage data, since the outcome variable of our regressions of interest is wages.

The measure we will use is value added net of the firm's *predicted* wage bill. Let $\bar{w}_{k,t}$ be the average wage of workers in occupation k in period t and let $WB_{j,t}$ denote the wage bill. Define also $\hat{w}_{jk,t} \equiv w_{jk,t}/\bar{w}_{k,t}$, where $\hat{w}_{jk,t}$ are firm j specific deviations from the occupation-wide average wage. Manipulating the expression $WB_{j,t} = \alpha_{j,t} Y_{j,t}$, we have

$$\underbrace{\log Y_{j,t} - \log \sum_k N_{jk,t} \bar{w}_{k,t}}_{\hat{Y}_{j,t}} = -\log \alpha_{j,t} + \log \frac{\sum_k N_{jk,t} \bar{w}_{k,t} \hat{w}_{jk,t}}{\sum_k N_{jk,t} \bar{w}_{k,t}}$$

The left hand side is our measures of log value added minus log predicted wage bill. On the right hand side is $-\log \alpha_{j,t}$ plus a measure that is the wage-bill weighted deviations of firm j wages from the occupation-specific average wage. By subtracting out the predicted wage bill, we are removing variation in output per worker due to different wages across occupations and differences across firms in their distribution of wages.

What remains in our residualized $\hat{Y}$ term is $\log \frac{\sum_k N_{jk,t} \bar{w}_{k,t} \hat{w}_{jk,t}}{\sum_k N_{jk,t} \bar{w}_{k,t}}$, which is the wage bill-weighted average deviation of wages from their occupational averages. The variation in this term will be much smaller than the variation in the log predicted wage bill $\log \sum_k N_{jk,t} \bar{w}_{k,t}$ because the (wage bill-weighted) variance of within-occupation average wages across firms is much smaller than the (employment weighted) variation of wages across occupations.

To calculate this residualized measure in the data , we first compute the predicted wage bill based on the occupational composition of hours worked in the firm:

$$\widetilde{WB}_{j,t} = \sum_k hours_{jk,t} \bar{w}_{j,t},$$

where $hours_{jk,t}$ is the total number of hours reported worked by workers in occupation j at firm k in year t, and $\bar{w}_{j,t}$ is the economy-wide average of wages of workers in occupation j in year t. We then run an OLS regression to obtain a predicted level of value added, given the firm's predicted wage bill and controls:

$$\tilde{Y}_{j,t} = \xi_1 \widetilde{WB}_{j,t} + \xi_2 Z_{j,t},$$

where $Z_{j,t}$ are controls including industry and year fixed effects, as well as the log number of hours worked by workers without an assigned occupation. Then residual value added is:

$$\widehat{Y}_{j,t} = \log Y_{j,t} - \log \tilde{Y}_{j,t}.$$

B.3 Construction of the Differentiation Score

ROCK Clustering Algorithm Once a threshold is defined, the ROCK algorithm computes the similarity of all points, and the points become neighbors if their similarity exceeds the threshold θ. Next, the ROCK algorithm creates *links*, defined as the number of common neighbors between points. Then, using this measure of links across points, the algorithm categorizes observations into clusters.

Alternative Specificity Measures We create two measures of position specificity using the Lightcast data. The first measure we construct is the average number of unique skill clusters listed in a job posting by occupation. While simple, we believe that the skill count measure will have informational content about the difficulty of replacing a worker, beyond what the wage level of the worker would imply. The first row of Table 2 provides summary statistics of the mean number of unique skills by occupation: a job posting will on average have 4.2 unique skill clusters posted, with a standard deviation of 1.49. Many low-wage services jobs fall near the bottom of this range: job postings for "Dishwashers" list a mean of 1.86 unique skill clusters, while "Database Administrator" job postings list a mean of 8.18 unique skill clusters.

Lightcast Job Clustering Examples (

Table 9: Examples from the Lightcast Skill Taxonomy

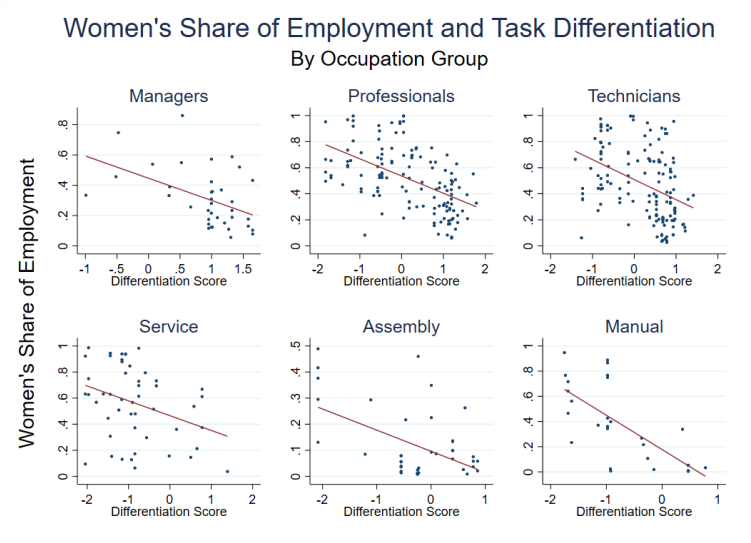
Skill Category	Skill Subcategory	Skill
Information Technology	Database Architecture and Administration	Database Administration
Information Technology	Operating Systems	Linux
Finance	General Finance	Financial Services
Media and Communications	Writing and Editing	Writing
Information Technology	Databases	Sybase (Software)
Information Technology	Query Languages	SQL (Programming Language)
Physical and Inherent Abilities	Critical Thinking and Problem Solving	Troubleshooting (Problem Solving)
Administration	Administrative Support and Clerical Tasks	Administrative Functions
Information Technology	Databases	SAP Sybase Adaptive Server Enterprise
Information Technology	Scripting	Scripting
Information Technology	Database Architecture and Administration	SQL Server Integration Services (SSIS)
Information Technology	Scripting Languages	Perl (Programming Language)
Engineering	Automotive Technologies	Automotive Service Excellence (ASE) Certification
Marketing and Public Relations	Digital Marketing	Email Marketing
Business	Business Management	Management
Administration	Office and Productivity Equipment and Technology	Microsoft Office
Sales	General Sales Practices	Sales
Administration	Office and Productivity Equipment and Technology	Microsoft PowerPoint
Marketing and Public Relations	Digital Marketing	Search Engine Marketing
Media and Communications	Communication	Communication Strategies
Administration	Office and Productivity Equipment and Technology	Microsoft Outlook
Customer and Client Support	Customer Service	Customer Service
Administration	Office and Productivity Equipment and Technology	Microsoft Excel
Marketing and Public Relations	Social Media	Social Media
Physical and Inherent Abilities	Critical Thinking and Problem Solving	Problem Solving
Marketing and Public Relations	Marketing Strategy and Techniques	Interactive Marketing
Sales	Account Management	Account Management
Business	Business Operations	Business Objectives
Physical and Inherent Abilities	Personal Attributes	Decisiveness
Marketing and Public Relations	Web Analytics and SEO	Search Engine Optimization
Media and Communications	Communication	Verbal Communication Skills
Business	Project Management	Project Management
Business	Project Management	Project Stakeholders
Media and Communications	Communication	Persuasive Communication
Business	Project Management	Project Schedules
Business	Business Continuity	Contingency Planning
Information Technology	System Design and Implementation	Systems Development Life Cycle

Figure B.2: Sample Skills in Lightcast Data

Medical and Health Services Managers											
bgtjobid	3.80E+10	3.78E+10	3.81E+10	3.78E+10	3.82E+10	3.81E+10	3.82E+10	1.03E+08	3.78E+10	2.38E+08	3.82E+10
soc	119111	119111	119111	119111	119111	119111	119111	119111	119111	119111	119111
Cluster	12	12	12	12	141	141	141	1180	9999	9999	9999
scheduling	1	1	0	0	1	1	1	1	0	0	0
emergency and intensive care	1	1	1	1	0	0	0	1	1	0	0
general medicine	1	1	1	1	0	0	0	1	0	0	0
advanced patient care	1	1	1	1	0	0	0	0	1	0	0
basic patient care	1	1	1	1	0	0	0	0	1	0	0
medical support	1	1	1	1	0	0	0	0	0	0	0
oncology	1	1	0	0	0	0	0	0	0	0	0
patient education and support	1	1	1	1	0	0	0	0	0	0	0
surgery	1	1	1	0	0	0	0	0	0	0	0
health information management	0	0	0	0	1	1	1	1	0	1	0
general administrative and clerical	0	0	0	0	1	1	1	1	0	0	0
medical billing and coding	0	0	0	0	0	0	0	1	0	1	0
medical records	0	0	0	0	0	0	0	1	0	1	0
data analysis	0	0	0	0	0	0	0	1	0	0	0
data management	0	0	0	0	0	0	0	1	0	0	0
data techniques	0	0	0	0	0	0	0	1	0	0	0
management information systems	0	0	0	0	0	0	0	1	0	0	0
medical documentation and abstrac...	0	0	0	0	0	0	0	1	0	0	0
process improvement	0	0	0	0	0	0	0	1	0	0	0
budget management	0	0	0	0	0	0	0	0	1	1	1
business process and analysis	0	0	0	0	0	0	0	0	0	1	0
people management	0	0	0	0	0	0	0	0	1	1	1
quality assurance and control	0	0	0	0	0	0	0	0	0	1	0
auditing	0	0	0	0	0	0	0	0	0	0	1
businessstrategy	0	0	0	0	0	0	0	0	0	0	1
contract management	0	0	0	0	0	0	0	0	0	0	1
healthcare procedure and regulation	0	0	0	0	0	0	0	0	0	0	1
occupational health and safety	0	0	1	0	0	0	0	0	1	0	0
	Direct Caregiving				Admininstrative			Admin. Analysis	Care Analysis	Management	Regulation

B.4 Task Differentiation and Women's Employment Share by Occupation

Figure B.3: Task Differentiation and Women's Employment Share by Occupation



This figure reports the share of workers in an occupation who are women and the occupation's differentiation score, separately for detailed 4-digit DISCO codes within broad 1-digit DISCO occupation groups. Managers, professionals, assembly, and manual occupations show the steepest negative relationship between task differentiation and the share of workers in each occupation who are women.