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

Justin Bloesch
Harvard University

Birthe Larsen
Copenhagen Business School

Bledi Taska
Burning Glass Technologies

August 15, 2022

Abstract

We argue that productive firms share rents with workers only in occupations where workers have individual hold-up power. We present a model of wage determination where firms produce using a novel generalization of Kremer (1993)'s O-ring production function. Workers have individual hold-up power if (i) labor is organized into distinct, differentiated positions (ii) the output of positions is individually complementary or "critical" in the production process, and (iii) skills are position-specific, i.e., skills are acquired on the job and are not transferable across positions or firms. If output losses from an unfilled position are larger at productive firms, incomplete contracts and on-the-job search incentivize productive firms to pay differentially high wages. We estimate individual worker hold-up power by occupation using the effect of worker deaths on firm profits in Danish administrative data and using a measure of within-firm, across-position task differentiation from US job posting data. High hold-up occupations exhibit both higher wage levels and higher long-run passthrough of permanent firm productivity innovations to wages, supporting the main model predictions. Accounting for heterogeneity in hold-up power across occupations has numerous implications for wage inequality: (1) greater employment of men in high hold-up occupations can account for one fifth of the Danish gender wage gap; (2) rising "superstar firms" increase wage inequality; (3) hold-up power decreases the responsiveness of wages to labor market slack.

*Corresponding author: Justin Bloesch (email: jbloesch@g.harvard.edu). Bloesch thanks his advisors Gabriel Chodorow-Reich, Lawrence Katz, Ludwig Straub, and James Stock for guidance and support. We thank Antoine Bertheau, Adrien Bilal, John Coglianese, Harris Eppsteiner, Jason Furman, Xavier Gabaix, Andrew Garin, Ed Glaeser, Fane Groes, Omeed Maghzian, Namrata Narain, Anna Stansbury, Jacob Weber and seminar participants at Harvard University, the 2021 Search and Matching Conference, the University of Copenhagen, and the Copenhagen Business School 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

As wage inequality has grown in a number of advanced economies, researchers have taken a renewed interest in understanding the role that firms play in determining wages. Recent research has explored whether firms have heterogeneous wage effects for different types of workers. Song et al. (2019) find that in prior decades in the United States, all types of workers earned more at large firms, but in recent decades only college graduates earn high wages at large firms while the firm-size wage premium has nearly disappeared for non-college workers. Other research focuses on passthrough of productivity or demand shocks to wages,[1] finding higher passthrough for ex-ante high-wage workers, men, and workers with higher attachment to the firm. In this paper, we will examine the effect of firm productivity on wages, where heterogeneity operates through occupations. Figure 1 shows the cross-sectional relationship between individual worker hourly wages in Denmark between 2008 and 2016, residualized for standard demographic controls, and firm output per worker, residualized for occupational composition, year, and industry. The cross-sectional productivity-wage elasticity is 0.2 for managers, a little under 0.1 for most middle-wage occupations, and nearly zero for low-wage service occupations.

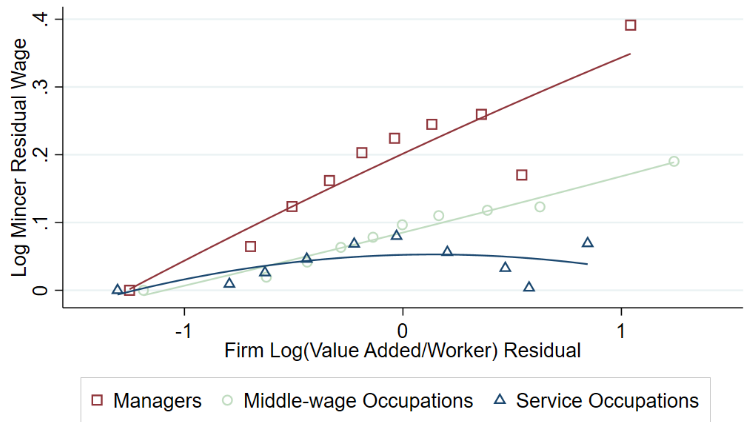
In this paper, we argue that at least half of the cross-sectional elasticity of wages to productivity in Figure 1 reflects wage *premia*, and we argue that the underlying source of different productivity-wage elasticities is heterogeneity in individual worker hold-up power across occupation groups. Workers have individual hold-up power when (i) labor in production is organized into distinct, differentiated positions, (ii) the output of positions is individually complementary or "critical" in the production process, and (iii) skills are position-specific, i.e., skills are acquired on the job and are not transferable across positions or firms. If output losses from an unfilled position 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. With opportunities for on-the-job search, incomplete contracts incentivize productive firms to pay higher wages to decrease differentially costly turnover.

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 an experienced cook 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 has position specificity, 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 this example is limited to a firm with only two positions, our more general model formalizes these intuitions in firms with a continuum of positions and diminishing returns to labor.

The main contributions of this paper are four-fold. First, we establish conditions under which an individual worker has hold-up power over inframarginal firm output in a multi-worker firm and develop a novel and tractable production function to generate these conditions. Second, we measure hold-up power by occupation group in two ways, both by estimating the effect of worker deaths on firm profits in Denmark and by constructing a separate measure of within-firm, across-position task differentiation in US job posting data. Third, we provide empirical evidence that hold-up power predicts measures of productivity-wage premia, estimated using long-run passthrough of permanent productivity innovations to wages. Fourth, we demonstrate that heterogeneity in hold-up power is useful for analyzing numerous dimensions of wage inequality,

[1] See Kline et al. (2019) in the United States, Garin and Silverio (2018) in Portugal, Friedrich et al. (2019) in Sweden, and Maibom and Vejlin (2021) and Chan et al. (2020) in Denmark.

Figure 1: Cross-Sectional Productivity-Wage Elasticity by Occupation



The wage and labor productivity measures are residuals of a regression of log wages or firm value added per worker on standard Mincer regression covariates: sex, years of education, potential experience, potential experience squared (with interactions on years of education), and industry, year, and 6-digit occupation code fixed effects. The sample is full-time workers in their "main" November job between 2008-2016, in private sector firms that are at least 5 years old.

including the gender wage gap, the effect of superstar firms on wage dispersion, and the responsiveness of wages and employment to labor market slack.

We begin the paper by introducing a production function that generalizes the O-Ring production in Kremer (1993), allowing for intermediate cases of complementarity in which a failed task or vacant position decreases output in proportion to average product, rather than total product as in the traditional O-Ring case. We introduce a key parameter that varies by occupation called the *degree of complementarity*, which governs how disruptive it is for a position to suddenly go unfilled. Mathematically, the key feature is that turnover generates output losses that are multiplicative in firm average labor productivity. Thus, the output loss from sudden turnover is not just "at the margin" but instead interacts with the firm's inframarginal productivity. This distinction allows us to introduce two separate concepts of marginal product, where the *extensive marginal product* is the marginal output of a firm being one position smaller or larger, and the *intensive marginal product* is the lost output due to an incumbent worker leaving after the production arrangement has been set. These notions of marginal product differ due to the multiplicative output and the requirement that the combination of positions needed to produce is set each period before turnover and production occur. This is a particular form of lower short-run substitutability of factors, incorporating the idea that labor inputs are indivisible after the firm has assigned tasks.

Next, in a simple two-period model with worker turnover, we show that both individual position production complementarities and position-specific skills are necessary for individual workers to have hold-up power, which generates wage premia at productive firms.[2] We derive a simple closed-form elasticity of wages to firm productivity as a function of worker hold-up power.[3] We show that with multi-worker firms, high average product arises from greater concavity in the revenue function with respect to the number of positions. While higher wages increase the probability of retaining a given worker, the firm does not need to pay

[2]We show that firm-specific skills raise the level of wages but not the slope of wages with respect to firm productivity.

[3]This setup implies that firms are worse off when workers have hold up power and may attempt to "despecify" workers or make production less fragile. For most of this paper, we assume that the complementarities and position specificity associated with employing a particular occupation is an exogenous feature of producing a given type of product. A more general model could allow firms to pay a cost to despecify positions, for example by training workers to learn each others' roles or to change production such that output is less fragile. We discuss an example in Appendix A.5.

higher wages to become larger, as a firm with twice as many positions but the same wage simply doubles the number of workers while maintaining a constant retention rate.

To obtain an empirical measure of individual hold-up power for each occupation, we use two sources of information. First, we estimate the effects of worker deaths on a measure of firm profits in Danish administrative data, measured as value added less wages and salaries. Profit losses are highest when the deceased worker is a manager, even relative to the worker's prior wages. When the deceased worker was employed in an occupation in the middle of the wage distribution, such as professionals, technicians, and skilled blue collar jobs, firm profit losses are moderate. Firm profit losses are the smallest relative to prior wages for low-wage service, administrative, and manual occupations. We further show that profit losses tend to be higher at high productivity firms, especially for managers, providing a direct test of our production function.

Second, we estimate occupation-level measures of how differentiated a job tends to be from other jobs within the firm based on the skill requirements in online job postings from Burning Glass Technologies. Following Lazear (2009)'s "skill-weights approach", skills are specific to a position because positions require unique combinations of general skills. Therefore we construct an index that measures, for a given occupation, how different are the task requirements in a given job from the task requirements of other jobs within the firm. We argue that this measure should proxy well for individual hold-up power: (i) position outputs are more likely to be complementary if they are differentiated, and (ii) skills are more likely to be position specific, with incumbent co-workers less able to do each other's jobs, if skills are differentiated across positions. While our task differentiation measure is correlated with traditional measures of skill such as years of education, there is significant variation in the degree of task differentiation among occupations of similar levels of education. We show that this measure of task differentiation is predictive of profit losses in the event of a worker death, providing evidence that this measure captures hold-up power well at the detailed occupation level.

We then estimate the effect of innovations to firm productivity on wages across occupation groups using a common instrumental variables strategy that regresses long-run changes of individual workers' wages on long-run changes in productivity, conditioning on workers who stay in the firm. We find the highest passthrough among managerial occupations, with an elasticity of wages in response to permanent shocks of approximately 0.14. Workers in craft and assembly occupations, as well as professionals and technicians, show passthrough elasticities of around 0.06-.07. Workers in administrative, sales, or low-wage service and manuals jobs exhibit passthrough elasticities of 0.04. Testing the effect of task differentiation on passthrough at the detailed occupation level, we find that passthrough of productivity changes to wages is higher in occupations with higher task differentiation, while average educational attainment in an occupation has an insignificant to negative effect on passthrough. Together, these results support the main prediction that individual hold-up power generates wage premia at productive firms, while also showing that hold-up power is not simply a function of traditional measures of skill.

One key feature of our model is that wages increase with firm productivity in way that is independent of firm size, in contrast to models of upward sloping labor supply that are often used to explain firm wage premia. Estimating the firm size wage effect from switchers to discriminate between models, we find that the elasticity of wages to firm size is very small (less than .01) and slightly negative after conditioning on firm productivity. The only exception to this finding is top executives: when workers change firms and are listed as top executive at both firms, their earnings increase when switching to a larger firm. We show that this can be easily reconciled if the complementarities of an executive are proportional to the size of the firm, consistent with the "size of stakes" hypothesis in Gabaix and Landier (2008).

Turning to implications, we show that accounting for heterogeneous individual hold-up power across occupations has numerous implications for interpreting wage inequality. First, we show that hold-up power, and higher turnover costs in general, raise the level of an occupation's wage. Because men tend to be employed in higher hold-up occupations, rents from hold-up power raise men's wages by 3-4 percentage points more than for women, accounting for approximately one fifth of the gender wage gap in Denmark.

Next, we explore the effect of rising firm productivity dispersion and the rise of "superstar firms" on wage inequality in the spirit of Autor et al. (2020). We show that in Denmark from 2001 to 2015, large firms experienced differential increases in productivity, while value added has become more concentrated in large firms. We show that wages rose in higher hold-up occupations in these large, productive firms, increasing both total wage inequality and within-occupation wage inequality.

We then show that hold-up power decreases the response of an occupation's wages to labor market slack. In a simple modification to the baseline model, we introduce turnover costs that depend on labor market tightness. Because high hold-up occupations also have turnovers costs that do not depend on labor market tightness (i.e., output losses from complementarities), total turnover costs are less cyclically sensitive, and so wages in these occupations are less cyclical. We show supportive evidence from the cyclicality of wages for leisure and hospitality workers in recent US business cycles.

Most of the results in this paper are derived from a stylized, two-period setting where workers' outside options are exogenously specified. To verify the predictions of the model when workers' outside options are endogenously determined, we derive an equilibrium labor market model with on-the-job search and heterogeneous firms. We confirm three main predictions from the partial equilibrium analytical model: (1) hold-up power increases the level of an occupation's wage; (2) hold-up power increases the elasticity of wages to firm productivity, and (3) wages in high-hold up occupations are less sensitive to changes in labor supply, which is the general equilibrium analogue to the partial equilibrium exercise on labor market slack. In this last exercise, we show that when the supply of workers increases, low hold-up jobs absorb the majority of the change, with wages falling and employment rising relative to the high hold-up jobs.[4]

1.1 Related Literature

A large literature has sought to explain heterogeneous wages for observably identical workers, based on the employer, industry, and institutional status of the worker. Blanchard and Summers (1986) and Lindbeck et al. (1989) explore insider-outsider theories in which insiders use market power or institutional advantages to exploit high turnover costs. Gibbons et al. (2005) and Gibbons and Katz (1992) explore whether inter-industry wage premia are explained by unobserved worker quality or rent-sharing. Krueger and Summers (1988) argue that the inter-industry wage differentials cannot be explained by competitive wage setting alone. That high wages and queuing may result from training costs and lost output is discussed in Stiglitz (1974). Katz (1986) explores various efficiency wage mechanisms, and Montgomery (1991) offers a search-theoretic explanation in which productive firms offer higher wages to find workers faster.

Card et al. (2018) provide the canonical model of rent sharing in multi-worker firms. Lamadon et al. (1994) use a related framework, where rents arise due to finite long-run extensive margin labor supply elasticities. Card et al. (2016) find heterogeneous labor supply elasticities by gender.

Manning (2006) shows that whether or not a firm is monopsonistic depends on whether recruiting costs are convex. In our model, recruiting costs are linear and training costs are linear respect to firm size in the long run, implying that larger firms do not pay higher wages in steady state. Manning (2011) provides a comprehensive overview of imperfect competition in the labor market. Our production function shares a

[4]One application of this result is that increasing the supply of workers with college-level skills may decrease the relative wage of high-education, lower hold-up jobs such as professionals and technicians, while the wages of managers with greater hold-up power would be less affected. Therefore, to the extent that top manager's salaries rise due to hold-up power at increasingly productive firms, expanding the supply of educated workers would be unlikely to reverse this trend.

feature with that in Manning (1994), in which firms operate with a 'blueprint,' and deviations in the firm's labor force away from the optimal level specified in the blueprint generate larger output losses than would be lost by operating with a smaller blueprint.

A range of papers suggest that workers have heterogeneous ability to bargain with employers. Hall and Krueger (2012) show that employers of high-wage occupations tend to bargain, while employers of low-wage workers tend to post wages. Caldwell and Harmon (2019) show that when a worker's knowledge of outside offer increases due to growth at firms of former co-workers, higher wage occupation see the largest percent gains in wages. Lachowska et al. (2021) show that when dual job-holders receive a raise in their secondary job, high-wage workers tend to get raises in their primary job, while low-wage workers tend to switch jobs. Our framework provides a microfoundation for why match surplus to the firm is larger for high-wage occupations.

This paper shares a feature with Stole and Zwiebel (1996) that inframarginal productivity affects wage setting. However, our model does so without relying on off-equilibria breakdowns in negotiations to affect wage determination. Additionally, our framework allows for estimating heterogeneity in how inframarginal productivity affects wages across occupations. Cahuc et al. (2006) estimates heterogeneous bargaining power accounting for on-the-job search.

This paper is related to research that explores how differentiation of tasks within a firm makes skills specific to the firm. Lazear (2009) argues that skills are firm specific because firms require particular combinations of general skills. Leping (2009) measures how task differentiation differs across occupations as a predictor of on-the-job training. Yamaguchi (2012) conceives of occupations as bundles of tasks to measure the similarity of and distance between occupations. Edmond and Mongey (2019) argue that low-education occupations have homogenized and that firms' inability to 'unbundle' labor gives workers in highly differentiated occupations rents due to comparative advantage. Lise and Postel-Vinay (2020) describe occupations as bundles of tasks and argue that cognitive skills take longer to acquire than manual skills. Haanwinckel (2018) models workers as assigned to tasks in different degrees, where a firm is a bundle of tasks and firms face upward sloping labor supply due to idiosyncratic preferences across firms. Goldin and Katz (2016) argue that increasing substitutability of workers within an occupation decreases the part-time hours premium and the gender wage gap.

Empirically, we follow Bertheau et al. (2021) and Jäger (2016) in the worker deaths literature. We find in common with Jäger (2016) that managers in particular exhibit high complementarities with other workers in the firm. Bennedsen et al. (2020) finds that hospitalization of CEOs also affects firm performance. Our study relates to the estimation of employee replacement costs as in Dube et al. (2010) and Muehlemann and Leiser (2018). These studies generally find that replacement costs are larger relative to wages in high-skill occupations, and that training costs relatively outweigh recruiting costs as a share of all turnover costs. Bertheau et al. (2021) estimate that turnover costs are much higher than in previous studies.

Cobb and Lin (2017), Bloom et al. (2018), and Mueller et al. (2017) point out that large firms continue to offer a premium to high-wage college educated workers, but non-college workers no longer earn a premium at large firms. While our study does not approch this question directly, we hypothesize that (i) some low-education occupations may have 'de-specified', and (ii) the occupational composition of low-education work has shifted toward low specificty occupations. In the absence of instutitions that impose wage compression within the firm, large, productive firms may show growing within-firm wage inequality. Gregory et al. (2020) demonstrates that firms have heterogeneous effects on workers' wage growth, arguing that firms provide heterogeneous learning environments for acquiring human capital.

In our model, workers may move jobs that have the same or even lower wages due to idiosyncratic preferences over different workplaces. This mechanism is explored theoretically in Arnott and Stiglitz (1985), and the variance of match values due to idiosyncratic preferences relative to wage variation are show to be large in Hall and Mueller (2018) and Sorkin (2018).

Various authors have also studied turnover costs and complementarities to understand the durability of employment relationships. Nagypál (2001) develops a related two-worker firm model, in which two types of labor are complements, but the quality of the match in the skilled job is learned over time. Oi (1962) argues that firms are less likely to layoff high "fixity" workers in recessions. In companion papers Jovanovic (1979a) and Jovanovic (1979b) argues that workers learn on the job about match quality over time and that the probability of turnover depends on the current beliefs of the quality of the match.

That workers may be complementary in production is explored in the organizations literature, such as Baldwin (2018). Eisfeldt and Papanikolaou (2013) argue that key talent that is specific to the firm is a key component of organization capital, and that better outside options of key workers renders firms with high organizational capital to be riskier, further implying that key talent has claims to firm's cash flows.

The rest of the paper is organized as follows. Section 2 develops the theory of individual hold-up power and derives a closed form elasticity of wages to average product that depends on workers' individual hold-up power. Section 3 describes the data used in this paper, and Section 4 estimates the degree of hold-up power by occupation. Section 5 estimates passthrough of changes in firm productivity to wages by occupation. Section 6 discusses implications for wage inequality. Section 7 derives an on-the-job search model that demonstrates the robustness of the results from the basic theory in a dynamic equilibrium setting. Section 8 concludes.

2 Theory

In this section, we derive a production function with production complementarities between individual positions within firm. We then embed firms that use this production technology into a simple frictional labor market and show that workers can extract higher wages from high productivity firms only if production exhibits complementarities *and* if workers are not perfectly replaceable, i.e., skills are position-specific.

It is important to note that we argue that individual hold-up power is a property of an occupation. In Sections 2.1 and 2.2, we will assume that any given firm produces using positions of only one type of occupation. In Section 2.3, we will generalize the model so that firms produce using multiple occupations, each with their own degree of complementarity and degree of hold-up power, and we show that the predictions are nearly identical as under one-occupation firms. Throughout the paper, occupations will be denoted with subscript j.

2.1 Large Firms with Positions and Individual Complementarities

Consider a firm that is producing in only one period and is exogenously endowed with a production function. To introduce the production function, we define two production inputs, the number of positions N, and the share of positions filled with fully specific, trained workers X. An unfilled position indicates that either a worker is unproductive or that the position is vacant. We take a stand on this distinction later in this section. We also define the degree of complementarity $m \geq 0$, which is a parameter that governs how disruptive an unfilled position is to the production process. In the simplest version of the model, all positions will have the same degree of complementarity.[5] Let output take the form:

$$Y = F(N)g(m, X).$$

[5] In Section 2.3, we will generalize the function to allow firms to utilize multiple occupations, each with their own degree of complementarity.

The general form of production has two parts: a familiar outer structure that determines the potential revenue product of the firm as a function of the number of positions $F(N)$, with $F'(N) > 0$ and $F''(N) \leq 0$.[6] The second term $g(m, X)$ is the share of potential output that is achieved due to the extent that the firm is producing with specific workers. $g(m, X)$ is defined to have three properties:

$$(1): \quad g(m, 1) = 1 \qquad (2): \quad g_X(m, X) > 0 \qquad (3): \quad \frac{\partial^2 \log(g(m, X))}{\partial X \partial m} > 0.$$

The first condition states that if all positions are filled with trained workers, then the firm achieves its potential output given N, regardless of degree of complementarity m in production. The second condition states that the share of potential output achieved is increasing in the share of positions filled X. The third states that the percent response of g to changes in the share of filled positions X is higher when m is higher. That is, when complementarities are high, filling positions with productive workers generates more output in percent terms, and losing productive workers is more costly to output in percent terms.

Generalized O-Ring We will briefly introduce a production function that generalizes the O-ring production function of Kremer (1993), and we show that its large firm approximation follows the form of $Y = F(N)g(m, X)$ as above.

The key intuition of the O-ring form of production is that labor inputs are divided into discrete positions, and that output losses are multiplicative when the tasks of a position are not performed. Let the firm produce using N tasks, and the maximum total output a firm can produce given N tasks is $F(N) = AN^\alpha$, where A is a productivity shifter and α governs the curvature of potential output with respect to the number of positions. Firm total output is equal to:

$$Y = AN^\alpha \prod_{i=1}^{N} \Big(1 - \frac{\alpha + m}{N}(1 - q_i)\Big), \tag{1}$$

where the second term is a cumulative product of N individual terms, $q_j \in \{0, 1\}$ is an indicator of whether tasks of position i were successfully performed, and m is the degree of complementarity. There are two main modifications to the original O-ring production. First, the size of the firm N appears in the denominator multiplying the $(1 - q_i)$ term. This means that the fraction of output lost from a non-performing position scales inversely with the size of the firm. Second, in the numerator of the multiplication are the terms $\alpha + m$ where α is the same as before and m is the degree of complementarity for position i. This production function reduces to the traditional O-ring when $m = N - \alpha$.

Large Firm Approximation For the rest of the paper, we will use a large firm approximation of generalized O-ring:

$$Y = AN^\alpha e^{-(\alpha+m)(1-X)}, \tag{2}$$

where N is now a continuous measure of the number of positions, and the function that determines the share of potential output is $g(m, X) = e^{-(\alpha+m)(1-X)}$.[7] Recall that X is the share of positions that are filled with trained workers with position specific skills, and we assume that the remaining $1 - X$ share of positions are filled with unproductive workers.[8] Intuitively, the exponential structure indicates that marginal unfilled positions decrease output by a constant proportion, and this proportion increases with the degree of complementarity m.

[6] We do not distinguish whether diminishing returns of of $F()$ come from declining marginal physical product or downward sloping demand, and in the appendix we show that both underlying models of concavity can generate the same functions forms.

[7] We discuss this discrete case in more detail in Appendix A.1 and show that the continuous version is a large firm approximation of the discrete generalized O-ring production.

[8] This will be important when defining firm average product.

We define two new terms that will be central to the following analysis: the *extensive marginal product* and *intensive marginal product*, denoted MPL^E and MPL^I, respectively. In words, the extensive marginal product is the additional output from the firm becoming larger, holding the share of positions filled constant:

$$MPL^E = \frac{\partial Y}{\partial N}\Big|_X = \alpha A N^{\alpha-1} e^{-(\alpha+m)(1-X)},$$

with $X \in [0,1]$.[9] To understand this expression, consider the case where all positions are filled and the firm adjusts its size: $MPL^E|_{X=1} = \alpha A N^{\alpha-1}$, which is the expression for marginal product under neoclassical production with Cobb-Douglas. If the share of positions filled X is less than 1, then the marginal output from the firm growing the number of positions, but holding the share of positions fixed, will be lower than if all positions were filled: $e^{-(\alpha+m)(1-X)} < 1$ if $X < 1$.

The intensive marginal product is the additional output gained by filling positions with productive workers, holding the number of positions N fixed:

$$\begin{aligned} MPL^I = \frac{\partial Y}{\partial X}\Big|_N \times \frac{1}{N} &= (\alpha+m) A N^{\alpha-1} e^{-(\alpha+m)(1-X)} && \text{if } X \in [0,1] \\ &= \underbrace{\alpha A N^{\alpha-1}}_{MPL^E|_{X=1}} + \underbrace{m A N^{\alpha-1}}_{\text{Multiplicative losses}} && \text{if } X = 1. \end{aligned}$$

Evaluating the intensive marginal product at all positions being filled ($X = 1$), we see that the intensive marginal product is the sum of two terms: the extensive marginal product and an additional term that is equal to m multiplying the firm's average product $mAN^{\alpha-1} = m \times \frac{Y}{N}$. Lastly, the expression for average product is:

$$\begin{aligned} APL = \frac{Y}{N} &= A N^{\alpha-1} e^{-(\alpha+m)(1-X)} && \text{if } X \in [0,1] \\ &= A N^{\alpha-1} && \text{if } X = 1. \end{aligned}$$

Together, these expressions generate convenient properties. First, similar to standard Cobb-Douglas, the ratio of average product to extensive marginal product is constant:

$$\frac{APL}{MPL^E} = \frac{1}{\alpha}.$$

Next, relating the two marginal product terms, the intensive marginal product MPL^I is weakly greater than the extensive marginal product MPL^E and is strictly greater if $m > 0$:

$$MPL^I = \big(1 + \frac{m}{\alpha}\big) MPL^E = (\alpha + m) \times APL.$$

The key features is that the output losses from losing productive workers interacts with inframarginal productivity only if $m > 0$.

Discussion While the production function describes firms with a continuum of positions, it is useful to consider a discrete case. Consider two firms that share identical production functions, but firm A has 100 positions and firm B has 99 positions. If all positions at both firms are filled with fully productive workers, then firm A will produce more than firm B. However, suppose that firm A loses a trained worker, and so

[9] We assume that the remaining $N(1-X)$ positions are occupied by workers who are unproductive in their roles. While in output terms there is no difference between a vacant position and one filled with an unproductive worker, this assumption allows for a more convient derivation of model properties and is consistent with the remainder of the paper. We relax the assumption that untrained workers are completely unproductive in the next section, Section 2.2.

has 100 positions but only 99 filled. If $m > 0$, then output will be *lower* at firm A than at firm B. This feature can be seen as a form of lower ex-post substitutability of labor inputs once production has been set, i.e. production follows a "blueprint".[10]

2.2 Position Specific Skills and Individual Hold-Up Power

In this section, we use a simple two period model to show that production complementarities alone are insufficient to generate productivity-wage premia in multi-worker firms with diminishing revenue returns to labor, and it is necessary that both the production process exhibits complementarities *and* that workers' skills in performing complementary tasks are position specific.

We additionally show that when firms choose the number of positions with which to produce, measured firm productivity in terms of output per worker is a function of the concavity of the revenue function with respect to the number of positions α.

Two Period Model There are two periods. Consider a firm production function with proportional complementarities as described above with a common m for all positions. At the beginning of period 1, the firm chooses the number of positions N with which it will produce. The firm sets a common wage policy for all workers w, which it must commit to pay in the later period to all workers in the firm. At the end of period 1, N workers are hired and trained in their specific positions. In period 2, workers draw from a distribution of outside wages $F(w')$. Workers care only about wages and change jobs if the competing firm offers a higher wage. Prior to production at the end of period 2, the firm can hire untrained workers to fill the vacant positions. New hires are less productive than trained incumbents, and let $d \in [0, 1]$ denote the gap in productivity in complementary tasks between fully trained and new workers; d is therefore the measure of position specificity. New hires are paid the same wage as incumbent, trained workers.[11] The firm's problem becomes:

$$\max_{N,w} AN^{\alpha} e^{-(\alpha+dm)(1-X)} - wN \tag{3}$$

subject to:

$$X = F(w),$$

where $F()$ is the CDF of the outside offer distribution. Let $F(w') = 1 - \left(\frac{w'}{\underline{w}}\right)^{-\gamma}$, for $w' \geq \underline{w}$. In this functional form, γ is the elasticity of the quit probability with respect to the wage. Solving for the firm's optimal choice of wages, we obtain:

$$w^* = \left(\gamma\Big(1 + \frac{dm}{\alpha}\Big)\right)^{\frac{1}{\gamma}} \underline{w}. \tag{4}$$

The firm's optimal choice for the number of positions is:

[10] The short-run rigidity of production, and its fragility to disruption from individual workers, raises the question of whether firms can adjust production to be less dependent on a single worker. Appendix A.3 outlines a dynamic model in which newly hired workers acquire position-specific skills each period with probability p. This model can also be reinterpreted as the firm being able to rearrange production to fit the skills of the newly hired worker with probability p each period. Another extension includes a choice margin where firms can expend resources to decrease the degree of complementarity m, either by paying a fixed or a flow cost. In Appendix A.5, we derive one such scenario where firms can expend flow resources to decrease m.

[11] This assumptions allows us to derive a closed-form solution for the elasticity of wages to average product. In Appendix A.4, we relax this assumption and allow the untrained workers to be paid a lower wage, specifically the minimum outside wage $\underline{w}$.

$$N^* = \left(\frac{\alpha A e^{-(\alpha+dm)(\frac{w^*}{\underline{w}})^{-\gamma}}}{w^*}\right)^{\frac{1}{1-\alpha}}. \tag{5}$$

There are a few important aspects to note about the firm's choice of wage. First, as the market minimum wage $\underline{w}$ increases, the firm's optimal wage increases as well: $\frac{\partial w^*}{\partial \underline{w}} > 0$. Second, as the degree of complementarity m increases, the optimal wage increases, as the losses of a worker leaving grow as m increases. $\frac{\partial w^*}{\partial m} > 0$. Third, the wage is increasing in position specificity (the gap of productivity between trained and untrained replacement workers): $\frac{\partial w^*}{\partial d} > 0$.[12]

We also have that $\frac{\partial w^*}{\partial \alpha} < 0$ if $dm > 0$. While in simple terms α is just the curvature of the production function, α^{-1} is also the ratio of average to extensive marginal product. Therefore, as α falls and output becomes more concave with respect to number positions, average product *rises* relative to extensive marginal product, and the product $dm > 0$ gives rise to a positive relationship between average product and wages. Under the assumption that heterogeneity in α the only source of heterogeneity in average product across firms (i.e., complementarities m and position specificity d are constant across firms), then the elasticity of wages to average product is:

$$\varepsilon_{w,apl} = \frac{\mathbf{h}}{\gamma\alpha + (1+\gamma)\mathbf{h}}, \tag{6}$$

where $\mathbf{h} = dm$ is the measure of individual hold-up power over firm average product, equal to the product of the degree of complementarity m and position specificity d.[13] This expressions shows that $\varepsilon_{w,apl} \to 0$ as $\mathbf{h} \to 0$ (i.e., $m \to 0$ or $d \to 0$), and $\varepsilon_{w,apl} \to \frac{1}{\gamma+1}$ as $\mathbf{h} \to \infty$. This says that if positions exhibit no complementarities *or* if skills are not position specific, i.e. $\mathbf{h} = 0$, then higher productivity firms in terms of observed average productivity pay no wage premium. As the product of complementarities and position specificity $\mathbf{h} = dm$ grows, the elasticity of wages to average product increases. Crucial to this setup is that output losses from turnover are multiplicative in average product, and that newly hired workers cannot fully offset the multiplicative losses.

Another result is that the parameter A, which typically stands in for total factor productivity, has no effect on a firm's optimal choice of wages. Instead, A only affects firms size: as A increases, the firm's optimal number of positions increases. Facing the same retention function, two firms with the same α, m, and d but different values of A will each choose expand until their extensive marginal products are equalized. Since the ratio of average to extensive marginal product is just $1/\alpha$, the two firms will have identical average products and wages. This result ultimately comes from the assumption of linear extensive margin wage costs: the marginal cost of another position is constant and independent of the size of the firm, unlike in standard monopsony models. Instead, the choice of the wage reflects the firm's tradeoff between per-position wage and turnover costs, and the firm's size is chosen independently given that optimal wage.

Lastly, a key endogenous result arises, that the wage will be equal to the extensive marginal product:

$$w^* = MPL^E.$$

This condition simply results from firm optimization. The extensive marginal product is the additional

[12] This problem is subject to a technical condition that replacement workers must be sufficiently productive that the firm finds it optimal to hire the replacement workers: $(1-d)m > 1$. That is, untrained workers must be at least as productive as w^* to justify hiring them. This requires both that (i) complementarities m are sufficiently large and replacement productivity $(1-d)$ is sufficiently high. However, this technical restriction is a function of the two-period setup, since in a dynamic setting, the value of untrained workers to the firm is their future productivity. In Appendix A.3, we derive a dynamic version of the firm's problem that yields an identical wage policy but has no technical restrictions on d and m.

[13] We provide a full derivation in the Appendix A.2.

output the firm gets from being one position larger, and the wage is the cost.[14] If a firm is optimizing, these two terms will be equalized if MPL^E is concave in the number of positions. This result will helpful for identifying hold-up power $\mathbf{h}$ when we turn to estimation in Section 4.

Complementarities without Position Specificity - If output exhibits production complementarities ($m > 0$), but skills are not specific to a position ($d = 0$), then workers will not have individual hold-up power over the rest of the firm's output. In this scenario, a firm can swap in workers as soon as workers leave and so never faces multiplicative losses. Firms may even over-hire if positions require on-the-job learning but workers are replaceable across positions ex-post. An example may be a catering firm that requires twenty staff to be present to staff an event, and the services are dysfunction if the team is understaffed. The firm may train more than twenty workers and compensate workers to be on call, so that if a workers suddenly leaves, the firm never falls below the critical threshold of twenty. While workers in this setting may have significant collective hold-up power over firm output, they have little hold-up power individually.

Position Specificity without Complementarities - If workers have position specific skills, but output is not individually complementary, then workers will have no hold-up power over the rest of the firm's output. Some production environments may include highly differentiated labor, but ultimately the outputs for the production unit are not interdependent. For example, consider a newspaper that loses an international reporter. This reporter's work may be highly specialized, but it is unlikely to affect the work of the newspaper's opinion columnists.

Sources of Firm Productivity Heterogeneity - We generate heterogeneity in firm productivity from differences in the concavity of the output with respect to the number of positions α. It is worth considering what kind of production environments would be characterized by a firm with steep concavity (low α) where workers may also high hold-up power (high $\mathbf{h}$). Consider a production arrangement where the firm is better off with a small team of workers who are unlikely to turn over, and where the firm benefits from limiting the number failure points in the production process. This may also be a situation where tasks are sufficiently complex that communication costs are high, and integrating additional positions into the production process generates more failure points with little additional productivity. Examples may include a limited number of engineers on a particular design project, or a limited number of investment bank team members working on an acquisition. In Appendix A.13 we also show that high average productivity can also result from high markups due low to product demand elasticities, and profitable firms share rents with high workers in high hold-up occupations.

2.3 Bridging the Model to Data

In Sections 2.1 and 2.2, we developed a stylized model to derive predictions relating individual hold-up power $\mathbf{h}$ to wages. In order take the model to the data, we must make two modifications: (i) properly account for turnover costs that are not due to production complementarities (i.e., training costs, search costs, etc.), and (ii) generalize to allow firms to produce using multiple occupations, each with their own degree of hold-up power.

"Marginal" Turnover Costs What differentiates our model is that position specific skills allow workers to disrupt firm inframarginal productivity. However, not all turnover costs originate from multiplicative output losses resulting from complementarities and specificity. In many models, including those with search costs, training costs, or firm specific skills, turnover costs are paid in wages or affect output at the margin.

[14] This also implies that if there are production complementarities, i.e., $m > 0$, the wage will be less than the intensive marginal product MPL^I. We discuss the intuition of wage markdowns from intensive marginal product in detail in Appendix A.6.

We will show that these marginal or "non-multiplicative" turnover costs increase the *level* of the wage, but only individual hold-up power increases the *slope* of wages to average productivity.

Define δ as the per-period cost of unfilled positions as a fraction of per-period wages of trained workers. Then the firm's output is:

$$Y = AN^{\alpha}e^{-(\delta\alpha+\mathbf{h})(1-X)}.$$

In the case of $\delta = 0$, all turnover costs come from multiplicative output losses. As δ increases above 0, the non-multiplicative turnover costs increase. For example, the model in the previous section is a special case with $\delta = 1$, resulting from the firms paying full wages to the new workers while new workers are unproductive in non-complementary tasks. Allowing for a flexible value of δ in the firm's problem from the prior section, the optimal wage is:

$$w^* = \left(\gamma\Big(\delta + \frac{\mathbf{h}}{\alpha}\Big)\right)^{\frac{1}{\gamma}} \underline{w},$$

The elasticity of wages to average product when varying α is then:

$$\varepsilon_{w,APL} = \frac{\mathbf{h}/\delta}{\gamma\alpha + (\gamma+1)\mathbf{h}/\delta}, \tag{7}$$

where the elasticity depends on $\mathbf{h}/\delta$. In practice, the passthrough elasticity will depend on what share of turnover costs come from disruptions to inframarginal product versus non-multiplicative "marginal" turnover costs. Because marginal turnover costs raise the level of the wage, the increase in wages that a worker gets from a lower α (and hence higher average product) when $h > 0$ is proportionally less.

Now that we have fully defined the replacement productivity of untrained workers, we can define the flow surplus to the firm of a match with a fully trained, specific worker, net of the productivity of an untrained worker in occupation j:

$$\begin{aligned} Surplus_j^{firm} = \Big(\frac{\partial Y}{\partial X}\frac{1}{N}\Big)^{net} &= \delta_j MPL_j^E + \mathbf{h}_j APL \\ &= w_j^*\big(\delta_j + \frac{\mathbf{h}_j}{\alpha}\big). \end{aligned} \tag{8}$$

Thus when firms are optimizing, the flow surplus of a match to the firm will be equal to the (optimal) wage times the sum of δ and $\frac{\mathbf{h}}{\alpha}$, which respectively correspond to the non-multiplicative or "marginal" turnover costs and the inframarginal turnover costs from individual hold-up power. This formulation will be useful when we empirically identify $\mathbf{h}$ using exogenous worker separations in Section 4.

Given the expressions for the level of the wage and the firm's surplus in the match, positions with high turnover costs $\delta + \frac{\mathbf{h}}{\alpha}$ will mechanically have high firm surplus and will subsequently have high wages. However, it's also possible that occupations with high turnover costs have also have better outside options $\underline{w}$. To the extent that turnover costs and outside options are correlated can provide a microfoundation for the framework in Lachowska et al. (2021), where high-skill jobs have higher match surplus in the firm. In Appendix A.8, we show that our model can rationalize the result in both Lachowska et al. (2021) and Caldwell and Harmon (2019), that high skill workers can extract more from employers in response to a change in individual outside options.[15]

[15] While higher levels of firm surplus may encourage firms and workers to bargain rather than just have the firm set wages, we show in Appendix A.8 that this doesn't not simply map into different level of bargaining power in a Nash bargaining framework. This is because high hold-up occupations have both higher passthrough of firm productivity to wages *and* higher passthrough of idiosyncratic outside options to wages.

The total match surplus to the firm will be the present discounted value of flow surpluses, which would require extending the model into a dynamic setting. In Appendix A.3, we derive a dynamic version of the model in which untrained workers become fully specific, productive workers with probability p each period, with an expected training time of $1/p$.[16] The total surplus of the match will then depend on the speed of learning p and the rate at which firms discount.

Firms with Multiple Occupations Previously we assumed that each firm utilizes positions of only one type of occupation. In this section, we demonstrate that this simplification is not restrictive, and that our mapping of hold-up power $\mathbf{h}_j = d_j m_j$ to wages is robust to firms with multiple occupations.

Consider a firm that produces using positions of two different occupations a and b, where the number of positions are N_a and N_b, respectively. The outer production structure is Cobb-Douglas, where σ_a and σ_b are the elasticities of output (within the parentheses) with respect to the quantity of positions a and b, and α is the same overall concavity. The firm maximizes revenue minus wages:

$$\max_{N_a, N_b, w_a, w_b} = \left(N_a^{\sigma_a} N_b^{\sigma_b}\right)^{\alpha} e^{-\sum_j \sigma_j(\delta_j \alpha + d_j m_j)(1-X_j)}, \quad j \in \{a, b\}$$

subject to:

$$X_j = F(w_j).$$

If the outside offer distribution for each occupation j is similar to before with the form $F(w_j') = 1 - \left(\frac{w_j'}{\underline{w}_j}\right)^{-\gamma}$, for $w_j \geq \underline{w}_j$, then it is simple to show that we end up with the same wage expression:

$$w_j^* = \left(\gamma\left(\delta_j + \frac{d_j m_j}{\alpha}\right)\right)^{\frac{1}{\gamma}} \underline{w}_j, \quad j \in \{a, b\}.$$

We show in Appendix A.9 that the elasticity of wages to measured average product $\varepsilon_{w_j, apl}$ is nearly identical to the expression in the single-occupation case.

2.4 Comparison with Alternative Models

Our model of individual worker hold-up power differs from other common models of imperfect labor market competition and firm wage premia.

Upward Sloping Labor Supply/Convex Hiring Costs - A common model applied to questions of firm premia is the upward sloping labor supply as in Card et al. (2018) and Lamadon et al. (1994) that use upward sloping labor supplies to generate firm wage premia. In these models, firms must pay higher wages in order to be large, dipping deeper into a pool of workers that are less willing or more difficult to be employed. The main distinguishing features of individual hold-up power is that the incentive for a firm to pay high wages need not be a function of firm size. A related model to the upward sloping labor supply is convex adjustment costs, where in response to shocks, firms increase wages more for types of labor that have more steeply convex adjustment costs, as in Kline et al. (2019). However, as argued by Kuhn (2004), if the convex costs are in percent terms relative to firm size, rather than in absolute terms, then wages should equilibrate after firms have grown to their desired size. Therefore, individual worker hold-up power can match the patterns of productivity wage premia that (i) do not depend on firm size, and (ii) exist both in response to shocks and in steady state, which neither the upwards sloping labor supply nor convex adjustment cost models can match. Further, upward sloping labor supply models predict a tight relationship between wage markdowns and firm size-wage elasticities, which we document to be empirically in conflict in Appendix A.7.

[16] We show that that the optimal wages is $w^* \approx \left(\frac{\gamma}{p}\left(\delta + \frac{\mathbf{h}}{\alpha}\right)\right)^{\frac{1}{\gamma}} \underline{w}$

Stole and Zwiebel (1996) - Another common model originates from Stole and Zwiebel (1996), in which multi-worker firms bargain bilaterally with each worker. In this setting, firm average productivity affects the bargained wage: the firm considers that if negotiations break down, the firm will have to bargain with a subsequent worker who has greater marginal product. Following this logic, the inframarginal productivity of hypothetical workers at deeper levels of bargaining affect the bargaining *position* of the firm and thus affects what a worker at the first layer of bargaining can extract, given a worker's bargaining power β. However, to the extent that there is heterogeneity in bargaining power among types of workers, and therefore the ability to extract higher wages from more productive firms, is an assumption in a Stole and Zwiebel (1996)-type setting. In our model of complementarities and position specific skills, the firm's outside option differs across types of workers due to heterogeneous ability to disrupt production across positions. With wage posting and imperfect contracts, heterogeneous wage premia at productive firms is a result. In a similar way that wage posting and incomplete contracts generates isomorphic outcomes as bargaining (i.e. Manning 2011), we generate results that look like heterogeneous bargaining power but with firm wage setting and incomplete contracts.

Search and Matching - A third common model is the standard Diamond-Mortensen-Pissarides search model, in which firms pay a vacancy posting cost, and after meeting, firms and workers bargain over the surplus of the match. Combined with a free entry condition, the surplus of a match is pinned down by the cost of a vacancy and the per-period job filling rate for the firm. Therefore in this benchmark model, the most that a worker can hold up is the expected flow of vacancy of costs. With complementarities and position specificity, workers can hold up a portion of firm output in addition to the expected cost of a vacancy, as a worker's sudden departure disrupts the output of other factors in the firm.

Firm Specific Skills - When skills are only firm specific, workers are internally substitutable. As we argued in Section 2.2, if workers are substitutable, firms are never at risk of losing inframarginal product from individual worker turnover. As such, wage setting decisions take into account only the productivity of labor at the margin, and therefore inframarginal productivity is irrelevant for wage setting, resulting in no wage premium at productive firms. However, as we showed in Section 2.3, non-multiplicative turnover costs do raise the level of the wage. In total, firm specific skills raise the level of an occupation's wage but do not generate a slope of wages with respect to productivity.

Rent Sharing Models - Many studies such as Card et al. (2016) put forward a simple rent sharing model of the form $w_{ikt} = w_0 + \gamma_j S_{kt}$, where S_{kt} is the surplus per worker at firm k and w_0 is an outside competitive wage. The rent sharing parameter γ_j may differ by occupation j (or by gender). Our study argues instead that the variation across occupations will come from variation in the surplus per worker S due to heterogeneous complementarities and position specificity, instead of differences in the rate at which surplus is shared γ. For example, low-wage service workers receive low passthrough not because their bargaining power is low, but because there is little surplus in the match to begin with.

Efficiency Wage Models - Our model of individual worker hold-up power is most closely related to efficiency wage models. In efficiency wage models, firms get more "effective" labor from the same number of workers by paying a higher wage. We achieve a similar result where firms have a higher share of trained workers by paying higher wages $X'(w) > 0$, where the mechanism is turnover and incomplete contracts. A model in which effort were increasing in the wage and interacts multiplicatively with average product would be isomorphic. The contribution of our paper to the efficiency wage literature is to isolate a particular set of conditions (i.e., individual complementarities and position specificity) where efficiency wage motivations (i) differ across occupations and (ii) interact with firm productivity heterogeneously across occupations.

Assignment Models and Sorting - In assignment models such as Tervio (2008) and Gabaix and Landier (2008), two main results emerge: (1) better quality workers are assigned to positions where the stakes are largest, and (2) small differences in worker quality can generate large differences in wages. While we abstract from ex-ante worker heterogeneity in this paper, our generalization of O-ring production provides a framework for extending assignment models to settings of multi-worker firms.

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 2017, and job posting data in the US from Burning Glass Technologies from 2010 to 2018.

3.1 Danish Administrative Data

We use administrative from Denmark on workers and firms from 2008-2017 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.[17] Beginning in 2008, the reported wage data 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 as IDAN but on a monthly level. We use the demographic registries to attain 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 that includes 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 with variable *cvrnr* rather than 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 two separate (though not mutually exclusive) sets of Danish firm and worker data. In the section estimating the effect of worker deaths, we use data from 1,218 firms who experience a worker death, all of which have between 15 and 100 full-time equivalent employees, as well as an equal number of matched placebo firms. The summary statistics for these firms can be found in first three rows of Table 1. The firms in the sample on average employ 45.9 full-time equivalent workers and produce approximate $4.3 million in value added.

In our passthrough exercises, we use a much larger sample of workers and firms, including all full-time workers employed at a private sector firm that reports value added and industry, whose firm employed an average of at least 15 full-time equivalent employees. 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$\times$year observations. There are approximately 1 million worker$\times$ firm observations per year, out of a labor force of about 2.5 million.[18]

Grouping Occupations Danish occupations are categorized using the Danish International Standard Classification of Occupations, or DISCO codes, which has ten categories. We exclude analysis on agricultural and armed forces occupations. Summary statistics on workers' wages, average tenure, differentiation score (which we will explain in Section 4.2), average years of schooling, and share of employment by DISCO group in the 2008-2016 passthrough sample are shown in Table 2.

[17]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.

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

Table 1: Firm Summary Statistics

	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		

	Passthrough 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 passthrough analyses. In the deaths analysis, we restrict our attention to firms with an average number of full time equivalent employees between 15 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 passthrough sample, we use larger firms, restricting attention to firms with an average of at least 15 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.

Throughout the paper, we divide workers into four broad occupation groups: managers (DISCO group 1), professionals/technicians (groups 2 and 3), crafts/assembly (groups 7 and 8) , and administrative/service/manual (groups 4, 5, and 9). 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. However, we treat these groups separately 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 policies, characterized by a generous welfare state and high union density, but with limited employment protections, high labor mobility, and industry-set minimum wages. The level of labor mobility, measured by the rate of hiring and separations, is high for OECD countries and is more similar to the US than other economies in continental Europe (Caldwell and Harmon (2019)). There is no national minimum wage, but minimum wages are set at the industry level through collective bargaining agreements between labor unions and employer associations. In recent decades, bargaining at the firm level has become more important in determining wages (Dahl et al. (2013)), and firms have considerable ability to raise individual wages above collectively agreed minimums.

3.2 Burning Glass Technologies (BGT) Job Posting Data

We use the vacancy data from Burning Glass Technologies, which has 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 BGT cleans and categorizes into broad skill categories. We will discuss merging the 6-digit SOC codes used by Burning Glass to Danish ISCO in Section 4.2.

Table 2: Mean Characteristics of 1-Digit Occupation Groups

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

This table reports summary statistics for the 1-digit occupation groups in Danish data, using the sample specified in Section 3. The data is from 2008-2016 including firms with on average at least 15 employees that also report value added. Values at the top of each cell are the mean, and values in parentheses report standard deviations.

4 Measuring Individual Hold-up Power

In this section, we measure individual hold-up power by occupation using two methods. First, we estimate the effect of worker deaths of firm profits, measured as value added less wage and salary payments, across occupations and across firms of different levels of productivity. This provides estimates of total turnover costs and hold-up power for four broad occupation groups. Second, we estimate a measure of within-firm, across-position task differentiation for each occupation from US job postings data and merge to Danish occupations, providing us a proxy of hold-up power at the detailed occupation level.

4.1 Identification of Parameters with Exogenous Worker Separations

In this section, we will show that we can identify the degree of hold-up power **h** if we observe exogenous separations of workers from firms by observing changes in firm-level value added less salaries. We are empirically estimating the surplus to the firm derived in Section 2.3. We put forward the following lemma:

Lemma 1 *If workers are exogenously separated and firms do not adjust the number of positions, then the per period total turnover costs* $\delta + \frac{h}{\alpha}$ *is identified by the change in reported value added following a worker separation:*

$$\Delta Y \approx \big(\Delta X \times N \times w_{\text{-}1}\big) \times \big(\delta + \frac{\boldsymbol{h}}{\alpha}\big)$$

If workers are exogenously separated and firms can adjust the number of positions N *holding the effective share of positions* X *fixed after separation has occurred, then total per period turnover costs* $\delta+\frac{\boldsymbol{h}}{\alpha}$ *are identified by the change in value added less wage and salary payments following an exogenous worker separation:*

$$\Delta(Y - wN) \approx \big(\Delta X \times N \times w_{\text{-}1}\big) \times \big(\delta + \frac{\boldsymbol{h}}{\alpha}\big). \tag{9}$$

***h** and δ are separately identified with an interaction term on firm productivity:*

$$\Delta(Y - wN) \approx \underbrace{(\Delta X \times N \times w_{\text{-}1})}_{\textit{Shock to firm}} \times \delta + \underbrace{(\Delta X \times N \times w_{\text{-}1} \times \frac{1}{\alpha})}_{\textit{Interaction with firm productivity}} \times \boldsymbol{h}.$$

Proof: see Appendix A.15.

It is worth noting that value added less salaries, rather than value added, is the outcome of interest. The main reason is that firms may downsize following the death of a worker, and the incentive to downsize may differ based on the type of worker. Since a lower share of positions filled X decreases the productivity of other workers in the firm, deaths of workers with higher hold-up power will cause larger declines in the productivity of the remaining workers. Therefore, we would expect that firms that experience a death of a particularly high hold-up worker will shrink the number of positions by more. With that said, firms are maximizing output net of costs, and if firms shrink their size, their wage costs fall as well.[19] Therefore, value added less salaries is the correct measure to relate to the firm's maximization problem and also addresses bias introduced if the outcome were value added due to differential incentive to downsize following a worker death.

The intuition for the third statement in Lemma 1 is that identifying **h** comes from the feature that for the same hold-up power **h**, output losses are larger at higher productivity firms. Using the exogenous separations of workers in the same occupation but across firms of different productivity, we can uncover the rate at which higher firm productivity generates higher output losses, thereby identifying **h**.

Sample Our sample uses deaths between 2008 and 2015 and firm outcomes from 2005 to 2017. We restrict our sample to deaths of non-owner workers[20] who were employed at the firm in the month prior to the death and who worked at least 100 hours in the prior month.[21] We select firms with 15 to 100 full time equivalent employees to focus on firms in which the effect of a death will be sufficiently large relative to the typical variance of output and profit. Due to our relatively short sample, the event window is six periods, denoted $s \in \{-3, 2\}$. The death occurs in period $s = 0$. Firms k are matched to five placebo firms on the exact year and 1-digit industry in period s = -1. The five selected control firms have the least sum of squares of differences in log firm size (measured in full time equivalent workers) and value added less salaries in periods s = -1 and s = -3. We average the outcomes of the control firms using equal weights. Treated firms are excluded from the pool of control firms.

Empirical Specification To estimate equation (9), we will run regressions of the form:

$$\tilde{Y}^j_{ks} = c^j + \psi^j_k + \sum_{s=-3}^{2} \beta^j_s \times \mathbb{1}\{s\} \times \tilde{w}^j_{k(i),\text{-}1} + \varepsilon_{ks}, \tag{10}$$

separately for each occupation j and pooling across occupations. The regression in equation (10) recovers a composite coefficient for total turnover costs: $(\beta^j_0 + \beta^j_1 + \beta^j_2)/3 = -(\delta + \frac{\mathbf{h}}{\alpha})$, where the units on the coefficients are the average annual flow loss in years of prior salary for three years. The outcome variables of interest Y is value added less salaries. We construct a within-match k, within-period s difference $\tilde{Y}_{ks} = (Y^{treated}_{ks} - Y^{placebo}_{ks})/N_{k,\text{-}1}$, where $Y^{placebo}_{ks}$ is the average outcome of the placebo firms that are matched to

[19] Value added less salaries then provides a measure of accounting profit which does not account for payments to capital.

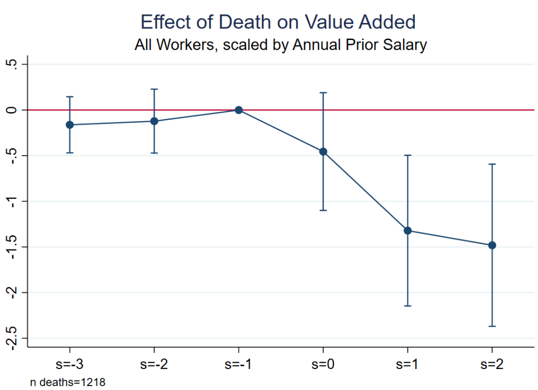
[20] Individuals in this sample are very unlikely to be employers: our deaths sample is dervied from the BFL database, where very few workers have a value of "A" for the variable *type*, indicating that the individual is the owner of a firm.

[21] 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. We show how the results in this sample differ in Appendix B.2.

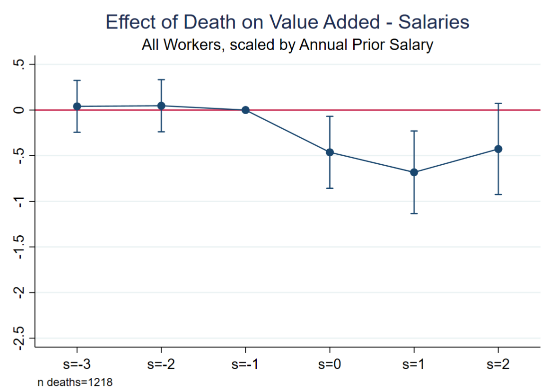
firm k, and $N_{k,\text{-}1}$ is average number of workers in the three years prior to the shock. $w_{k(i),\text{-}1}$ is the prior annual wages of the the deceased worker i at firm k, and $\tilde{w}_{k(i),\text{-}1} = w_{k(i),\text{-}1}/N_k$ scales the size of the shock to the firm by the prior size fo the firm. Monetary values are deflated into 2005 Danish kroner and converted to dollars using 6 DKK/USD. We winsorize the value added less salaries differences $\tilde{Y}$ at the 10th and 90th percentiles separately for pre- and post-treatment observations. Standard errors are clustered at the match level, where each cluster has six observations of a treatment-placebo matched pair. Both the outcome variable and the wage variable are annual, allowing us to interpret coefficients as losses to value added less salaries in years of average prior wages.

Figure 2: Effect of a Worker Death on Firm Value Added and Value Added less Salaries

(a) Value Added



(b) Value Added - Salaries



These figures report the event study for the specification pooling all occupations. Both wages and the outcome variables are annual, so the coefficients can be interpreted as flow losses per year in years of prior wages. Panel (a) reports the estimates when the outcome variable is value added. Panel (b) reports the estimates when the outcome variable is value added less salaries. Standard error bars report 95% confidence intervals.

Results Figure 2 reports the results of equation (10), pooling all occupation groups, and using both value added and value added less salaries as the outcome variable. On average over three years, firms that experience a worker death see a cumulative decline in value added and value added less salaries equivalent to 3.4 and 1.7 years of workers' wages, respectively. The results differ dramatically when the outcome variable is value added or value added less salaries, reflecting that some firms shrink considerably after a worker death.[22]

Figure 3a reports the estimates of value added less salaries following a death for each occupation group, measured in years of prior wages per year, averaging the coefficients from the the treatment years $\delta_j + \frac{\mathbf{h}_j}{\alpha} = -(\beta_0^j + \beta_1^j + \beta_2^j)/3$ using the event study specification in equation (10). The coefficients reflect losses in years of prior wages per year. For example, the point estimate on managers is -.93, indicating that the firm's losses of value added less salaries are equal to .93 years of the managers' prior wages in each year, implying a total cumulative loss nearly 3 years worth of prior wages.

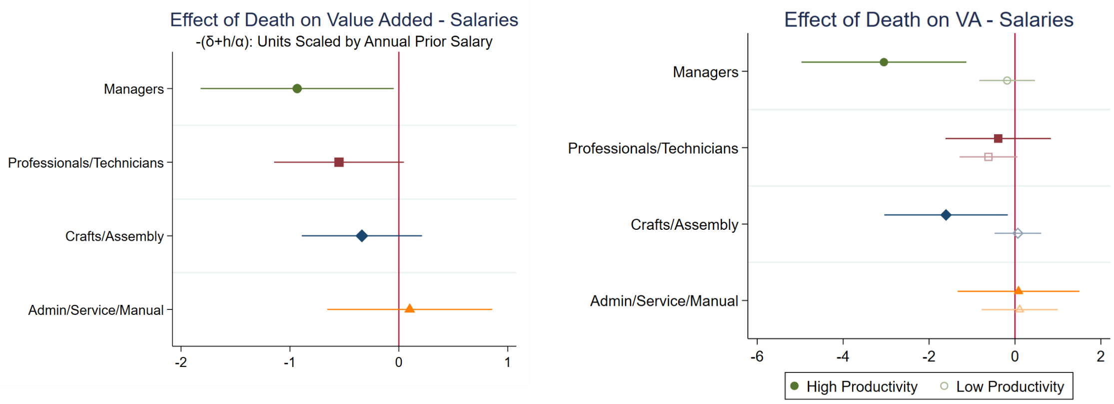
In Figure 3b, we estimate the effect of deaths at high and low productivity firms. For each occupation, high productivity firms have roughly double the output per worker compared to low productivity firms. We

[22]The gap between the change value added and the change value added and less salaries is largest when deceased worker was a manager or professional, suggesting that the incentive for firms to downsize is indeed larger for occupations with higher complementarities or hold-up power.

Figure 3: Effect of Death on Firm Value Added net of Wages and Salaries, Years of Prior Wages per Year

(a) by Occupation

(b) by Occupation and Productivity



This figure shows the value added losses in years of prior salary per year. Panel (a) reports the the number year of value added salaries lost, in units of the average prior wage of workers, averaged across 3 years from period $s = 0$ to $s = 2$: $(\beta_0^j + \beta_1^j + \beta_2^j)/3$. A point estimate of -.55, for example for crafts/assembly, indicates that the estimated per-year losses in value added less salaries is equal to .55 years of average prior pay of workers in that occupation, implying a cumulative profit loss of over a year and a half worth of prior wages. These estimates come from 118 deaths of workers classified as managers; 315 deaths of workers classified as professionals/technicians, 452 deaths of workers classified as crafts/assembly, and 333 deaths of workers classified as administrative/service/manual. Panel (b) reports losses of value added less salaries, in years of prior wages, at high and low productivity firms for each occupation group, where firms are split into two evenly size groups. Standard errors report 95% confidence intervals.

find that deaths at productive firms generate significantly higher profit losses when the deceased worker was in a managerial or crafts/assembly occupation, and the magnitudes of profit losses at these high productivity firms are large: in the case of a manager death, the point estimate is nearly 3 years worth of prior salary per year.

Next we estimate the interaction regression specified in equation (11) to separately estimate regular turnover costs δ from hold-up power $\mathbf{h}$. To simplify the interpretation, we will modify the notation. On the right hand side, we pool the post-treatment periods into single indicators. We include an interaction term that multiplies the size of the treatment $\tilde{w}^j_{k(i),\text{-}1}$ by the inverse of the firm's labor share $\hat{\alpha}^{-1}_{k,\text{-}1}$ in the prior periods, where regression coefficient $\mathbf{h}^{est}_j$ directly corresponds to our estimate of the model parameter $\mathbf{h}_j$:

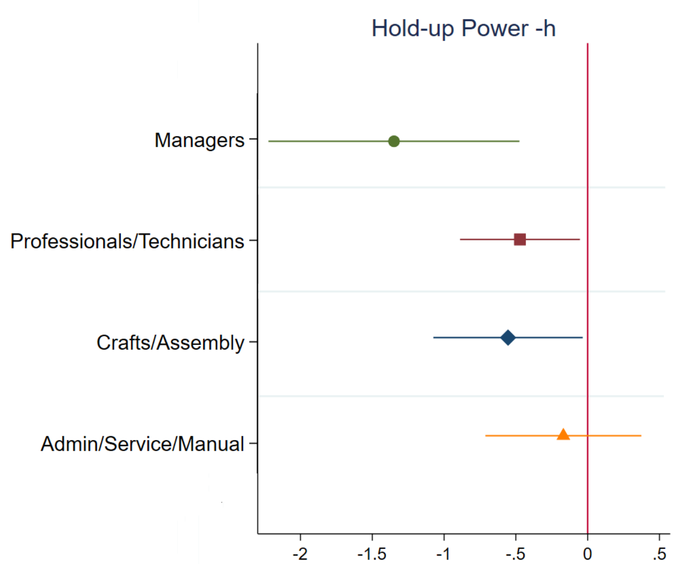
$$\tilde{Y}^j_{ks} = c^j + \psi^j_k + \sum_{s=-3}^{-1} \beta^j_s \times \mathbb{1}\{s\} \times \tilde{w}^j_{k(i),\text{-}1} - \delta^{est}_j \times \underbrace{\mathbb{1}\{s \geq 0\} \times \tilde{w}^j_{k(i),\text{-}1}}_{\text{baseline}} - \mathbf{h}^{est}_j \times \underbrace{\mathbb{1}\{s \geq 0\} \times \tilde{w}^j_{k(i),\text{-}1} \times \hat{\alpha}^{-1}_{k,\text{-}1}}_{\text{interaction}} + \varepsilon_{ks}, \tag{11}$$

The pooled term on the right hand side that is not interacted with the inverse of the firm's labor share has a regression coefficient that directly corresponds to δ_j. This can be interpreted as a kind of intercept: how large would profit losses be a "zero" productivity firm. In practice in the model, labor productivity measured in value added per wage bill is bounded below by 1: as firm the firm's production converges to being linear in labor, the wage bill share of value added converges to 1, and the wage is equal to average product.

To estimate this equation, we use a non-linear regression to restrict $\delta^{est}_j > 0$. We show in Appendix B.1 that when we run the standard linear regression with OLS, we estimate negative values of δ_j for each occupation. This is because the difference in estimated profit losses between high and low productivity firms is larger than the theory would predict given restrictions on the model parameters, particularly for managers and crafts/assembly.

Figure 4 shows the results of the non-linear estimation. We estimate the largest hold-up power for managers with $\mathbf{h} > 1$. Hold-up power in low-wage administrative/service/manual jobs is small and not significantly different from zero. The middle wage occupations show an intermediate level of hold-up power $\mathbf{h}$ around 0.5.

Figure 4: Estimate of Hold-up Power $\mathbf{h}$ by Occupation



The figure reports the estimates of hold-up power $\mathbf{h}_j$ by occupation using the non-linear estimation of equation (11). Managers are estimated to have the highest hold-up power, and administrative/service/manual are estimated to have nearly none. Estimates of regular turnover costs δ_j are approximately zero for each occupation.

External Validity We will be using the estimates from this section to predict passthrough of firm productivity shocks to wages. One concern is that the individual hold-up power of workers in a given occupation at a small firm may not be externally valid when applied to larger firms. When firms are larger, there may be more workers who can substitute in for any given co-worker, and larger firms may develop promotion structures that train and elevate particular workers when higher level positions become open. We partially address this by using firms up to 100 workers in our analysis, which is larger than is standard for the literature (Bertheau et al. (2021), Jäger (2016)).[23]

A further concern is that deaths may be a more severe shock to the firm than the typical shock that determines firm wage setting, namely that a worker is poached by another firm. In this case, we would be overestimating the degree to which turnover costs affect wage setting decisions.

[23] A further theoretical extension that could justify similar levels of hold-up even at large firms could be that even in large firms, workers are organized into teams, where individual workers' output is likely complementary with that of other team members.

4.2 Task Differentiation in US Job Postings

To construct an estimate of worker hold-up power for more detailed occupations, we construct a "differentiation score" for each occupation group using online job postings from Burning Glass Technologies (henceforth BGT). The measure is designed to capture how different a typical job posting's skills requirements are from the skill requirements of other jobs *within the firm.* Recall the two necessary conditions for individual hold-up power: (1) the output of positions are complementary, and (2) skills are position specific. We argue that within-firm task dissimilarity is a good proxy for both of these conditions. First, the output of positions is less likely to be complementary if their tasks are homogenous, as task homogeneity tends to imply output substitutability. Second, both incumbent co-workers and outside hires are less likely to be able to replace tasks if the required combination of tasks of a position is uncommon.

To construct the differentiation score, we apply a clustering algorithm that allocates job postings for a given establishment to clusters based on the set of skills listed in the job posting. The algorithm is called ROCK (Guha et al. (2000)), or "a RObust Clustering using linKs", and is used for clustering categorical data.[24] In the job posting data, BGT cleans and categorizes the skills found in job postings into skill groups and skill group families.[25] The intermediate level, skill groups, is the level of categorization that we use in the ROCK clustering algorithm. Examples of how skills are categorized are provided in Appendix Figure B.4; we use the middle column of skills used, of which there are over 700 in total. An example allocation of job postings to clusters can be seen in Appendix Figure B.5.

To compute the differentiation score by detailed (6-digit soc) occupation, we first compute two statistics: (i) what share of job postings are successfully put into a cluster, and (ii) conditional on being in a cluster larger than one, the log of the average size of that cluster. Taking the first principal component of these two metrics gives us our differentiation score. Both components of this score are intended to measure how often and to what degree a given job has a very different set of tasks from other jobs in the firm. The two measure are moderately correlated (ρ=0.42), and the first principle component removes noise in each of the two measures.

Summary statistics for the differentiation score by occupation, aggregated to the 1-digit DISCO occupation level, can be found in column 3 of Table 2. We compute these by merging US soc occupations to the DISCO codes in the Danish administrative data,[26] and then taking a weighted average of these scores within broad DISCO occupation groups. Managerial and professional jobs are the most differentiated, followed by technicians. Crafts and assembly occupations have moderately high levels of differentiation relative to their low levels of average education. Service and manual jobs have the lowest differentiation scores. Some examples of differentiation scores at the detailed occupation level include: "Computer and Information Systems Managers" 1.67, "Pharmacists" -0.27, "Construction Equipment Operators" .49, "Cooks, fast food" -2.21.

There are two concerns regarding the representativeness of job postings data to measure the within-firm task differentiation of employed workers. First, online job posting data tends to overrepresent white collar and STEM jobs that require a bachelor's degree (Carnevale et al. (2014)). Overrepresentation of these high-wage occupations in the job postings would tend to bias down the differentiation scores of these occupations, as there should be a greater number of similar postings to be put into these jobs' clusters. The second concern is that low-wage jobs will tend to have a higher turnover, meaning that the actual number of employed workers who are high-wage will be higher given the number of job postings. This will tend to offset the first bias, and the overall direction of the bias in ambiguous.

[24] A common application is to cluster transaction data in which sets of goods are often purchased together.

[25] BGT refers to these groups as "skill clusters" and "skill cluster family", but we will use the term "skill group" to avoid using the word cluster with two different meanings.

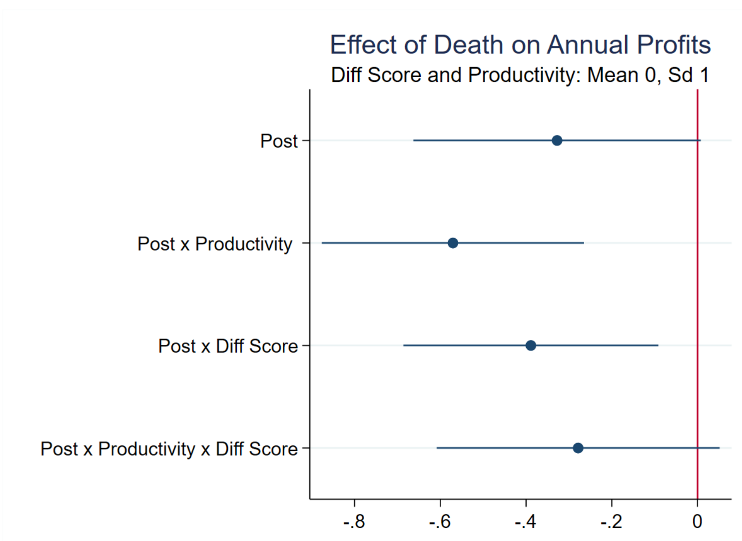
[26] A special thanks to Eskil Heinesen to providing the crosswalk of US SOC occupations to Danish ISCO codes.

Death Estimates by Task Differentiation Next we will combine our measures of hold-up power by estimating the extent to which high task differentiation occupations generate larger profit losses after separations and if task differentiation further predicts higher profit losses at productive firms. Equation (12) includes a pooled post-treatment term, as well as interactions on the inverse of firm k's labor share, the differentiation score of the occupation j, and the double interaction including these last two terms:

$$\begin{aligned}\tilde{Y}_{ks} = \alpha + \gamma_k + \sum_{s=-3}^{-1} \beta_s \mathbb{1}(period_s) \times \tilde{w}_{k(i),\text{-}1} \\ + \beta_{post} \underbrace{\mathbb{1}\{post\} \times \tilde{w}_{k(i),\text{-}1}}_{\text{baseline}} + \beta_{prod} \underbrace{\mathbb{1}\{post\} \times \tilde{w}_{k(i),\text{-}1} \times labsh_k^{-1}}_{\text{productivity}} \\ + \beta_{diff} \underbrace{\mathbb{1}\{post\} \times \tilde{w}_{k(i),\text{-}1} \times \text{diff}_j}_{\text{Diff Score}} + \beta_{int} \underbrace{\mathbb{1}\{post\} \times \tilde{w}_{k(i),\text{-}1} \times \text{diff}_j \times labsh_k^{-1}}_{\text{interaction}} + \epsilon_{ks}. \qquad (12)\end{aligned}$$

Figure 5 shows the results of a pooled regression containing all occupations. Both firm productivity and the task differentiation score are normalized to have a mean of 0 and standard deviation of 1. "Post" indicates that the year observed is in the treated periods. The "Post × Productivity" coefficient indicates that in general, profit losses are larger when a death occurs at a higher productivity firm. The "Post × Diff score" coefficient indicates that deaths of workers in occupations with high task differentiation scores generate larger profit losses relative to prior wages, and the "Post × Productivity × Diff score" coefficient indicates that the slope of profit losses with respect to firm productivity is increasing in task differentiation. This final coefficient is the strictest test that our task differentiation measure is proxying for hold-up power, and therefore the task differentiation measure should be predictive of passthrough.

Figure 5: Effect of Death On Profit with Task Differentiation Interaction



This figure shows the effect of a one standard deviation increase in productivity, one standard deviation effect of task differentiation, and their interaction on firm profits following a worker death. Task differentiation increases the cost of a death relative to prior wages significantly, such that an occupation with a differentiation score one standard deviation will generate approximately zero effect on firm profits, while an occupation with task differentiation one standard deviation above average will have yearly profit losses near -0.7 years of prior wages. While the interaction term on productivity and task differentiation is not significant at the 5% level, this figure indicates that the higher profit losses at productive firms generated by deaths of workers in high differentiation score occupations is economically significant.

5 Productivity-Wage Elasticity Estimation

In this section, we estimate the elasticity of wages to firm average product. We already derived a theoretical formula for this elasticity as a function of model parameters:

$$\varepsilon^{j}_{w,apl} = \frac{\mathbf{h}_j/\delta_j}{\gamma\alpha + (1+\gamma)\mathbf{h}_j/\delta_j}. \tag{13}$$

If this model is the underlying data generating process, we should be able to infer estimates of hold-up power $\mathbf{h}$ from the elasticity of wages to firm average product, as high $\mathbf{h}$ occupations will exhibit higher passthrough elasticities. We estimate passthrough for four broad occupation groups as well as estimate the effect of the task differentiation score on passthrough. Lastly, we estimate the effect of firm size on wages using switchers, which will be important for differentiating our model from other common models.

Measuring Firm Productivity for Passthrough Estimation In our model, we are interested in labor productivity: output per hour worked. However, a simple measure of value added per worker would include variation simply based on the occupational composition of the firm: firms with high wage workers should produce more output per worker, or else the firm would not be profitable. We account for the occupational composition of firms by constructing a residualized value added per worker measure. In Appendix A.9, we show that residualizing for occupational composition is the correct empirical measure if changes in observed average product are due to changes in underlying curvature α or shifts in occupational composition. As a first step, we compute the predicted wage bill based on the occupational composition of hours worked in the firm:

$$\widehat{\text{wage bill}}_{kt} = \sum_j \text{hours}_{jkt}\bar{w}_{jt},$$

where hours_{jkt} is the total number of hours reported worked by workers in occupation j at firm k in year t, and $\bar{w}_{jt}$ 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 log value added, given the predicted wage bill in the firm and controls:

$$\widehat{\log(VA)}_{kt} = \xi_0 + \xi_1 \log(\widehat{\text{wage bill}})_{kt} + \xi_2 Z_{kt},$$

where Z_{kt} are controls including industry and year fixed effects, as well as the log number of hours worked by workers without an assigned occupation. Our residualized productivity measure is then:

$$\hat{Y}_{kt} = \log(VA)_{kt} - \widehat{\log(VA)}_{kt},$$

where $\log(VA)_{kt}$ is the observed log value added reported by firm k in year t.

5.1 Passthrough Estimates

Sample We include firms that report value during the years 2008-2016 and have an average of at least 15 full-time equivalent employees. Summary statistics on the firms can be found in Section 3.1.

Specification and Results In this section, we estimate the passthrough of firm productivity changes to wages for stayers. In our main specification, we use a similar strategy as Card et al. (2018), Maibom and Vejlin (2021), and Juhn et al. (2018), regressing three year changes in wages on three year changes in our firm productivity measure $\hat{Y}_k$, using seven year changes in our valued added worker residual $\hat{Y}_k$ as instruments.

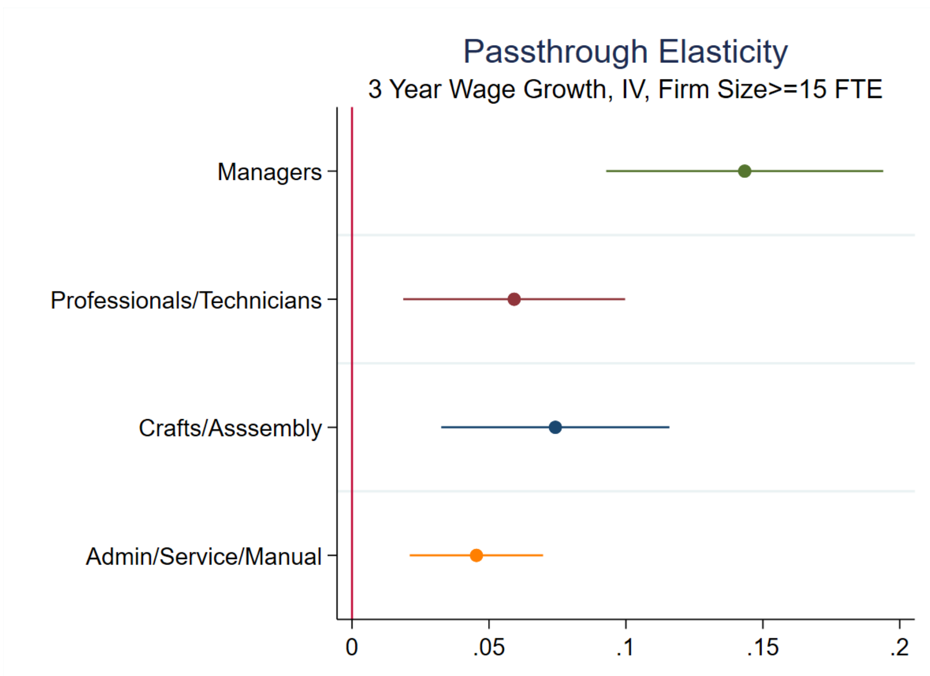
This technique is used to extract permanent changes to productivity if productivity evolves according to a combination of permanent and transitory shocks. While we use a similar specification as Card et al. (2018), the interpretation is different. Through the lens of the model, long-run changes in TFP A (often interpreted as demand shocks) result in changes in firm size but not wages. The assumption is that firms are able to fully adjust to a new level of employment within two years of the shock: that is, if demand shocks increase the output per worker, the firm is able to hire a sufficient number of workers within two years to reach the same extensive marginal product that existed prior to the shock. Changes in output per worker that last longer than two years are due to changes in α, which governs the long-run ratio of average to extensive marginal product.

Our main specification includes an interaction term with dummies for each broad occupation, with a corresponding instrument for each interaction (suppressing the time subscript t):

$$\Delta \log(\mathrm{w}_{ijk,\text{-}1,s}) = \tau_1 \Delta \hat{Y}_{ik,\text{-}1,s} + \tau_2 \sum_j \Delta \hat{Y}_{ik,\text{-}1,s} \times \mathbb{1}\{j\} + X_i \gamma + Z_k \tau + F_j \theta + e_{ijk}, \tag{14}$$

where X_i is a vector of person controls, Z_k is a vector of firm controls, F_j is the set of detailed-level occupation fixed effects, and $s = 2$. In our main specification, we divide occupations into the same four categories as in the deaths analysis: managers, professionals/technicians, crafts/assembly, and administrative/service/manual. Standard errors are clustered at the firm level. Figure 6 shows the main results.

Figure 6: Passthrough Elasticity of Productivity to Wages, Four Occupation Groups



This figure reports the results of the IV passthrough estimation using equation (14) separately for each occupation. Managers have the highest passthrough at .143, followed by professionals/technicians and crafts/assembly at .059 and .074, respectively, and administrative/service/manual occupations have the lowest passthrough at .045.

As shown in Figure 6, managers have the highest estimated passthrough, with a point estimate of .143. Professionals/technicians and crafts/assembly have the next highest passthrough, estimated at .059 and .074, respectively. Administrative/service/manual occupations have the lowest passthrough, with a point estimate of .045. To address concerns of reverse causality where changes in worker productivity affects both

Table 3: Cross-Sectional and Passthrough Estimates by Occupation Group

	$\log(w)_t$			$d\log(w)_{t\text{-}1,t+2}$		
		OLS		OLS	IV	
	Main	Main	Main	Main	Main	Size≥50
	(1)	(2)	(3)	(4)	(5)	(6)
Productivity $\hat{Y}$	.071**	.	.	.	.	.
	(.005)					
$\log(FTE)$	.	.010**	-.009**	.	.	.
		(.003)	(.003)			
$\hat{Y} \times \mathbb{1}$ {Managers}	.	.	.224**	.029**	.143**	.103**
			(.018)	(.003)	(.026)	(.022)
$\hat{Y} \times \mathbb{1}$ {Professional/Technician}	.	.	.089**	.014**	.059**	.049*
			(.007)	(.003)	(.021)	(.023)
$\hat{Y} \times \mathbb{1}$ {Crafts/Assembly}	.	.	.091**	.021**	.074**	.064**
			(.008)	(.002)	(.021)	(.025)
$\hat{Y} \times \mathbb{1}$ {Admin/Service/Manual}	.	.	.061**	.014**	.045**	.034**
			(.007)	(.002)	(.012)	(.013)
Occ FE Digit	4	4	4	4	4	4
Observations (Thous.)	5,989	5,989	5,989	2,441	625	492
R-squared	.575	.569	.579	.119	.071	.073
Clustered SE's	Firm	Firm	Firm	Firm	Firm	Firm

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

This table reports estimates for equation (14) in both levels and changes. "Main" indicates the full worker sample. Columns 1, 2, and 3 show the OLS levels regression, where the firm size and productivity variables are averages over the entire sample period. The first column reports only firm productivity, and the second column reports only firm size. The standard deviations for average $\hat{Y}$ and the log of FTE are 0.63 and 1.96, respectively. The third column includes firm productivity interacted with dummies for each broad occupation group, as well as firm size. The levels regressions include fixed effects at the 4-digit occupation level. Column 4 shows the OLS first-differences over three year changes, and columns 5 and 6 show the IV estimates, where three year changes in log value added residuals $\Delta\hat{Y}_{k,t\text{-}1,t+2}$ are instrument with five year changes $\Delta\hat{Y}_{k,t\text{-}3,t+4}$. The sample in column 6 restricts to sample to firms that had at least 50 full time equivalent employees in year $t-1$. All regressions include individual worker controls include the 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.

the worker's own wages and firm productivity, we also report results after restricting the sample to firms with at least 50 full time equivalent workers. The passthrough estimates are similar, although the point estimate on managers is slightly attenuated.

Table 3 shows the same results as Figure 6, in addition to some simple specifications for comparison. Columns 1, 2, and 3 show OLS cross-sectional specifications that estimate the relationship between firm productivity, size, and wages. In column 1, the cross-sectional elasticity between productivity and wages for all occupations is 0.071. With a standard deviation of firm productivity of 0.63, this suggests that workers at firms that are one standard deviation above average earn 9% higher hourly wages than workers at a firm one standard deviation below average. When looking at the results for firm size in column 2, a worker at a large firm earns only 4% more than at a small firm, with an elasticity of 0.01 and standard deviation of 1.96. Column 3 includes both firm size and productivity, but firm productivity is interacted with a dummy for each occupation group. Managers have the highest cross-sectional elasticity of productivity to wages at .224, and administrative/service/manual jobs have the lowest value at .061. Notably, the coefficient on firm size is now negative at -.009, suggesting that in the cross-section, firm size is not predictive of wages after conditioning on productivity.

Column 4 of Table 3 reports the estimates of an OLS specification that regresses three year changes in wages on three year changes in productivity, and columns 5 and 6 show the results reported in Figure 6. Compared to the IV estimates, the coefficients in the OLS specification are much smaller, with magnitudes of one fourth to one third of the size. This reflects that the IV strategy addresses measurement error that will still be present in the OLS specification, as well as the OLS regression includes some components of transitory shocks that are likely to have lower passthrough.

Comparing columns 5 and 6 with column 3, we can infer what share of the cross-sectional elasticity of wages to productivity is due to rents, with the remainder due to sorting of higher wage workers to productive firms within occupation groups. For both administrative/service/manual and crafts/assembly occupations, the passthrough elasticity is approximately three-quarters of the cross-sectional elasticity. For professionals/technicians and managers, rents account for about two-thirds of the cross-sectional elasticity, leaving a larger role for sorting.

Comparing Estimates of $\hat{\mathbf{h}}_j$ to Passthrough Elasticities In Section 4, we directly estimated our two turnover cost parameters δ_j and $\mathbf{h}_j$. However, as is evident from the passthrough elasticity in equation (13), there is a one-to-one mapping between the passthrough elasticity and the ratio $\mathbf{h}_j/\delta_j$. Given estimates of total turnover costs $\widehat{\delta + \frac{\mathbf{h}}{\alpha}}$, and our passthrough elasticity, we can back out separate estimates $\mathbf{h}$ and δ. Rearranging equation (13), the ratio of $\mathbf{h}_j/\delta_j$ is:

$$\widehat{\mathbf{h}/\delta} = \frac{\gamma\alpha\hat{\varepsilon}_{w,apl}}{1-(\gamma+1)\hat{\varepsilon}_{w,apl}}, \tag{15}$$

which we can evaluate given assumptions on γ and α.

Table 4 performs this exercise. Column 1 reports the passthrough elasticities from the main specification shown in Figure 6. Column 2 reports the implied values of $\widehat{\mathbf{h}/\delta}$ using the passthrough elasticities and equation (15) when assuming $\gamma = 4$ and $\alpha = .6$. Column 3 reports the 3-year estimates of average turnover costs per year $\widehat{\delta + \frac{\mathbf{h}}{\alpha}}$ for each occupation group. Columns 4 and 5 reports the implied value of hold-up power $\mathbf{h}$ and non-multiplicative turnover costs δ. This resulting value of $\mathbf{h}$ is .51 for managers and .09 and .07 for professional/technicians and crafts/assembly, respectively, and .02 for administrative/service/manual occupations. This relative ordering is consistent with the findings in Figure 3b that the managers exhibit the largest difference betwen the effect of deaths at high versus low productivity firms.

Table 4: Implied Breakdown of δ and $\mathbf{h}$ from Estimated Turnover Costs and $\hat{\varepsilon}_{w,APL}$

	(1)	(2)	(3)	(4)	(5)
	$\hat{\varepsilon}_{w,APL}$	$\mathbf{h}/\delta$	$\widehat{\delta+\frac{\mathbf{h}}{\alpha}}$	$\mathbf{h}^{implied}$	$\delta^{implied}$
Managers	.143	1.20	.93	.51	.42
Professional/Technician	.059	.20	.55	.09	.46
Crafts/Assembly	.074	.28	.34	.07	.27
Admin/Service/Manual	.045	.14	-	-	-

This table computes the implied values of marginal turnover costs δ and hold-up power $\mathbf{h}$, using the equation $\widehat{\mathbf{h}/\delta} = \frac{\gamma\alpha\hat{\varepsilon}_{w,apl}}{1-(\gamma+1)\hat{\varepsilon}_{w,apl}}$ and the estimates of total turnover costs $\widehat{\delta+\frac{\mathbf{h}}{\alpha}}$ from Section 4. We assume that $\alpha = .6$ and $\gamma = 4$.

5.2 Levels and Passthrough: Task Differentiation vs. Education

To estimate heterogeneity in hold-up power using a more detailed level of occupation, we use a similar IV estimation strategy as in Section 5.1 but interact changes in firm productivity with the task differentiation score computed in Section 4.2. In addition, we will estimate a levels regression to estimate the effect of task differentiation on the level of wages.

In the levels regression, the regressors are the level of average firm productivity $\bar{Y}_k$, a measure of occupational average education, the task differentiation score, and covariates including individual education:

$$\log(w_{ijkt}) = \eta_1\bar{Y}_{ik} + \eta_2\overline{\text{Educ}}_j + \eta_3\text{Diff}_j + X_i\gamma + Z_k\tau + u_{it} \tag{16}$$

In the passthrough estimation, the regressors are three year changes in the firm residualized output per worker $\Delta\hat{Y}_{k,\text{-}1,2}$, as well as with its interactions with the differentiation score and average occupational education attainment. The instruments are the seven year changes and corresponding interactions, and we include occupation fixed effects. We estimate:

$$\Delta\log(\text{w}_{ijk,\text{-}1,2}) = \chi_1\Delta\hat{Y}_{ik,\text{-}1,2} + \chi_2\Delta\hat{Y}_{ik,\text{-}1,s}\times\overline{\text{Educ}}_j + \chi_3\Delta\hat{Y}_{ik,\text{-}1,2}\times\text{Diff}_j + X_i\gamma + Z_k\tau + F_j\theta + e_{it} \tag{17}$$

Table 5 reports the results. The differentiation score is normalized to have a mean of 0 and standard deviation of 1, while years of education is only normalized to have a mean of 0. Columns 1, 2, and 3, report the cross-sectional levels regression, which together show that after controlling for individual demographic controls including education, both more highly educated and high task differentiation occupations pay a higher level of wages. Even after controlling for both individual and average occupation education, an occupation that is one standard deviation above average in task differentiation will pay over 5 percent higher wages than an occupation with an average level of task differentiation. Column 4 reports passthrough results only including the average years of education by occupation as an interaction term and shows that high education occupations do not have significantly higher passthrough than low education occupations. Column 5 reports passthrough results that include only the differentiation score. An occupation with a one standard deviation above average higher differentiation score will have a predicted passthrough of .073, and an occupation one standard deviation below average will have a passthrough of .039. Column 6 includes passthrough results with both interaction terms. The coefficient on education turns insignificantly negative, and the point estimate on the differentiation score increases substantially.

These results suggest that it is the degree of task differentiation, rather than education, that is the most predictive of higher passthrough. Within high education occupations, managers have high task differentiation but moderate education relative to professionals/technicians and have higher passthrough. Within low education occupations, crafts and assembly have higher task differentiation but similar education as

Table 5: Level and Passthrough Estimates by Average Education and Differentiation Score

	$\log(w)_t$ Levels - OLS			$d\log(w)_{t\text{-}1,t+2}$ Passthrough - IV		
	(1)	(2)	(3)	(4)	(5)	(6)
$\hat{Y}_t$	.077***	.080***	.073***	.062***	.056***	.057***
	(.013)	(.014)	(.012)	(.013)	(.013)	(.013)
$\overline{\text{Educ Years}}_j$	.076***		.052***	.005		-.008
	(.002)		(.003)	(.006)		(.007)
Diff. Score$_j$		.099***	.053***		.017*	.027**
		(.003)	(.004)		(.010)	(.011)
Occupation FE	N	N	N	Y	Y	Y
Observations (Thous.)	4,740	4,740	4,740	555	555	555
R-squared	.509	.503	.515	.082	.081	.080
Clustered SE's	Firm	Firm	Firm	Firm	Firm	Firm

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

Columns 1, 2, and 3 report OLS results for equation (16). Sample sizes are smaller than in Table 3 because this table only includes occupations with successfully merged differentiation scores. Columns 4, 5, and 6 report estimates for equation (17). Column 4 reports results including only the interaction with average years of education within an occupation. Column 5 reports only the interaction with an occupation's differentiation score. Column 6 reports the full specification. Instruments include seven year changes in firm residual productivity, as well as interactions with the occupational variable(s) included in each specification. 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.

administrative/service/manual occupations. We interpret these results as evidence that hold-up power is not a function of only traditional measure of skill: workers in highly educated occupations need not exhibit high hold-up power if the production process does not have complementarities or if skills are not position specific.

5.3 Productivity vs. Size

Using the wage changes of job switchers, we can also test if firm size is predictive of wages. This provides a test of hold-up power as a source of wage premia, where wages should be a function of productivity but not firm size, against models that predict wage premia at large firms, regardless of productivity.[27] We regress changes in wages of switchers on changes in the average productivity and average size of the worker's two employers, using the specification with s=3:

$$\Delta\log(\mathrm{w}_{ij,t,t+s}) = \chi_0^j + \chi_1^j \Delta\bar{Y}_{i,k(t),k(t+s)} + \chi_2^j \Delta\log(\overline{FTE})_{i,k(t),k(t+s)} + X_i^j\gamma + e_{it}, \tag{18}$$

[27] In Appendix A.7, we show that models of upward sloping labor supply, where the convexity of labor supply may differ across occupations, can also generate heterogenous effects of worker deaths on firm profits (due to wage markdowns) and heterogeneous cross-sectional relationships between firm productivity and wages across occupations. However, in these models, the wage markdown will be equal to the firm size-wage elasticity, and our empirical estimate suggest that these values are different by at least an order magnitude. We discuss why these results are difficult to reconcile with extensions to the upward sloping labor supply model.

Table 6: Firm Productivity and Size Wage Effects, Levels and Switchers

Dependent:	$log(w)_t$ Cross-Sectional			$d\log(w)_{k(t),k(t+3)}$ Switchers			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Workers	All	All	All	All	All	All	Executives
$\bar{Y}$	.071**		.090**	.028**		.034**	.015
	(.005)		(.006)	(.002)		(.002)	(.035)
$\log(\overline{FTE})$		.010**	-.010*		.004**	-.003**	.041**
		(.003)	(.003)		(.001)	(.001)	(.021)
Observations (Thous)	5,989	5,989	5,989	71	71	71	.326
R-squared	.575	.569	.579	.050	.043	.051	.063

Robust standard errors in parentheses. $corr(\overline{Y}, \log(\overline{FTE}))$=.69.

** p<0.01, * p<0.05. $sd(\overline{Y})$ = .59, $sd(\log(\overline{FTE}))$ = 1.96.

This table reports estimate of the effect of firm productivity and firm size on wages. Columns 1, 2, and 3 report OLS estimates in the cross-section. Columns 4, 5, and 6 report esimates from switchers. The switcher sample includes workers who switch firms between year t and year $t+1$. Therefore, the wage reading at the leaving firm is from the final November of the workers tenure. The wage reading at the arrival firm is in the third November, so the worker has at least two full years of tenure in the firm.

We use this specification to make use of the common assumption that observing the same worker across firms will net out worker fixed effects. The coefficients in the regression will be unbiased under the assumption of exogenous mobility.

Table 6 shows the results of the above regression, both in the cross-section and for switchers. Focusing on column 5, we find that workers who switch into larger firms receive only very small increases in wages, with an elasticity of less than 1 percent. This presents a challenge for models of long-run upward sloping labor supply as a source of heterogeneity in firm premia, but it is consistent with a model where rents are shared by productive firms to workers with hold-up power, independently of firm size.[28]

In the last column of Table 6, we report the effect of firm size and productivity on workers who are listed as a top executive in both firms before and after the switch. We estimate that when an executive moves to a firm that is 100 log points larger, their pay increases by 4.1 log points. This is consistent with the idea that production complementarities are increasing in firm size for executives, as outlined in Appendix A.11.

6 Implications

In the previous sections, we present evidence that individual hold-up power differs across occupation groups. In this section, we will show that accounting for this heterogeneity has implications for the gender wage gap, the effect of superstar firms on wage inequality, and the difference across occupations in the responsiveness of wages to labor market slack.

6.1 Wage Level Premia from Hold-up Power and the Gender Wage Gap

While the main tests of our model relate to the slope of wages to firm productivity, our model also predicts that higher turnover costs and hold-up power increase the *level* of the wage in an occupation. In this section, we show that men's differential employment in higher hold-up occupations can account for approximately

[28] A.7 discuss this challenge in more detail.

one fifth of the gender wage gap.[29]

Table 7: Hold-up Rents by Occupation and Gender

			Employment Shares		Rents	
	(1)	(2)	(3)	(4)	(5)	(6)
Occupation	$\widehat{\delta_j + \frac{\mathbf{h}_j}{\alpha}}$	Premium	Women	Men	Women	Men
Manager	.93	.14	2.6	6.7	.004	.009
Professional/Tech	.55	-	45.3	35.7	-	-
Crafts/Assembly	.34	.08	5.7	34.3	.004	.027
Admin/Serv/Manual	0	-	46.4	23.3	-	-
Total			100	100	.008	.037

This table calculates the level of rents that women and men earn in each occupation due to turnover costs and hold-up power by occupation. Column 1 reports the point estimates of turnover costs from worker deaths. Column 2 reports the wage level premium relative to admin/service/manual workers using the equation $w_j^* = \left(\gamma(\delta_j + \frac{\mathbf{h}_j}{\alpha})\right)^{\frac{1}{\gamma}} \underline{w}_j$. Columns 3 and 4 report employment shares, and columns 5 and 6 add up total rents. The employment shares are calculated using all workers, not just workers at private sector firms.

One simple way to calculate the premia that men earn in higher hold-up occupations is to plug in the estimated turnover costs from Section 4 into our wage equation and weight the rents by employment shares. Recall that for a given outside options distribution characterized by γ and $\underline{w}$,[30] wages will be equal to:

$$w_j^* = \left(\gamma(\delta_j + \frac{\mathbf{h}_j}{\alpha})\right)^{\frac{1}{\gamma}} \underline{w}_j,$$

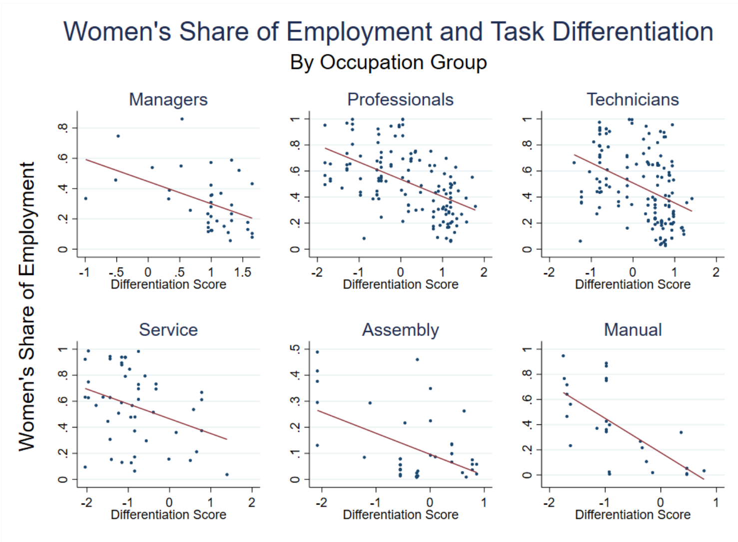
where the wage level increases in the total turnover costs $\delta_j + \frac{\mathbf{h}_j}{\alpha}$. As a simple back-of-the-envelope calculation, we will compute the wage premium earned by managers and crafts/assembly workers relative to professional/technicians and admin/manual/service workers, respectively, as these two broad groups largely split workers into high- and low-education occupations. Table 7 reports the results from this exercise. Column 1 reports the point estimates of turnover costs from worker deaths. Column 2 reports the wage level premium using the equation $w_j^* = \left(\gamma(\delta_j + \frac{\mathbf{h}_j}{\alpha})\right)^{\frac{1}{\gamma}} \underline{w}_j$.[31] Columns 3 and 4 report employment shares, and columns 5 and 6 add up total rents. In total, this exercise would predict that men earn 1.037/1.008 = 2.8% higher wages due to employment in occupations with higher turnover costs.

[29] The ratio of average women's to men's average hourly wage in our sample at private sector firms is approximately .81. The coefficient on the gender dummy in a standard mincer regression in our sample of private sector firms is .15 log points.

[30] Throughout the paper, this we focus on heterogeneity in *firm* outside options across workers in different types of positions, holding worker outside options constant wherever possible. Heterogeneity in worker outside options and its application to the gender wage gap has been explored in studies including Caldwell and Danieli (2021) and Caldwell and Harmon (2019).

[31] A technical condition in the firm's problem is that the firm must pay at least $\underline{w}$ in order to recruit new workers. Therefore, the optimal wage equation is better described as $w_j^* = \min\{\left(\gamma(\delta_j + \frac{\mathbf{h}_j}{\alpha})\right)^{\frac{1}{\gamma}} \underline{w}_j, \underline{w}_j\}$

Figure 7: 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.

Using information on differentiation scores at the detailed occupation level, we can look with greater detail if men work high in hold-up occupations, even within broad occupation groups. Figure 7 plots the share of workers in each occupation who are women on the vertical axis and the measure of task differentiation for selected occupation groups on the horizontal axis. The figure starkly shows that occupations with higher within-firm task differentiation are overwhelmingly male. Across all individuals, the average differentiation score for men is 0.40 and -0.41 for women. Using our estimate of the level wage benefit of working in higher hold-up occupations in column (3) of Table 5, this suggests that men earn (.40 - (-.41))*5.3% = 4.2% higher wages from working in higher task differentiation occupations. This is moderately larger but broadly consistent with our estimate using coarse occupation groups. In total, greater representation of men in occupations with high hold-up power and turnover costs contributes to approximately 3-4% higher relative wages for men, which can account for about one fifth of the gender wage gap.[32]

6.2 Superstar Firms and Wage Inequality

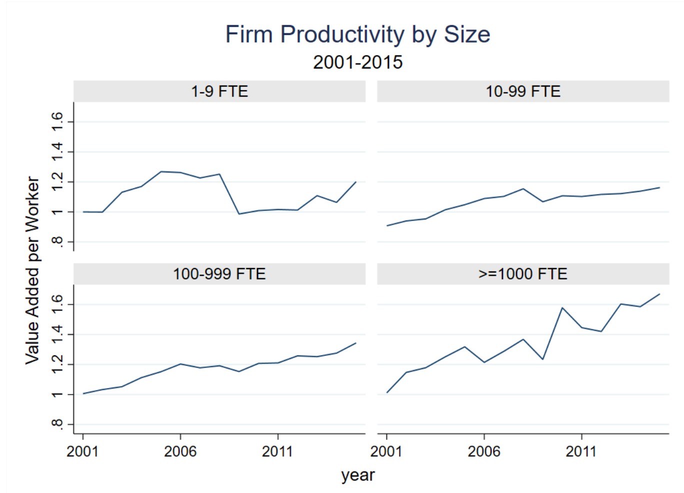
Many advanced economies experienced significant divergence in productivity between larger leader firms and smaller firms, while also experiencing declines in the aggregate labor share of income. A range of studies have argued that these two trends can be reconciled by the rise of "superstar" firms, where sales or value added is increasingly concentrated in a small set of large, high-productivity, and low-labor share firms, and sales has shifted away from low-productivity, high-labor share firms.[33] We show that these trends describe the changes in the firm distribution in Denmark quite well, and we show that wages in large firms have

[32]In Appendix B.6, we explore within-occupation gender wage gaps, and we find that occupations with high task differentiation have larger within-occupation gaps, even controlling for the average education level of an occupation.

[33]Kehrig and Vincent (2017) document this shift in manufacturing firms in the US. Gouin-Bonenfant et al. (2018) shows a similar change in all firms in Canada. Andrews et al. (2015) reports that these changes are present in services and are global. Autor et al. (2020) emphasize the reallocation from high- to low-labor share firms, driven by increasing competitive advantage of productive firms due to "winner take most" dynamics.

grown only in occupations that exhibit some degree of hold-up power.

Figure 8: Changes in Average Labor Productivity by Firm Size, Denmark 2001-2015



This figure shows changes in value added per worker (measured as value added per full time equivalent) from 2001 to 2015. Value added per worker is deflated by the Danish consumer price index and normalized relative to the level of the smallest firms (1-9 FTE) in 2001. Productivity growth over this period is monotonically related to firms size, and value added per worker at the largest firms grows approximately 65%.

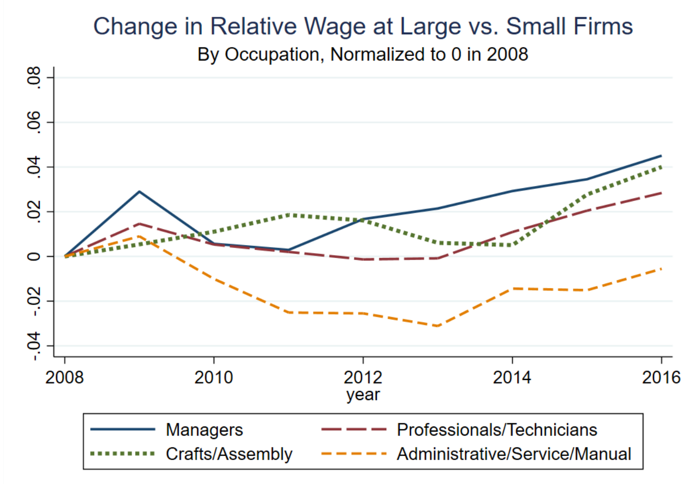
Figure 8 reports the inflation-adjusted average value added per worker in firms of different sizes from 2001 to 2015, relative to the productivity of the smallest firms (1-9 full time equivalents) in 2001. This figure shows that larger firms have seen faster productivity growth, with firms between 100-999 FTE increasing output per worker by 38% and the largest firms with over 1000 FTE increasing productivity by 65%. The shares of employment across these groups has been relatively constant across this time period, implying that an increasing share of value added is produced by these larger firms. The trend in labor shares (not shown) mirrors the changes in productivity: the labor share in the largest firms, those with at least 1000 FTE, fell from .71 to .55 between 2001 and 2015, while the aggregate labor share in all private sector firms with at least one employee fell from .73 to .64.

Next, we will show that relative wages have increased for workers in large and productive firms, but only for occupations that we estimated to have some degree of hold-up power. In Figure 9, we split workers into two evenly sized groups within each occupation according to the size of their firm.[34] Then we take the ratio of average wages of workers at the larger firms divided by the average wage of workers in smaller firms. From 2008-2016, wages for the three higher-wage occupation groups (managers, professionals/technicians, crafts/assembly) increased in above-average sized firms relative to wages in smaller firms.[35] The ratio of average productivity (measured in simple value added per worker) at large firms over the average productivity at small firms rose from 1.11 in 2008 to 1.31 by the end of the sample. The trends in productivity and wages across firm size are therefore consistent with a story that the rise of superstar firms increased within-occupation wage inequality, but only for occupations that exhibit individual worker hold-up power.

[34]The cutoff between large and small firms each year is approximately 150 full time equivalent employees.

[35]Our data provides information on occupation beginning in 2008.

Figure 9: Changes in Relative Wages at Large vs. Small Firms by Occupation



This figure shows the ratio wages of workers at "top half" firms relative to "bottom half" firms, where firms are split into two groups with equal total employment, and the top half of firms are the largest firms. Relative wages for workers at large firms rise for most occupation group by 3-4% from 2008 to 2016, except for workers in administrative/service/manual occupations.

6.3 Wages and Labor Market Tightness

In this section, we show that if some turnover costs depend on the level of labor market slack or tightness, then occupations with a lower share of fixed turnover costs (i.e., ones that do not depend on labor market slack, such as output disruptions from complementarities), will have wages that are more sensitive to changes in labor market tightness.

Let firms face three kinds of replacement costs: (1) costs that vary with labor market tightness θ, where $\theta = v/s$ is the ratio of vacancies to searchers (2) a replacement cost δ that interacts with extensive marginal product, which we will assume to be training costs, and (3) multiplicative losses from complementarities and position specific skills. The first two kinds of costs are equivalent to paying incumbent employees to divert their time from productive activities to search and training costs. The production function is then:

$$Y = N^{\alpha} e^{-(\alpha(c*\theta+\delta)+\mathbf{h})(1-X)},$$

yielding an optimal wage equation of:

$$w^* = \left(\gamma\left(c\theta + \delta + \frac{\mathbf{h}}{\alpha}\right)\right)^{\frac{1}{\gamma}}.$$

The semi-elasticity of wages to labor market tightness will be equal to:

$$s_{w,\theta} = \frac{\partial ln(w^*)}{\partial \theta} = \frac{1}{\gamma}\frac{c}{c\theta + \delta + \frac{\mathbf{h}}{\alpha}}, \qquad \text{with} \quad \frac{\partial s_{w,\theta}}{\partial(\delta + \frac{\mathbf{h}}{\alpha})} < 0, \tag{19}$$

where the semi-elasticity is decreasing in fixed turnover costs $\delta + \frac{\mathbf{h}}{\alpha}$, meaning that wages will be less sensitive to labor market slack when fixed turnover costs are larger.

A Simple Calibration We will calibrate our model to match the US labor market from 2001-2019. Our vacancy measure v is the JOLTS vacancy rate, and our measure of searchers s is .85 minus the employment to population ratio of prime age (25-54) workers,[36] with tightness $\theta = v/s$. Average tightness over this period is $\bar{\theta} = 0.6$. We calibrate turnover cost parameters to align with the numbers in Muehlemann and Leiser (2018) who shows that approximately 21% of hiring costs are search costs. To be consistent with our estimates of total turnover costs, we assume that vacancy cost parameter c=.15 for low-wage and mid/high-wage occupations, generating a higher share of turnover costs for low hold-up jobs that depend on tightness. We assume that fixed training costs δ are .2 for mid/high-wage occupations but only .1 for low-wage occupations. Lastly, we calibrate hold-up power $\mathbf{h}$ to be .1 in mid/high-wage occupations but 0 in low-wage occupations. In total, this generates a semi-elasticity of wages to tightness of .09 and .19 for mid/high-wage and low-wage occupations, respectively.

Table 8: Calibration of Turnover Costs

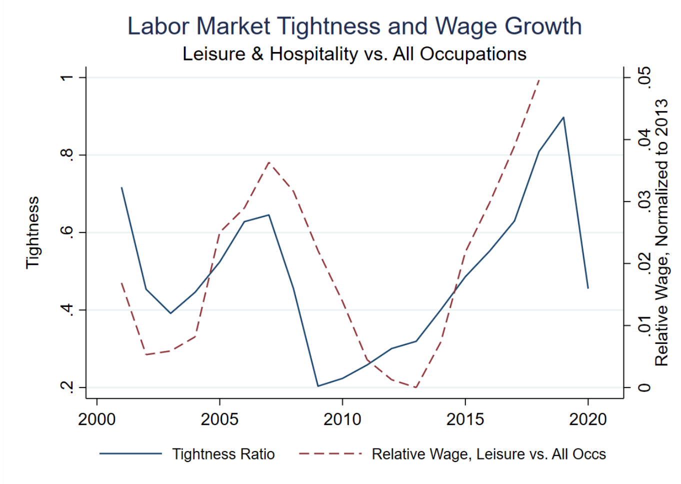
	Low	Mid/High
Recruiting Costs c	.15	.15
Training Costs δ	.1	.2
Hold-up $\mathbf{h}$	0	.1
$s_{w,\theta}$	.19	.09

This table reports the calibrated values for recruiting costs c, training costs δ, and hold-up power $\mathbf{h}$, reported values separately for low turnover cost occupations and mid/high turnover cost occupations. As usual, we assume $\gamma = 4$ and $\alpha = .6$.

We apply our calibrated model to the US labor market since 2000. Focusing on the change between the weakest part of the recovery from the Great Recession to the peak of recovery in 2019, our measure of tightness measure rises from 0.2 to approximately 0.9, implying a change of 0.7. Plugging in this change into our semi-elasticity, this would imply a compression of wages by (.19-.09)$\times$ .7 = .07, or 7 log points. Turning to the data, looking at the ratio of the employment cost index of leisure and hospitality workers to all workers in Figure 10, the relative wage increases by 5 percent from 2013 to 2019. In total, a reasonably calibrated version of our model can match the excess cyclicality of wages in low-wage service sectors, and higher wage occupations with a greater proportion of turnover costs are more insulated from changes in labor market slack.

[36] Cajner et al. (2021) show that the labor force participation rate is cyclical with a significant lag. Therefore employment to population ratios will be a more reliable indicator of labor market slack when secular trends in partcipation are muted.

Figure 10: Labor Market Tightness and Relative Wages in Leisure & Hospitality



This figure plots the labor market tightness against the relative wage in leisure & hospitality workers, which mostly closely proxies for low-wage service work at the industry level, to all private sector workers in the United States. The numerator of labor market tightness is the vacancy rate from the Job Openings and Labor Turnover Survey (JOLTS). The denominator of labor market tightness is 0.85 minus the employment to population ratio for 25-54 year old workers. The relative wage measure is the employment cost index of private sector leisure and hospitality workers over the employment cost index of all private sector workers. The relative wage index is pulled forward by two years to reflect that the effect of market tightness on wages likely operate with a lag.

7 Hold-up Power in Equilibrium with On-the-Job Search

Throughout the paper, we use a stylized two-period model with an exogenous outside offer distribution to generate predictions about wages by occupation and firms. In this section, we show that the predictions from the stylized model still hold in a dynamic setting where the outside offer distribution is endogenously determined. We show that (i) high hold-up positions pay higher wages in steady state, (ii) wages increase more at higher productivity firms when individual hold-up power **h** is higher, and (iii) wages in occupations with higher hold-up power are less sensitive to changes in labor supply.

Overview of Modeling Decisions This section develops the on-the-job search model to quantitatively assess the implications of the model with complementarity and position specificity in an equilibrium labor market. The model is designed to study firm's retention problem in the face of heterogeneous worker turnover costs to the firm. This is most similar in spirit to Burdett and Mortensen (1998), henceforth BM, however, we make some major modifications to highlight the conceptually important aspects of the model while maintaining tractability. The most important of these is that workers have time-varying idiosyncratic utility over workplaces, which is important to achieve meaningful wage dispersion from realistic levels of firm productivity. In a model like BM where workers care only about wages, the most productive firms need to pay only marginally higher than other firms to recruit workers. Therefore to achieve realistic levels of wage dispersion, the distribution of firm productivity must have extremely high variance. With idiosyncratic workplace preferences, high-wage firms still must compete with firms paying slightly lower wages, requiring that more productive firms offer meaningfully higher wages without unrealistically large productivity dispersion. That the variance of idiosyncratic workplaces preferences is large is supported by Hall and Mueller (2018) and Sorkin (2018). Also as shown by Albrecht et al. (2018), introducing idiosyncratic amenities also allows for a degenerate wage distribution with on-the-job search, drastically simplifying the equilibrium.

Now that workers may in theory find it optimal to switch to jobs offering both higher and lower wages than their current job, we cannot use local approximation methods as in Fukui (2020) to solve for a steady state, as the entire distribution of workers' outside options enters the firm's wage setting problem. Therefore, to ensure tractability, we restrict our attention to environments in which workers are ex-ante identical and there are at most two "types" of jobs. In the first exercise, the two types will be jobs will result from firms having different underlying curvature α but similar degree of hold-up power $\mathbf{h}$. In the second exercise, firms will have the same underlying curvature α, but will differ in the hold-up power of workers.

Time Time is discrete, and this model will later be calibrated to a quarterly frequency.

Heterogeneity and Notation In what follows, we will consider only equilibria in which all jobs of type j pay the same wage w_j. Because we will restrict the environments we consider to have two types of jobs, we can without loss of generality define the possible values of j as $j \in \{L, H\}$, where L indicates low wage jobs, and H indicates the high wage jobs.

Workers Workers have per-period utility

$$U_{it} = w_{ijt} + \iota_{ikt},$$

where ι_{ikt} is an idiosyncratic, i.i.d. preference of worker i for workplace k in period t. Workers discount at rate β^W. The population of workers is normalized to 1.

Firms There is an exogenous mass M of firms, and there is no entry. Firms are denoted with subscript k. Firms can employ positions of different hold-up powers h_j, and firms may differ in their concavity of output with respect to size α. The number of positions for each type j is denoted N_{jk}. We restrict the set of employment contracts such that the firm chooses a single wage for a given position type j, regardless of the worker occupying that position is trained or untrained.[37] Firms cannot charge workers for leaving the firm and can post vacancies to recruit workers at a flow cost of c.

Matching Function Each period, matches between workers and firms may end, either exogenously or endogenously, and new matches are created in a frictional matching market. Let the measure of searchers be denoted S, and the aggregate number of vacancies be v. Let the rate at which employed workers are allowed to search be λ (unemployed workers' search probability is 1). Then the measure of searchers is $S = \lambda(N^H + N^L) + u = \lambda(1-u) + u$, where u is the unemployment rate and N^H and N^L are the masses of workers employed in high and low hold-up jobs, respectively, with $N^j = \int_k N_{jk} dk$. The standard tightness variable θ is defined as $\theta = \frac{v}{S}$. Workers do not pay a cost to search.

Workers' Problem and Value Functions The value of a state to a worker is always denoted by V. At the beginning of the period, workers can either be employed or unemployed, and the values of each are denoted V_E^j and V_U, with $j \in \{L, H\}$. Employed workers search with probability λ, and unemployed workers search every period. After the worker has searched, the worker encounters an offer with probability $f(\theta)$. Thus, the value of being employed or unemployed is:

[37] In the section, we will examine the properties of the model in steady state, and so the firm's choice of the wage each period will be simply be the time-invariant wage policy of the firm.

$$
\begin{aligned}
V_E^H &= \lambda f(\theta) E[V^H_{\text{two offers}}] + (1-\lambda f(\theta)) E[V^H_{\text{one offer}}] \\
V_E^L &= \lambda f(\theta) E[V^L_{\text{two offers}}] + (1-\lambda f(\theta)) E[V^L_{\text{one offer}}] \\
V_U &= f(\theta) E[V_{\text{one offer}}] + (1-f(\theta)) E[V_Q],
\end{aligned}
$$

where $E[V^j_{\text{two offers}}]$ is the expected value of having a competing offer while employed in state j, $E[V^j_{\text{one offer}}]$ is the value of having one offer on hand (either because the worker is unemployed and found a job, or a currently employed worker does not have a competing outside offer this period), and $E[V_Q]$ is the expected value of being unemployed and having found no offer (which will be equal to expected value of quitting a job).

Next, a worker who was employed and successfully found a competing offer has to choose among three outcomes: stay, change jobs, or quit:

$$
\begin{aligned}
V^L_{\text{two offers}} &= p^H \max\{V_k^L, V_l^H, V_Q\} + (1-p^H) \max\{V_k^L, V_l^L, V_Q\} \\
V^H_{\text{two offers}} &= p^H \max\{V_k^H, V_l^H, V_Q\} + (1-p^H) \max\{V_k^H, V_l^L, V_Q\},
\end{aligned}
$$

where firm k is their current firm, firm l is the competing firm, p^H is the probability that the outside offer will be a high wage job, and probability $1-p^H$ is the probability of that the outside offer is a low wage job.[38] A worker with one offer, on the other hand, can choose between taking that offer (keeping their job if employed, taking the job if unemployed) and unemployment:

$$
\begin{aligned}
V^H_{\text{one offer}} &= \max\{V^H_{prod}, V_Q\} \\
V^L_{\text{one offer}} &= \max\{V^L_{prod}, V_Q\}.
\end{aligned}
$$

Lastly, the value of being in each state during production is:

$$
\begin{aligned}
V^H_{prod} =& w_k^H + \iota_k + \beta\big((1-s)V_E^{H'} + sV_U'\big) \\
V^L_{prod} =& w_k^L + \iota_k + \beta\big((1-s)V_E^{L'} + sV_U'\big) \\
V_Q =& b + \iota_u + \beta V_U',
\end{aligned}
$$

where w_k^H is the high wage w_k^L is the low wage, s is an exogenous separation rate, and the worker is in their same employed state at the beginning of the next period with probability of $(1-s)$ and unemployed with probability s.

Firm's Problem Firms produce using positions with two levels of hold-up power: high and low, denoted $j \in \{H, L\}$. The firm chooses separate wages policies w_H and w_L that all workers in each respective position earn. For positions with hold-up power, δ_j is the gap in productivity between trained and untrained workers in non-multiplicative tasks, and $\mathbf{h}_j$ is the hold-up power. The production function with two occupations is the same as in Section 2.3. Firms post vacancies for both kinds of positions, and vacancy costs are c per vacancy for both kinds of position. Per-period profits (suppressing the firm k subscript) are:

$$
A(N_{Ht}^{\sigma_H} N_{Lt}^{\sigma_L})^{\alpha} e^{-\sum_j \sigma_j(\alpha\delta_j + \mathbf{h}_j)(1-X_{jt})} - \sum_j w_j N_{jt} - \sum_j cV_{jt},
$$

[38]Note that due to the assumption of random search, workers in high hold-up jobs, low hold-up jobs, and unemployed workers are all equally likely to encounter a high hold-up job.

where X_t is the share of H positions filled with trained, specifically skilled workers.

Each type of position j will have a retention probability $r_j(w)$ and job filling probability $\phi_j(w)$. The retention probability $r_j(w)$ of a given worker is itself the weighted sum of three separate probabilities, depending on the search status of the worker. Let $r_j^H(w)$ be the probability that a worker is retained by while in a position paying w who encounters a high-up job through on-the-job search, let $r_j^L(w)$ be the probability that a worker is retained by a firm paying w who encounters a low hold-up job through on-the-job search, and let $r_j^U(w)$ be the retention probability of a worker whose only outside option this period is unemployment. Then we have:

$$r_j(w) = \lambda f(\theta)\Big(p^H r_j^H(w) + (1-p^H) r_j^L(w)\Big) + \Big(1-\lambda f(\theta)\Big) r_j^U(w).$$

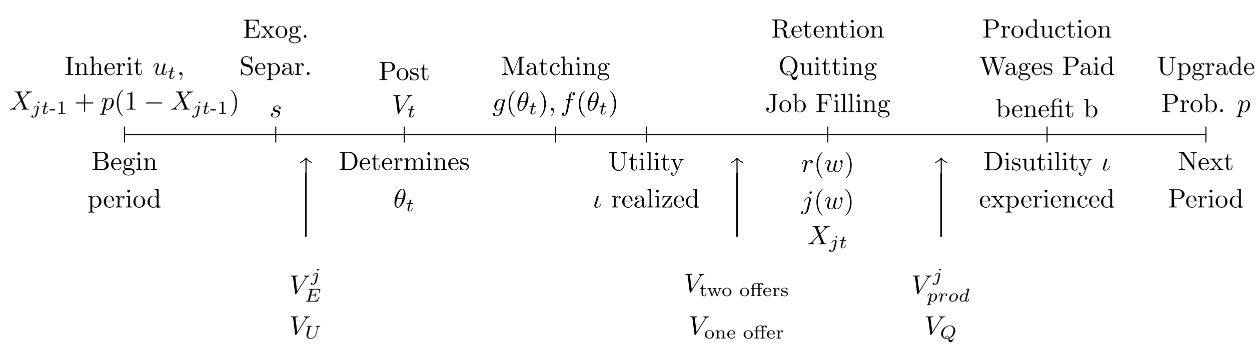
The job filling probability will be an analogous weighted function of the probability of recruiting workers whose prior state is a high hold-up job, low hold-up job, or unemployment: $\phi_j^H(w)$ $\phi_j^L(w)$, $\phi_j^U(w)$, respectively. Let $\frac{\lambda N^H}{S}$ be the share searchers who are currently employed in high-wage jobs, let $\frac{\lambda N^L}{S}$ be the searchers who a currently employed in low-wage jobs and $\frac{u}{S}$ be the share of sear chers who are unemployed. The firm's job filling probability is then:

$$\phi_j(w) = \Big(\frac{\lambda N^H}{S}\phi_j^H(w) + \frac{\lambda N^L}{S}\phi_j^L(w) + \frac{u}{S}\phi_j^U(w)\Big),$$

where $\lambda N^H + \lambda N^L + u = S$.

To simply the problem, we will make two assumptions. First, we will only consider steady states, and so we will restrict firms to make a indefinite choices of size N_j and wage policy w_j for each occupation j. Second, we will assume that firms do not discount, and so solving the firm's problem is equivalent to solving a static maximization problem subject to constraints that maintain constant endogenous variables.

Figure 11: Model Timing



With these assumptions and the definitions of the worker's value functions and the retention and job-filling probabilities in hand, we can explicitly write out the sequence of events within each period, as shown in Figure 11. At the beginning of the period, the firm inherits the share of positions filled X_j end of the previous period plus a share p of untrained workers who upgrade to become specifically trained. Matches are exogenously destroyed with probability s, after which workers enter unemployment and the positions become vacant. Firms then post vacancies, which determines the aggregate labor market tightness and matches occur. Workers then see the realizations of idiosyncratic utility of their job options and choose which job to take or decide whether to be employed or unemployed. After quit and hiring decisions have been finalized, untrained workers are allocated to unfilled vacancies, production occurs, and wages and unemployment benefits are paid. After production is finished, a fraction p of untrained workers who now have experience in their respective positions become fully productive specific workers.

This creates a law of motion for the share of positions filled X_{jt}:

$$X_{jt} = (1-s)r_j(w_j)(X_{jt-1} + p(1-X_{jt-1})). \tag{20}$$

The firm's problem is then:

$$\max_{w_j, N_j, V_j, j\in\{L,H\}} A(N_H^{\sigma_H} N_L^{\sigma_L})^\alpha e^{-\sum_j \sigma_j(\alpha\delta_j + \mathbf{h}_j)(1-X_{jt})} - \sum_j w_j N_j - \sum_j cV_j,$$

$$\text{s.t.} \quad X_j = \frac{p(1-s)r_j(w_j)}{1-(1-p)(1-s)r_j(w_j)} \tag{21}$$

$$V_j = \frac{1-(1-s)r_j(w_j)}{g(\theta)\phi_j(w_j)} N_j. \tag{22}$$

The first constraint simply rearranges equation (20) when $X_{jt} = X_{j,t-1} = X_j$. The second equation specifies how many vacancies the firm must post given the number of positions N_j, the share of workers who leave each period $1-(1-s)r_j(w_j)$, and the number of vacancies need to fill a given number of positions, which is a function of the matching rate $g(\theta)$ and the probability that a match turns into a hire $\phi_j(w_j)$.

Equilibrium A steady-state equilibrium consists of a tightness θ, an aggregate mass of vacancies v, shares of vacancies for H and L positions, unemployment rate u (and implied mass of searchers $S = u + \lambda(1-u)$), a pair of wages $\{w_L, w_H\}$, employment sizes N_L and N_H, and mobility decision for workers with and without outside offers, such that (i) workers maximize utility, (ii) firms maximize profits, and (iii) labor market flows balance.

Calibration We calibrate the model to have a quarterly period length. The only non-standard calibration is that the discount factor β^W is significantly lower than is the standard for a quarterly model. This helps us attain a separation elasticity more in line with estimates in the data, which tend to be quite low: Manning (2011) typically finds separation elasticities in the range of -2, and Card et al. (2018) assumes a slightly higher value of -4.

Table 9: Calibration

parameter	Value	Meaning/Reason	parameter	Value	Meaning/Reason
β^{worker}	.97	Worker Discount Factor	α	.6	Curvature
β^{firm}	1	Firm Discount Factor	σ_H	.3	High hold-up weight
s	.025	Exog. Separation	σ_L	.7	Low hold-up weight
λ	.06	OTJ Search Probability	p_H	.125	Upskilling probability
c	1.4	Vacancy Cost	b	0	Unemp Benefit

In each of the following exercises, we report comparative steady states.

Exercise 1: Steady State Characterization and Comparison of Wage Levels In this first exercise, we assume that firms employ H and L type positions, and all firms are identical with $\alpha = .6$. We assume high hold-up powers have $\mathbf{h}_H = .1$, $\delta_H = .3$, and $p_H = .125$. Low hold-up positions are equivalent to neoclassical production, with $\mathbf{h}_H = 0$, $\delta_H = 0$; this means that there are no turnover costs except vacancy costs, and new hires are immediately as productive as experienced incumbents.

Table 10: Steady State Endogenous Outcomes

Outcome	Value	Outcome	Value
w_L	1.36	employment share$_H$	.29
w_H	1.53	vacancy share$_H$	.23
w_H/w_L	1.13	retention $r(w_L)$	.958
u	.045	retention $r(w_H)$	.994
vacancies v	.10	quit elasticity$_L$	-4.1
tightness θ	.99	quit elasticity$_H$	-5.3

This table reports endogenous variables for a calibration where firms use high and low hold-up positions in production. High hold-up postions have hold-up power $\mathbf{h} = .1$, regular turnover costs $\delta_H = .3$, and probability of upskilling of .125, implying two full years to reach full productivity. Low hold-up positions have $\mathbf{h}_L = 0$ and $\delta_L = 0$, making these workers equivalent to workers in neoclassical production.

Table 10 shows the values of the endogenous objects in steady state. Workers in high hold-up positions earn 13% higher wages than workers in positions with no hold-up power. The share of employed workers in high hold-up positions is .29, even though the share of vacancies that are for high hold-up positions is .23. This reflects the fact that there is a "wage ladder", where workers are more likely to remain in a high hold-up job, generating fewer vacancies per position. This is further reflected in the retention rates (excluding exogenous separations): each quarter, the probability that a worker chooses to stay in a high hold-up position $r(w_H)$ is .994, suggesting that less than 1% of workers voluntarily quit each quarter. In contrast, workers in low hold-up jobs stay in their job with probability .958, as both the likelihood of switching jobs and quitting into unemployment are higher. Testing perturbations around the optimal wage policy in both high and low hold-up jobs, the elasticity of total separations to wages is -4.1 in low hold-up jobs and -5.3 in high hold-up jobs.

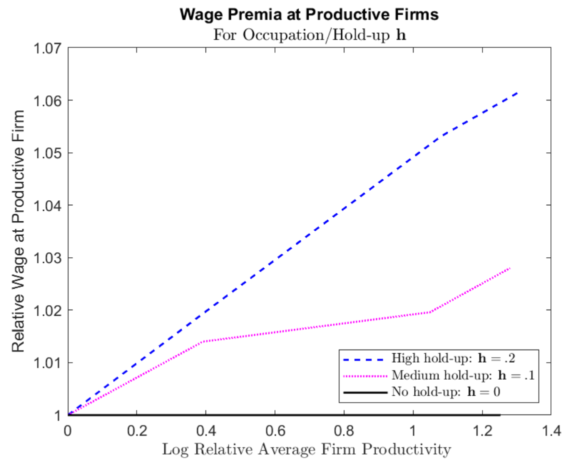
In total, this exercise generates separation elasticities that similar to the values that we assume in Section 2 and generate wage premia that are similar in magnitude as in the discussion on the effect of hold-up and the level of wages and the gender wage gap in Section 6.1.

Exercise 2: Wage Premium at Productive Firms In the second exercise, we will show that workers at higher productivity firms earn more when hold-up power $\mathbf{h}$ is higher. We assume that the economy is populated with two kinds of firms with different underlying curvature α, generating heterogeneity in firm average productivity, and we compare the wages of workers in positions with the same hold-up power $\mathbf{h}$ across firms of different productivity.

Figure 12 reports the relative wage of workers in high hold-up positions at high versus low productivity firms, comparing the outcomes in different steady states when we vary hold-up power $\mathbf{h}$ and firm curvature α. We vary hold-up power between $\mathbf{h} \in \{0, .1, .2\}$, and we vary α from .7 (low productivity) to .2 (high productivity). Consistent with the predictions from the analytical model in Section 2, the slope of wages with respect to firm productivity is higher when hold-up power $\mathbf{h}$ increases, and there is no premium for working at high productivity firms if hold-up power $\mathbf{h} = 0$.[39]

[39] The elasticies in Figure 12 are approximately half the size as in passthrough formulas in Section 2. This primarily reflects the fact that the equilibrium results in a degenerate wage distrubution, so matching a quit elasticity at individual points in an equilibrium model does not generate the same results as a constant quit elasticity for any wage that is assumed in Section 2.

Figure 12: Wage Premia at Productive Firms by Hold-up Power $\mathbf{h}$



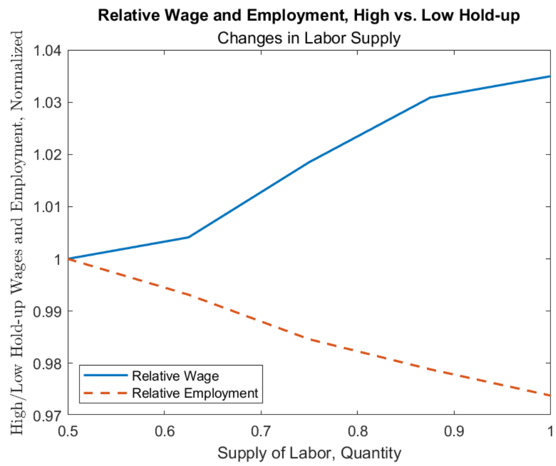
This figure reports the relative wage in a high productivity firm relative to the wage in a low productivity firm for different values of worker hold-up power $\mathbf{h}$. For example, when $\mathbf{h} = .2$, a firm with averge product that is 100 log points higher will pay a 5 percent wage premium. These estimates imply passthrough elasticities of approximately .05 for $\mathbf{h} = .2$ and .025 for $\mathbf{h} = .1$.

Exercise 3: Wages and Labor Supply In this third exercise, we estimate the relative response of wages and employment in high and low hold-up occupations in response to a change in labor supply. While similar in spirit to the exercise in Section 6.3 where we change labor market tightness, in the equilibrium model labor market tightness is an endogenous outcome. Therefore we will perform the following thought experiment: suppose that labor markets across skill groups are segmented, and so we can solve for submarket-level wages and employment shares independent of other submarkets. If there is no firm entry, how do changes in the supply of workers affect the relative wage in high and low hold-up positions?[40]

To estimate how changes in labor supply affects employment and wages, we vary the mass of available workers from 0.5 to 1, and we compute the relative employment and relative wages at high hold-up firms, normalized to 1 when the supply of workers is equal to 0.5. High hold-up jobs have hold-up $\mathbf{h}_H = .2$, $\delta_H = .3$, and $p_H = .125$. Low hold-up positions have $\mathbf{h}_L$ and $\delta_L = 0$. Figure 13 shows the results. As labor supply increases, the relative wage of low hold-up jobs falls, meaning that the relative wage of high hold-up jobs increases. While the absolute level of wages falls for both groups, the results mean that wages fall less in percent terms in high hold-up jobs. This is because high hold-up positions have turnover cost that are insensitive to the supply of workers, which incentivizes firms to keep wages for high hold-up positions high. Since the marginal cost of a position remains high when wages are high, firms with high hold-up postions expand the number of positions by less in response to the increase in labor supply. In contrast, wages fall more and employment expands by more in low hold-up jobs. In total, the wage and employment levels of high hold-up positions are less responsive to changes in labor supply, and workers in these high hold-up positions are more insulated from changes in the outside labor market.

[40]For example, suppose the supply of college educated workers increases. The occupations that tend to employ college graduates are managers, professionals, and technicians. If the new labor market entrants could in theory fill positions in any of these occupations, how would relative wages and employment change across occupations? A second scenario is to consider a low-education labor market, where there may be a sudden increase in migration by low-education workers, comparing the response in moderate hold-up blue collar jobs versus low hold-up service jobs.

Figure 13: Response of Employment and Wages to Labor Supply Changes



This figure reports the wage and employment levels at high hold-up positions, relative to the wage and employment levels at low hold-up positions. The ratios are normalized to 1 when the supply of workers is 0.5. As labor supply increases, wages in the low hold-up firms fall by more, and the relative wage at high hold-up firms increases. Relative employment has the opposite response: low hold-up firms expand the number of positions more aggressively as wages fall, while high hold-up firms increase the number of positions by a limited amount.

8 Conclusion

In this paper, we show that the combination of individual production complementarities and position specific skills generate hold-up power for individuals workers, and firms are incentivized to share rents with workers in high hold-up positions. We measure hold-up power across occupations and estimate the passthrough of firm productivity innovations to wages. We find that managers exhibit both the highest hold-up power and passthrough, while low-wage service, manual, and administrative jobs have the smallest measure on both dimensions. We show that accounting for this heterogeneity in individual worker hold-up power has implications for the gender wage gap, the effect of superstar firms on the distribution of wages, and the responsiveness of wages to occupational labor supply.

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 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 framwork 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.

One interesting implication of our model and results is that, despite hold-up power being detrimental to firms' profitability, we still observe significant levels of hold-up power and task differentiation across a wide range of occupations. If firms have some ability to expend resources to limit the degree of specificity and hold-up, then we should observe these attempts to "despecify" workers; standardization of reporting procedures; rotation of workers across roles to keep knowledge of production processes general within the firm; promotional structures to ease transitions when a co-worker leaves. These questions motivate further research to the understand if and why some occupations are costly to despecify.

Individual hold-up power may relate to the hours premium studied in Goldin (2014), as jobs that reward long hours may also have the properities of individual hold-up power. If it is costly for the firm to subdivide bundles of tasks that are assigned to a single position, then that type of position will likely reward long hours and also be fairly crucial in the production process. In order to achieve the flexible schedules that Goldin (2014) proposes, we may have to understand what prevents firms from despecifying positions.

While we address sorting only briefly in this paper, a natural next direction for research would be to allow ex-ante worker heterogeneity and study the predictions for sorting. Allowing sorting across individual positions within multi-worker firms may provide further insights into the distribution of earnings, particularly in the upper and upper-middle quantiles of the income distribution.

References

Albrecht, James, Carlos Carrillo-Tudela, and Susan Vroman. 2018. "On-the-Job Search with Match-Specific Amenities." *Economics Letters* 162 15–17.

Andrews, Dan, Chiara Criscuolo, and Peter N Gal. 2015. "Frontier Firms, Technology Diffusion and Public Policy: Micro Evidence from OECD Countries." OECD Future of Productivity Background Paper, 2015.

Arnott, Richard J, and Joseph E Stiglitz. 1985. "Labor Turnover, Wage Structures, and Moral Hazard: The Inefficiency of Competitive Markets." *Journal of Labor Economics* 3 (4): 434–462.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *The Quarterly Journal of Economics* 135 (2): 645–709.

Baldwin, Carliss Y. 2018. "Design Rules, Volume 2: How Technology Shapes Organizations." *Harvard Business School Research Paper Series* 19–042.

Bennedsen, Morten, Francisco Pérez-González, and Daniel Wolfenzon. 2020. "Do CEOs Matter? Evidence from Hospitalization Events." *The Journal of Finance* 75 (4): 1877–1911.

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

Blanchard, Olivier J, and Lawrence H Summers. 1986. "Hysteresis and the European Unemployment Problem." *NBER macroeconomics annual* 1 15–78.

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.

Burdett, Kenneth, and Dale T Mortensen. 1998. "Wage Differentials, Employer Size, and Unemployment." *International Economic Review* 257–273.

Cahuc, Pierre, Fabien Postel-Vinay, and Jean-Marc Robin. 2006. "Wage Bargaining with On-the-Job Search: Theory and Evidence." *Econometrica* 74 (2): 323–364.

Cajner, Tomaz, John M Coglianese, and Joshua Montes. 2021. "The Long-Lived Cyclicality of the Labor Force Participation Rate."

Caldwell, Sydnee, and Oren Danieli. 2021. "Outside Options in the Labor Market." Working Paper.

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.

Carnevale, Anthony P, Tamara Jayasundera, and Dmitri Repnikov. 2014. "Understanding Online Job Ads Data." *Georgetown University, Center on Education and the Workforce, Technical Report (April).*

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

Cobb, J Adam, and Ken-Hou Lin. 2017. "Growing Apart: The Changing Firm-Size Wage Premium and its Inequality Consequences." *Organization Science* 28 (3): 429–446.

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.

Di Addario, Sabrina L, Patrick M Kline, Raffaele Saggio, and Mikkel Sølvsten. 2021. "It Ain't Where You're From, It's Where You're At: Hiring Origins, Firm Heterogeneity, and Wages." National Bureau of Economic Research Working Paper no. 28917.

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

Dube, Arindrajit, Eric Freeman, and Michael Reich. 2010. "Employee Replacement Costs." UC Berkeley Working Paper Series.

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

Eisfeldt, Andrea L, and Dimitris Papanikolaou. 2013. "Organization Capital and the Cross-Section of Expected Returns." *The Journal of Finance* 68 (4): 1365–1406.

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.

Fukui, Masao. 2020. "A Theory of Wage Rigidity and Unemployment Fluctuations with On-the-Job Search." Job Market Paper, Massachusetts Institute of Technology.

Gabaix, Xavier, and Augustin Landier. 2008. "Why has CEO Pay Increased so Much?." *The Quarterly Journal of Economics* 123 (1): 49–100.

Garin, Andrew, and Filipe Silverio. 2018. "How Does Firm Performance Affect Wages? Evidence from Idiosyncratic Export Shocks." University of Illinois at Urbana-Champaign Working Paper.

Gibbons, Robert, and Lawrence Katz. 1992. "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?." *The Review of Economic Studies* 59 (3): 515–535.

Gibbons, Robert, Lawrence F Katz, Thomas Lemieux, and Daniel Parent. 2005. "Comparative Advantage, Learning, and Sectoral Wage Determination." *Journal of Labor Economics* 23 (4): 681–724.

Goldin, Claudia. 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4): 1091–1119.

Goldin, Claudia, and Lawrence F Katz. 2016. "A Most Egalitarian Profession: Pharmacy and the Evolution of a Family-Friendly Occupation." *Journal of Labor Economics* 34 (3): 705–746.

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.

Gouin-Bonenfant, Emilien et al. 2018. "Productivity Dispersion, Between-Firm Competition, and the Labor Share." In *Society for Economic Dynamics. Meeting Papers*, Volume 1171.

Gregory, Victoria et al. 2020. "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search." Job Market Paper, New York University.

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

Haanwinckel, Daniel. 2018. "Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution." Job Market Paper, University of California, Berkeley.

Hall, Robert E, and Alan B Krueger. 2012. "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search." *American Economic Journal: Macroeconomics* 4 (4): 56–67.

Hall, Robert E, and Andreas I Mueller. 2018. "Wage Dispersion and Search Behavior: The Importance of Nonwage Job Values." *Journal of Political Economy* 126 (4): 1594–1637.

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

Jovanovic, Boyan. 1979a. "Firm-Specific Capital and Turnover." *Journal of Political Economy* 87 (6): 1246–1260.

Jovanovic, Boyan. 1979b. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (5, Part 1): 972–990.

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.

Kehrig, Matthias, and Nicolas Vincent. 2017. "Growing Productivity Without Growing Wages: The Micro-Level Anatomy of the Aggregate Labor Share Decline." *Economic Research Initiatives at Duke (ERID) Working Paper* (244): .

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.

Kuhn, Peter. 2004. "Is Monopsony the Right Way to Model Labor Markets? A Review of Alan Manning's Monopsony in Motion." *International Journal of the Economics of Business* 11 (3): 369–378.

Lachowska, Marta, Alexandre Mas, Raffaele Saggio, and Stephen A Woodbury. 2021. "Do Workers Bargain over Wages? A Test Using Dual Jobholders." National Bureau of Economic Research Working Paper no. 28409.

Lamadon, Thibaut, Magne Mogstad, and Bradley Setzler. 1994. "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.

Leping, Kristjan-Olari. 2009. “Measuring the Specificity of Human Capital: a Skill-based Approach.” *Baltic Journal of Economics* 9 (1): 39–54.

Lindbeck, Assar, Dennis J Snower et al. 1989. “The Insider-Outsider Theory of Employment and Unemployment.” *MIT Press Books* 1.

Lise, Jeremy, and Fabien Postel-Vinay. 2020. “Multidimensional Skills, Sorting, and Human Capital Accumulation.” *American Economic Review* 110 (8): 2328–76.

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

Manning, Alan. 1994. “How Robust is the Microeconomic Theory of the Trade Union?.” *Journal of Labor Economics* 12 (3): 430–459.

Manning, Alan. 2006. “A Generalised Model of Monopsony.” *The Economic Journal* 116 (508): 84–100.

Manning, Alan. 2011. “Imperfect Competition in the Labor Market.” In *Handbook of Labor Economics*, Volume 4. 973–1041, Elsevier.

Montgomery, James D. 1991. “Equilibrium Wage Dispersion and Interindustry Wage Differentials.” *The Quarterly Journal of Economics* 106 (1): 163–179.

Muehlemann, Samuel, and Mirjam Strupler Leiser. 2018. “Hiring Costs and Labor Market Tightness.” *Labour Economics* 52 122–131.

Mueller, Holger M, Paige P Ouimet, and Elena Simintzi. 2017. “Wage Inequality and Firm Growth.” *American Economic Review* 107 (5): 379–83.

Nagypál, Éva. 2001. “Match-Specific Learning and Unemployment.” Working Paper.

Oi, Walter Y. 1962. “Labor as a Quasi-Fixed Factor.” *Journal of Political Economy* 70 (6): 538–555.

Postel-Vinay, Fabien, and Jean-Marc Robin. 2002. “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity.” *Econometrica* 70 (6): 2295–2350.

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.

Stiglitz, Joseph E. 1974. “Alternative Theories of Wage Determination and Unemployment in LDC’s: The Labor Turnover Model.” *The Quarterly Journal of Economics* 88 (2): 194–227.

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

Tervio, Marko. 2008. “The Difference that CEOs make: An Assignment Model Approach.” *American Economic Review* 98 (3): 642–68.

Yamaguchi, Shintaro. 2012. “Tasks and Heterogeneous Human Capital.” *Journal of Labor Economics* 30 (1): 1–53.

A 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 $q_j \in \{0,1\}$ be an indicator of whether task j was successfully performed. Then total firm output takes the form of:

$$Y = N^{\alpha} \prod_{j \in 1}^{N} q_j,$$

with $\alpha \in (0,1]$. 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:

$$\frac{\Delta Y}{\Delta q_j}\Big|_{q_i=1 \forall i \neq j} = N^{\alpha}.$$

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 in that scales inversely with the size of the firm, rather than total product. Consider a function of the form:

$$Y = AN^{\alpha} \prod_{j \in 1}^{N} \Big(1 - \text{`}\frac{\alpha + m_j}{N}(1 - q_j)\Big).$$

There are two main modifications to the original O-ring production. First, the size of the firm N appears in the denominator multiplying the $(q_j - 1)$ term. This means that the fraction of output lost from a non-performing worker scales inversely with the size of the firm, which delivers convenient properties that we will explore shortly. Second, in the numerator of the multiplication are the terms $\alpha + m_j$ where α is the same as before and m_j is the degree of complementarity in production for position j.

Define the extensive marginal product MPL^E to be:

$$MPL^E\big|_{q_j=1\forall j} = \frac{\Delta Y}{\Delta N}\bigg|_{q_j=1\forall j} \approx A\alpha N^{\alpha-1},$$

evaluated at a sufficiently large N. The intensive marginal product is the change in output from a position going unperformed or unfilled, holding N constant: $\frac{\partial Y}{\partial q_j}$. Consider the intensive marginal product of labor MPL^I evaluated at $q_j = 1$ for all except a single j:

$$MPL^I\big|_{q_i=1\ \forall i \neq j} = \frac{\Delta Y}{\Delta q_j}\bigg|_{q_i=1\ \forall i \neq j} = A(\alpha + m)N^{\alpha-1} = \underbrace{A\alpha N^{\alpha-1}}_{MPL^E} + \underbrace{AmN^{\alpha-1}}_{\substack{\text{Multiplicative} \\ \text{losses}}}.$$

We can see that there are two components of intensive marginal product: the first is just equal to the extensive marginal product: the level of ouptut loss that is similar to as when a firm gets smaller. But crucially, there is a second term governed by m_j, which is fixed for position type j and does not vary with α. This says that the absence of a worker in occupation j creates an output loss of m_j/N share of output, regardless of the productivity or the size of the firm. This is the channel through which output losses of specific workers interacts with inframarginal product.

To find a larger firm approximation of our firm with a discrete number of positions, we will use the limit definition of e:

$$\lim_{N\to\infty}\Big(\frac{N-1}{N}\Big)^{N}\to e^{-1}.$$

With this in hand, consider a firm in which 20% of positions are unfilled. How much output is lost relatively to the benchmark of the firm having 100% of its positions filled and productive? For each incremental position that is unfilled, output is multiplied by $\frac{N-1}{N}$, but only $.2N$ times. Therefore, the total multiplicative factor multiplying the potential output of the firm is $\left(\frac{N-1}{N}\right)^{.2N}=\left(\frac{N-1}{N}\right)^{(1-.8)N}=\left(\left(\frac{N-1}{N}\right)^{N}\right)^{1-.8}\xrightarrow[N\to\infty]{} e^{-1(1-.8)}=e^{(.8-1)}$. Therefore, as we take the large firm limit and N represents a mass, we have a new important variable X, the share of positions filled. We show in the appendix that for a general m, the production function becomes

$$Y=AN^{\alpha}e^{-(m+\alpha)(1-X)}.$$

A.2 Deriving the Elasticity of Wages to Average Product $\varepsilon_{w,apl}$

We begin with the continuous production function:

$$Y=N^{\alpha}e^{-(\alpha+m)(1-X)}$$

$$\frac{Y}{N}=APL=N^{\alpha-1}e^{-(\alpha+m)(1-X)}$$

$$\frac{\partial Y}{\partial N}=MPL^{E}=\alpha N^{\alpha-1}e^{-(\alpha+m)(1-X)}$$

$$\frac{\partial Y}{\partial X}\times\frac{1}{N}=MPL^{I}=(\alpha+m)N^{\alpha-1}e^{-(\alpha+m)(1-X)}.$$

Thus we still have:

$$MPL^{I}=\big(1+\frac{m}{\alpha}\big)MPL^{E}$$

$$APL=\frac{MPL^{E}}{\alpha}.$$

From the firms problem, we have

$$\max_{N,w}N^{\alpha}e^{-(\alpha+md)(1-X)}-wN\text{ s.t.}$$

$$X=F(w).$$

Using $F(w)=1-w^{-\gamma}$, the problem is written in one line as:

$$\max_{N,w}N^{\alpha}e^{-(\alpha+md)w^{-\gamma}}-wN.$$

Taking first order conditions:

$$FOC_{N}:\underbrace{\alpha N^{\alpha-1}e^{(\alpha+md)(w)^{-\gamma}}}_{MPL^{E}}-w=0$$

$$FOC_{w}:\gamma w^{-\gamma-1}(\alpha+dm)N^{\alpha}e^{-(\alpha+md)(w)^{-\gamma}}-N=0.$$

Notice that we have

$$MPL^E = w^*.$$

Rerranging the two first order conditions, we get the optimal wage expression:

$$w^* = \left(\gamma\left(1 + \frac{dm}{\alpha}\right)\right)^{\frac{1}{\gamma}}.$$

In logs:

$$\log(w^*) = \frac{1}{\gamma}\left(\log(\gamma) + \log\left(1 + \frac{dm}{\alpha}\right)\right).$$

Taking

$$\begin{aligned}\frac{\partial \log(w^*)}{\partial \alpha^{-1}} =& \frac{1}{\gamma}\frac{dm}{1 + \frac{dm}{\alpha}} \\ \frac{1}{\alpha}\frac{\partial \log(w^*)}{\partial \alpha^{-1}} =& \frac{1}{\gamma}\frac{dm}{1 + \frac{dm}{\alpha}}\frac{1}{\alpha}.\end{aligned}$$

Simplifying, we get:

$$\varepsilon_{w,\alpha^{-1}} = \frac{1}{\gamma}\frac{dm}{\alpha + dm}.$$

Returning to the results that:

$$APL = \frac{MPL^E}{\alpha} = \frac{w^*(\alpha)^{-1}}{\alpha} = w^*(\alpha^{-1})\alpha^{-1}$$

In logs, then

$$\log(APL) = \log(w^*(\alpha^{-1})) + \log(\alpha^{-1})$$

$$\begin{aligned}\frac{\partial \log(APL)}{\partial \alpha^{-1}} &= \frac{\frac{\partial w^*}{\partial \alpha^{-1}}}{w^*} + \frac{1}{\alpha^{-1}} \\ \alpha^{-1}\frac{\partial \log(APL)}{\partial \alpha^{-1}} &= \left(\frac{\frac{\partial w^*}{\partial \alpha^{-1}}}{w^*} + \frac{1}{\alpha^{-1}}\right)\alpha^{-1}.\end{aligned}$$

Thus yielding:

$$\varepsilon_{apl,\alpha^{-1}} = \varepsilon_{w^*,\alpha^{-1}} + 1.$$

What's going on economically, at this point? Average product APL and α^{-1} are clearly related. If extensive marginal products were equalized across firms, then APL is exactly α^{-1}. But, of firms with different α^{-1} , wages are not exactly equalized, as higher α^{-1} will lead to higher wages. Hence the elasticity of APL to α^{-1} will be $\varepsilon_{w^*,\alpha^{-1}} + 1$.

For the last step, we have $\varepsilon_{w^*,\alpha^{-1}}$ and $\varepsilon_{apl,\alpha^{-1}}$, and we want $\varepsilon_{w,apl}$. Thus, we will have

$$\begin{aligned}\varepsilon_{w,apl} =& \varepsilon_{w^*,\alpha^{-1}}\varepsilon_{\alpha^{-1},apl} \\ \varepsilon_{w,apl} =& \varepsilon_{w^*,\alpha^{-1}}\frac{1}{\varepsilon_{apl,\alpha^{-1}}} \\ \varepsilon_{w,apl} =& \frac{\varepsilon_{w^*,\alpha^{-1}}}{\varepsilon_{w^*,\alpha^{-1}} + 1}.\end{aligned}$$

Plugging in

$$\varepsilon_{w,\alpha^{-1}} = \frac{1}{\gamma}\frac{dm}{\alpha + dm},$$

our expression of interest becomes

$$\varepsilon_{w,apl} = \frac{\frac{1}{\gamma}\frac{dm}{\alpha+dm}}{\frac{1}{\gamma}\frac{dm}{\alpha+dm} + 1} = \frac{dm}{dm + \gamma(\alpha + m)} = \frac{dm}{\gamma\alpha + (\gamma + 1)dm}.$$

A.3 Firm's Dynamic Problem with Slow Learning On the Job

The cost of worker turnover costs may also depend on the lenght of time it takes for a replacement worker to be fully productive. In the main body of the text, the assumption is that a replacement worker is less productive for only one period. This section generalizes the model so that each period, untrained workers have some probability p that they become productive. To do so, we consider a firm in a dynamic setting.

Consider a firm that is maximizing the flow of future profits:

$$\max_{\{N_t\},\{w_t\}} \sum_{t=0}^{\infty} \beta^t \Big(N_t^{\alpha} e^{-(\alpha+md)(1-X_t)} - w_t N_t \Big), \text{ s.t.}$$

$$X_t = r(w)\Big(X_{t-1} + p(1 - X_{t-1}) \Big),$$

where p is the probability that a worker who is in a specific position but is not yet skilled uprades to a fully productive worker. As $\beta \to 1$, the choice of $\{N_t\}, \{w_t\}$ is the level of N and w that maximizes the per-period profit.

If firms are maintaining a constant choice of N and w, then the state variable X, the share of positions filled with trained workers, will be in steady state. To find the steady-state value, we can rearrange the the law of motion for X assuming that $X_t = X_{t-1} = X$:

$$X(r(w)) = \frac{pr(w)}{1 - r(w)(1-p)}.$$

Examining the case of $\beta = 1$, we can just solve the firm's static maximization problem:

$$\max_{N,w} N^{\alpha} e^{-(\alpha+dm)(1-X(w))} - Nw \text{ s.t.}$$

$$X = \frac{pr(w)}{1 - r(w)(1-p)}.$$

$$\begin{aligned} FOC_N : & \qquad \alpha N^{\alpha-1} e^{-(\alpha+dm)(1-X(w)} - w = 0 \\ FOC_w : \; X'(r(w))\gamma w^{-\gamma-1}(\alpha+dm) & \qquad N^{\alpha} e^{-(\alpha+dm)(1-X(w))} - N = 0. \end{aligned}$$

Since

$$X'(w) = \frac{-pr'(w)}{\big(1 - r(w)(1-p)\big)^2},$$

rearranging the first order conditions yields:

$$\frac{-pr'(w)}{\big(1 - r(w)(1-p)\big)^2} \gamma(\alpha+dm)w = \alpha.$$

It is worth noting that in a monthly model, $r(w)$ will be close to 1. If we take the approximation of this equation by setting $r(w) = 1$, we get

$$\begin{aligned} \frac{-pr'(w)}{\big(1 - (1-p)\big)^2}(\alpha+dm)w &\approx \alpha \\ \frac{-r'(w)}{p}(\alpha+dm)w &\approx \alpha. \end{aligned}$$

Rearranging and substituting back in that $r'(w) = \gamma w^{-\gamma-1}$ we get our optimal wage expression:

$$w^* \approx \left(\frac{\gamma}{p}\Big(1 + \frac{dm}{\alpha}\Big)\right)^{\frac{1}{\gamma}}.$$

Now the probability p that an untrained worker becomes a fully productive position-specific worker enters into the wage setting in a very similar way as the productivity of replacement workers. As p falls, and the expected training duration $1/p$ increases, the cost of turnover increases, incentivizing the firm to pay higher wages.

Notice, however, that our function for the elasticity of wages to average product will be the same: $\varepsilon_{w,apl} = \mathbf{h}/(\gamma\alpha + (\gamma+1)\mathbf{h})$. That is, training time $1/p$ affects only the level of wages but not the long-run passthrough of productivity to wages.

Invariance to the Unit of Time Suppose the base unit of time is a quarter. How would we change in the model if we want to the unit of time to be months? Let p^q be the quarterly probability of a worker becoming productive, and p^m is the probability in a monthly model, with $p^m = \frac{1}{3}p^q$.

$$F^m(w) = 1 - \frac{1}{3}w^{-\gamma}; \quad F^q(w) = 1 - w^{-\gamma}.$$

Recall that the steady state share of filled positions is:

$$X(r(w)) = \frac{pr(w)}{1 - r(w)(1-p)}.$$

Recall the case where $\beta = 1$, and the firm's maximization problem yields:

$$\frac{pr'(w)}{\big(1 - r(w)(1-p)\big)^2}\gamma(\alpha + (1-d)m)w = \alpha.$$

Again using the approximation that $r^q(w)$ and $r^m(w)$ are close to 1, we have:

$$\begin{aligned}\alpha \approx& \frac{pr'(w)}{\big(1-(1-p)\big)^2}\gamma(\alpha + (1-d)m)w^{-\gamma}\\ =&\frac{r'^q(w)}{p^q}\gamma(\alpha + (1-d)m)w^{-\gamma} \approx \frac{r'^m(w)}{p^m}\gamma(\alpha + (1-d)m)w^{-\gamma}.\end{aligned}$$

Since we can now fairly simply adjust the unit of time, we can compare the replacement times across different occupations. Suppose occupations a and b have identical hold-up power $\mathbf{h}$ and outside offer distributions $F(w)$ but different replacement times, with $p_a > p_b$. Then:

$$\frac{w_a^*}{w_b^*} = \Big(\frac{p_b}{p_a}\Big)^{\frac{1}{\gamma}}.$$

A.4 Relaxing Equal Wages between Trained and Untrained Workers

In Section 2.2 of the main text, we assume that new hires who are untrained receive the same wage as fully trained workers with position specific skills. We make this assumpution to derive a closed form passthrough elasticity of average product to wages. In this section, we relax that assumption and show that main results of the passthrough elasticities are unaffected. Consider a firm facing the same problem as in Section 2.2, where the firm chooses the number of positions N and the wage of trained workers w, but can rehire untrained workers for some exogenous wage w_u. The firm's problem is then:

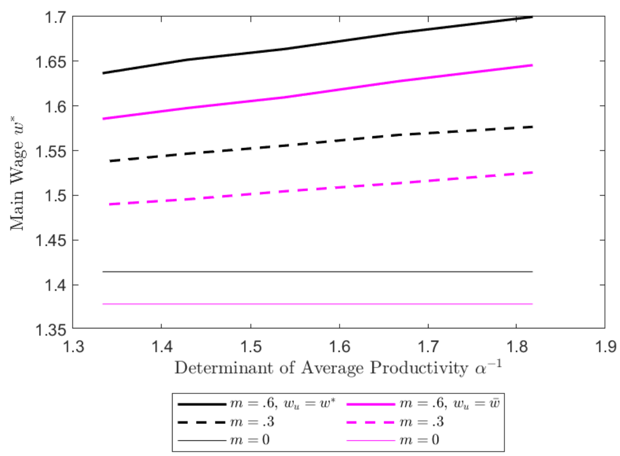
$$\max_{N,w} N^{\alpha} e^{-(\delta\alpha+h)(1-X)} - wNX - w_u N(1-X)$$

s.t.

$$X = F(w),$$

where w is the wage paid to trained workers and w_u is the wage paid to untrained workers. Using the functional form $F(w) = 1 - (w/\bar{w})^{-\gamma}$ and setting $\bar{w} = 1$. The firm pays the trained wage w to NX workers and pays the untrained wage w_u to $N(1-X)$ workers.

Figure A.14: Optimal Wage under $w_u = w^*$ and $w_u = \bar{w}$



This figure plots the optimal wage for trained workers w^* under different values of the degree of complementarity m, the underlying curvature that determines average product α, 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 magenta (lighter shade) lines indicate that untrained workers are paid the outside minimum wage $\bar{w} = 1$. For any value of m, firms pay a slightly lower wage level to trained workers if the untrained workers get the low wage. However, the slope of wages with respect to α^{-1} are functionally identical across the assumptions on wages for untrained workers.

Figure A.14 shows the optimal wage for trained workers w^* under different values of m, α^{-1}, and the wage setting assumptions of untrained workers. The black (darker shade) lines show the optimal wage when untrained workers earn the same wage as trained workers. The magenta (lighter shade) lines show the optimal wage for trained workers w^* when untrained workers are paid $w_u = \bar{w} = 1$. As the figure shows, the wage policies for trained workers are quite similar across the two scenarios, but the level is shifted down when untrained workers are paid the minimum wage $\bar{w}$. Since untrained workers are cheaper when they are paid $\bar{w}$, turnover becomes less expensive for the firm, and the firm can pay trained workers less. However, the important result from this exercise is that the slope of wages with respect to changes in the underlying concavity α^{-1} is nearly identical across different values of the degree of complementarity m.

A.5 Endogenous Hold-up Power

Suppose that firms can hire a flow of consulting services who are able to help the firm despecify workers, thereby making **h** a choice variable. The consultants charge the firm a fraction of the wage bill based on the resulting level of h, in total charging $wN(\mathbf{h}^{\frac{-1}{\kappa}} - 1)$.

$$\max_{N,w,m} N^{\alpha}e^{-(\alpha+\mathbf{h})(1-X)} - wN(1+\mathbf{h}^{\frac{-1}{\kappa}}-1)$$
$$=N^{\alpha}e^{-(\alpha+\mathbf{h})(1-X)} - wN\mathbf{h}^{\frac{-1}{\kappa}}$$

subject to $X = 1 - w^{-\gamma}$.

$$FOC_N : \alpha N^{\alpha-1}e^{-(\alpha+\mathbf{h})w^{-\gamma}} - w\mathbf{h}^{\frac{-1}{\kappa}} = 0$$
$$FOC_w : \gamma(\alpha+h)w^{-\gamma-1}N^{\alpha}e^{-(\alpha+\mathbf{h})w^{-\gamma}} - N\mathbf{h}^{\frac{-1}{\kappa}} = 0$$
$$FOC_m : - w^{-\gamma}N^{\alpha}e^{-(\alpha+\mathbf{h})w^{-\gamma}} - \kappa wN\mathbf{h}^{\frac{-1}{\kappa}-1} = 0.$$

Combining the first order conditions on the wage w and hold-up power $\mathbf{h}$ yields:

$$\mathbf{h}^* = \frac{\kappa\gamma\alpha}{1-\kappa\gamma},$$

with a technical restriction that $0 < \kappa < \gamma^{-1}$. Note that firms with higher average product will choose lower values of $\mathbf{h}$, all else equal: $\varepsilon_{apl,\alpha} < 0$ and $\partial\mathbf{h}^*/\partial\alpha > 0$. This arises naturally from the fact that firms with steeper concavity (low α) and hence higher average product have more to lose from turnover and will pay higher costs to despecify workers.

Given that we typically calibrate γ to be 4, it's worth considering the interpretation of κ when $\kappa < \gamma^{-1} = .25$. When κ is very low, it is very cheap for the firm to choose an $\mathbf{h}$ close to 0. Therefore, $\mathbf{h}^*$ falls as κ falls. The optimal wage equation is:

$$w^* = \left(\frac{\gamma}{1-\kappa\gamma}\right)^{\frac{1}{\gamma}},$$

which also indicates that wages are lower when κ is small and $\partial w^*/\partial\kappa > 0$. Therefore, in total, firms that face higher costs of despecifying workers κ settle on higher values of hold-up $\mathbf{h}^*$ and pay higher wages. Higher productivity (low α) firms end up choosing lower hold-up $\mathbf{h}^*$ to counteract the higher hold-up power. Interestingly, low α firms do not pay higher wages, though their per-worker costs are higher because choosing a lower $\mathbf{h}$ is costly, though this result is likely due to the particular choice of functional form.

A.6 Wage Markdowns from Intensive Marginal Product

In Section 2.2, we show that the wage is equal to the extensive marginal product, $w^* = MPL^E$. This means there is a gap between the wage and the intensive marginal product MPL^I when production exhibits complementarities, i.e., when $m > 0$: $MPL^I = (1 + \frac{m}{\alpha})MPL^E = (1 + \frac{m}{\alpha})w^* > w^*$. Rearranging this expression to look like a markdown, we get:

$$\frac{MPL^I - w^*}{MPL^I} = \frac{m}{\alpha + m}.$$

This shows that the wage markdown from intensive marginal product MPL^I is increasing in complementarities m. However, wages are an increasing function of hold-up power $w^* = (\gamma(1 + \frac{\mathbf{h}}{\alpha}))^{\frac{1}{\gamma}}$, with $\mathbf{h} = dm$, which is increasing in complementarities m if workers are imperfectly replaceable, i.e., $d > 0$. It thus appears that greater complementarities both increase the markdown of wages from intensive marginal product *and* increase worker wages. How can both of these statements be true? The key is that the intensive marginal product MPL^I is endogenous, and high values of complementarities increases the total productivity of a match of a worker to a position, i.e., MPL^I.

First consider the case where workers have no position specificity, $d = 0$. The productivity of the match MPL^I is increasing in complementarities m. However, the firm never faces these losses if the firm refills the position. Therefore, as complementarities m grow, the productivity of the match MPL^I grows, and the firm keeps all of the additional match productivity, as workers cannot hold up the match productivity due to $d = 0$.

This changes if workers are position specific, i.e., $d > 0$. As complementarities m grow, turnover becomes more costly to the firm, and the firm finds it optimal to pay workers higher wages. However, this increases the cost of a position, so the firm chooses to have fewer positions and therefore a higher extensive marginal product MPL^E. Because the ratio of intensive and extensive marginal product is determined by production parameters α and m, intensive marginal product MPL^I also increases. Therefore when workers have some position specificity $d > 0$, increasing complementarities both raises the level of the wage and the intensive marginal product MPL^I. In total, when hold-up power $\mathbf{h} = dm$ is high, the worker enjoys rents while the firm captures significant surplus from the match, and these are jointly made possible by limiting the number of positions.[41]

This discussion demonstrates the importance of multi-worker firms. Because firms choose the number of positions, and because the extensive marginal product MPL^E is declining in the number of positions, both extensive marginal product MPL^E and intensive marginal product MPL^I are endogenous. As hold-up power increases, extensive marginal product endogenously increases, which would be impossible if we assume an exogenous marginal product.

A.7 The Wage Markdown/Employer Size-Wage Premium Puzzle

In our model of individual worker hold-up power, we make a strong distinction between the effects of productivity and firm size on wages. It may appear that a more common model of neoclassical production and upward sloping labor supply may be able to explain some of the empirical facts we present in this paper if (i) occupations differ in their labor supply elasticities to the firm, and (ii) high observed average productivity results from firms with high TFP A being further along their labor supply curve, thus being constrained to have high wages and high productivity. If heterogeneous labor supplies where the underlying model, we would observe that (a) occupations with low labor supply elasticities will have larger markdowns, measurable using profit losses from sudden worker separations, and (b) occupations with low labor supply elasticities have high firm size-wage elasticities. We will now show that the model of upward sloping labor supply makes joint predictions that are strongly rejected by our empirical estimates.

Consider a firm that produces with a neoclassical concave production technology $Y = (AN)^{(1-\beta)}$ and pays wage costs wN.[42] The firm faces upward sloping labor supply, so $N = w^{\psi}$ is the labor supply curve, implying $w = N^{\frac{1}{\psi}}$. The firm maximizes:

$$\max_{w} = (AN)^{(1-\beta)} - wN, \qquad \text{s.t. } N = w^{\psi}.$$

This gives us:

$$w^* = \frac{\psi}{1+\psi} MPL, \tag{23}$$

[41] It is worth noting that the surplus of the match to the worker is not well defined because the worker has a random distribution of outside offers, rather than a single outside option. However, in the dynamic model in Section 7 the worker's share of surplus is well-defined.

[42] suppose from the decreasing marginal product of comes from downward sloping product demand, where $1 - \beta = \frac{\sigma-1}{\sigma}$ and $-\sigma$ is the product demand elasticitiy.

where $\frac{1}{1+\psi}$ is the standard markdown from marginal product. The elasticity of wages to the number of workers N is the inverse of the labor supply elasticity:

$$\varepsilon_{w,N} = \frac{1}{\psi}.$$

These two equations show there should be a tight link between the losses in value added less salaries if a worker is exogenously separated scaled by the prior wage, $\frac{MPL-w^*}{w^*}$, and the firm size-wage elasticity $\varepsilon_{w,N}$: $\frac{MPL-w^*}{w^*} = \varepsilon_{w,N} = \frac{1}{\psi}$. However, we show that this link is broken in our empirical estimates. From manager deaths, where annual losses in value added less salaries are nearly a year of wages, implying a markdown of 1/2, would yield ψ=1. However, this would imply a firm size-wage elasticity of 1, when in the data we observe only 0.05, more than an order of magnitude less. For other occupations where the profit losses are approximately 1/3 of prior wages over the same time horizon implying $\psi = 3$ should have a firm size-wage elasticity of 1/3, while in the data it is approximately 0.01.[43]

Manning (2011) shows that the low estimated firm size wage premium can be reconciled by introducing recruiting costs, and as the convexity of recruiting costs per worker decreases, the firm size wage premium declines as well. Therefore, occupations with steeply increasing marginal recruiting costs experience the largest markdown but also lowest responsiveness of employment to wages. This is the most similar to our model, but the turnover costs are coming from high marginal recruiting costs rather than production disruptions. They key differentiating prediction form our model of individual worker hold-up power is the output losses should be higher in proportion to prior wages when exogenous separations occur at high average productivity firms.

A.8 Response of Wages to Idiosyncratic Outside Options

A range of papers suggest that workers have heterogeneous ability to bargain with employers. Hall and Krueger (2012) show that employers of high-wage occupations tend to bargain, while employers of low-wage workers tend to post wages. Caldwell and Harmon (2019) show that when a worker's knowledge of outside options increases due to growth at firms of former co-workers, higher wage occupation see the largest percent gains in wages. Lachowska et al. (2021) show that when dual job-holders receive a raise in their secondary job, high-wage workers tend to get raises in their primary job, while low-wage workers tend to switch jobs.

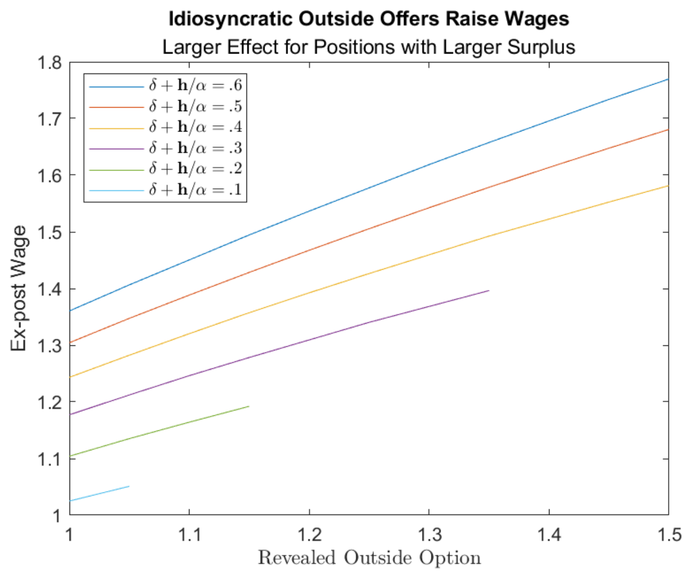
All of these studies suggests that workers in high-wage occupations have more to gain from bargaining. In this section, we will show that this is consistent in a model with individual hold-up power, where high hold-up jobs are the ones where there is more surplus to hold up. This generates predictions that differ from standard bargaining models, as high hold-up power both increases the passthrough of productivity to wages and increases the passthrough of worker's idiosyncratic outside options to wages.

Consider the problem from Appendix A.4 where untrained hires are paid the outside minimum wage $\underline{w}$, which for simplicity we set equal to 1: $\underline{w} = 1$. After the firm has made decisions for optimal size and wages, suppose that a single worker i can credibly signal that their outside offer distribution is different. In particular, suppose that their outside offer ω is distributed $F(\omega) = 1 - (\frac{\omega}{\underline{w}'_i})^{-\gamma}$, where $\underline{w}'_i$ denotes worker i's outside minimum wage. With this new knowledge, the firm may find it optimal to update the wage of the worker. Figure A.15 shows the numerical results for different values of outside option base wage $\underline{w}'_i$ and turnover costs $\delta + \mathbf{h}/\alpha$. As turnover costs and therefore total match surplus grow, the firm is more willing to update the worker's wage. For low levels of turnover costs, the firm is only willing to raise the wage a small amount, as beyond a certain level, avoiding turnover costs is not worth the additional wage

[43]It should be noted as well that the elasticity of wages to average product is always 1 in the model of upward sloping labor supply, which fails empirically.

cost. Therefore a worker in a job with low turnover costs that receives a better distribution of outside offers is more likely to leave the firm. On the other hand, workers in jobs with high turnover costs are able to extract larger raises from their current employers when revealing better outside options. Since we find that high-wage occupations tend to have proportionately higher turnover costs, our model can rationalize the results in Caldwell and Harmon (2019) and Lachowska et al. (2021) that workers in high-wage jobs extract more from current employers when their outside options improve.

Figure A.15: Wage and Idiosyncratic Outside Options



This figure plots the reoptimized wage for an individual worker who has revealed an idisoyncratic outside offer distribution with minimum outside wage $\underline{w}_i'$. The basline outside options is $\underline{w} = 1$. Each line plots the reoptimized wage for each level of turnover costs $\delta + \mathbf{h}/\alpha$. Comparing across lines vertically, firms with higher turnover costs pay higher wages for a given outside offer distribution. As idiosyncratic outside offers of a single worker improve, the firm's optimal wage for that worker increases. However, at the 45° line, the worker's outside minimum wage exceeds the incumbent firm's optimal wage, so the worker will always switch. Matches with greater turnover costs have greater surplus, so wages in high turnover cost positions can respond more to greater increases in a worker's idiosyncratic outside options.

The result that high hold-up (and high turnover cost) jobs can extract higher wages from idiosyncratic outside options generates distinct predictions from the standard model of Nash bargaining. For workers with high bargaining power β, wages depend more on productivity and less on outside options, and the wages of low β workers depend less on productivity and more outside options. In our setting, higher hold-up power predicts *both* higher passthrough of productivity to wages *and* higher response of wages to changes in workers individual, idiosyncratic outside options. This is because high production complementarities mean that the total match productivity is large relative to wages, i.e., there is a markdown from intensive marginal product. If workers are imperfectly replaceable ($d > 0$), then a firm raises wages when the firm becomes more productive, but because there is significant markdown of wages from intensive marginal product, the firms have room to raises wages when a worker presents better outside options.

Our theoretical result that workers in high-surplus matches are more likely to extract raises from current employers and less likely to switch jobs is reminiscent of the three-party bargaining model of Postel-Vinay and Robin (2002). However, our model differs in two respects. First, in our model, high-productivity employers preemptively pay workers higher wages to avoid costly turnover, so workers do not necessarily need realizations of outside offers to earn premia at productive firms. This would be consistent with the findings in

Di Addario et al. (2021) that firm wage effects of the destination firm explain at least an order of magnitude more variance in starting wages than origin effects. The second difference is that we achieve variation in marginal product within an occupaiton (both intensive and extensive) as an endogenous outcome from a model with diminishing returns to labor, while productivity dispersion in Postel-Vinay and Robin (2002) is imposed, by construction of linear production technology. This allows us to explain in which occupations does productivity dispersion across firms endogenously arise, i.e., occupations with high production complementarities exhibit greater equilibrium productivity dispersion.

A.9 Firms with Workers of Different Degrees of Complementarity

In the body of the text, we assume that every firm employs workers with only one degree of complementarity. In this section, we outline a production function where the firm employs multiple types of workers. Consider a firm with two types of labor, where M is the number of positions of occupation type a and N is the number of positions of occupation type b.

$$Y = (N_a^{\sigma_1} N_b^{\sigma_2})^{\alpha} \prod_{i=1}^{N_a} \left(\frac{\sigma_1(\alpha + m_a)}{N_a}(q_{i,a} - 1) \right) \prod_{i=N_a+1}^{N_b} \left(\frac{\sigma_b(\alpha + m_b)}{N_b}(q_{i,b} - 1) \right),$$

with $\sigma_a + \sigma_b = 1$. We arrive at analogous expressions for the ratio of intensive to extensive marginal product:

$$\frac{MPL_a^I}{MPL_a^E} = \frac{\alpha + m_a}{\alpha}, \qquad \frac{MPL_b^I}{MPL_b^E} = \frac{\alpha + m_b}{\alpha}.$$

The continuous version of output takes the form:

$$Y = (N_a^{\sigma_a} N^{\sigma_b})^{\alpha} e^{-\sum_j \sigma_j(\delta_j \alpha + h_j)(1 - X_j)},$$

with $j \in \{a, b\}$. Given a labor retention function $X_j = 1 - \left(\frac{w_j}{\underline{w}_j}\right)^{-\gamma}$, we solve for the optimal wage for each occupation $j \in \{a, b\}$:

$$w_a^* = \left(\gamma(\delta + \frac{h_a}{\alpha}) \right)^{\frac{1}{\gamma}} \underline{w}_a$$
$$w_b^* = \left(\gamma(\delta + \frac{h_b}{\alpha}) \right)^{\frac{1}{\gamma}} \underline{w}_b.$$

The elasticity of optimal wages to α^{-1} is:

$$\varepsilon_{w_j, \alpha^{-1}} = \frac{1}{\gamma} \frac{h_j}{\delta \alpha + h_j},$$

yielding an total passthrough elasticity:

$$\varepsilon_{w_a, apl} = \frac{h_a}{\gamma(\delta \alpha + h_a) + \sigma_a h_a + \sigma_b h_b \frac{\alpha + h_a}{\alpha + h_b}},$$

where σ_j is the wage bill share of occupation j. Therefore, this expression does show that the elasticity of wages to measured average product will depend on the wage share of each occupation. This is because the firm's average product responds endogenously to the optimal wages: if workers have a lot of hold-up power and increases in α^{-1} lead to an increase in wages, the firm will pull back on employment, thereby raising average product. Therefore, the shares of wages going to occupations of different levels of hold-up power matters for how much total average product responds to underlying changes in α.

Proof When firms have multiple occupations workers, we need to redefine average productivity. We define average product to be output divided by wage-weighted employment, where the weights on the employment of each occupation are the average wages of each occupation in the economy as a whole, $\bar{w}_a$ and $\bar{w}_b$. Average product is then:

$$\widetilde{APL} = \frac{Y}{N_a\bar{w}_a + N_b\bar{w}_b}.$$

From the firm's first order conditions, we also have that the total wage bill is only fraction α of total output:

$$Y = \frac{N_a w_a + N_b w_b}{\alpha}.$$

Combining the last two expressions and rewriting all of the firm's endogenous variables as functions of α^{-1} ($\bar{w}_a$ and $\bar{w}_b$ are exogenous to the firm), we have:

$$\widetilde{APL}(\alpha^{-1}) = \frac{\big(N_a(\alpha^{-1})w_a(\alpha^{-1}) + N_b(\alpha^{-1})w_b(\alpha^{-1})\big)\alpha^{-1}}{N_a(\alpha^{-1})\bar{w}_a + N_b(\alpha^{-1})\bar{w}_b}.$$

In logs, this is:

$$\log(\widetilde{APL}(\alpha^{-1})) = \log\big(N_a(\alpha^{-1})w_a(\alpha^{-1}) + N_b(\alpha)w_b(\alpha^{-1})\big) + \log(\alpha^{-1}) - \log\big(N_a(\alpha^{-1})\bar{w}_a + N_b(\alpha^{-1})\bar{w}_b\big).$$

Deriving with respect to α^{-1} and suppressing the (α^{-1}) of the endogenous variables, we have:

$$\frac{\partial\log(\widetilde{APL}(\alpha^{-1}))}{\partial\alpha^{-1}} = \frac{N_a\frac{\partial w_a}{\partial\alpha^{-1}} + w_a\frac{\partial N_a}{\partial\alpha^{-1}} + N_b\frac{\partial w_b}{\partial\alpha^{-1}} + w_b\frac{\partial N_b}{\partial\alpha^{-1}}}{N_a w_a + N_b w_b} + \alpha - \frac{\bar{w}_a\frac{\partial N_a}{\partial\alpha^{-1}} + \bar{w}_b\frac{\partial N_b}{\partial\alpha^{-1}}}{N_a w_a + N_b w_b}.$$

Thus, we have an expression for the elasticity average product to α^{-1} for firm paying average wages. Evaluating this term at $w_a = \bar{w}_a$ and $w_b = \bar{w}_b$, terms will cancel, yielding:

$$\frac{\partial\log(\widetilde{APL}(\alpha^{-1}))}{\partial\alpha} = \frac{N_a\frac{\partial w_a}{\partial\alpha^{-1}} + N_b\frac{\partial w_b}{\partial\alpha^{-1}}}{N_a w_a + N_b w_b} + \alpha.$$

$$\varepsilon_{\widetilde{APL},\alpha^{-1}} = \frac{N_a w_a\varepsilon_{w_a,\alpha^{-1}} + N_b w_b\varepsilon_{w_b,\alpha^{-1}}}{N_a w_a + N_b w_b} + 1.$$

Taking the inverse of this expression, we have:

$$\varepsilon_{\alpha^{-1},\widetilde{APL}} = \frac{N_a w_a + N_b w_b}{N_a w_a(\varepsilon_{w_a,\alpha^{-1}} + 1) + N_b w_b(\varepsilon_{w_b,\alpha^{-1}} + 1)}.$$

We can simplify by dividing by the wage bill and using the fact that the wage bill share parameters σ_h and σ_l are constant: $\sigma_j = \frac{N_j w_j}{N_a w_a + N_b w_b}$:

$$\varepsilon_{\alpha^{-1},\widetilde{APL}} = \frac{1}{\sigma_a(\varepsilon_{w_a,\alpha^{-1}} + 1) + \sigma_b(\varepsilon_{w_b,\alpha^{-1}} + 1)}.$$

Using again that $\varepsilon_{w_j,\alpha^{-1}} = \frac{1}{\gamma}\frac{h_j}{\delta\alpha + h_j}$:

$$\varepsilon_{w_a,\widetilde{APL}} = \varepsilon_{w_a,\alpha^{-1}}\varepsilon_{\alpha^{-1},\widetilde{APL}} = \frac{h_a}{\gamma(\delta\alpha + h_a) + \sigma_a h_a + \sigma_b h_b + \frac{\delta_a\alpha + h_a}{\delta_b\alpha + h_b}}.$$

This does imply that the inference of h_j from a given elasticity of wages to measured average product will depend on the occupational composition of the firm. However, in the following table, we will show that this effect is small. Consider a firm with two occupations, a and b, where the degrees of complementarity for the two occupations are $h_a = .5$ and $h_b = .1$, respectively. Then varying the wage bill share θ_a of occupation a (which is itself a function of h_a, h_b, and σ_a), we can see how the passthrough elasticity is affected.

Table 11: Passthrough Elasticity as Occupation Composition Changes, $h_a = .5$, $h_b = .1$

θ_a	0	.5	1
$\varepsilon_{w_a,apl}$	.114	.106	.099
$\varepsilon_{w_b,apl}$	.031	.033	.036

A.10 Firms with Capital

Consider a firm with the production function:

$$Y = K^{1-\alpha} N^{\alpha} e^{-(\alpha+h)(1-X)}.$$

The firm can rent capital elastically at rate r and faces the retention function $X = 1 - w^{-\gamma}$. The firm maximizes profits:

$$\max_{K,N,w} K^{1-\alpha} N^{\alpha} e^{-(\alpha+h)w^{-\gamma}} - rK - wN.$$

The first order conditions are as follows:

$$\begin{aligned}
FOC_K: &\quad (1-\alpha)K^{-\alpha}N^{\alpha}e^{-(\alpha+h)(1-X)} - r = 0 \\
FOC_N: &\quad \alpha K^{1-\alpha}N^{\alpha-1}e^{-(\alpha+h)(1-X)} - w = 0 \\
FOC_w: &\quad \gamma w^{-\gamma-1}(\alpha+h)K^{1-\alpha}N^{\alpha}e^{-(\alpha+h)(1-X)} - N = 0.
\end{aligned}$$

Combining the second and third expressions yields the familiar expression:

$$w^* = \left(\gamma(1 + \frac{h}{\alpha})\right)^{\frac{1}{\gamma}}.$$

Now the firm's size is indeterminate, but the capital to labor ratio is a function of parameters and the wage:

$$\frac{K}{N} = \frac{w^*}{r}\frac{1-\alpha}{\alpha}.$$

Thus firms with lower α will be more capital intensive.[44] Note that from the first order condition on N, we have

$$\begin{aligned}
&\frac{\alpha}{N}Y = w \\
\Rightarrow &\frac{wN}{Y} = \alpha.
\end{aligned}$$

The constant labor shares as in Cobb-Douglas are maintained. As the labor share falls (i.e., lower α), the wage increases if $m > 0$. This confirms the idea that what matters for wages is average labor productivity, and under labor complementarities à la O-ring, workers with $h > 0$ at capital intensive firms will have greater hold-up power and therefore higher wages.

[44] This section reverses the common notation of α being the capital share in Cobb-Douglas. In the rest of this paper, we consider production with only labor and no capital, using the notation with α in the exponent of N. We maintain that convention in this appendix section for consistency.

A.11 Complementarities and Firm Size

In our original exposition, we constructed the production function such that the degree of complementarity m_j for occupation group j would be invariant to firm size N. However, complementarities in production could in theory be increasing or decreasing in firm size, depending on the context. For example, if firms are more easily able to find replacements for workers in large firms, thereby mitigating disruption when a worker leaves, then complementarities would be decreasing in firm size. Alternatively, if the number of other workers whose productivity is affected by a top manager increases with firm size, then the complementarity of the top managers may be increasing in firm size. Consider a further generalization of the O-ring production function such that the degree of complementarity m is multiplied by a function of firm size N:

$$Y = AN^{\alpha} \prod_{i=1}^{N} \frac{\alpha + mN^{\beta}}{N}(q_i - 1),$$

where $\beta > 0$ indicates increasing complementarities with firm size and $\beta < 0$ indicates decreasing complementarities with firm size.

Using the static firm's problem and plugging in $1 - X = w^{-\gamma}$, the firm's problem is:

$$\max_{N,w} N^{\alpha} e^{-(\alpha + dmN^{\beta})w^{-\gamma}} - wN$$

$$\begin{aligned} FOC_N :& \Big(\alpha AN^{\alpha-1} - w^{-\gamma} \quad md\beta N^{\beta-1}AN^{\alpha}\Big)e^{-(\alpha+dmN^{\beta})w^{-\gamma}} - w = 0 \\ FOC_w :& \qquad \gamma w^{-\gamma-1} dmAN^{\alpha+\beta} e^{-(\alpha+dmN^{\beta})w^{-\gamma}} - N = 0. \end{aligned}$$

Rearranging and solving for w yields:

$$w^* = \left(\gamma + \frac{dm}{\alpha}N^{\beta}(\gamma + \beta)\right)^{\frac{1}{\gamma}}.$$

We can see that the optimal wage is going to be increasing in N if $\beta > 0$, where complementarities are increasing in firm size. On the other hand, if complemetarities are decreasing in firm size, i.e., $\beta < 0$, then the optimal wage is decreasing in N. Note that wages will be a function of TFP A if $\beta \neq 0$, as A will affect the choice of firm size N. In the case that $\beta = 0$, we have the same expression as before:

$$w^* = \left(\gamma(1 + \frac{dm}{\alpha})\right)^{\frac{1}{\gamma}}.$$

Complementarities and Firm Size with Multiple Worker Types Now consider a firm with two worker types M and N, where the degree of complementarity of type N workers depends on the number of both M and N type workers. This can be thought of as N type workers are top managers.

$$\max_{N,w} (M^{\sigma_a} N^{\sigma_b})^{\alpha} \prod_{i=1}^{M} \left(1 - \frac{\sigma_a(\alpha + m_a)}{M}(1 - q_{i,a})\right) \prod_{i=M+1}^{N} \left(1 - \frac{\sigma_b(\alpha + m_b M^{\beta_1} N^{\beta_2})}{N}(1 - q_{i,b})\right),$$

With the large firm approximation in the static setting, and imposing the outside offer distribution $F(w)$ such that $1 - X = w^{-\gamma}$, the firm solves:

$$\max_{M,N,w_M,w_N} = (M^{\sigma_a} N^{\sigma_b})^{\alpha} e^{-\left(\sigma_a(\alpha + d_a m_a)w_M^{-\gamma} + \sigma_b(\alpha + d_b m_b M^{\beta_1} N^{\beta_2})w_N^{-\gamma}\right)} - w_M M - w_N N,$$

We end up with the optimal wage for type N workers is:

$$w_N^* = \left(\gamma + \frac{h_b}{\alpha} M^{\beta_1} N^{\beta_2}(\gamma + \beta_2)\right)^{\frac{1}{\gamma}}.$$

Most plausibly, we would have $\beta_1 > 0$, where the degree of complementarities of N type workers with respect to total output of the firm is increasing in the number of M type non-supervisory workers. You may even have $\beta_2 = -\beta_1$, such that the degree of complementarity of N type workers depends on the ratio of M to N workers in the firm.

A.12 Extreme Firm-Size Complementarities: CEOs

Consider the above example where N type workers are top executives, and for simplicity $\beta_2 = 0$. Then we have:

$$w_N^* = \left(\gamma(1 + \frac{d_b m_b}{\alpha} M^{\beta_1})\right)^{\frac{1}{\gamma}}.$$

Taking the elasticity of w_N^* with respect to the number or non-supervisory workers M yields:

$$\varepsilon_{w_N^*,M} = \frac{\beta_1}{\gamma} \frac{dm}{\frac{\alpha}{M^{\beta_1}} + dm}.$$

As M becomes large and $\beta_1 \to 1$, then $\varepsilon_{w_N^*,M} \to 1/\gamma$.

Thus, for workers for whom firm size increases complementarities, the formula for firm-size wage premia will be very similar to the productivity-wage elasticity. This captures the importance of the "size of stakes" for CEO's discussed in Gabaix and Landier (2008).

A.13 Curvature from Finite Product Demand Elasticities

In the main text, we derive heterogeneous firm average product as resulting from heterogeneity in the curvature of "potential" revenue with respect to the number of positions $F''(N)$, and in the baseline case we assume this function is Cobb-Douglas with curvature parameter α. In this section we will show that high firm average product can come from low product demand elasticities and high markups.

Consider the following utility function, where consumers have a modified CES utility over a continuum of products, where y_i is the quantity of good i and q_i is the quality of good i:

$$U = \left(\int_i y_i^{\frac{\sigma-1}{\sigma}} q_i\right)^{\frac{\sigma}{\sigma-1}}.$$

A consumer maximizes utility given prices $\{p_i\}$ and qualities $\{q_i\}$, choosing quantities $\{y_i\}$:

$$\max_{y_i} = \left(\int_i y_i^{\frac{\sigma-1}{\sigma}} q_i\right)^{\frac{\sigma}{\sigma-1}} - p_i y_i.$$

Taking first order conditions for products i and $j \neq i$ yields:

$$\left(\frac{y_i}{y_j}\right)^{\frac{1}{\sigma}} = \left(\frac{q_i}{p_i}\right) \Big/ \left(\frac{q_j}{p_j}\right).$$

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

$$y_i^{\frac{1}{\sigma}} = \frac{q_i}{p_i}.$$

The seller of good i produces with a linear technology, where the quantity y_i is produced as a linear function of the number of positions $y_i = A_i N_i$, and the quality of product is a function of the share of positions that are filled with skilled, trained workers: $q_i = e^{-(\alpha+m)(1-X_i)}$.

$$\max_{y_i,p_i,N_i,w_i} p_i y_i - w_i N_i$$

subject to

$$\begin{aligned} y_i^{\frac{1}{\sigma}} &= \frac{q_i}{p_i} \\ y_i &= N_i \\ q_i &= e^{-(\alpha+\mathbf{h})(1-X_i)} \\ X_i &= 1 - w_i^{-\gamma}. \end{aligned}$$

Substituting in the constraints and consolidating the maximization, the firm's problem can be recast as:

$$\max_{N_i w_i} A_i^{\frac{\sigma-1}{\sigma}} N_i^{\frac{\sigma-1}{\sigma}} e^{-(\alpha+\mathbf{h})(w_i^{-\gamma})} - w_i N_i,$$

yielding:

$$w_i^* = \Big(\gamma\big(1 + \frac{h}{\frac{\sigma-1}{\sigma}}\big)\Big)^{\frac{1}{\gamma}},$$

which is identical to the main wage equation if $\frac{\sigma-1}{\sigma} = \alpha < 1$ as in equation (4) in the main text.

A.14 Relation to Jäger (2016)

Jäger (2016) finds that in response to worker deaths, the earnings of workers in similar occupations increases, suggesting that workers in similar occupations are substitutes. This evidence may seem at odds with our model, where every worker's ouptut is complementary with the output of the entire rest of the firm. However, these need not be in conflict: all that matters is that incumbent workers of the same occupation of the deceased are the *best* substitute for the deceased workers in the short run. Consider a firm with concavity α and occupation with complementarity degree m, giving $MPL^I = \frac{m+\alpha}{\alpha} MPL^E$ and $MPL^E = w^*$. Suppose the firm has two options: do not response to a worker separation (beyond hiring an untrained replacement worker, who is not yet productive), or offering incumbent workers with related skills to replace part of the separated worker's output. Suppose to do so, however, the firm has to pay an overtime rate $w^{ot} = w^*(1+\psi^{ot})$, with $\psi^{ot} > 0$. Suppose a coworker's replacement productivity is $\delta = d^{coworker} \in (0,1)$. The firm is strictly better off paying the co-workers to replace lost output if:

$$\begin{aligned} d^{coworker} MPL^I &> w^*(1+\psi^{ot}) \\ d^{coworker} MPL^I / w^* &> (1+\psi^{ot}) \\ d^{coworker} (1 + \frac{m}{\alpha}) &> (1+\psi^{ot}). \end{aligned}$$

Thus, as long as output is sufficiently complementary (high m), co-worker productivity $d^{coworker}$ is sufficiently high, and the overtime rate ψ^{ot} is not too high, the firm will find it optimal to pay overtime to the co-workers of the deceased worker to partially replace lost ouput. However, paying workers overtime is not a profitable long term solution, as fully trained replacement workers are more cost effective than paying overtime to imperfectly substitutable coworkers from other positions:

$$\frac{w^*}{MPL^I} < \frac{w^*(1+\phi^{ot})}{d^{coworker}MPL^I},$$

by $\phi^{ot} > 1$ and $d^{coworker} < 1$. Thus higher MPL^I (and therefore higher complementarity m) is consistent with larger output losses and increased wages for co-worker in response to a worker death.

A.15 Proof of Identification

In this section, we prove that if firms do not choose to change the number of positions in response to worker deaths, the correct outcome measure is the change in value added. If firms systematically decrease employment after a separation shock based on the hold-up power of a worker, then using value added to estimate underlying parameters will be biased, and the correct outcome measure to use is value added less wage and salary costs.

Firms Hold Constant Number of Positions N - Suppose that in response to an exogenous separation of a worker, the firm's share of fill positions X falls by ΔX:

$$\begin{aligned}\Delta Y =& N^\alpha e^{-(\delta\alpha+\mathbf{h})(1-X)} - N^\alpha e^{-(\delta\alpha+\mathbf{h})(1-(X-\Delta X))}\\ =& N^\alpha e^{-(\delta\alpha+\mathbf{h})(1-X)} - N^\alpha e^{-(\delta\alpha+\mathbf{h})(1-X)}e^{-(\delta\alpha+\mathbf{h})\Delta X}\\ =& \big(N^\alpha e^{-(\delta\alpha+\mathbf{h})(1-X)}\big)\big(1-e^{-(\delta\alpha+\mathbf{h})\Delta X}\big).\end{aligned}$$

Using the first order approximation that $e^{-z} \approx 1 - z$ for small z, we get that

$$\begin{aligned}\Delta Y =& \big(N^\alpha e^{-(\delta\alpha+\mathbf{h})(1-X)}\big)\big(1-e^{-(\delta\alpha+\mathbf{h})\Delta X}\big)\\ \approx& \frac{N}{\alpha}\big(\alpha N^{\alpha-1} e^{-(\delta\alpha+\mathbf{h})(1-X)}\big)\big((\delta\alpha+\mathbf{h})\Delta X\\ =& \frac{N}{\alpha} w^*\big((\delta\alpha+\mathbf{h})\Delta X,\end{aligned}$$

where the last line uses the first order condition on the number of positions prior to the shock: $w^* = \alpha N^{\alpha-1}e^{-(\delta\alpha+\mathbf{h})(1-X)}$. Rearranging, we end up with:

$$\frac{\big(\frac{\Delta Y}{\Delta X}\big)\times\frac{1}{N}}{w^*} = \delta + \frac{\mathbf{h}}{\alpha}.$$

Therefore, if in response to the shock, the firm does not change the number of positions N and hires a replacement worker that offsets δ of regular losses and d of multiplicative losses, then the outcome variable used to identify these parameters is change in value addd.

Firms Reoptimizing the Number of Positions N - If a worker suddenly leaves, and the firm's share of positions filled with productive workers X falls, then productivity of other workers is negatively affected and the firm may find it optimal to decrease its size. Suppose that after a separation shock occurs, the firm is able to change N, but maintains a new lower level of the share of filled positions $X_2 = X_1 - \Delta X$:

We will show that if firms (a) hire a replacement worker who partially offsets the losses of the worker who left, and (b) the firm can freely adjust N down post-shock while holding post-shock X fixed, then the correct dependent variable of a separations regression in the data is value added less wage and salary payments.

$$\Delta(Y - wN) = N_1^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-X_1)} - N_2^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-(X_1-\Delta X))} - (w_1 N_1 - w_2 N_2). \quad (24)$$

Before going further, we can show that

$$FOC_N : \alpha N^{\alpha-1} e^{-(m+\alpha)(1-d)(1-X+\Delta X)} - w = 0$$
$$FOC_w : \gamma w^{-\gamma-1}(m+\alpha)(1-d) N^{\alpha} e^{-(m+\alpha)(1-d)(1-X+\Delta X)} - N = 0.$$

Like in the standard problem, the inclusion of ΔX in the exponent does not matter and cancels out, so we have that the pre- and post-separation optimal wage for remaining workers is identical

$$w_1^* = w_2^* = \Big(\gamma(\delta + \frac{\mathbf{h}}{\alpha})\Big)^{\frac{1}{\gamma}}.$$

Next, we can show how the optimal number of positions N changes now that the firm faces a lower share of postions filled X. From the first order condition on N above, we have:

$$N^* = \left(\frac{\alpha e^{-(\delta\alpha+\mathbf{h})(1-X+\Delta X)}}{w^*}\right)^{\frac{1}{1-\alpha}}.$$

Taking the ratio from before and after the shock and putting this expression in logs, we have:

$$\frac{N_1}{N_2} = \left(\frac{e^{-(\delta\alpha+\mathbf{h})(1-X)}}{e^{-(\delta\alpha+\mathbf{h})(1-X+\Delta X)}}\right)^{\frac{1}{1-\alpha}}$$
$$\log\left(\frac{N_1}{N_2}\right) = \frac{1}{1-\alpha}\Big((\delta\alpha + \mathbf{h})\Delta X\Big)$$
$$\Rightarrow \frac{\Delta \log(N)}{\Delta X} = \frac{\delta\alpha + \mathbf{h}}{1-\alpha}.$$

Thus, when a worker is lost, the productivity of the remaining workers declines. The firm still finds it optimal to pay the same wage to minimize wage and turnover costs, and so the firm wants to lower employment to raise the extensive marginal product MPL^E in order to make the marginal position worth the cost. This force becomes stronger if workers' absence is more disruptive due to a high δ or $\mathbf{h}$. However, since employment and therefore additional value added declines will change depending on the parameter values, we no longer can use changes in value added as the outcome variable.

Before going on with the proof, we know that incremental changes in X yield incremental changes in N. We can approximate the previous line with

$$\frac{\Delta \log(N)}{\Delta X} = \frac{\Delta \log(N)\frac{1}{N}}{\Delta X \times \frac{1}{N}} \approx \frac{\Delta N}{\Delta X \times \frac{1}{N}}.$$

Thus, both the numerator and denominator are in terms of individual workers, since X is scaled by $1/N$. Returning now to equation (24), we can plug in the common wage w:

$$\begin{aligned}\Delta(Y - wN) =& N_1^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-X_1)} - N_2^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-(X_1-\Delta X))} - w(\Delta N)\\ =& (N_1^{\alpha} - N_2^{\alpha}) e^{-(\delta\alpha+\mathbf{h})(1-X_1)} + N_2^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-X_1)} - N_2^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-X_1)} e^{-(\delta\alpha+\mathbf{h})\Delta X} - w(\Delta N)\\ =& (N_1^{\alpha} - N_2^{\alpha}) e^{-(\delta\alpha+\mathbf{h})(1-X_1)} + N_2^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-X_1)}(1 - e^{-(\delta\alpha+\mathbf{h})\Delta X}) - w(\Delta N)\\ \approx& \Delta N \alpha N_1^{(\alpha-1)} e^{-(\delta\alpha+\mathbf{h})(1-X_1)} + N_2^{\alpha} e^{-(\delta\alpha+\mathbf{h})(1-X_1)}((\delta\alpha + \mathbf{h})\Delta X) - w(\Delta N).\end{aligned}$$

Again using the feature that $w^* = \alpha N^{\alpha-1} e^{-(\delta\alpha+\mathbf{h})(1-X)}$, we can replace the first term:

$$\Delta(Y-wN) = \Delta N w + w\frac{N_2}{\alpha}(-(\delta\alpha + \mathbf{h})\Delta X) - w(\Delta N)$$
$$\Delta(Y-wN) = wN_2(-(\delta + \frac{\mathbf{h}}{\alpha})\Delta X)$$
$$\Rightarrow \frac{\frac{\Delta(Y-wN)}{\Delta X \times N}}{w} = \delta + \frac{\mathbf{h}}{\alpha}.$$

Thus, losses of value added less salaries, relative to the prior wage, identify $\delta + \mathbf{h}/\alpha$.

Separating h from δ - Now suppose we run an estimation of value added minus salary losses at high and low productivity firms and yield the following estimates:

$$\hat{\beta}_h = \delta_j + \frac{\mathbf{h}_j}{\alpha_h}, \qquad \hat{\beta}_\ell = \delta_j + \frac{\mathbf{h}_j}{\alpha_\ell}.$$

where $\hat{\beta}_k = \frac{losses_k^{VA-w}}{w_k^*}$ (k indicates firm type, j indicates occupation). Solving both equations for δ_j, we have

$$\hat{\beta}_h - \frac{\mathbf{h}_j}{\alpha_h} = \hat{\beta}_\ell - \frac{\mathbf{h}_j}{\alpha_\ell}.$$

Rearranging, we get

$$\mathbf{h}_j = \frac{\hat{\beta}_h - \hat{\beta}_\ell}{\alpha_h^{-1} - \alpha_\ell^{-1}}.$$

Using that $APL_k = \frac{w_k}{\alpha_k}$, we have:

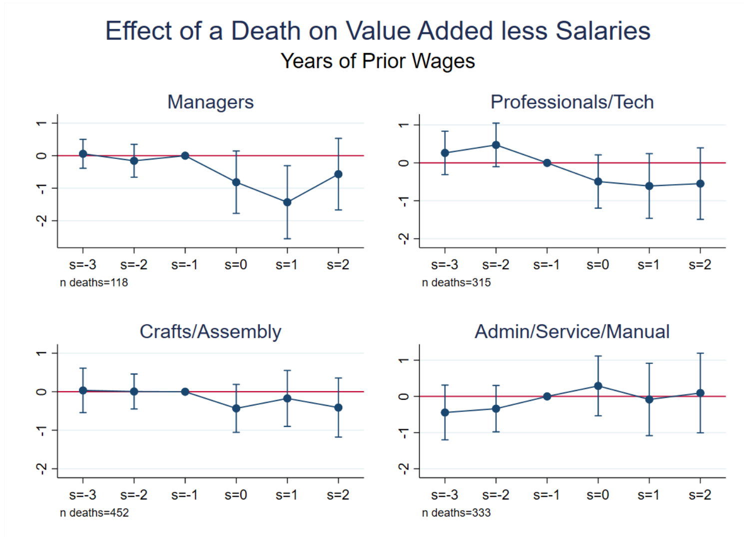
$$\mathbf{h}_j = \frac{\hat{\beta}_h - \hat{\beta}_\ell}{\frac{APL_h}{w_h} - \frac{APL_\ell}{w_\ell}}.$$

Therefore, we can identify **h** from the additional output losses relative to prior wages $\hat{\beta}$ at higher productivity firms.

B Empirical Appendix

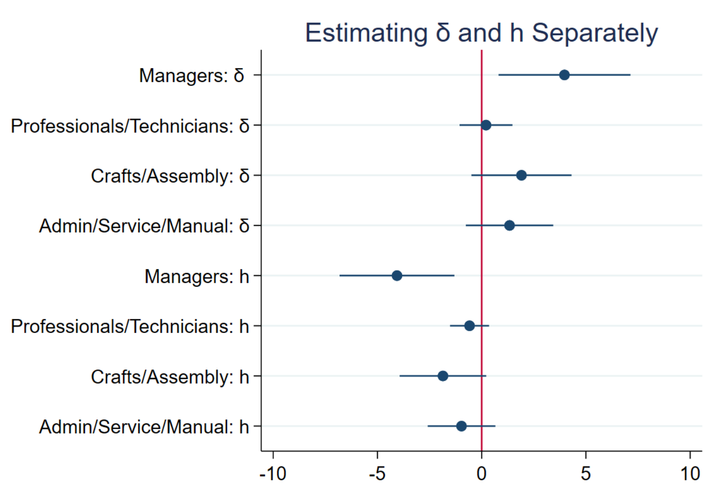
B.1 Additional Figures for Death Estimates

Figure B.1: Effect of a Worker Death on Value Added - Salaries by Occupation



This figure reports the estimates of the dynamic difference-in-difference from equation (10) reported in Figure **??**.

Figure B.2: Worker Deaths Unrestricted Interaction Regression

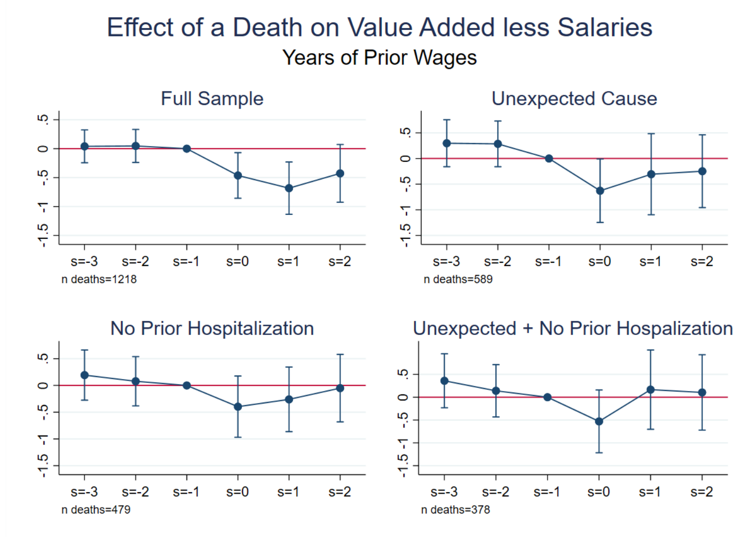


This figure reports the unrestricted estimates of equation (11), with the sign of the coefficients reversed. A negative value of $\mathbf{h}^{est}$ and positive value of δ^{est} in this figure indicate that the slope of profit losses with respect to firm productivity (measured as the inverse of the labor share) is very steep. The theory predicts that δ cannot be negative, which would predict both non-positive values for both regression coefficients in this figure.

B.2 Unexpected Deaths

In our data, we have access to hospital records as well as medical cause of death. It is common to restrict to workers who (i) had an unexpected cause such as a heart attack or an accident, and (ii) did not have a hospital admission in a prior time window. If we make these restrictions our sample shirnks dramatically and in general decreases the size of our estimates, as shown in Figure B.3.

Figure B.3: Effect of "Unexpected" vs. All Deaths on Value Added - Salaries



Restricting to unexpected deaths may produce different estimates from including all deatsh for various reasons. First, if the death is expected, the firm may have more time to adjust, thereby making unexpected deaths have larger effects. However, workers who die unexpected tend to be younger and may be in less critical roles, thereby lowering the effect of unexpected deaths relative to all deaths. Lastly, concerns remain about reverse causality, such as poor performance of the firm inducing stress and negative health outcomes for workers.

B.3 Measurement of Differention Scores

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 Burning Glass 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.

B.4 Burning Glass Job Clustering Examples

Figure B.4: Sample Skills in Burning Glass Data

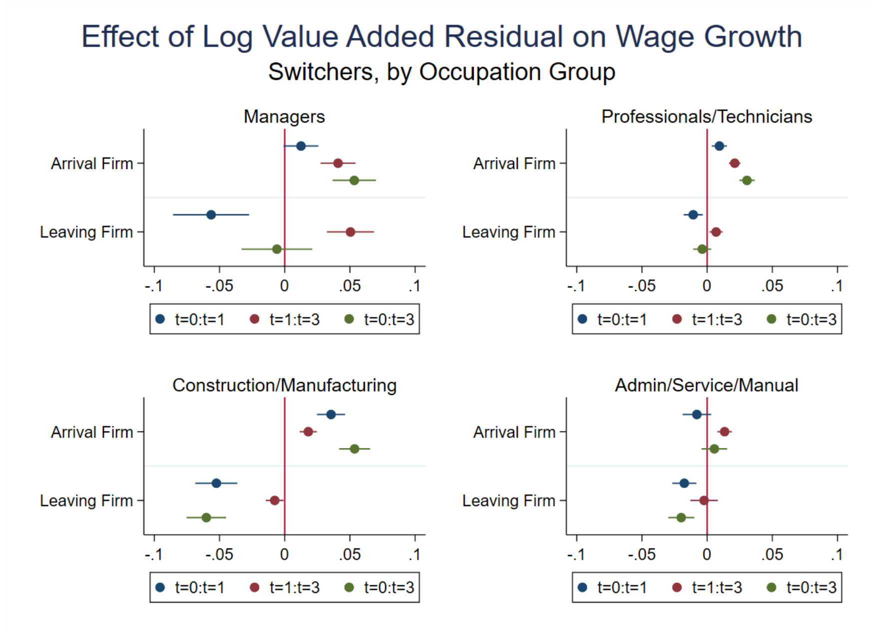
	skillclusterfamily	skillcluster	skill
1	Analysis	Natural Language Processing (NLP)	Speech Recognition
2	Analysis	Statistics	Multivariate Testing
3	Architecture and Construction	Carpentry	Basic Tools
4	Engineering	Engineering Practices	Prototype Design Development
5	Environment	Water Testing and Treatment	Water Quality Modeling
6	Finance	Financial Risk Management	Strategic Risk Management
7	Finance	Financial Trading	Exchange-Traded Products (ETP)
8	Finance	Mortgage Lending	Mortgage Loan Origination
9	Finance	Tax	CorpTax
10	Health Care	Alternative Therapy	Therapeutic Procedures
11	Health Care	Basic Patient Care	Telemedicine
12	Health Care	Cardiology	Electrocardiogram (EKG) Equipment
13	Health Care	Clinical Data Management	Clinical Programming
14	Health Care	Clinical Research	Clinical Trial Design
15	Health Care	Emergency and Intensive Care	Critical Care Nursing
16	Health Care	Medical Records	Medical Records Management
17	Health Care	Oncology	Bone Marrow Transplant
18	Health Care	Physical Therapy	Physical Disability
19	Health Care	Routine Examination Tests and Procedures	Treatment Recommendation
20	Human Resources	Occupational Health and Safety	Health and Safety Compliance
21	Industry Knowledge	Supply Chain and Logistics Industry Knowledge	Export Industry Knowledge
22	Industry Knowledge	Transportation Industry Knowledge	Shipping Industry Knowledge
23	Information Technology	Data Management	IBM InfoSphere Optim
24	Information Technology	Database Administration	Shareplex
25	Information Technology	Database Administration	Oracle Database Administration (DBA)
26	Information Technology	Extraction, Transformation, and Loading (ETL)	Data Transformation
27	Information Technology	Networking Hardware	DSLAM
28	Information Technology	Programming Principles	Functional Programming
29	Information Technology	SAP	SAP CRM
30	Information Technology	Scripting Languages	JRuby
31	Information Technology	Systems Administration	Backup administration
32	Information Technology	Test Automation	Test Development
33	Manufacturing and Production	Computer-Aided Manufacturing	Siemens Plc
34	Marketing and Public Relations	Online Advertising	Internet Advertising
35	Sales	General Sales	Cold Calling
36	Sales	Prospecting and Qualification	Product Promotion
37	Sales	Specialized Sales	Furniture Sales
38	Supply Chain and Logistics	Facility Management and Maintenance	Facility Repair
39	Supply Chain and Logistics	General Shipping and Receiving	Shipping Methods
40	Supply Chain and Logistics	Material Handling	Forklift Operation

Figure B.5: Sample Skills in Burning Glass 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.5 Wage Elasticities Estimated off of Switchers

Figure B.6: Productivity-Wage Elasticity Estimates by Occupation, Switchers



To supplement our estimates from productivity shocks, we estimate the effect on a worker's wages from switching jobs across firms of different productivity levels, conditional on staying in the same occupation group over the course of the move. We use the following specification, running the regression separately for each occupation group j:

$$d\log(\mathrm{w}_{ij,t+r,t+s}) = \chi_0^j + \chi_1^j \log(\overline{VA}_{i,k(t)}^{resid}) + \chi_2^j \log(\overline{VA}_{i,k(t+1)}^{resid}) + X_{it}^j \gamma + e_{it}. \tag{25}$$

The timing assumption is that the worker switches jobs between time t and $t+1$. We estimate this wage growth regression over three different time horizons: t to $t+1$, $t+1$ to $t+3$, and cumulatively from t to $t+3$. This allows us to decompose the timing of when firm wage premia are accumulated if wage premia are backloaded, which we do indeed find.

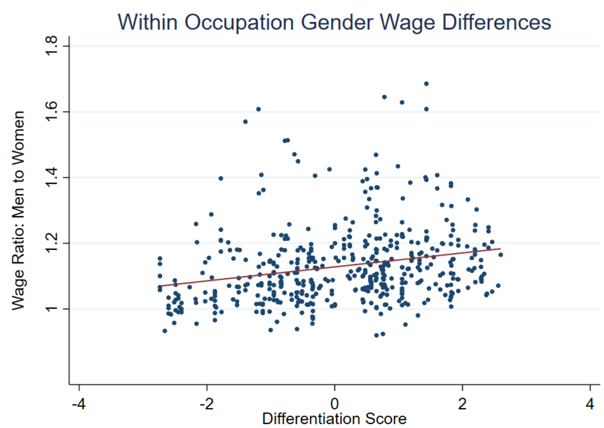
We present the results in figure B.6. For each occupation group, there are six plotted coefficients. The top three coefficents plot χ_2^j, which shows the effect on wage growth from $t+r$ to $t+s$ of the average productivity level of the *arrival* firm. The bottom three coefficents plot χ_1^j, which shows the effect on wage growth from $t+r$ to $t+s$ of the average productivity level of the *leaving* firm. To interpret these coefficients, consider the case for managers in the top left panel. In the year of the switch, between time t and $t+1$, workers who leave a firm 10 log points more productive than the average will see wage growth that is 0.5 log points lower than a worker who leaves an average firm, suggesting an elasticity of .05. However, the productivity of the arrival firm seems to matter little - the elasticity is below .02 and is on the edge of statistical significance. Next, consider the dot second from the top in the managers panel: this is the coefficient on wage growth from $t+1$ to $t+3$ due to the productivity of the arrival firm. This coefficient being positive means that workers who switch into productive firms see faster wage *growth* in the years following the switch to higher productivity firms.

Finally, the 3rd and 6th dots in each panel plot the cumulative effect on wage growth from t to $t+3$ from arriving at a more productive firm and leaving a more productive firm, respectively. In general, the estimate coefficents follow a similar pattern as for the passthrough coefficients: wages for switchers in managers and construction/manufacturing occupations appear to be more affected by the productivity of the firm than for professionals/technicians and administrative/service/manual workers. Wages for managers and professionals/technicians are particularly backloaded.

B.6 Task Differentiation and Within-Occupation Gender Wage Gap

An even larger portion of the gender wage gap is within, rather than across, occupations. Figure B.7 shows the raw hourly wage gap between men and women within each detailed 4-digit occupation, where on the x-axis is the 4-digit occupation's task differentiation score. The average within-occupation gap is 13 percent, and a typical occupation with differentiation scores one standard deviation above and below will have within-occupation wage gaps of 10.5 percent and 15.5 percent, respectively. Table 12 reports a regression of the within-occupation gender wage gap on the differentiation score of an occupation and the average years of education of workers in the occupation. When both occupation characteristics are included, we find that only the differentiation score predicts the within-occupation gender wage gap.

Figure B.7: Task Differentiation and Within-Occupation Gender Wage Gap



This figure plots the ratio of unadjusted average hourly of men divided by average hourly earnings of women for each 4-digit DISCO occupation on the y-axis. On the x-axis is the measure of differentiation from section

Table 12: Within-Occupation Gender Wage Gap: Differentiation vs. Education

Outcome Variable	Gap	Gap
Differentiation Score	.028**	.025***
	(.006)	(.006)
Mean Educ Years		.005
		(.004)
Observations	525	525
R-squared	.042	.045

Robust standard errors in parentheses.

** p<0.01, * p<0.05.

This table reports a regression where the unit of observation is the occupation and the outcome variable is the ratio of raw hourly wages of men to women in each 4-digit DISCO code. The dependent variables are the differentiation score of an occupation and the average years of education of workers in the occupation. The differentiation scores standardized to mean 0 and standard deviation 1. We restrict to occupations with at least 1000 person×years of employment in the occupation for both men and women from 2008-2016. The unweighted within-occupation mean gender wage gap is 13% among occupations in this sample.