Understanding High-Wage and Low-Wage Firms*

Horng Chern Wong†

Latest version here.

March 2021

Abstract

Some firms pay higher wages than others for identical workers. To unpack the firm wage premium distribution, I develop and implement a new structural decomposition using datasets that cover the universe of employers and employees in France. Existing research shows that firm wage premia depend on firms' *labor productivity* and *wage-setting power*. This paper shows that they also depend on firms' *product market power* and *labor share of production*. I find these firm characteristics to be correlated with each other, implying that: (i) without taking the latter set of characteristics into account, workhorse models that generate firm wage premia overestimate the contributions of firms' labor productivity and wage-setting power; (ii) conventional measures of input misallocation overestimate the degree of labor misallocation; (iii) high productivity firms have low labor shares of revenue not only because of greater product market power, but also greater labor market power and low labor shares of production.

Keywords: firm heterogeneity, wage inequality, market power, production technology, labor share
JEL codes: D24, D33, E2, J3, J42

---

*First version: August 2019. Special thanks to my supervisors, Thijs van Rens and Roland Rathelot, for their invaluable guidance, support, and patience. I am very grateful to Dennis Novy and Vincent Sterk for numerous constructive feedback and generosity with their time. For very helpful comments, I thank Treb Allen, Jesper Bagger, Pablo Beker, Dan Bernhardt, Paula Bustos, Christine Braun, Luis Candelaria, Danilo Cascaldi-Garcia, Emma Duchini, Dita Eckardt, Jan Eeckhout, James Fenske, Joaquin García-Cabo, Monika Gehrig-Merz, Basile Grassi, Daniel Habermacher, Simon Hong, Omiros Kouvavas, Francis Kramarz, Alice Kuegler, Thomas Le Barbanchon, Eui Jung (Jay) Lee, Sang Yoon (Tim) Lee, Timo Leidecker, Norman Loayza, Bogdan Marcu, Isabelle Mejean, Andrew Oswald, Roberto Pancrazi, Ayush Pant, Carlo Perroni, Josep Pijoan-Mas, Franck Portier, Fabien Postel-Vinay, Camilla Roncoroni, Federico Rossi, Alessandro Ruggieri, Ayşegül Şahin, Marta Santamaria, Sylvia Sarpietro, Jesse Shapiro, Ben Smith, Isaac Sorkin, Douglas Staiger, Liliana Varela, Marija Vukotic, Jonathan Yeo, and the many seminar and conference participants at Warwick, UCL, World Bank DECRG (KL), Bocconi, Brown, CEMFI, CERGE-EI, CREST, Dartmouth, Manchester University, Stockholm University, IEE Geneva, IESR Jinan, SMYE, RES, SOLE, EEA-ESEM, Queen Mary, and Royal Holloway. I thank the Department of Economics at University College London, where part of this work was completed, for their hospitality. I gratefully acknowledge financial support from the ESRC and data access from DARES, DGFiP, INSEE, and CASD.

†CREST, École Polytechnique. Email: horngchernwong@gmail.com

1 Introduction

Some firms pay higher wages than others for identical workers. This is known as the *firm wage premium*. Following the pioneering work of Slichter (1950) and Abowd, Kramarz, and Margolis (1999), a large body of empirical research confirms this finding in a number of countries.[1] The firm wage premium distribution plays an important role in explaining a range of labor market phenomena, from classic questions such as the long-term wage loss of displaced workers (Schmieder, Von Wachter, and Heining, 2018) and the gender wage gap (Card, Cardoso, and Kline, 2015), to recent questions about how globalization (Dauth, Findeisen, and Suedekum, 2018) and the rise of "superstar" firms (Song et al., 2019) impact the wage distribution. Firm wage premia also affect aggregate productivity by reallocating workers from low-wage firms to high-wage firms (Haltiwanger, Hyatt, Kahn, and McEntarfer, 2018; Bilal, Engbom, Mongey, and Violante, 2019).

What determines the firm wage premium distribution? This paper offers a new structural decomposition of firm wage premia to better understand the firm-level drivers. To interpret standard regression-based estimates of firm wage premia, I build a structural framework in which (i) labor market frictions prevent firm wage premia from being competed away, and (ii) firm heterogeneity determines how much a firm is willing to pay to hire a given worker, compared to other firms.[2]

I use the model to show that firm wage premia depend on firms' labor productivity, wage-setting power, product market power, and labor share of production.[3] Existing work on firm wage premia largely emphasizes the role of differences in firms' *labor productivity* and *wage-setting power*.[4] Yet, recent research on the macroeconomics of resource allocation and labor's share of income document important differences in firms' *product market power* and *labor share of production*, which are key determinants of firms' labor demand.[5] Since these firm characteristics are often studied separately, little is known about their interrelationships or their relative importance for the firm wage premium distribution. Using rich administrative datasets on the universe of employers and employees in France, I estimate each of these firm characteristics. I then combine the model with my estimates to unpack the firm wage premium distribution.

My central finding is that differences in product market power and the labor share of production are quantitatively important, accounting for 11% and 29% of the firm wage premium

[1]See, for example, Card, Heining, and Kline (2013), Song, Price, Guvenen, Bloom, and Von Wachter (2019), and Alvarez, Benguria, Engbom, and Moser (2018).

[2]See, for example, Burdett and Mortensen (1998), Postel-Vinay and Robin (2002), and Bagger, Christensen, and Mortensen (2014).

[3]For now, "labor share of production" refers to the labor exponent in a Cobb-Douglas production function.

[4]See, for example, Card, Cardoso, Heining, and Kline (2018) and Caldwell and Harmon (2019).

[5]For example, Karabarbounis and Neiman (2014) argues that aggregate capital-labor substitution drives the U.S. labor share decline, while De Loecker, Eeckhout, and Unger (2020) argue that product market power is key.

distribution. These dimensions have received little attention in the firm wage premium literature so far.[6,7] Without taking them into account, workhorse models of frictional labor markets overestimate the role that firms' labor productivity and wage-setting power differences play in explaining firm wage premia. This is because these firm characteristics are correlated with firms' product market power and labor share of production. Differences in firm wage premia tend to reallocate workers from low-wage firms to high-wage firms (Haltiwanger et al., 2018). To the extent that the role of labor productivity differences is overestimated, the extent to which workers reallocate from less productive firms to more productive firms would be overstated – a process that is key for aggregate efficiency. On the other hand, dispersion in wage-setting power leads to misallocation of labor (Azkarate-Askasua and Zerecero, 2020; Berger, Herkenhoff, and Mongey, 2020). Overestimating the role of wage-setting power differences would imply more room for labor market policy interventions than is warranted.

The correlations between the estimated firm characteristics have implications for understanding labor misallocation and the low labor share of revenue among high productivity firms. In particular, the negative correlation between labor productivity and the labor share of production: (i) implies that conventional measures of labor misallocation (Hsieh and Klenow, 2009) based on revenue per hour overstate the extent of aggregate productivity gains from removing labor market frictions, because these measures do not account for firms' ability to sidestep these frictions by substituting labor with other inputs; (ii) provides a new explanation for why more productive firms have lower labor shares of revenue besides market power (De Loecker et al., 2020).

To develop a structural decomposition of firm wage premia, I begin by building a structural model to interpret regression estimates of firm wage premia. In the model, labor market frictions sustain firm wage premia and firms are endogenously different from each other along multiple characteristics.[8] As in workhorse frictional labor market models, firms differ in labor productivity and wage-setting power (Burdett and Mortensen, 1998). Wage-setting power is defined as the fraction of marginal revenue product of labor paid as wages. I refer to this fraction as the *wage markdown*.[9] Compared to these models, the new features of my framework are differences in product market power and the labor share of production. Product market power refers to firms' *price-cost markups* and the labor share of production refers to the firm-specific *output*

[6] See Manning (2011) and Card et al. (2018) for surveys of the literature.

[7] Indeed, since research on the labor share tend to work with perfectly competitive labor markets, firm wage premia do not exist in those settings.

[8] Since these differences in firm characteristics are equilibrium outcomes, I also refer to them as "channels of firm heterogeneity".

[9] In a perfectly competitive labor market, wages equal the marginal revenue product of labor, therefore there are no wage markdowns. In a frictional labor market, the wage markdown can vary across firms due to differences in labor supply elasticities, outside options, or relative bargaining positions (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Berger et al., 2020).

elasticity with respect to labor inputs.[10] These new features introduce firm-specific downward-sloping labor demand curves into a standard frictional labor market framework. I then obtain a structural equation linking firm wage premia to these firm characteristics.

To estimate these firm characteristics, particularly the distribution of wage markdowns across firms, I develop a new approach by combining empirical methods from industrial organization and labor economics. Wage markdowns have received increasing attention but their cross-sectional properties are not yet well-documented.[11,12] I do so by building on the production-based approach of De Loecker and Warzynski (2012) to accommodate imperfectly competitive labor markets and worker heterogeneity, which involves panel data methods commonly used in labor economics (Abowd et al., 1999; Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler, 2020). This approach has the advantage that it does not require the researcher to specify particular market structures in a wide array of product and labor markets. To separately estimate output elasticities from productivity, I estimate production functions using a control function approach (Ackerberg, Frazer, and Caves, 2015), in which I use firms' past input choices to instrument for their current choices under the following timing assumption: firms' past input choices are orthogonal to current productivity shocks. To separately identify firms' wage markdowns from price-cost markups, I exploit the fact that labor market power is a distortion only on labor demand, while product market power is a common distortion on the demand for each input.[13]

I use large administrative datasets from France, covering the population of employers and employees between 1995 and 2014. Estimating firm wage premia and firm heterogeneity requires detailed information about workers and firms. I estimate the former using matched employer-employee panel data, which includes key information on hourly wages and employer identifiers for over 25 million workers per year. I estimate the latter using firm balance sheet panel data, which contains information on gross production, employment, capital, and intermediate inputs for over 2 million firms per year. The main advantages of these distinct datasets from France are that they are jointly available and are not limited to manufacturing or large firms.[14,15]

Understanding the extent of firms' wage-setting power is first-order for assessing its welfare implications (Azkarate-Askasua and Zerecero, 2020; Berger et al., 2020). Even with a high

[10] With a Cobb-Douglas production function $Y = K^{1-\alpha}L^{\alpha}$, the labor elasticity of output is α.

[11] A growing literature documents monoposonistic labor market competition. See Azar, Marinescu, Steinbaum, and Taska (2018), Rinz (2018), and Lamadon, Mogstad, and Setzler (2019).

[12] For recent model-based estimates of wage markdowns, see Berger et al. (2020), Webber (2015), and Tortarolo and Zarate (2018).

[13] De Loecker and Warzynski (2012) developed the insight that markups can be identified from the fact that it is a common distortion on each of the firm's input demand.

[14] Examples of the few countries with both types of data include Brazil, Denmark, Norway, and Sweden.

[15] Balance sheet data are often only available for large firms (e.g. Compustat) or manufacturing firms (e.g. Germany, Mexico, and Colombia). This is an important concern since manufacturing employment is declining in many countries.

national minimum wage by international standards and near universal coverage of collective bargaining agreements, my estimates suggest that French firms hold considerable wage-setting power – the median firm marks down wages by approximately 20%. These wage markdowns are far from uniform. Firms at the 75^{th} percentile of the wage markdown distribution pay 97% of the marginal revenue product of labor as wages, while those at the 25^{th} percentile pay 67%.

Firms also display large differences in the labor elasticity of output. The interquartile range is 0.16. Consistent with existing sector-level estimates, I find moderate dispersion in sectoral output elasticities (Basu, Fernald, Fisher, and Kimball, 2013; Oberfield and Raval, 2020); most of the dispersion occurs within sectors, with an average within-sector interquartile range of 0.14. Similarly, my estimates of large productivity and price-cost markup dispersion across firms are also in line with the existing literature (Syverson, 2011; De Loecker et al., 2020).

With these estimated firm characteristics in hand, I use the structural firm wage premium equation to decompose its empirical distribution. To maximize interpretability, my decomposition allocates each dimension its marginal contribution. The decomposition asks: "how much can differences in a particular firm characteristic account for firm wage premia, holding other characteristics constant?" I find that wage markdowns, labor productivity, price-cost markups, and labor elasticities of output contribute 25%, 35%, 11%, and 29% to firm wage premium dispersion. These results indicate that firm characteristics at the center of the labor share debate – price-cost markups and labor elasticities of output – are quantitatively important drivers of the firm wage premium distribution. Without taking them into account, the model suggests that the explanatory power of firms' labor productivity and wage-setting power would be overstated, accounting for up to 53% and 47%, of the firm wage premium distribution.

Yet, the dispersion of firm wage premia in the data is moderate compared to the large differences in estimated firm characteristics. Firm wage premia account for 4.5% of total wage dispersion while existing work typically finds a number between 10% and 20% (Card et al., 2013; Alvarez et al., 2018; Song et al., 2019). This difference is a result of more precise estimates of firm wage premia upon addressing a well-known estimation bias due to the lack of worker mobility (Andrews, Gill, Schank, and Upward, 2008), consistent with recent work by Bonhomme, Lamadon, and Manresa (2019) and Lamadon et al. (2019). Nevertheless, the distribution of firm wage premia is quantitatively important; the 90-10 ratio of firm wage premia of 1.25 is comparable to the gender wage gap in Japan, which is among the highest in OECD countries.

The main explanation for the coexistence of substantial firm heterogeneity and relatively moderate firm wage premium dispersion is the negative correlation between labor productivity and labor elasticities of output. At the same time, more productive firms appear to have higher intermediate input and capital elasticities of output than less productive firms. Through the lens of the structural framework, this empirical pattern suggests that more productive firms are more likely to substitute labor with other factor inputs.[16] The reason is that more productive firms have higher labor demand, but the presence of labor market frictions imply that firms face an upward-sloping labor supply curve: firms must pay higher wages to attract more workers. If labor and other inputs are (imperfect) substitutes, then more productive firms tend to substitute away from labor inputs to avoid a higher relative cost of labor. The ability of more productive firms to substitute labor with other inputs partially offsets their higher labor demand relative to less productive firms. This reduces their willingness to pay higher wage premia to compete in hiring workers.

This negative correlation between firms' labor productivity and labor elasticities of output has important implications for measuring the allocative efficiency of labor inputs. The variance of the marginal revenue product of labor is a sufficient statistic for labor misallocation (Hsieh and Klenow, 2009), commonly proxied for by the average revenue product of labor (value-added or revenue per worker or per hour). While this proxy is widely available, it overstates the variance of the marginal revenue product of labor by about three times. This mismeasurement overstates the aggregate productivity and output gains of removing labor market power distortions. Revenue per worker is an accurate proxy of the marginal revenue product of labor only when labor elasticities of output and price-cost markups are constant across firms within a given sector. However, the inverse relationship between labor productivity and labor elasticities of output suggests that firms that are more constrained by labor market frictions can circumvent these frictions by substituting labor with other inputs.

High productivity firms also matter for the labor market because they play a key role in driving the aggregate labor income share (Autor, Dorn, Katz, Patterson, and Van Reenen, 2020; Kehrig and Vincent, 2020). My previous finding offers a new explanation for their low labor shares of revenue in addition to product and labor market power: lower labor elasticities of output. However, consistent with the existing literature, I also find that such firms charge higher price-cost markups (De Loecker et al., 2020). My estimates also provide empirical support for the hypothesis that high productivity firms have considerable labor market power (Gouin-Bonenfant, 2020). Quantitatively, a decomposition exercise suggests that low labor intensity production processes explain much of why highly productive firms have low labor shares of revenue, followed by important contributions from wage markdowns and price-cost markups. In an extention with high and low-skill labor, I show that the wage markdowns and labor elasticities of output among low-skill, rather than high-skill labor explain firm level labor shares.

Contributions to related literature. A large literature in labor economics estimates the

[16] For example, Goldschmidt and Schmieder (2017) show that firms that pay a wage premium relative to other firms outsource part of their production process. As a result, outsourced workers receive lower wages as they lose the wage premium paid by their previous employer.

separate contribution of workers and firms to the wage distribution (Abowd et al., 1999). The finding that different firms pay identical workers differently has been replicated in a number of countries, such as Brazil (Alvarez et al., 2018), Denmark (Bagger et al., 2014; Lentz, Piyapromdee, and Robin, 2019), Germany (Card et al., 2013), Portugal (Card et al., 2018), USA (Song et al., 2019; Sorkin, 2018; Lamadon et al., 2019), and Sweden (Bonhomme et al., 2019). While the estimated firm fixed effects, known as the firm wage premium, do not entail a structural interpretation, a few recent papers provide one (Bagger et al., 2014; Lamadon et al., 2019). These studies provide fully microfounded models to study counterfactual scenarios. My paper differs by imposing just enough structure on the data to unpack the firm wage premium distribution. This approach allows me to include a richer variety of firm heterogeneity.

My paper also relates to the literature by showing that the structural firm wage premium decomposition speaks to broader recent work on the impact of productivity dispersion (Berlingieri, Blanchenay, and Criscuolo, 2017), labor market power (Berger et al., 2020; Azar et al., 2018), product market power (De Loecker et al., 2020), and the aggregate production technology (Karabarbounis and Neiman, 2014) on wages and the aggregate labor share. I discuss each below. The most closely related paper is Mertens (2019), who studies how manufacturing firms' production technology and market power explain the decline of the German manufacturing labor share. In contrast, my paper studies the role of multiple channels of firm heterogeneity in determining firm wage premia and highlights the importance of the relationships between each channel.

A large strand of work examines firm productivity and rent-sharing as a driver of wage inequality between firms. There is ample evidence that firms share rents with employees, therefore firm productivity determines the wages they pay (Katz and Summers, 1989; Blanchflower, Oswald, and Sanfey, 1996; Hildreth and Oswald, 1997; Carlsson, Messina, and Skans, 2014; Kline, Petkova, Williams, and Zidar, 2017; Bell, Bukowski, and Machin, 2018; Garin and Silverio, 2019). Recent papers study the link between widening firm productivity distribution and rising between-firm wage inequality (Faggio, Salvanes, and Van Reenen, 2007; Berlingieri et al., 2017). My paper contributes by quantifying the extent to which rent-sharing explains firm wage premia.

Motivated by high levels of concentration in labor markets, a growing number of researchers study the effect of labor market power on wages. Dube, Jacobs, Naidu, and Suri (2018), Naidu, Nyarko, and Wang (2016) provide evidence for labor market monopsonistic competition. Recent papers by Azar et al. (2018), Benmelech, Bergman, and Kim (2018), and Abel, Tenreyro, and Thwaites (2018) study the effects of labor market concentration on wages. Berger et al. (2020), Jarosch, Nimczik, and Sorkin (2019), Brooks, Kaboski, Li, and Qian (2019) each provide a structural framework that links labor market concentration to wages. Gouin-Bonenfant (2020) studies how the firm productivity distribution affects the aggregate labor share through labor market imperfect competition in a wage-posting model. Caldwell and Danieli (2019) and Caldwell and Harmon (2019) study the effects of relative bargaining positions on wages. This paper complements the work of Hershbein, Macaluso, and Yeh (2020), who independently develop a closely related methodology to measure wage markdowns among U.S. manufacturing firms. My paper adds to this literature by (i) documenting the economy-wide distribution of firm-specific wage markdowns that apply to a subset of wage-posting and wage-bargaining models, (ii) quantifying the importance of wage markdowns for firm wage premia, and (iii) showing that highly productive firms tend to markdown wages more than other firms.

Recent work on the labor share of national income focuses on the role of product market power (price-cost markups) and the labor intensity of the aggregate production technology (labor elasticity of output). Elsby, Hobijn, and Şahin (2013) discuss the role of outsourcing in the US labor share decline, while Karabarbounis and Neiman (2014) and Hubmer (2019) focus on capital-labor substitution. Barkai (2020) makes the case for growing product market power as a key driver. Autor et al. (2020) and Kehrig and Vincent (2020) show that the falling US labor share is due to the rising market share of high productivity firms with low labor shares of revenue. De Loecker et al. (2020) show that low labor share firms charge high markups. My contribution is to show that wage markdowns and labor elasticities of output also account for signifcant shares of firm level differences in labor shares.

Road map. In Section 2, I present the structural framework for firm wage premia. In Section 3, I describe how the structural firm wage premium equation is estimated. Section 4 provides information on the French administrative datasets. In Section 5, I present the main findings. Section 6 discusses the implications. The conclusion is in Section 7.

2 A Framework to Decompose Firm Wage Premia

I present a wage-posting framework with frictional labor markets from which I derive an equation for the firm-specific wage premium. The framework is a dynamic version of the Manning (2006) generalized model of monopsony augmented with imperfectly competitive product markets and a general production function with capital, labor, and intermediate inputs. The model is set in partial equilibrium. I impose just enough structure on this framework to allow a number of endogeneously determined channels of firm heterogeneity, but leave unspecified many of the primitives governing the equilibrium outcome of the model, such as parametric distribution functions for productivity or the product market structure.

There are two main ingredients in this framework – labor market frictions and firm heterogeneity. Labor market frictions imply that workers cannot instantaneously find another job and hiring is costly for firms, allowing a distribution of firm-specific wage premium to survive. Firm heterogeneity then determines the wage premium a firm is willing to pay to hire workers of a given skill level.

In Appendix D, I show that the firm wage premium equation can be derived from a wage-bargaining protocol and under a few distinct microfoundations for imperfectly competitive labor markets.[17] I pursue some degree of flexibility because both wage-setting protocols are used by firms in reality (Hall and Krueger, 2010). Wage-setting throughout this paper is contemporaneous.

2.1 Departures from standard frictional labor market models

The first departure is that the goods market is imperfectly competitive. Firms face downward-sloping demand curves and are able to set their own prices. Each firm j faces an inverse product demand curve:

$$P_{jt} = \tilde{D}_s(Y_{jt}, D_{jt}) \tag{1}$$

where P_{jt} denotes the price charged by firm j in sector s at time t, Y_{jt} denotes the firm's output, and D_{jt} denotes the firm's idiosyncratic demand. The demand function is twice differentiable, with $\tilde{D}_{s,y} < 0$ and $\tilde{D}_{s,yy} > 0$. The firm's idiosyncratic demand D_{jt} can be a function of aggregate, sectoral, or firm-specific demand shifters. The assumption of imperfectly competitive goods markets generates a distribution of firm-specific price-cost markups, an important determinant of firms' labor demand in the macroeconomics of resource allocation and the labor share of income (Edmond, Midrigan, and Xu, 2015; Peters, 2020; De Loecker and Eeckhout, 2018).

The second departure is that firms operate a general production function with diminishing marginal returns to each input, instead of a constant-returns-to-labor production function:

$$Y_{jt} = X_{jt} F_{st}(K_{jt}, H_{jt}, M_{jt}) \tag{2}$$

I assume that this production function is sector-specific and twice differentiable. X_{jt} is the Hicks neutral productivity term, which is subject to the following autoregressive process $\ln X_{jt} = G(\ln X_{jt-1}) + \epsilon^x_{jt}$ where ϵ^x_{jt} is a random productivity shock. K_{jt}, H_{jt}, and M_{jt} denote capital, efficiency units of labor, and intermediate inputs, at firm j at time t. Efficiency units of labor can be written as $H_{jt} = \bar{E}_{jt} L_{jt}$, where $\bar{E}_{jt}$ denotes average efficiency and L_{jt} denotes amount of labor. By allowing diminishing marginal returns to labor and not restricting the elasticity of substitution between any pair of factor inputs, I allow the elasticity of output with respect

[17] I show that this framework can be microfounded by a random search or directed search model of frictional labor markets. I also derive the firm wage premium from a monopsonistic model based on workplace differentiation (Card et al., 2018).

to each input to differ across firms.[18] I discuss what these output elasticities and price-cost markups depend on in the next subsection.

2.2 Deriving the firm-specific wage premium equation

Time is discrete. Capital and intermediate input markets are perfectly competitive. Firms can hire more workers by paying higher wages, as in monopsony models such as Robinson (1933) and Burdett and Mortensen (1998). In addition, firms can also increase recruitment effort, as in job search models such as Diamond (1982), Mortensen (1982), and Pissarides (1985). Each firm j posts piece-rate wages per efficiency unit of labor (Barlevy, 2008; Engbom and Moser, 2018; Lamadon et al., 2019), denoted Φ_{jt}. A worker i with efficiency E_{it} obtains a wage $W_{it} = E_{it}\Phi_{jt}$. Taking logs, this wage equation maps into the classic two-way fixed effect ("AKM" henceforth) regression model due to Abowd et al. (1999), $w_{jt} = e_{jt} + \phi_{jt}$, where lowercase letters denote variables in logs. This regression is estimated in section 3.2. The piece-rate wage (Φ) is therefore the *firm-specific wage premium.*

Firm j's effective labor is subject to the following law of motion:

$$H_{jt} = (1 - s_{jt})H_{jt-1} + R_{jt} \tag{3}$$

with:

$$s_{jt} = s(\Phi_{jt}, A_{jt}) \tag{4}$$

$$R_{jt} = R(\Phi_{jt}, A_{jt}, V_{jt}) \tag{5}$$

where s_{jt} denotes its worker separation rate, which is allowed to depend on the firm-specific wage premium Φ_{jt} and non-wage characteristics A_{jt}. I assume that $s(.)$ is twice differentiable in Φ, $s_{\Phi}(.) < 0$ and $s_{\Phi\Phi}(.) > 0$. Firms' recruitment size in efficiency units (R_{jt}) depends on its posted wage, its non-wage characteristics, and its recruitment effort (V_{jt}). I assume that the recruitment function $R(.)$ is twice differentiable and monotonically increasing in its wages, value of non-wage characteristics, and recruitment effort, with diminishing marginal returns. Therefore, all else equal, firms that offer higher wages and better non-wage amenities have a higher recruitment rate and lower separation rate.

The assumption that firm-specific separation and recruitment rates depend on the wages offered is informed by models of on-the-job search such as Burdett and Mortensen (1998) and Mortensen (2010), or directed search models such as Kaas and Kircher (2015). I also allow recruitment and separation to depend on non-wage amenities, as there is evidence that non-wage

[18] Moreover, diminishing marginal returns implies that the notion of firms in this framework is based on optimal firm sizes. In contrast, firms are a collection of jobs with the same productivity in standard frictional labor market models with linear production functions in labor.

amenities are important determinants of worker flows between firms (Sorkin, 2018). Together, equations (3), (4), and (5) form the firm-specific upward-sloping labor supply curve.

Firms' recruitment efforts are subject to recruitment costs $c(V_{jt})$. I assume that the recruitment cost function is twice differentiable, and that $c_V(.) > 0$ and $c_{VV}(.) > 0$, so that the marginal cost of recruitment effort is increasing in recruitment.

Firm j's profit maximization problem can be written as:

$$\Pi(X_{jt}, D_{jt}, A_{jt}; K_{jt-1}, H_{jt-1}) = \max_{P_{jt}, I_{jt}, M_{jt}, \Phi_{jt}, V_{jt}} P_{jt}Y_{jt} - R_t^K K_{jt} - P_t^M M_{jt} - \Phi_{jt}H_{jt} - c(V_{jt})V_{jt}$$
$$+ \beta E_t[\Pi(X_{jt+1}, D_{jt+1}, A_{jt+1}; K_{jt}, H_{jt})]$$

subject to (1), (2), (3), (4), and (5). Let R_t^K and P_t^M denote the competitive price of capital and intermediate inputs. The timing of events is as follows. First, firms obtain an idiosyncratic draw of productivity and demand. Then, firms post wages, exert recruitment effort, and employ workers and other inputs. Finally, firms produce. This timing assumption is consistent with the recent class of multiworker firm models (for example, Kaas and Kircher (2015), Elsby and Michaels (2013), and Schaal (2017)).

Solving for the first-order condition with respect to Φ gives the firm-specific wage premium equation:

$$\Phi_{jt} = WM_{jt} \cdot ARPH_{jt} \cdot PM_{jt}^{-1} \cdot LEO_{jt} = WM_{jt} \cdot MRPH_{jt} \tag{6}$$

which is a log-linear function of four channels of firm heterogeneity. The last three components of this equation form the marginal revenue product of labor ($MRPH$). I discuss each component of the equation below.

Wage markdown (WM). This component is the fraction of marginal revenue productivity of labor paid as wages. It measures the wage-setting power of firms and it can be written as:

$$WM_{jt} = \frac{\epsilon_{jt}^H}{1 + \epsilon_{jt}^H - \beta E_t \left(\frac{(1-s_{jt+1})J_{jt+1}}{c_{V,jt}V_{jt} + c(V_{jt})} \right) R_{V,jt}} \tag{7}$$

where $\epsilon_{jt}^H = \epsilon^H(\Phi_{jt}, a_{jt}, V_{jt})$ is the firm-specific labor supply elasticity, $c_{V,jt}V_{jt} + c(V_{jt})$ is the marginal recruitment cost, and J_{jt+1} is the marginal profit to the firm of having an additional worker next period. Equation (7) shows that firms facing lower labor supply elasticities possess stronger wage-setting power, and therefore post wages further below the marginal revenue product of labor. The firm-specific labor supply elasticity (ϵ_{jt}^H) can be further decomposed into:

$$\epsilon_{jt}^H = \frac{R_{jt}}{H_{jt}} \epsilon_{\Phi,jt}^R - \frac{s_{jt}H_{jt-1}}{H_{jt}} \epsilon_{\Phi,jt}^s > 0$$

which is a function of the wage elasticity of recruitment ($\epsilon^R_{\Phi,jt} > 0$) weighted by the share of new recruits in the firm, net of the wage elasticity of separations ($\epsilon^s_{\Phi,jt} < 0$) weighted by the employee share of separated workers. The second component in the denominator is the expected discounted marginal profits to the firm of an additional worker next period relative to recruitment costs. This component shows that firms expecting a high marginal value of a worker next period are willing to pay a higher wage markdown in the current period.

Equation (7) nests static monopsony models in the tradition of Robinson (1933), in which firms use wages as the sole instrument for hiring workers. In this case, firms' hiring is constrained by their labor supply curves.[19] The wage markdown then reduces to:

$$WM_{jt} = \frac{\epsilon^H_{jt}}{1 + \epsilon^H_{jt}}$$

which is simply a function of labor supply elasticities.

The specific functional form for labor supply elasticities (ϵ^H_{jt}) depends on the microfoundation for firm-specific labor supply curves (formed by equations (3), (4), and (5)) pursued by the researcher. In appendix D, I show in a random search and a directed search wage-posting model with on-the-job search that this elasticity depends on the elasticity of the job-filling and separation rates with respect to wages. In an oligopsonistic model in which upward-sloping labor supply curves are microfounded by workplace differentiation, I show that the firm-specific labor supply elasticity depends on the firm's labor market share. Finally, in a random search model with wage-bargaining, I show that the labor supply elasticity in the wage markdown replaced by a function of relative bargaining power and workers' value of outside options.

Although I do not take a stance on the joint distribution of the heterogeneity in primitives (idiosyncratic productivity (X), demand (D), and non-wage amenities(A)), I now discuss how these primitives map into wage markdowns. Consider two firms that are identical along all dimensions, but one has higher productivity than the other. Then the firm with the higher productivity will have a higher labor demand and pay higher wages (Φ). Since the more productive firm pays higher wages, it locates itself at the part of the labor supply curve where the labor supply elasticity (ϵ^H) is lower: it faces less labor market competition locally compared to the less productive firm. The lower labor supply elasticity reflects the lower recruitment (ϵ^R_Φ) and separation elasticity with respect to wages (ϵ^s_Φ): the high-wage firm cannot raise the recruitment rate and reduce the separation rate by much if it offers yet higher wages, since it already pays the highest wages. The wage markdown is therefore lower. The same is true in a comparison of two firms which are identical along every dimension except idiosyncratic demand

[19] As shown by Manning (2006), one can think of this as a case in which any firm j faces no recruitment costs if it wishes to hire a number of workers below or at the level supplied at a given wage premium Φ_{jt}, but faces an infinite recruitment cost should it wish to hire more than that.

(D). The prediction that more productive firms have lower wage markdowns is standard in monopsonistic or oligopsonistic models, such as Burdett and Mortensen (1998).

Next, consider two firms that are identical along every dimension except non-wage amenities (A). Then the high-wage firm is the one with less desirable non-wage amenities (lower A). In the model, non-wage amenities act as a labor supply shifter. The firm with less desirable amenities has a labor supply curve that is shifted inwards compared to the firm with better amenities. The former firm therefore faces a higher marginal cost of hiring a worker relative to the latter. As such, the firm with less desirable amenities pays higher wages and hires less workers, locating itself at the more elastic part of the labor supply curve. The firm with less desirable amenities therefore has a higher wage markdown.

It is worth noting that this structural framework nests a workhorse model of frictional labor markets - the Burdett and Mortensen (1998) model. This model will be a useful benchmark for interpreting some of the decomposition results in Section 5. To obtain the Burdett-Mortensen model from this framework, the following additional assumptions are needed:

- The labor market is characterized by search frictions and workers search on-the-job;
- The goods market is perfectly competitive and the production function is linear in labor;
- Firms attract new workers by posting wages only;
- Firms are in their steady state.

The first assumption takes a stand on the source of firms' monopsony power in the labor market. As Burdett and Mortensen (1998) show, the combination of search frictions and on-the-job search implies a non-degenerate wage distribution, even when workers and firms are homogenous.[20] The second assumption ensures that the firms' revenue functions exhibit constant marginal returns to labor. This assumption implies that the output elasticity with respect to labor is equal to 1 across all firms. The third assumption is standard in traditional monopsony models.[21] The fourth assumption implies that the wage markdown is only a function of the firm-specific labor supply elasticities. Under these assumptions, the firm's profit-maximization problem reduces to:

$$\Pi_j = \max_{\Phi_j}(X_j - \Phi_j)H(\Phi_j)$$

Therefore, the firm chooses a wage premium by trading off profits per worker and firm size. The firm wage premium is then $\Phi_j = WM_j \cdot ARPH_j$, where $WM_j = \frac{\epsilon^H(\Phi_j)}{1+\epsilon^H(\Phi_j)}$. This gives the

[20]For a proof of this classic result, I refer the reader to the original paper.

[21]One distinction between monopsonistic wage-posting models (Robinson, 1933) and search models with wage-bargaining (Diamond, 1982; Mortensen, 1982; Pissarides, 1985) is that in the former, wages are the only instrument firms use to hire workers, while in the latter, vacancy-posting is the sole instrument. Wages are decided before a match is formed (ex-ante) in wage-posting models, while in the latter, wages are set ex-post. In my framework, firms use both wages and vacancies (recruitment effort) to hire workers.

Burdett-Mortensen model.

Average revenue product of labor (ARPH). This is the theory-consistent measure of productivity for the firm wage premium.[22] It can be written as:

$$ARPH_{jt} = \frac{P_{jt}Y_{jt}}{H_{jt}}$$

which is the ratio of sales revenue over efficiency units of labor. The firm wage premium equation (6) shows that, all else equal, more productive firms pay a higher wage premium. This is because more productive firms make larger profits from an employment relationship due to labor market frictions. This is a standard prediction of models of imperfect labor market competition.

Since total revenue is increasing in firms' idiosyncratic productivity (X) and demand (D), all else equal, the average revenue product of labor is increasing in these underlying firm primitives. Moreover, the average revenue product of labor is decreasing in the value of non-wage amenities (A), all else equal. This is because the firm with less desirable amenities will have to pay higher wages to hire a given number of workers, reducing its total number of recruits. Since the production function satisfies diminishing marginal returns to labor, the average revenue product is higher for firms with less desirable amenities.

Price-cost markup (PM). This component captures firms' price-setting power. It can be written as:

$$PM_{jt} = \frac{\epsilon^G_{jt}}{\epsilon^G_{jt} - 1}$$

where which ϵ^G_{jt} is the firm-specific price elasticity of demand. The specific functional form for the price elasticity of demand depends on the researcher's microfoundation for the product demand curve (1). For example, with an oligopolistic competition market structure and a nested constant elasticity of substitution (CES) demand system, it depends on the firm's market share of sales (Edmond et al., 2015).[23] Equation (6) shows that, all else equal, firms with higher markups pay a lower wage premium. The intuition is that firms that are able to charge positive markups maximize profits by producing less than they would in the perfectly competitive benchmark, which reduces their labor demand and the wage premium they are willing to pay.

The price-cost markup is increasing in firms' idiosyncratic productivity (X) or demand (D). Consider two firms with different productivity X, but are otherwise identical. Then, the more

[22] This component is also commonly called "labor productivity".

[23] Since the demand function in equation (1) is static, the price-setting problem is also static. This is the most common formulation of product demand. However, there are increasingly used dynamic formulations, in which firms' price-setting decisions affect future demand, for example, due to customer accummulation (Gourio and Rudanko, 2014).

productive firm is able to produce with lower marginal cost and charges a lower price. The more productive firm therefore locates itself on the part of the product demand curve where the price elasticity of demand is lower: it faces less product market competition locally, since it charges lower prices than its competitors. The more productive firm therefore has a higher price-cost markup. Similarly, firms with higher idiosyncratic demand (D), all else equal, charge higher price-cost markups. This is because firms facing a higher demand for a given price faces lower price elasticity of demand.

If two firms are identical except for the value of their non-wage amenities (A), then the firm with less desirable amenities (lower A) will have lower price-cost markups. This is because the firm with less desirable amenities must pay comparatively higher wages to attract workers, implying a higher marginal cost of producing a given amount of goods. This firm therefore produces and sells less output at higher prices, locating itself on the part of the product demand curve where the price elasticity of demand is higher.

Labor elasticity of output (LEO). This component measures a firm's percentage increase in output from a one percent increase in labor inputs:

$$LEO_{jt} = \frac{\partial \ln Y_{jt}}{\partial \ln H_{jt}}$$

Equation (6) shows that firms for which output is highly elastic with respect to labor inputs pay a higher wage premium, all else equal. This is because firms with a higher labor elasticity of output have a higher labor demand.

To see what the firm-specific labor elasticity of output depends on, compare a sector-specific Cobb-Douglas and CES production function. For simplicity, assume that firms produce with only capital and labor inputs. The Cobb-Douglas production function is:

$$Y_j = H_j^{\alpha_s^H} K_j^{\alpha_s^K}$$

where α^H is the weight on labor inputs (which captures the rate of diminishing marginal returns in the Cobb-Douglas case). The labor elasticity of output in this case is sector-specific rather than firm-specific:

$$LEO_s = \alpha_s^H$$

The CES production function is:

$$Y_j = (\alpha_s^H H_j^{\sigma_s} + \alpha_s^K K_j^{\sigma_s})^{\frac{1}{\sigma_s}}$$

where σ_s is the elasticity of substitution between inputs. The labor elasticity of output is now firm-specific:

$$LEO_j = \frac{\alpha_s^H}{\alpha_s^H + \alpha_s^K (K_j/L_j)^{\sigma_s - 1}}$$

This comparison shows that the firm-specific labor elasticity of output depends on the (i) sector-specific input weights, (ii) sector-specific elasticity of substitution between any pair of inputs, and (iii) the firm-specific factor intensities (which depends on their relative cost). If capital and labor are substitutes ($\sigma > 1$), then the labor elasticity of output is decreasing in the capital-labor ratio, implying a faster rate of diminishing returns to labor.

Consider two firms that have different idiosyncratic productivities (X). Then the more productive firm has a higher labor elasticity of output if the elasticity of substitution between labor and other inputs is less than one (complements), while the opposite is true if the elasticity of substitution is greater than one (substitutes). As shown in Section 5, the empirically relevant case is the latter. The more productive firm wants to hire more workers to produce higher output. However, because of labor market frictions, firms must pay higher wages to hire more workers. Therefore, the more productive firm faces a higher relative cost of labor compared to the less productive firm. Since labor and other inputs are substitutes, the more productive firm substitutes labor with other inputs, increasing the capital-labor ratio and intermediate-input-labor ratio, reducing the labor elasticity of output. The same is true when two firms have different idiosyncratic demand (D), but are otherwise identical.

Similarly, if two firms have different values of non-wage amenities (A), but are otherwise identical, then the firm with the less desirable value of amenities (lower A) will have a lower labor elasticity of output. This is because this firm faces a higher relative cost of labor. If labor and other inputs are substitutes, then this firm will substitute labor with other inputs, reducing the labor elasticity of output.

Marginal revenue product of labor (MRPH). The last three components of the firm wage premium equation (6) form the marginal revenue product of labor. This component has two interpretations. In wage-posting models, such as the one presented here, wages are determined *before* forming an employment relationship. Therefore, the firm wage premium reflects a firm's willingness to pay for a worker of a given efficiency and the marginal revenue product of labor reflects the firm's labor demand. In wage-bargaining models, such as the one presented in Appendix D, wages are determined through bargaining over the total match surplus *after* matching. Since, all else equal, the total match surplus is larger for high marginal revenue product firms, bargained wages are also higher. Therefore, dispersion of the firm wage premium in a bargaining model due to differences in the marginal revenue product of labor reflects surplus sharing, holding wage markdowns constant across firms.

The fact that the marginal revenue product of labor depends on the average revenue product of labor, price-cost markup, and labor elasticity of output has important implications for its measurement. The dispersion of MRPH is important not only for wages, but also the efficiency of the allocation of labor across firms (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). Under the standard assumptions in frictional labor market models that the product market is perfectly competitive and production technologies exhibit constant returns to labor, the MRPH is equal to the ARPH. This simplifies measurement as the ratio of sales or value added per worker (hour) can be directly measured in most firm balance sheet or matched employer-employee datasets.[24] I explore these implications further in Section 5.

2.3 Discussion

The firm wage premium equation (6) has a few advantages. First, it features channels of firm heterogeneity related to the broader literature on between-firm wage inequality (Barth, Bryson, Davis, and Freeman, 2016; Song et al., 2019; Faggio et al., 2007; Bell et al., 2018; Berlingieri et al., 2017) and the labor share of income (Autor et al., 2020; Kehrig and Vincent, 2020; De Loecker and Eeckhout, 2018; Berger et al., 2020; Gouin-Bonenfant, 2020; Karabarbounis and Neiman, 2014; Hubmer, 2019) in a simple and transparent way. Second, the log-linear structure substantially simplifies a decomposition of the distribution of firm wage premia without requiring the researcher to fully specify and estimate the underlying primitives of the model, such as the joint distribution of firms' intrinsic productivity and non-wage amenities.

However, the following caveats apply. First, I consider only wage-setting protocols with a static nature: contemporaneous wage-posting and wage-bargaining. In doing so, I abstract from important wage-setting mechanisms such as the sequential auctions mechanism (see Postel-Vinay and Robin (2002) for a model in which firms Bertrand-compete in wages). This is because the introduction of diminishing returns to labor in a frictional labor market model comes with additional modelling complications on the wage-setting front. In particular, one will need to take into account the fact that the marginal product of labor changes when a worker leaves or joins a firm, which potentially triggers a renegotiation between the firm and other incumbent employees. This is also known as the Stole and Zwiebel (1996) problem. Moreover, on the empirical front, the sequential auctions wage-setting mechanism would violate the AKM identifying assumption of random mobility conditional on worker and firm fixed effects, since mobility would then also depend on the previous employer.[25] This restriction implies that I do not consider within-firm wage differentials due to within-firm worker heterogeneity in outside options. However, within-firm wage dispersion due to differences in human capital is allowed for.

[24] Under the weaker assumptions of constant markups and a Cobb-Douglas production function, the ARPH is proportional to MRPH. This is not true when markups and output elasticities vary across firms.

[25] This is also known as "history dependence" or "state dependence" in Bonhomme et al. (2019)

Second, implicit in the efficiency units specification of the production function, I assume that worker types are perfect substitutes (within sectors), although the average worker efficiency and firm productivity are complements. This restrictive assumption implies that the production function is not log-supermodular (or submodular) in worker and firm productivity, and thus abstracts from worker-firm sorting based on production complementarities (Eeckhout and Kircher, 2011; Bagger and Lentz, 2019).[26] In return, this assumption (i) delivers a mapping between the widely-estimated two-way fixed effect regressions (Abowd et al., 1999; Bonhomme et al., 2019) and the structural firm wage premium equation; and (ii) keeps the firm heterogeneity estimation procedure computationally affordable and data requirements feasible. This is because the estimation strategy involves estimating flexible production functions without restrictions on the elasticity of substitution between pairs of factor inputs. Relaxing this assumption by introducing multiple worker types exponentially increases the number of parameters to be estimated and quickly renders the estimation procedure infeasible. In Appendix E, I extend the analysis to include more and less skill intensive occupations.

3 Estimating the Structural Firm Wage Premium Equation

3.1 Empirical approach

To use the structural firm wage premium equation to decompose the empirical distribution of firm wage premia, I first estimate firm wage premia, then estimate firm-specific measures of the wage markdown, average revenue product of labor, price-cost markup, and labor elasticity of output. This second step will require information about firm wage premia.

One approach to estimate each channel of heterogeneity would be to estimate a fully-specified structural framework. However, this requires the researcher to specify the market structure in each product and labor market. Alternatively, a common approach to measure firm-specific price-cost markups is the cost share approach (Foster, Haltiwanger, and Syverson, 2008). This approach measures firms-specific markups using firm-specific sales-to-total-cost ratios. However, a key assumption required to implement the cost share approach is that all input markets are perfectly competitive, which precludes the estimation of wage markdowns.

To overcome these challenges, I adapt the production-based markup estimation approach by De Loecker and Warzynski (2012) and De Loecker et al. (2020) to accommodate imperfectly competitive labor markets. In the original approach, one first estimates the output elasticities, then computes price-cost markups from a variable input's expenditure share of revenue. I show that when labor markets are imperfectly competitive, estimating output elasticities requires

[26]However, one can extend the framework to allow sorting based on non-wage amenities (Lamadon et al., 2019), or worker and firm productivity in which the firm screens for workers above a productivity threshold (Helpman, Itskhoki, Muendler, and Redding, 2017) within the framework.

knowledge of the firm-specific wage premium.[27] Then, once output elasticities obtained, I show that price-cost markups and wage markdowns can be disentangled by exploiting the fact that price-cost markups distort each input demand, while wage markdowns distort only labor demand.

3.2 Estimating firm wage premia

A common way of estimating firm wage premia is to estimate firm effects from an AKM regression – a Mincerian regression with worker and firm effects. The firm effects are identified from worker mobility between firms. In practice, a key issue in estimating firm effects is the lack of such worker mobility, which leads to noisy firm effects estimates that tend to upward-bias the variance of firm effects. To address this *limited mobility bias*, I first classify firms into groups using a k-means clustering algorithm, then estimate a version of the AKM regression replacing firm effects with firm-group effects, following Bonhomme et al. (2019) (BLM henceforth).[28] When there are as many firm-groups as there are firms, this regression converges to the classic AKM regression. The firm-group fixed effects are then identified by workers who switch between firm-groups. Relative to the AKM regression, this procedure has the advantage that it substantially increases the number of switchers used to identify firm-group effects, which enables firm wage premia to be precisely estimated.

Specifically, I estimate the following regression:

$$\ln W_{it} = X'_{it}\beta + a_i + \phi_{g(j(i,t))} + \nu_{it}$$

where i denotes the individual, j denotes the firm, $g(j)$ denotes the group of firm j at time t, a_i are worker fixed effects, $\phi_{g(j(i,t))}$ are firm-group fixed effects, and X_{it} is a vector of time-varying worker characteristics, including age polynomials, part-time status, and 2-digit occupation indicators. Occupation fixed effects in this regression are identified by workers who switch occupations, but not employers. This helps capture some of the wage effects of changes in human capital.

To classify firms with similar firm wage premia into the same group, I group firms based on

[27]Methodologically, this is closely related to Dobbelaere and Mairesse (2013), who estimate price-cost markups and monopsony power at the firm-level. My approach differs by: (i) allowing a labor-augmenting technology component in the production function, (ii) estimating a flexible translog production function, (iii) using a more flexible control function production function estimation procedure (Ackerberg et al., 2015), and (iv) showing the firm wage premia are required in the control function when firms have labor market power. Morlacco (2019) exploits a similar idea to estimate firms' market power in foreign intermediate input markets.

[28]Bonhomme et al. (2019) develop two flexible frameworks for estimating worker and firm fixed effects: (a) A static framework that allows interactions betweens worker and firm effects, and (b) A dynamic framework that allows endogenous worker mobility and first-order Markovian wage dynamics. In this paper, I use a linear BLM framework for several reasons: (i) The log additive wage regression appears to be a good first-order approximation of the structure of wages (consistent with findings in Bonhomme et al. (2019)), (ii) a fine classification of firms into clusters, important for the purpose of this paper, quickly renders BLM estimation computationally intractable.

the similarity of their internal wage distributions. The idea is that, conditional on the AKM regression, firms with a similar firm effects and worker effects should have similar internal wage distributions. If two firms have internal wage distributions of very similar shapes, but their average wages differ significantly, then the AKM wage equation suggests that they have very different firm effects. If two firms have very similar average wages, but the shape of their internal wage distributions differ substantially, they are clustered into different groups.

Specifically, let $g(j) \in \{1, 2, ..., G\}$ denote the cluster of firm j, and G the total number of clusters. The k-means algorithm then finds the partition of firms such that the following objective function is minimized:

$$\min_{g(1),...,g(J),H(1),...,H(G)} \sum_{j=1}^{J} N_j \int \left(\hat{F}_j(\ln W_{ij}) - H_{g(j)}(\ln W_{ij})\right)^2 d\gamma(\ln W_{ij})$$

where $H(g)$ denotes the firm-group level cumulative distribution function for log wages at group g, $\hat{F}_j$ is the empirical CDF of log wages at firm j, and N_j is the employment size of firm j. The total number of groups G is the choice of the researcher. I choose sector-specific G such that the variance of log wages between firm-groups captures at least 95% of the total between-firm variance.[29] This choice is motivated by the following consideration: having a coarse classification of firms into fewer groups leads to many more workers who switch between firm-groups, which substantially improves the precision of firm wage premium estimates. However, this comes at the cost of potentially averaging away considerable amounts of multidimensional firm heterogeneity within firm-groups. In practice, I apply the clustering algorithm by 2-digit sectors for the following intervals of time: 1995-1998, 1999-2002, 2003-2005, 2006-2008, 2009-2011, 2012-2014. The time intervals are chosen to keep the number of observations similar across estimation samples.

AKM regressions rely on the assumption that worker mobility is as good as random conditional on observed worker characteristics, worker fixed effects, and firm fixed effects. Formally, $E(\nu_{it}|X_{it}, a_i, \phi_{g(j(i,t))}) = 0$. This assumption rules out worker mobility based on wage realizations due to the residual component of wages. In addition, AKM regressions impose log additivity of the worker and firm components of wages. In Appendix C, I assess the AKM regression specification and find it to be a reasonable approximation to French wage data. In particular, I follow Bonhomme et al. (2019) and estimate a version of the AKM regression that relaxes both of these restrictions find a small gain in R^2 of 0.01.

[29] An alternative way of selecting the number of firm-groups is to use network connectivity in terms of switchers between firms (Jochmans and Weidner, 2019; Bonhomme et al., 2019). However, because the French administrative data, DADS Postes, is required for this classification step and this dataset does not track worker mobility across firms, a measure of network connectivity cannot be constructed.

3.3 Estimating the channels of firm heterogeneity

My estimation approach for the four channels of firm heterogeneity has three steps. First, I compute the average revenue product of labor in efficiency units $ARPH = \frac{PY}{EL}$. To do so, I first compute the average labor productivity $\frac{PY}{L}$ as the total revenue per hour, and then compute the model-consistent average efficiency of workers per hour as the difference between the firm's average wage and the firm wage premium, $\bar{E} = \frac{\bar{W}}{\Phi}$. The log of the firm-specific average worker efficiency is normalized to have a mean of 0 in the cross-section.

The second and third steps extend the production-based approach of De Loecker and Warzynski (2012) and De Loecker et al. (2020). In the second step, I estimate a production function to obtain firm-specific output elasticities. I estimate the following sector-specific translog production function, which is a second-order approximation of any well-behaved production function:

$$\begin{aligned} y_{jt} = \; & \beta_{h,s} h_{jt} + \beta_{k,s} k_{jt} + \beta_{m,s} m_{jt} + \beta_{o,s} o_{jt} \\ & + \beta_{hh,s} h_{jt}^2 + \beta_{kk,s} k_{jt}^2 + \beta_{mm,s} m_{jt}^2 + \beta_{oo,s} o_{jt}^2 \\ & + \beta_{hk,s} h_{jt} k_{jt} + \beta_{hm,s} h_{jt} m_{jt} + \beta_{ho,s} h_{jt} o_{jt} \\ & + \beta_{km,s} k_{jt} m_{jt} + \beta_{ko,s} k_{jt} o_{jt} + \beta_{mo,s} m_{jt} o_{jt} + x_{jt} + \epsilon_{jt} \end{aligned} \tag{8}$$

where lowercase letters represent the natural log counterparts of variables written in uppercase letters. With a slight change of notation, let m_{jt} now denote firm j's use of raw materials and o_{jt} denote its use of services. As I explain in the next section, the French administrative data sets measure these intermediate inputs separately. Define ϵ_{jt} as the error term orthogonal to firms' input choice, which can be measurement error.

As Gandhi, Navarro, and Rivers (2019) show, the control function approach does not generally identify the production function parameters when considering a gross output production function. In essence, returns-to-scale and markups generally cannot be separately identified. To address this issue, I follow Flynn, Gandhi, and Traina (2019) in imposing constant returns-to-scale *on average*, while allowing returns-to-scale to depend on firms' input choices besides the proxy variable input.[30] This entails the following parameter restrictions:

$$2\beta_{mm,s} = -(\beta_{hm,s} + \beta_{km,s} + \beta_{mo,s})$$

$$E_s \left[LEO_{jt} + KEO_{jt} + MEO_{jt} + OEO_{jt}\right] = E_s \left[RTS(k_{jt}, h_{jt}, o_{jt})\right] = 1$$

where RTS denotes returns-to-scale, and LEO_{jt}, KEO_{jt}, MEO_{jt}, and OEO_{jt} denote the labor, capital, material, and service elasticities of output.

[30] Flynn et al. (2019) show that constant returns-to-scale is a good approximation.

The production function cannot be estimated by ordinary least squares, as there are three potential sources of bias to the production function parameters - an endogeneity bias, an output price bias, and an input price bias (De Loecker and Goldberg, 2014).

Firms' input demand is an endogenous choice of the firm and depends on the firm's productivity realization x_{jt}. This is likely to bias the production function parameters upwards. To address this endogeneity issue, I follow a control function approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015). This approach allows the reseacher to "observe" the firms' idiosyncratic productivity by inverting their optimal input demand function for a fully flexible variable input. My choice of the variable input used to construct the control function is services as it is by far the most strongly correlated with contemporaneous output growth at the firm level in French balance sheet data, as shown in Table 7 in appendix A.

Using the first-order conditions for each factor input, I obtain the following optimal service demand function:

$$o_{jt} = o(x_{jt}, k_{jt}, h_{jt}, m_{jt}, \mathbf{Z}_{jt}, \phi_{jt})$$

where $\mathbf{Z}_{jt}$ is a vector of exogenous firm characteristics that can affect its input demand, which includes location fixed effects, sector fixed effects, and year fixed effects. Since firm-specific input unit prices, especially for intermediate and capital inputs, are unobserved in most existing datasets, my estimation operates under the assumptions that firms are price-takers in intermediate and capital input markets, and firms within a given sector and location face the same input prices.[31] However, because I observe hourly wages at the individual worker level, my datasets enable me to extend the estimation procedure to allow imperfectly competitive labor markets. This extension entails augmenting the control function to include firm-specific wage premia ϕ. This inclusion controls for the fact that in this environment firms have some degree of wage-setting power, which distorts relative input prices, hence, relative input demand.

To obtain the control function, I invert the optimal intermediate input demand function and express idiosyncratic total factor productivity as a function of observed variables:

$$x_{jt} = x(h_{jt}, k_{jt}, m_{jt}, o_{jt}, \mathbf{Z}_{jt}, \phi_{jt}) \tag{9}$$

The key assumption for invertibility is that, conditional on the variables in the control function, service input demand is monotonically increasing in firm productivity x_{jt}.

The production function can then be estimated following the two-step GMM approach described in Ackerberg et al. (2015). In step one, I combine (8) and (9) and estimate the following

[31] This assumption is standard in the production function literature due to unobserved input prices (De Loecker and Goldberg, 2014). Relative to standard datasets, my dataset includes wages at the worker level. I can therefore control for differences in firm-specific input demands due to differences in wages, which can arise due to differences in worker composition and market power.

by OLS:

$$y_{jt} = \Psi(h_{jt}, k_{jt}, m_{jt}, o_{jt}, \mathbf{Z}_{jt}, \phi_{jt}) + \epsilon_{jt} \tag{10}$$

approximating $\Psi(.)$ with a high-order polynomial in its arguments. This step estimates and removes the residual term ϵ_{jt}, capturing measurement error and unobserved productivity shocks that are orthogonal to input choices, from output. Specify law of motion for the log of Hicks neutral productivity x as:

$$x_{jt} = g(x_{jt-1}) + \zeta_{jt} \tag{11}$$

where $g(.)$ is a flexible function and ζ_{jt} is a productivity shock. In step two, estimate the production function parameters. Combining the control function (9), the predicted output from (10), and the law of motion for productivity (11), I form the following moment conditions:

$$E[\zeta_{jt}(\beta)\mathbf{X}_{jt}] = \mathbf{0}$$

where $\mathbf{X}_{jt}$ is a vector of current and lagged variables:

$$\begin{aligned}\mathbf{X}_{jt} =[&o_{jt-1} \;\; o^2_{jt-1} \;\; m_{jt-1} \;\; m^2_{jt-1} \;\; h_{jt-1} \;\; h^2_{jt-1} \;\; k_{jt} \;\; k^2_{jt} \\ &k_{jt-1}h_{jt-1} \;\; k_{jt-1}m_{jt-1} \;\; k_{jt-1}o_{jt-1} \;\; h_{jt-1}m_{jt-1} \;\; h_{jt-1}o_{jt-1} \\ &\phi_{jt-1} \;\; \phi_{jt-1}\mathbf{F}_{jt-1} \;\; \mathbf{Z}_{jt}]'\end{aligned}$$

$\mathbf{F}_{jt-1}$ is a vector of the firm's factor inputs. This moment condition is consistent with the timing assumption of the structural framework in the previous section. Firms' input demand and posted wages in the current period are orthogonal to future productivity shocks. In addition, capital inputs are assumed to be dynamic and pre-determined, so firms' current capital input demand are orthogonal to current productivity shocks. I combine the two steps into one and implement Wooldridge (2009).

A common challenge in the production function estimation literature is that output prices are rarely observed (De Loecker and Goldberg, 2014).[32] In typical firm-level balance sheet data, output is usually measured in terms of sales revenue or the nominal value of production. Under the assumptions above, the estimated production function is then:

$$p_{jt} + y_{jt} = f(h_{jt}, k_{jt}, m_{jt}, o_{jt}) + p_{jt} + x_{jt} + \epsilon_{jt}$$

where $p_{jt} + x_{jt}$ is the revenue TFP (Foster et al., 2008). The control function is therefore for revenue-TFP rather than quantity-TFP. The potential negative correlation between output prices and input demand could lead to a downward output price bias. The intuition is that, all

[32] When output prices are observed, they are typically for specific industries, e.g. beer brewing (De Loecker and Scott, 2016), or for the manufacturing industry, such as the US Manufacturing Census.

else equal, firms that set higher prices tend to sell less output, which in turn requires less inputs to produce. It is therefore important to discuss the conditions under which unobserved output prices do not bias estimates of output elasticities.

If firm heterogeneity in prices (markups over marginal costs) is driven by differences in production costs due to productivity x, the firm wage premium ϕ, or regional or sectoral differences in capital or intermediate input prices, these are controlled for in the control function. However, differences in idiosyncratic demand uncorrelated with TFP could still drive markup (hence, price) variation *beyond* what is controlled for in the control function. Therefore, I additionally include controls for markup heterogeneity. Informed by oligopolistic competition and trade models such as Atkeson and Burstein (2008) and Edmond et al. (2015), I include export status and market shares as additional controls. Informed by models of customer capital (Gourio and Rudanko, 2014), which predict that firms accummulate customers over time, I also include firm age. The lags of these additional controls therefore also appear in the vector $\mathbf{X}$ in the moment conditions of the estimation procedure $E[\zeta_{jt}(\beta)\mathbf{X}_{jt}] = \mathbf{0}$. The key assumption here is that these additional controls sufficiently capture variation in markups uncorrelated with TFP. This assumption rules out a role for differences price elasticities of demand due, for example, to product quality differences, conditional on firms' TFP.[33]

I then compute the labor (LEO_{jt}) and service input (OEO_{jt}) elasticities of output as follows:

$$LEO_{jt} = \beta_h + 2\beta_{hh}h_{jt} + \beta_{kh}k_{jt} + \beta_{hm}m_{jt} + \beta_{ho}o_{jt}$$

$$OEO_{jt} = \beta_o + 2\beta_{oo}o_{jt} + \beta_{ho}h_{jt} + \beta_{ko}k_{jt} + \beta_{mo}m_{jt}$$

I estimate production functions by 2-digit sectors within three time intervals: 1995-2000, 2001-2007, 2008-2014. Therefore, production function coefficients differ across sectors and time.

In the third step of the estimation of firm heterogeneity, I exploit the fact that price-cost markups are common distortions to the demand of each input while wage markdowns distort only labor demand to separately identify price-cost markups and wage markdowns. Under the assumption that service intermediate inputs are variable inputs and firms take their prices as given, their prices are equal to their marginal revenue products (De Loecker and Warzynski, 2012). Therefore, markups represent the only wedge between service input prices and their marginal products. One can then express price-cost markups as a function of the service input share and service input elasticity of output:

$$PM_{jt} = OEO_{jt}\frac{P_{jt}Y_{jt}/\exp(\hat{\epsilon}_{jt})}{P_t^O O_{jt}}$$

[33] As discussed in De Loecker and Goldberg (2014), this assumption can be relaxed by (i) imposing particular demand systems, such as a nested CES demand system, or (ii) obtaining output price data, which tend to be available for a subset of manufacturing firms in customs trade data or manufacturing censuses.

where $\exp(\hat{\epsilon}_{jt})$ removes measurement error or any other variation orthogonal to the firm's input choice from revenue shares, with $\hat{\epsilon}_{jt}$ the residual from the first stage when estimating the production function. I apply this correction to all revenue shares and average revenue products of labor.

I now obtain wage markdowns using the wage bill to service input expenditure ratio and the output elasticities:

$$WM_{jt} = \frac{\Phi_{jt} H_{jt}}{P_t^O O_{jt}} \cdot \frac{OEO_{jt}}{LEO_{jt}} = \frac{\bar{W}_{jt} L_{jt}}{P_t^O O_{jt}} \cdot \frac{OEO_{jt}}{LEO_{jt}}$$

Since the price-cost markup is a common input distortion, it cancels out and therefore does not feature in this equation. Further, under the assumption that services are flexible inputs, the only remaining distortion in this equation is the wage markdown.

Finally, I obtain the marginal revenue product of effective labor as follows:

$$MRPH_{jt} = PM_{jt}^{-1} LEO_{jt} \frac{P_{jt} Y_{jt} / \exp(\hat{\epsilon}_{jt})}{H_{jt}}$$

4 Data Description

4.1 Administrative datasets from France

Estimating the structural firm wage premium equation using the approach described above requires two types of datasets. The empirical distribution of the firm wage premium is estimated with matched employer-employee datasets, which follow workers over time and employment spells at different firms. The four channels of firm heterogeneity in the model are estimated with firm balance sheet panel datasets. While both types of datasets have become increasingly accessible, they are typically not jointly available.[34] To the extent that firm balance sheet datasets are available, most cover only a set of large firms or manufacturing firms, or do not contain a panel structure.[35] I therefore use matched employer-employee and firm balance sheet panel data from France covering the population of firms and workers in the private sector between 1995 and 2014.

My sources for firm balance sheet information are the *Fichier de comptabilité unifié dans SUSE* (FICUS) and *Fichier approché des résultats d'Esane* (FARE) datasets, jointly available from 1995 to 2014. FICUS and FARE are compiled by the fiscal authority of France, *Direction Générale des Finances Publiques* (DGFiP), from compulsory filings of firms' annual accounting

[34] Countries for which both datasets are available to researchers, at the discretion of the statistical authorities, include Brazil, Denmark, Norway, Sweden, and France.

[35] Example dataset (large firms): Compustat database. Example (non-panel): US Census of Manufacturers, Census of Retail Trade, Census of Agriculture. Example (manufacturing only): Colombia and Mexico.

information. These datasets contain balance sheet information for all firms in France without restriction on the size of firms. There are over 2 million firms per year. From these datasets, I obtain information on variables such as sales, nominal value of production, employment, intermediate input and capital expenditure. I provide details of how each variable that enters the production function is measured in appendix A.

I also use annual French administrative data on employed workers, from 1995 to 2014, under the umbrella *Déclarations Annuelles de Données Sociales* (DADS). The DADS datasets are compiled by the national statistical institute of France, *Institut National de la Statistique et des Études Économiques* (INSEE), from compulsory reports of employee information to the French authorities. They contain information at the job level, such as age, gender, earnings, hours, and occupational category. One advantage of the DADS datasets is that work hours are observed, allowing researchers to construct and study variation in hourly wages. This addresses concerns that variation in earnings simply reflect variation in hours worked. They also include employer identifiers, called SIREN, which enables linking with firm balance sheet data. One disadvantage is that information about workers' education is not available.

The first DADS dataset is the DADS-Panel, which provides information on all employed workers in the private sector born in October in a panel structure (only October-workers born in even years are observed prior to 2002). Because workers are followed over time and their employer identifiers are observed, I use this dataset to estimate the AKM-BLM regressions described in the previous section to estimate firm wage premia.

The second DADS dataset is the DADS-Postes, which contains information on all existing jobs in France. Unlike the DADS-Panel, this is not a proper panel dataset. It is organized in an overlapping structure – each observation appears in the dataset under the same identifier for at most two periods (if the job exists for at least two periods). Therefore, this dataset cannot be used to estimate firm wage premia directly. Instead, to maximize the number of firms for which firm wage premia are estimated using the DADS-Panel, I first use the DADS-Postes to k-means cluster firms into groups of similar firms, as far as wages are concerned, prior to running the AKM-BLM regression. This approach has the advantage that firm wage premia can be estimated for firms that exist in the firm balance sheet data but not in the DADS-Panel because they do not have an employee who is born in October.

4.2 Analysis sample

I restrict firm level observations from the FICUS-FARE balance sheet data to several broad industries: construction, manufacturing, non-financial services, transportation, and wholesale and retail. The public administration, education, and utilities sectors are excluded. I include only firms with at least 5 employees. I harmonize all 2-digit and 5-digit industry codes to the latest available version (Nomenclature d'activités Française – NAF rév. 2). I drop 2-digit sectors with less than 500 observations within each 7-year interval (1995-2000, 2001-2007, 2008-2014). This is important when estimating production functions, especially flexible specifications such as the translog, as this procedure would be demanding on small sample sizes, and could lead to imprecise estimates of the production function parameters. In practice, few two-digit sectors have less than 500 observations in this time interval. I also drop firms within each 7-year interval that only appear once since estimating production functions requires at least two consecutive years of data.

For both of the DADS datasets, I focus on workers between the age of 16 to 65, who hold either a part-time or full-time job principal job (side jobs are dropped). I apply the same restrictions on the broad industries included as I do for the FICUS-FARE datasets. I keep workers in the following one-digit occupational categories: (a) Top management, such as chief executive officers or directors; (b) senior executives, such as engineers, professors, and heads of human resources; (c) middle management, such as sales managers; (d) non-supervisory white-collar workers, such as secretarial staff and cashiers; and (e) blue-collar workers, such as foremen and fishermen. All 1-digit, 2-digit, and 4-digit occupation codes are harmonized and updated to the latest version provided by INSEE (PCE-ESE 2003). Observations whose hourly wages fall outside three standard deviations of the mean are excluded.

Firm wage premia in the AKM-BLM regression are only identified for the sets of firms connected by worker mobility. I therefore focus on the largest connected set of firms. In practice, due to the clustering of firms into firm-groups using the DADS-Postes, my analysis pertains to the largest connected set of firm-groups, of which very few firms are not a part. This group consists of 158,163,180 people-year observations, an average of 7,908,159 per year. After clustering firms into groups, I link the DADS-Postes and DADS-Panel via the firm identifier (SIREN) to allocate each firm-year observation in the panel data a firm-group identifier and construct the estimation sample for firm wage premia. I implement the AKM-BLM regression on this sample.

After estimating firm wage premia, I collapse the dataset to the firm level and link it to the FICUS-FARE firm balance sheet data to construct the estimation sample for each dimension of firm heterogeneity. I implement the production function estimation routine on this sample. There are 4,907,010 firm-year observations in total and an average of 245,351 firms per year in this sample. Summary statistics for worker and firm characteristics are reported in Table 8.

5 Which Firm Characteristics Matter for Firm Wage Premia?

5.1 Product market power and labor elasticities of output matter

Having estimated firm wage premia and the underlying firm characteristics, this section shows that two firm characteristics that have received relatively little attention in the literature so far – product market power and labor elasticities of output – account for sizable shares of the variation in firm wage premia. Without taking them into account, standard models of frictional labor markets overstate the explanatory power of other firm characteristics – labor productivity and wage markdowns.[36]

Recall that the structural firm wage premium equation is a log-linear function of the four channels of heterogeneity:

$$\phi_{jt} = wm_{jt} + arph_{jt} - pm_{jt} + leo_{jt} \tag{12}$$

where lowercase letters are variables in logs. Because each channel of firm heterogeneity is exactly identified in my estimation approach, my decomposition of the empirical firm wage premium distribution is also exact.

Firm characteristics	Marginal contribution to the R^2
Wage markdown (wm)	0.25
Average revenue product of labor ($arph$)	0.35
Price-cost markup (pm)	0.11
Labor elasticity of output (leo)	0.29
Number of firms	4,907,010

Table 1: Shapley decomposition of the firm wage premium distribution, 1995-2014.

To maximize interpretability, my preferred variance decomposition method is a Shapley decomposition (Shorrocks, 1982, 2013).[37,38] I implement this variance decomposition by running equation (12) as a linear regression and then decomposing the R^2 into four components. Each component represents the marginal contribution of a channel of firm heterogeneity to the cross-sectional firm wage premium variation. Each component of this decomposition answers the question: "How much does variable X matter for firm wage premia, holding the other three variables constant?" Relative to a standard variance decomposition, the Shapley decomposition is easier to interpret because (i) it is more parsimonious, and (ii) the marginal contributions take

[36] Appendix E introduces an extension to findings in this section and Section 6 that includes skill intensive and less skill intensive occupations.

[37] Appendix B discusses the Shapley decomposition in detail.

[38] An alternative decomposition method is the ensemble decomposition (Sorkin, 2018). It can be written as: $1 = \frac{CV(\ln WM, \phi)}{V(\phi)} + \frac{CV(-\ln PM, \phi)}{V(\phi)} + \frac{CV(\ln LEO, \phi)}{V(\phi)} + \frac{CV(\ln ARPH, \phi)}{V(\phi)}$. I show this in Table 20 in Appendix D.

values between 0 and 1, and they sum up to 1.[39] Table 1 presents the Shapley decomposition results.

The first row of Table 1 shows that wage markdowns account for 25% of the firm wage premium distribution. Theoretically, wage markdowns can differ across firms for a number of reasons. As discussed in section 2, wage markdowns in the dynamic frictional wage-posting framework are a function of the firm-specific labor supply elasticity and the discounted expected marginal profits of keeping the worker next period. In static monopsonistic and oligopsonistic wage-posting models, wage markdowns depend on the firm-specific labor market shares of employment or wage bill (Boal and Ransom, 1997; Berger et al., 2020; Azar, Marinescu, and Steinbaum, 2019).[40] The structural framework in section 2 also nests the workhorse Burdett and Mortensen (1998) model of frictional labor markets. In its simplest form, in which workers and firms are homogenous, a well-known prediction of the Burdett-Mortensen model is that the wage distribution is non-degenerate. This is also known as "frictional wage dispersion" (Hornstein, Krusell, and Violante, 2011) and it shows up in the form of heterogeneous wage markdowns.[41]

Alternatively, as shown in Appendix F, wage markdown dispersion in wage-bargaining models of the labor market reflect heterogeneity in workers' share of the match surplus (relative bargaining power), outside options (captured by reservation wages), and the discounted expected marginal profits of retaining the worker. Estimating heterogeneous outside options and their effect on wages and the labor share of income is a growing literature (Caldwell and Danieli, 2019; Caldwell and Harmon, 2019; Schubert, Stansbury, and Taska, 2019; Jarosch et al., 2019).

Commonly used models of frictional labor markets often feature heterogeneous firm productivity in the form of the average revenue product of labor (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Mortensen, 2010; Bagger et al., 2014; Elsby and Michaels, 2013; Kaas and Kircher, 2015; Engbom and Moser, 2018; Gouin-Bonenfant, 2020; Lamadon et al., 2019). In these models, firm productivity determines the extent of wage premium a firm pays relative to other firms for an identically skilled worker, with more productive firms paying workers of a given skill higher wages. It is well-known that firm productivity is highly dispersed (Syverson, 2004), that changes in productivity dispersion are correlated with changes in between-firm wage dispersion (Faggio et al., 2007; Berlingieri et al., 2017), and that firm level productivity shocks pass through to wages even conditional on worker ability (Card et al., 2018; Kline et al., 2017). The third row of Table 1 shows that heterogeneous average revenue productivity of labor accounts for 35% of the firm wage premium distribution. This result implies

[39] I present results from a standard variance decomposition in Table 4.

[40] For a microfoundation that shows this, I refer the interested reader to Appendix F.

[41] This is because firms, trading off profits per worker and firm size, locate at different points on the labor supply curve and thus face different labor supply elasticities. In equilibrium, all firms make the same profits.

that a substantial share of the variation is due to the two additional channels of heterogeneity – price-cost markups and labor elasticities of output. Therefore, without taking them into account, workhorse models of frictional labor markets would overestimate the explanatory power of firm productivity and wage markdowns for firm wage premia.

Firm-level differences in price-cost markups often do not feature in frictional labor market models. However, this is a theoretically and quantitatively important determinant of labor demand and firm size in the macroeconomics of resource misallocation (Edmond et al., 2015; Peters, 2020). There is increasing evidence that markups vary significantly across firms (De Loecker and Eeckhout, 2018; Edmond, Midrigan, and Xu, 2018). At the same time, these models do not speak to the distribution of firm wage premia due to the assumption of perfectly competitive labor markets. As discussed in Section 2, price-cost markups affect firm wage premia through firms' labor demand in my structural framework. The fourth row of Table 1 shows that price-cost markups account for 11% of the firm wage premium distribution.

Firm-level differences in labor elasticities of output also appear to be quantitatively important for firm wage premia, as the fifth row of Table 1 shows. The labor elasticity of output is a key component of labor demand in macroeconomic models of the labor share (Karabarbounis and Neiman, 2014; Oberfield and Raval, 2020; Hubmer, 2019), but often does not feature in frictional labor market models. I find that heterogeneous labor elasticities of output account for 29% of the firm wage premium distribution.

The findings in Table 1 indicate that the marginal revenue product of effective labor (MRPH), which is equal to the sum of the last three components of equation (12), is the main driver of firm wage premium dispersion. Since firm wage premia are often estimated from statistical regressions, this result provides a quantitative structural interpretation. In wage-posting models, the contributions of this component can be thought of as reflecting differences in firms' labor demand, and hence, firms' willingness to pay for a given worker. This is because in wage-posting models wages are determined before commencing an employment relationship. Alternatively, in wage-bargaining models, wages are decided ex-ante through a bargaining process, in which case the contribution of the MRPH can be interpreted as arising from surplus sharing.

If one were to estimate a standard frictional labor market model without taking into account price-cost markups and labor elasticities of output, the model would overestimate the explanatory power of firms' labor productivity and wage markdowns for firm wage premia. To obtain an upper bound for which the role of differences in labor productivity and wage markdowns could be overestimated, I first re-estimating wage markdowns under the assumption that price-cost markups and labor elasticities of output are homogenous across firms within each sector, then implement the Shapley decomposition. This approach attributes all variation in price-cost markups and labor elasticities of output in a two-digit sector to wage markdowns. It is meant to approximate the setting of estimating a Burdett-Mortensen type model with heterogeneous firms using data on the joint distribution of wages and the average labor productivity. In this case, my decomposition suggests that up to 53% and 47% of firm wage premia can be accounted for by differences in firms' labor productivity and wage markdowns.

This result matters for two reasons. First, if models overestimate the extent to which the firm wage premium distribution reflects heterogeneous wage markdowns, then they also overestimate the role of wage markdowns in the misallocation of labor inputs across firms. This would overstate the extent to which labor market policies can address distortions due to labor market power. For example, when firms' wage markdowns are quantitatively important distortions to labor demand, the minimum wage can be an effective tool to correct such distortions and lead to welfare improvements (Berger et al., 2020).

Second, if models overestimate the role of firm productivity in driving firm wage premia, then they also overestimate the extent to which firm wage premia reallocate workers from less productive firms to more productive firms, as workers search on-the-job for better-paying firms. This worker reallocation role of wage dispersion is a key driver of aggregate productivity and wage growth in workhorse models of frictional labor markets (Haltiwanger et al., 2018; Bilal et al., 2019).

5.2 There are large difference in the estimated firm characteristics

This section documents the empirical distribution of each of the four channels of firm heterogeneity. I start by discussing new estimates of wage markdowns and labor elasticities of output. I then confirm the well-documented existence of large productivity dispersion and the more recently documented price-cost markup dispersion across firms (Syverson, 2004; De Loecker et al., 2020). Table 2 summarizes the empirical moments of each estimated dimension of firm heterogeneity in 2014. Table 9 in Appendix D shows moments related to the within-sector distribution of each dimension. Tables 10, 11, 12, and 16 in Appendix D report the variances of each channel of firm heterogeneity by broad industry categories.

The wage markdown has received increasing attention as a potentially important driver of wage inequality (Azar et al., 2018; Schubert et al., 2019; Caldwell and Danieli, 2019) and the distribution of labor shares (Berger et al., 2020; Gouin-Bonenfant, 2020; Jarosch et al., 2019; Brooks et al., 2019). Despite its theoretical relevance, its empirical distribution is not yet well-documented. The first row of Table 2 describes the distribution of wage markdowns (WM). I find substantial dispersion of wage markdowns across firms. My estimates show that firms at the 75^{th} percentile of the wage markdown distribution pay a wage 3% below the marginal revenue product of labor. At the 25^{th} percentile, workers obtain approximately two-thirds of their marginal revenue productivity (67%). As discussed in Section 2, this large dispersion of the wage markdown reflects differences in labor supply elasticities in a wage-posting model, or bargaining power and outside options in a wage-bargaining model. In either model, it could also reflect differences across firms in the future value of a worker, if the employment relationship remains intact.

Firm characteristics	Mean	Median	Variance	25th Pct	75th Pct
Wage markdown	0.85	0.80	0.15	0.67	0.97
Price-cost markup	1.28	1.24	0.13	1.11	1.41
Capital elasticity of output	0.05	0.05	0.00	0.02	0.07
Labor elasticity of output	0.26	0.24	0.02	0.17	0.33
Material elasticity of output	0.41	0.41	0.05	0.26	0.55
Service elasticity of output	0.28	0.26	0.02	0.19	0.36
Average revenue product of labor (log)	4.52	4.44	0.37	4.08	4.90
Marginal revenue product of labor (log)	2.83	2.82	0.10	2.64	2.99
Number of firms	243,453				

Table 2: Summary statistics for estimated firm characteristics in 2014.

I also find that most firms possess significant wage-setting ability – half of the firms in my sample pay less than 0.80 of the marginal revenue product of labor as wages. This suggests that there is ample room for wage increases at the typical firm. One way to assess firms' wage-setting ability is to compare it to firms' price-setting ability. I do so by inverting the wage markdown and then comparing it to the price-cost markup. This gives an inverted median wage markdown that is similar to median price-cost markups at the median firm (1.25 compared to 1.24).

Since there is little systematic documentation of the distribution of wage markdowns, it is not straightforward to compare my estimates with existing work. I start by comparing my estimates to those of Hershbein et al. (2020) and Mertens (2019), who find that the median US and German manufacturing firm pay 0.73 and 0.68 of the marginal revenue product of labor as wages. For the median French manufacturing firm the wage markdown is 0.77, implying less wage-setting power than their US or German counterparts. Another way to do such a comparison is to assume that my wage markdown estimates are generated by a static wage-posting model. As discussed in Section 2, wage markdowns in this case are entirely determined by labor supply elasticities. I consider a Burdett-Mortensen model, in which the wage markdown is $\frac{\epsilon^H}{1+\epsilon^H}$, and back out the implied labor supply elasticities.[42] This gives firm-specific labor supply elasticities of 2.03, 4, 32.33 at the 25^{th}, 50^{th}, and 75^{th} percentiles. This is higher than estimates for the US

[42]In Section 2, I show how the Burdett-Mortensen model can be obtained from my structural framework.

based on the Burdett-Mortensen model by Webber (2015), who find firm-specific labor supply elasticities of 0.44, 0.75, 1.13, at the 25^{th}, 50^{th}, and 75^{th} percentiles. Berger et al. (2020) find firm-specific labor supply elasticities driven by differences in market shares in an oligopsonistic model between 0.76 and 3.74 in the US. Relative to Webber (2015) and Berger et al. (2020), my wage markdown estimates for France imply, on average, a higher labor supply elasticity than the US.

The labor elasticity of output is a central part of the debate about the causes of the U.S. aggregate labor share decline. Existing estimates for labor elasticities of output (LEO) are usually at an aggregated level, for example, at the sector level or for the entire macroeconomy (Basu et al., 2013; Karabarbounis and Neiman, 2014; Oberfield and Raval, 2020). My estimates for firm-specific labor elasticities of output reported in the fourth row of Table 2 display substantial heterogeneity across firms, particularly within sectors. The interquartile range for labor elasticities of output is 0.16. However, my estimates are consistent with existing estimates that find moderate dispersion of labor elasticities of output across broad sectors: removing differences across 2-digit sectors reduces the interquartile range slightly to 0.14.[43] These findings suggest that labor elasticities of output are potentially important determinants of the distribution of firm wage premia and labor shares within sectors. This is explored further in the next section.

I now confirm that, consistent with existing findings, price-cost markups (PM) and firm productivity are highly dispersed across firms. The second row of Table 2 reports the summary statistics for price-cost markups. The median markup is 1.24. This is in the ballpark of existing estimates. De Loecker and Warzynski (2012) estimate markups using Slovenian manufacturing firm data and find median markups between 1.10 and 1.28. De Loecker and Eeckhout (2018) use Compustat data and find median markups in the US in 2014 of about 1.20. De Loecker et al. (2020) find markups at the 75^{th} percentile between 1.3 and 1.5 in 2014 in the US, while my estimates for France is 1.41 in 2014. Also using Compustat data, Edmond et al. (2018) report markups at the 75^{th} percentile of 1.31 and at the 25^{th} percentile of 0.97, an interquartile range for markups of 0.34. My estimates for the interquartile range is 0.30.

The second-to-last row of Table 2 reports the distributional statistics for the average revenue product of labor ($ARPH$) in logs. The dispersion of firm productivity is well-documented (Foster et al., 2008; Syverson, 2011) and a key feature of models of heterogeneous firms (Melitz, 2003). I find that the average revenue product of labor (in efficiency units) has an interquartile range of $\exp(4.90 - 4.08) = 2.27$. Most of the dispersion in productivity occurs within sectors, consistent with existing work (Syverson, 2011). The average interquartile range within two-digit sectors is 1.98.

[43]The reason that the average labor elasticity of output is lower than commonly used calibration targets of around 0.7 is because I estimate these from a gross output production function rather than value-added production function.

Compared to the average revenue product of labor, however, the marginal revenue product of labor is significantly less dispersed. Table 2 shows that the variance of log $MRPH$ is less than a third of the variance of log $ARPH$. This difference is not driven by outliers – the interquartile range for log $MRPH$ is $\exp(2.99 - 2.64) = 1.42$, substantially smaller than the corresponding number for log $ARPH$. As I discuss in Section 6, this difference leads to an overestimate of the extent of labor misallocation in a standard exercise to quantify the gains from removing dispersion in labor market power.

5.3 Moderate firm wage premium dispersion despite vast firm heterogeneity

While each channel of firm heterogeneity in equation (12) is highly dispersed and accounts for important shares of the firm wage premium distribution, this section shows that they do not translate into a highly *dispersed* firm wage premium distribution. This is despite the finding that the typical firm has significant wage-setting ability in Section 5.1, with half of the firms paying less than 80% of the marginal revenue product of labor as wages. The main message of this section is that the modest firm wage premium distribution masks substantial underlying differences in firm characteristics, and their negative correlations offset their direct effects on the firm wage premium distribution.

Table 3 reports statistics about the dispersion of firm wage premia in 2014. The variance of firm wage premia (ϕ) is modest (0.008), accounting for 4.5% of the wage distribution, similar to the numbers for the United States, Sweden, Austria, Norway, and Italy from Bonhomme et al. (2020). At the same time, the relevant firm characteristics in equation (12) are orders of magnitude more dispersed than firm wage premia, as the diagonals of Table 4 show.

	Firm-Specific Wage Premium (ϕ)	
Method	BLM	AKM
Variance	0.009	0.017
Fraction of Total Variance	4.5%	7.7%
90-10 ratio	1.25	1.45
90-50 ratio	1.13	1.19
50-10 ratio	1.11	1.22
Number of firms	243,453	78,268
Number of firm-groups	3,954	78,268
Number of workers	8,178,435	5,497,277

Table 3: Dispersion of firm wage premia in 2014.

Nevertheless, the dispersion of firm wage premia is a quantitatively important deviation from the law of one wage. Table 3 shows that a firm at the 90th percentile of the firm wage premium distribution pays a given worker a wage that is on average 25% more than a firm at the 10th percentile of the distribution. This gap is almost twice as large as the average gender wage gap in OECD countries and it is comparable to the gender wage gap in Japan, among the highest in OECD countries.[44]

These results suggest that tcorrelations between channels of firm heterogeneity in equation (12) offset each other's direct impact on the firm wage premium distribution.

5.4 More productive firms have lower labor elasticities of output

I now discuss which pair of correlations between the channels of firm heterogeneity compress the firm wage premium distribution. The main finding in this section is the following: firm productivity and labor elasticity of output are strongly negatively correlated. More productive firms tend to substitute labor with other inputs. This negative relationship implies that more productive firms generally have a less-than-proportionately higher labor demand compared to less productive firms, therefore they pay a less-than-proportionately higher wage premium.

Using equation (12), a standard variance decomposition of the firm wage premium distribution can be written as:

$$\begin{aligned} V(\phi) = \; & V(wm) + V(pm) + V(leo) + V(arph) \\ & + 2CV(wm, -pm) + 2CV(wm, leo) + 2CV(wm, arph) \\ & + 2CV(-pm, leo) + 2CV(-pm, arph) + 2CV(leo, arph) \\ = \; & V(wm) + V(mrph) + 2CV(wm, mrph) \end{aligned}$$

The variance terms are discussed in Section 5.3. The covariance terms show the importance of the relationships between each pair of firm characteristic. These are shown in Tables 4 and 5.

More productive firms have a lower labor elasticity of output. The third rows of the second column of Tables 4 and 5 present this main finding. This negative relationship is the most quantitatively important among the set of cross-terms that offset the effects of firm heterogeneity on the firm wage premium distribution (corr$(leo, arph) = -0.90$). At the same time, the correlation between firm productivity and the material elasticity of output is large and positive (0.51), while the correlation between firm productivity and the service elasticity of output is negative (0.28). The correlation between firm productivity and the capital elasticity of output is also negative but weaker (-0.17). While labor productivity and labor elasticities of

[44] According to OECD estimates, the average gender wage gap in OECD countries, defined as the median wage of males relative to the median wage of females, is 13.8% in 2017.

output are negatively correlated across sectors, there are interesting differences between sectors, as reported in Table 26. In manufacturing, labor productivity is positively correlated with material (0.47) and service (0.18), but negatively correlated with capital elasticities of output (-0.14). In non-financial services, labor productivity is positively correlated with material (0.28), service (0.43), and capital elasticities of output (0.12). Among wholesale and retail firms, however, labor productivity is positively correlated with material elasticities of output (0.63), but negatively correlated with service elasticities of output (-0.46), and uncorrelated with capital elasticities of output (0.00).

	wm	*arph*	*pm*	*leo*	*mrph*
wm	0.102				
arph	-0.051	0.367			
pm	-0.036	0.004	0.038		
leo	-0.010	-0.289	-0.017	0.279	
mrph	-0.096	0.057	-0.034	0.008	0.100

Table 4: Firm heterogeneity variance-covariance matrix in 2014.

	wm	*arph*	*pm*	*leo*	*mrph*
wm	1				
arph	-0.263	1			
pm	0.576	0.175	1		
leo	-0.057	-0.902	-0.167	1	
mrph	-0.956	0.299	-0.560	0.047	1

Table 5: Firm heterogeneity correlation matrix in 2014.

Conditional on my production function estimates, the structural framework of Section 2 contains two channels through which more productive firms can have a lower labor elasticity of output: (i) non-homotheticities in the production function; (ii) labor input substitution.

The non-homotheticity channel posits that, as firms become larger, their production process becomes more intensive in certain factor inputs – in this case, factor inputs other than labor. Therefore, as more productive firms wish to grow larger, their labor demand grows by less-than-proportionately compared to less productive firms. In my estimates, the prevalence of this channel appears to be limited as I find moderate deviation from constant returns to scale – 98% of firms fall within 5 percentage points of constant returns to scale. Recall that the production function estimation imposes constant returns to scale within sectors *on average*, but does not otherwise preclude firms from departing from constant returns to scale.

The labor input substitution channel works as follows. Since firms face upward-sloping labor supply curves due to labor market frictions, firms that want to hire more workers must offer higher wages. Because more productive firms wish to grow larger than less productive firms, the former face a higher cost of labor relative to other inputs. If labor and other inputs are imperfect substitutes, more productive firms substitute labor with other inputs to avoid higher relative costs of employing labor. In this case, the labor elasticity of output is decreasing in the firm's input intensity of other inputs, reducing the firm's labor demand and offered wage premium.

5.5 Other correlations between pairs of firm characteristics

Negative correlation between wage markdowns (wm) and the marginal revenue product of labor ($mrph$). Wage-posting models that allow wage markdowns to vary across firms predict that this pair of variables is negatively correlated across firms (Burdett and Mortensen, 1998; Gouin-Bonenfant, 2020; Berger et al., 2020), consistent with the model in Section 2. The intuition is that since firms with high $mrph$ (labor demand) pay higher wages, they face a locally less elastic labor supply curve, reflecting less labor market competition locally. Therefore, high $mrph$ firms have less incentives to pay a high fraction of $mrph$ as wages. Wage-bargaining models that allow outside options or bargaining power to vary by $mrph$ also share this prediction (Postel-Vinay and Robin, 2002; Jarosch et al., 2019). This prediction finds support in the last row of the first column in Tables 4 and 5. The covariance between wm and $mrph$ of -0.07 is large relative to most other covariance terms, and the correlation is -0.95. Therefore, given the distribution of marginal revenue productivity of labor across firms, wage markdowns are quantitatively important mechanisms that compress the distribution of firm wage premia.

However, under the assumptions that product markets are perfectly competitive and production technologies are linear in labor in standard frictional labor market models, the marginal revenue product of labor is equal to the average revenue product of labor ($mrph = arph$).[45] This implies that the correlation and covariance between wm and $arph$ is the same as that between wm and $mrph$. Table 5 shows that these correlations are far from being identical. In particular, the correlation between wm and $arph$ (-0.26) is considerably weaker than the covariance between wm and $mrph$ (-0.96).

Positive correlation between wage markdowns (wm) and price-cost markups (pm). This positive correlation suggests that firms with more market power in product markets are

[45] Under the weaker but common assumption of constant price-cost markups and sector-specific Cobb-Douglas production technologies, we have $mrph \propto arph$ instead.

generally not the same firms as those with more market power in labor markets. This has important implications for the aggregate productivity gains of equalizing product and labor market power distortions across firms. I explore the implications further in the next section.

Negative correlation between wage markdowns (wm) and labor elasticities of output (pm). This negative correlation suggests that firms that use labor intensive production technologies tend to have stronger wage-setting power. In a model with frictional labor markets and firms experience random opportunities to automate or outsource production processes such as Arnoud (2018), one rationale for this correlation could be that more labor intensive firms have a stronger bargaining position relative to their employees as they can threaten to substitute capital or intermediate inputs for labor.

Positive correlation between product market power (pm) and firm productivity ($arph$). The second rows, second columns of Tables 4 and 5 show that the relationship between firm productivity and price-cost markups is positive. This is consistent with models of variable markups, such as (Edmond et al., 2015). However, there is some heterogeneity between sectors. While markups and firm productivity are positively correlated in the manufacturing and non-financial service sectors, the correlation is weakly negative in the wholesale and retail sector. Overall, these results imply that price-cost markups compress the firm wage premium distribution.

Negative correlation between product market power (pm) and labor elasticities of output (leo). The fourth row in the third column of Tables 4 and 5 show that firms that charge higher markups (pm) tend to have lower labor elasticities of output (leo). Therefore, labor elasticities of output and price-cost markups generally reinforce each other's role in compressing the firm wage premium distribution.

6 Implications for Labor Misallocation and Labor Shares

Role of input substitution in labor misallocation across firms. The cross-sectional dispersion of the marginal revenue product of labor indicates misallocation of labor inputs (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). Conventional measures of labor misallocation overstate the variance of the marginal revenue product of labor and, hence, the degree of labor misallocation. This is because conventional measures of the variance of the marginal revenue product of labor do not account for firms' ability to substitute labor with other inputs in the presence of labor market frictions. However, my finding of a negative correlation between firm productivity and the labor elasticity of output shows that more productive firms tend to substitute labor with other inputs.

The firm wage premium equation (12) shows that the marginal revenue product of labor consists of the average revenue product of labor ($ARPH$), price-cost markups (PM), and labor elasticity of output (LEO). If PM and LEO are constant across firms – a common assumption – then the $MRPH$ is proportional to, and can be measured by, the $ARPH$. However, the finding in Section 5.4 that $ARPH$ and LEO are strongly negatively correlated implies that the $MRPH$ is considerably less dispersed than the $ARPH$. Indeed, in 2014 the variance of (log) $ARPH$ is over three times larger than the variance of (log) $MRPH$, $\frac{V(arph)}{V(mrph)} = 3.34$. The correlation coefficient between the two is 0.30. Figure 6 plots the de-meaned $mrph$ and $arph$. This mismeasurement stems from the fact that the when the elasticities of substitution between labor and other inputs are different from one, the $arph$ and leo are correlated. When this elasticity is greater than one, leo declines as more productive firms substitute labor with other inputs to circumvent labor market frictions. Labor market frictions generate upward-sloping labor supply curves that require firms to pay higher wages to hire more workers. Labor substitution therefore partially offsets more productive firms' higher labor demand and firm wage premium.

To get a sense of the extent to which aggregate efficiency gains from removing labor market frictions could potentially be overstated, I perform a Hsieh and Klenow (2009) exercise to compare the implied efficiency gains from using the conventional measure and my estimated measure of $MRPH$ dispersion. Let s denote the sector. As in Hsieh and Klenow (2009), assume that the sector-specific CES aggregator over firm-level output is $Y_s = \left(Y_{sj}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$. To derive closed-form solutions for aggregate sectoral efficiency (TFP_s), I impose the assumption that firms operate sector-specific Cobb-Douglas constant returns-to-scale production functions $Y_{sj} = X_{sj} K_{sj}^{\alpha_s^K} H_{sj}^{\alpha_s^H} M_{sj}^{\alpha_s^M}$. As in Section 2, firms face firm-specific labor supply curves. I assume that labor market frictions are the only distortions present. In Appendix B, I show that under these assumptions the sectoral TFP gains from removing labor market frictions is given by:

$$\ln TFP_s^* - \ln TFP_s = \frac{\rho}{2} V_s \left(\ln(MRPH_{sj}^{\alpha_s^H}) \right)$$

where TFP^* denotes the sectoral efficiency (total factor productivity) in a world without labor market frictions. Let $M\tilde{R}PH$ denote the measure of the marginal revenue product of labor that does not account for the negative correlation between $ARPH$ and LEO, while $M\hat{R}PH$ denotes the measure that does. Then, the average relative sectoral efficiency gains from removing labor market frictions is:

$$E\left[\frac{\ln T\tilde{F}P_s^* - \ln TFP_s}{\ln T\hat{F}P_s^* - \ln TFP_s} \right] = E\left[\frac{V_s \left(\ln(M\tilde{R}PH_{sj}) \right)}{V_s \left(\ln(M\hat{R}PH_{sj}) \right)} \right]$$

In 2014, on average the relative sectoral efficiency gains ratio is 2.93. This implies that the conventional measure of labor misallocation on average overstates the efficiency gains of removing labor market frictions by almost 3 times, relative to the measure of labor misallocation that takes the negative correlation between $ARPH$ and LEO into account.

New explanation for lower labor shares of revenue among more productive firms. The literature proposes two main explanations for why highly productive firms have low labor revenue shares: product market power (De Loecker et al., 2020) and labor market power (Gouin-Bonenfant, 2020). My finding of a negative correlation between firm productivity and the labor elasticity of output provides a new explanation.

The decline of the U.S. aggregate labor share of income has attracted significant academic attention. Earlier studies make the case for changes in the aggregate production technology, either through capital-labor substitution (Karabarbounis and Neiman, 2014; Oberfield and Raval, 2020) or intermediate-input-labor substitution (Elsby et al., 2013). However, recent research shows that the U.S. labor share decline is explained by the reallocation of sales towards highly productive "superstar" firms, which have low labor shares (Autor et al., 2020; Kehrig and Vincent, 2020). Hypotheses based on changes in the aggregate production technology do not account for this pattern.

Why do more productive firms have lower labor shares? To look into this, I categorize firms by size (sales revenue) into ten equal-sized groups within each two-digit sector. Consistent with De Loecker et al. (2020), Figure 1 shows that high productivity, low labor share firms charge higher price-cost markups. they tend to mark down wages more, which provides empirical support for the hypothesis that superstar firms reduce aggregate labor shares because they have more labor market power (Gouin-Bonenfant, 2020). Figure 1 also provides a new explanation for superstar firms' low labor shares: high productivity firms hav a low labor elasticity of output. Therefore, while the aggregate production technology cannot account for the fact that reallocation of sales towards low labor share firms drives the U.S. labor share decline, at the micro level, production technologies can play an important role.

To get a sense of which channel of heterogeneity matters most for firm-level labor revenue shares, I decompose firm-level differences in labor shares. Write the labor revenue share in logs (ls):

$$ls_j = wm_j - pm_j + leo_j$$

Then, labor shares can be decomposed as such:

$$V(ls) = CV(ls, wm) - CV(ls, pm) + CV(ls, leo)$$

Table 6 reports the labor share decomposition overall and by sectors. Overall, the decomposition suggests that labor elasticities of output account for the bulk of variation in labor shares, followed by wage markdowns and price-cost markups, with each playing an important role.

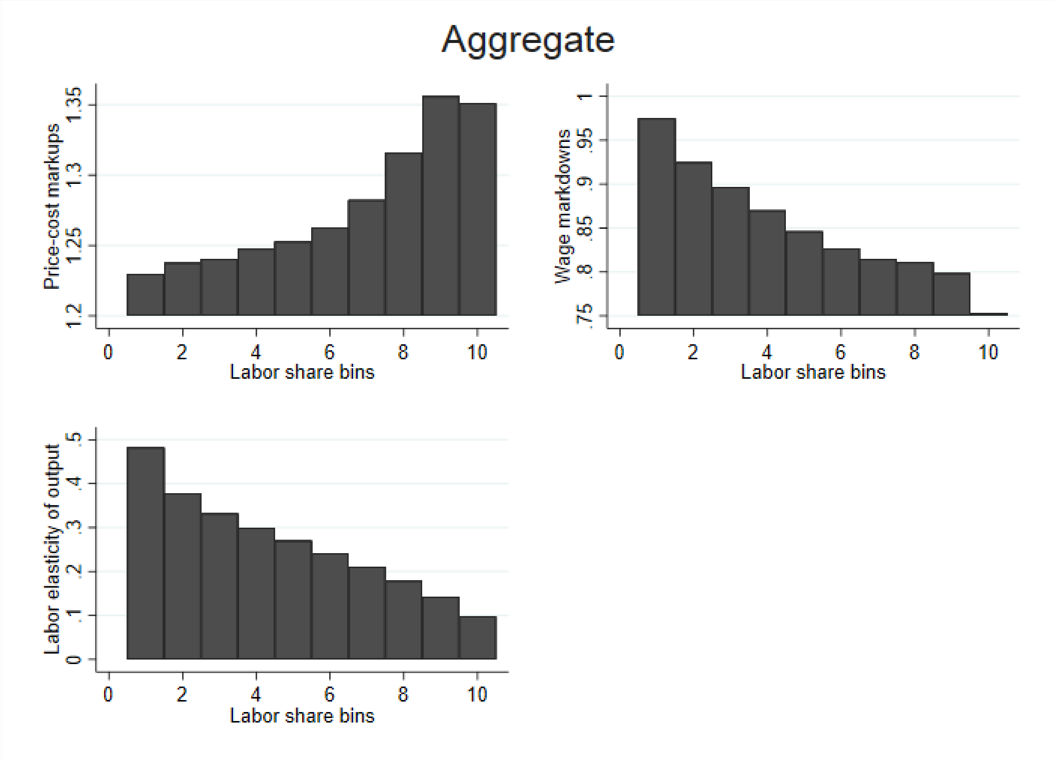

Figure 1: Labor shares and firm characteristics among French firms in 2014.

	Overall	CN	MN	NFS	W	T
$V(ls)$	0.386	0.186	0.231	0.255	0.360	0.166
$CV(ls, wm)$	0.061	0.037	0.041	0.030	0.115	0.005
$CV(ls, pm)$	-0.020	0.004	-0.018	-0.016	0.009	0.009
$CV(ls, leo)$	0.305	0.153	0.172	0.210	0.254	0.152

Table 6: Labor share decomposition in 2014.

Note: CN: Construction, MN: Manufacturing, NFS: Non-financial services, W: Wholesale and retail, T: Transportation.

However, there are some clear differences between sectors in the relative importance of these channels of heterogeneity for labor shares. Consider three large sectors: manufacturing, non-financial services, and wholesale and retail. Figure 9 shows that, in the manufacturing sector, the labor revenue shares are decreasing in firms' price-cost markups and labor elasticities of output. However, wage markdowns are U-shaped – manufacturing firms with lower labor shares mark down wages more, but those with the lowest labor shares (and highest productivity) do not mark down wages by much. The non-financial service sector displays patterns similar to the manufacturing sector, as shown in Figure 10. In the wholesale and retail sector, however, Figure 11 and Table 6 show that low labor elasticities of output and large wage markdowns entirely explain low labor revenue shares. Across all sectors, wage markdowns appear to be particularly relevant for labor shares among wholesale and retail firms.

Role of product and labor market power in labor misallocation across firms. I now show that my finding of a negative cross-sectional correlation between product and labor market power implies that these channels of heterogeneity partially dampen each other's effect on labor input misallocation. However, as shown in Figure 1, on average, the most productive firms tend to have the greatest product and labor market power. Therefore, among these firms, product and labor market power amplify each other's effect on labor misallocation.

Market power in the product (Edmond et al., 2015; Peters, 2020) and labor markets (Berger et al., 2020) have been separately shown to distort the allocation of labor across firms. However, whether they amplify or dampen each other's effects on labor misallocation depends on their cross-sectional correlation. Since both product and labor market power reduce firm size below the perfect competition benchmark, these distortions amplify each other's effects when they are positively correlated, as they tend to distort labor demand of the same firms. When they are negatively correlated, the opposite is true.

To show this, I set up a simple illustrative model in Appendix B and follow the methods of Hsieh and Klenow (2009) to derive the total factor productivity (TFP) of a given sector s:

$$\ln TFP_s \approx \gamma_s^a - \gamma_s^b V_s(\ln(PM_j \cdot WM_j^{-\gamma_s^c}))$$

where $\gamma_s^a > 0$, $\gamma_s^b > 0$, and $\gamma_s^c \in [0, 1]$ are constants. The higher the inverted wage markdown (WM^{-1}), the stronger the firm's labor market power. This equation shows that if product and labor market power are perfectly negatively correlated, so that $PM_j \cdot WM_j^{-\gamma_s^c} = \text{constant } \forall j \in s$, then not only are there are no TFP gains to equalizing market power distortions within sector s, but policies that generate dispersion in the joint market power component $PM_j \cdot WM_j^{-\gamma_s^c}$ lead to input misallocation and TFP losses. As Table 5 shows, the correlation between price-cost markups and inverted wage markdowns is -0.58, implying that TFP gains to equalizing both product and labor market power across firms are partially offset by their negative correlation.

To see the intuition, imagine that all firms have the same productivity draw, but they have different product and labor market power (price-cost markups and inverted wage markdowns).

Suppose that product and labor market power are negatively correlated and perfectly offset each other. In this case, the marginal revenue product of labor is constant across firms and there is no misallocation: firm sizes are the same in the cross-section. Next, suppose that we equalize markups across firms. Then, the only source of distortion to allocative efficiency is labor market power. Now, high labor market power firms are too small, and low labor market power firms are too large, generating a non-degenerate firm size distribution entirely due to misallocation.

7 Concluding Remarks

I investigate how firm characteristics determine the wage premium a firm pays relative to other firms for identical workers. To do so, I develop and implement a novel structural decomposition of the firm wage premium distribution. While a large literature emphasizes the importance of firms' labor productivity and wage-setting power in a frictional labor market, this paper highlights the role of firms' product market power and the labor share of production. My decomposition suggests that, without taking into account the role of firms' product market power and the labor share of production, workhorse models that generate a firm wage premium distribution overestimate the role of firm-level differences in labor productivity and wage-setting power. The decomposition also uncovers important correlations between these firm characteristics. First, there is a negative relationship between firm productivity and the labor share of production. Second, product and labor market power are negatively correlated in the cross-section.

These findings have important implications. First, my findings show that exceptionally productive superstar firms are different from other firms in several ways (Autor et al., 2020). I confirm that superstar firms charge disproportionately higher price-cost markups (De Loecker et al., 2020), but also provide empirical support for the hypothesis that these firms pay markedly stronger labor market power (Gouin-Bonenfant, 2020), and offer a new explanation for their low labor shares of revenue: low labor share of production. Second, the negative relationship between firm productivity and the labor share of production implies that conventional measures of the variance of the marginal revenue product of labor, a sufficient statistic for labor misallocation (Hsieh and Klenow, 2009), overstates the degree of labor misallocation across firms. Third, while the effects of product and labor market power on input misallocation are often studied separately (Edmond et al., 2018; Berger et al., 2020), their cross-sectional relationship decides whether they amplify or dampen each other's effects on misallocation.

The structural decomposition framework also has a number of potential applications. One application could be to study the extent to which the long-term wage loss from losing the firm wage premium for outsourced (Goldschmidt and Schmieder, 2017) or displaced workers

(Schmieder et al., 2018; Lachowska, Mas, and Woodbury, 2018) is due to the loss of bargaining power or to moving to a less productive firm. Other applications could be to use this framework to understand the rising dispersion of the firm wage premium in countries such as Germany (Card et al., 2013), or to decompose the fall in the US aggregate labor share into contributions of each dimension of firm heterogeneity (Autor et al., 2020; Kehrig and Vincent, 2020).

Bibliography

Abel, W., S. Tenreyro, and G. Thwaites (2018). Monopsony in the UK. mimeo LSE.

Abowd, J., F. Kramarz, and D. Margolis (1999). High wage workers and high wage firms. *Econometrica 67*(2), 251–333.

Ackerberg, D., K. Frazer, and G. Caves (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Alvarez, J., F. Benguria, N. Engbom, and C. Moser (2018). Firms and the decline in earnings inequality in Brazil. *American Economic Journal: Macroeconomics 10*(1), 149–189.

Andrews, M., L. Gill, T. Schank, and R. Upward (2008). High wage workers match with high wage firms: Clear evidence of the effects of limited mobility bias. *Economics Letters 117*(3), 824–827.

Arnoud, A. (2018). Automation threat and wage bargaining. *working paper*.

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031.

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics 135*(2), 645–709.

Azar, J., I. Marinescu, and M. Steinbaum (2019). Measuring labor market power two ways. *AEA Papers & Proceedings 109*, 317–329.

Azar, J., I. Marinescu, and M. I. Steinbaum (2017). Labor Market Concentration. NBER Working Papers 24147, National Bureau of Economic Research, Inc.

Azar, J., I. E. Marinescu, M. Steinbaum, and B. Taska (2018, March). Concentration in US Labor Markets: Evidence from Online Vacancy Data. IZA Discussion Papers 11379, Institute for the Study of Labor (IZA).

Azkarate-Askasua, M. and M. Zerecero (2020). The aggregate effects of labor market concentration. *Working paper*.

Bagger, J., B. Christensen, and D. Mortensen (2014). Productivity and wage dispersion: Heterogeneity or misallocation. *working paper*.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J. Robin (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review 104*(6), 1551–1596.

Bagger, J. and R. Lentz (2019). An empirical model of wage dispersion and sorting. *Review of Economic Studies 86*(1), 153–190.

Barkai, S. (2020). Declining labor and capital shares. *Journal of Finance 75*(5), 2421–2463.

Barlevy, G. (2008). Identification of search models using record statistics. *Review of Economic Studies 75*(1), 29–64.

Barth, E., A. Bryson, J. Davis, and R. Freeman (2016). It's where you work: Increases in earnings dispersion across establishments and individuals in the US. *Journal of Labor Economics 34*(52), 67–97.

Basu, S., J. Fernald, J. Fisher, and M. Kimball (2013). Sector-specific technical change. *working paper*.

Bell, B., P. Bukowski, and S. Machin (2018). Rent-sharing and inclusive growth. *CEP Discussion Paper No. 1584*.

Benmelech, E., N. Bergman, and H. Kim (2018). Strong employers and weak employees: How does employer concentration affect wages? Working Paper 24307, National Bureau of Economic Research.

Berger, D., K. Herkenhoff, and S. Mongey (2020). Labor market power. Working paper.

Berlingieri, G., P. Blanchenay, and C. Criscuolo (2017). The great divergence(s). *OECD STI Policy Paper No. 39*.

Bilal, A., N. Engbom, S. Mongey, and G. Violante (2019). Firm and worker dynamics in a frictional labor market.

Blanchflower, D., A. Oswald, and P. Sanfey (1996). Wages, profits and rent-sharing. *Quarterly Journal of Economics 111*, 227–252.

Boal, W. M. and M. Ransom (1997). Monopsony in the labor market. *Journal of Economic Literature 35*(1), 86–112.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020). How much shoould we trust estimates of firm effects and worker sorting. *Working paper*.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer-employee data. *Econometrica 87*(3), 699–738.

Brooks, W., J. Kaboski, A. Li, and W. Qian (2019). Exploitation of labor? classical monopsony and labor's share. *working paper No. 25660, National Bureau of Economic Research*.

Burdett, K. and D. Mortensen (1998). Wage differentials, employer size, and unemployment. *International Economic Review 39*(2), 257–273.

Caldwell, S. and O. Danieli (2019). Outside options in the labor market. *working paper*.

Caldwell, S. and N. Harmon (2019). Outside options, bargaining, and wages: Evidence from coworker networks. *working paper*.

Card, D., A. Cardoso, J. Heining, and P. Kline (2018). Firms and labor market inequality: Evidence and some theory. *Journal of Labor Economics 36*(1), 13–70.

Card, D., A. Cardoso, and P. Kline (2015). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics 131*(2), 633–686.

Card, D., J. Heining, and P. Kline (2013). Workplace heterogeneity and the rise of West German wage inequality. *Quarterly Journal of Economics 128*(3), 967–1015.

Carlsson, M., J. Messina, and O. Skans (2014). Wage adjustment and productivity shocks. *Economic Journal 126*(595), 1739–1773.

Chetty, R., J. Friedman, and J. Rockoff (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review 104*, 2593–2632.

Dauth, W., S. Findeisen, and J. Suedekum (2018). Adjusting to globalization in Germany. *working paper*.

De Loecker, J. and J. Eeckhout (2018). The rise of market power and the macroeconomic implications. *working paper*.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J. and P. Goldberg (2014). Firm performance in a global market. *Annual Review of Economics 6*, 201–227.

De Loecker, J. and P. Scott (2016). Estimating market power: Evidence from the us brewery industry. *working paper*.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–2471.

Diamond, P. (1982). Aggregate demand management in search equilibrium. *Journal of Political Economy 90*(5), 881–894.

Dobbelaere, S. and J. Mairesse (2013). Panel data estimates of the production function and product and labor market imperfections. *Journal of Applied Econometrics 28*(1), 1–46.

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2018). Monopsony in Online Labor Markets. NBER Working Papers 24416, National Bureau of Economic Research, Inc.

Edmond, C., V. Midrigan, and D. Xu (2015). Competition, markups, and the gains from international trade. *American Economic Review 105*(10), 3183–3221.

Edmond, C., V. Midrigan, and D. Xu (2018). How costly are markups? *working paper No. 24800, National Bureau of Economic Research*.

Eeckhout, J. and P. Kircher (2011). Identifying sorting – in theory. *Review of Economic Studies 78*(3), 872–906.

Elsby, M., B. Hobijn, and A. Şahin (2013). The decline of the us labor share. *Brookings Papers on Economic Activity 2*, 1–63.

Elsby, M. and R. Michaels (2013). Marginal jobs, heterogeneous firms, and unemployment flows. *American Economic Journal: Macroeconomics 5*(1), 1–48.

Elsby, M., R. Michaels, and M. Ratner (2018). Vacancy chains. *working paper*.

Engbom, N. and C. Moser (2018). Earnings inequality and the minimum wage: Evidence from Brazil. *working paper*.

Faggio, G., K. Salvanes, and J. Van Reenen (2007). The evolution of inequality in productivity and wages: Panel data evidence. *working paper 13351, National Bureau of Economic Research*.

Flynn, Z., A. Gandhi, and J. Traina (2019). Measuring markups with production data. *working paper*.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Gandhi, A., S. Navarro, and D. Rivers (2019). On the identification of gross output production functions. *Journal of Political Economy Forthcoming*(921-947).

Garin, A. and F. Silverio (2019). How responsive are wages to demand within the firm? evidence from idiosyncratic export demand shocks. *working paper*.

Goldschmidt, D. and J. Schmieder (2017). The rise of domestic outsourcing and the evolution of the German wage structure. *Quarterly Journal of Economics 132*(3), 1165–1217.

Gouin-Bonenfant, E. (2020). Productivity dispersion, between-firm competition, and the labor share. *working paper*.

Gourio, F. and L. Rudanko (2014). Customer capital. *Review of Economic Studies 81*(3), 1102–1136.

Hall, R. and A. Krueger (2010). Evidence on the determinants of the choice between wage posting and wage bargaining. *working paper No. 17159, National Bureau of Economic Research*.

Haltiwanger, J., H. Hyatt, L. Kahn, and E. McEntarfer (2018). Cyclical job ladders by firm size and firm wage. *American Economic Journal: Macroeconomics 10*(2), 52–85.

Helpman, E., O. Itskhoki, M. Muendler, and S. Redding (2017). Trade and inequality: From theory to estimation. *Review of Economic Studies 84*, 357–405.

Hershbein, B., C. Macaluso, and C. Yeh (2020). Monopsony in the U.S. labor market. *Working paper*.

Hildreth, A. and A. Oswald (1997). Wages and rent-sharing: Evidence from company and establishment panels. *Journal of Labor Economics 15*, 318–337.

Hornstein, A., P. Krusell, and G. Violante (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review 101*(7), 2873–2898.

Hsieh, C. and P. Klenow (2009). Misallocation and manufacturing tfp in China and India. *Quarterly Journal of Economics 124*(4), 1403–1448.

Hubmer, J. (2019). The race between preferences and technology. *working paper*.

Jarosch, G., J. Nimczik, and I. Sorkin (2019). Granular search, market structure, and wages. *working paper*.

Jochmans, K. and M. Weidner (2019). Fixed effect regressions on network data. *working paper*.

Kaas, L. and P. Kircher (2015). Efficient firm dynamics in a frictional labor market. *American Economic Review 105*, 3030–3060.

Karabarbounis, L. and B. Neiman (2014). The global decline of the labor share. *Quarterly Journal of Economics 129*(1), 61–103.

Katz, L. and L. Summers (1989). Industry rents: Evidence and implications. *Brookings Papers on Economic Activity: Microeconomics*, 209–290.

Kehrig, M. and V. Vincent (2020). The micro-level anatomy of the labor share decline. *working paper*.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2017). Who profits from patents? evidence from innovative firms. *working paper*.

Lachowska, M., A. Mas, and S. Woodbury (2018). Sources of displaced workers' long-term earnings losses. *working paper*.

Lamadon, T., M. Mogstad, and B. Setzler (2019). Imperfect competition, compensating differentials, and rent sharing in the u.s. labor market. *working paper*.

Lentz, R., S. Piyapromdee, and J. Robin (2019). On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data. *working paper*.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*(2), 317–341.

Lopes de Melo, R. (2018). Firm wage differentials and labor market sorting: Reconciling theory and evidence. *Journal of Political Economy 126*(1), 313 – 346.

Manning, A. (2006). A generalised model of monopsony. *Economic Journal 116*(508), 84–100.

Manning, A. (2011). Monopsony in motion: Imperfect competition in labor markets. *Princeton University Press*.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*, 1695–1725.

Mertens, M. (2019). Micro-mechanisms behind declining labor shares: Market power, production processes, and global competition. *working paper*.

Moen, E. (1997). Competitive search equilibrium. *Journal of Political Economy 105*(2), 385–341.

Morlacco, M. (2019). Market power in input markets: Theory and evidence from French manufacturing. *working paper*.

Mortensen, D. (1982). Property rights and efficiency in mating, racing, and related games. *American Economic Review 72*(5), 968–979.

Mortensen, D. (2010). Wage dispersion in the search and matching model. *American Economic Review: Papers and Proceedings 100*(2), 338–342.

Naidu, S., Y. Nyarko, and S.-Y. Wang (2016). Monopsony power in migrant labor markets: Evidence from the United Arab Emirates. *Journal of Political Economy 124*(6), 1735–1792.

Oberfield, E. and M. Raval (2020). Micro data and macro technology. *Econometrica (forthcoming)*.

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Peters, M. (2020). Heterogeneous markups, growth and endogenous misallocation. *Econometrica (forthcoming)*.

Pissarides, C. (1985). Short-run equilibrium dynamics of unemployment, vacancies, and real wages. *American Economic Review 75*(4), 676–690.

Postel-Vinay, F. and J. Robin (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica 70*(6), 2295–2350.

Restuccia, D. and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics 11*(4), 707–720.

Rinz, K. (2018, September). Labor Market Concentration, Earnings Inequality, and Earnings Mobility. CARRA Working Papers 2018-10, Center for Economic Studies, U.S. Census Bureau.

Robinson, J. (1933). The economics of imperfect competition. *Palgrave Macmillan*.

Schaal, E. (2017). Uncertainty and unemployment. *Econometrica 85*(6), 1675–1721.

Schmieder, J., T. Von Wachter, and J. Heining (2018). The costs of job displacement over the business cycle and its sources: Evidence from Germany. *working paper*.

Schubert, G., A. Stansbury, and B. Taska (2019). Getting labor markets right: Outside options and occupational mobility. *working paper*.

Shorrocks, A. (1982). Inequality decomposition by factor components. *Econometrica 50*(1), 193–211.

Shorrocks, A. (2013). Decomposition procedures for distributional analysis: A unified framework based on the shapley value. *The Journal of Economic Inequality 11*(1), 99–126.

Slichter, S. (1950). Notes on the structure of wages. *Review of Economics and Statistics 32*(1), 80–91.

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

Sorkin, I. (2018). Ranking firms using revealed preference. *Quarterly Journal of Economics 133*(3), 1331–1393.

Stole, L. and J. Zwiebel (1996). Intra-firm bargaining under non-binding contracts. *Review of Economic Studies 63*(3), 375–410.

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy 112*(6), 1181–1222.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

Tortarolo, D. and R. Zarate (2018). Measuring imperfect competition in product and labor markets: An empirical analysis using firm-level production data. *working paper*.

Webber, D. (2015). Firm market power and the u.s. earnings distribution. *Labor Economics 35*, 123–134.

Wooldridge, J. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

Wright, R., P. Kircher, B. Julien, and V. Guerrieri (2018). Directed search: A guided tour. *working paper*.

Online Appendix

Understanding High-Wage and Low-Wage Firms

Horng Chern Wong
CREST - Ecole Polytechnique and Stockholm University

Appendix A: Data and Measurement

To estimate production functions using FICUS (1995-2007) and FARE (2008-2014) firm balance sheet data and the DADS matched employer-employee data from France, I use the following measures:

- Sales revenue (PY): measured by the variable CATOTAL in FICUS, and REDI_R310 in FARE.
- Efficiency units of labor ($H = \bar{E}L$): the DADS provides the number of hours worked for each worker under NBHEUR, which enables the researcher to measure total hours (L) at a given firm. The average efficiency of workers ($\bar{E}$) is then measured as the difference between the unconditional mean wage and the firm wage premium, according to the theory.
- Capital (K): measured as total fixed physical assets under variable names IMMOCOR in FICUS, and IMMO_CORP in FARE.
- Materials (M): the French balance sheet data provides a breakdown of intermediate inputs into three components – materials purchased to be used as inputs in production (ACHAMPR in FICUS, REDI_212 in FARE), goods purchased to be resold (ACHAMAR in FICUS, REDI_210 in FARE), and purchase of services (details provided next). I correct for changes in inventory for materials to be used in production (using VARSTMP in FICUS, REDI_213 in FARE) and for goods purchased to be resold (VARSTMA in FICUS, REDI_211 in FARE). I measure M as the sum of these variables, except services.
- Services (O): measured as AUTACHA in FICUS, and REDI_214 in FARE. These variables include the costs of outsourcing and advertising.

	%ΔSales
%ΔLabor	0.27
%ΔCapital	0.19
%ΔMaterials (transform)	0.28
%ΔMaterials (resale)	0.34
%ΔServices	0.69

Table 7: Contemporaneous correlations between factor input growth rates and sales revenue growth rates, 1995-2014.

Appendix B: Shapley Decomposition Method

Recall the firm wage premium equation, written in natural logs:

$$\phi = wm + arph - pm + leo$$

This structural wage equation can now be seen as a linear regression. To assess the relative importance of each dimension of firm heterogeneity on the cross-sectional variation in firm wage premia, I implement the Shapley Decomposition of R^2 (Shorrocks, 1982, 2013). With four dimensions of heterogeneity, an analysis of variance approach generates ten variance and covariance terms, with potential negative contributions of certain variables, depending on the joint distribution of the explanatory variables. The Shapley approach offers simplicity in terms of the interpretation of the contribution of each dimension of heterogeneity, as it partitions the total R^2 into the marginal contributions of each variable. This gives four partial R^2's, one for each dimension of firm heterogeneity. Moreover, the partial R^2's never take negative values.

In cooperative game theory, the Shapley value is the unique solution to distributing the total surplus generated by a coalition of players. The idea is to view each variable (dimension of firm heterogeneity) as a player in a coalition, and the total R^2 as the total surplus. The Shapley decomposition then applies the Shapley value to partition the total R^2, based on each variable's marginal contribution. It is based on the following axioms, under which the Shapley value is derived:

- Efficiency: the entire surplus is distributed.
- Symmetry: any two players (variables) with same marginal contribution to the total surplus obtains the same share.
- Monotonicity: the total surplus is non-decreasing in the number of players.
- Null player: the null player does not obtain a share of the surplus.

The partial R^2 of a variable $X_j = \{wm, arph, pm, leo\}$ can then be written as:

$$R^2(x_j) = \sum_{T \subseteq V \setminus \{X_j\}} \frac{k! \cdot (p-k-1)!}{p!} \left(R^2(T \cup \{X_j\}) - R^2(T) \right)$$

where p denotes the number of variables, which is equal to four in this case; T is a regression with k number of variables, and V is the set of all combinations of regressor variables excluding X_j.

Appendix C: Assessing the AKM Regression Specification

Conditional Exogenous Mobility & Symmetry

The conditional exogenous mobility assumption underlying AKM regressions requires that worker mobility is as good as random conditional on observed worker characteristics, worker fixed effects, and firm fixed effects. If this assumption is a reasonable approximation, then one should observe systematic worker mobility up and down the firm effect quartiles. Moreover, workers should experience approximately symmetric wage changes as they move along the firm effect quartiles, given the log additive regression specification. On the other hand, in structural models of worker-firm sorting based on comparative advantage (Eeckhout and Kircher, 2011; Lopes de Melo, 2018), worker mobility is based on the match-specific component of wages, which is captured by the residual component of wages in the AKM regression. In this class of models the AKM regression is misspecified in the sense that the wage gains depend on value of the particular worker-firm match, for example, if highly skilled workers have a comparative advantage in high productivity firms. In the event-study exercise show in figure 2, I estimate the AKM regression with clustered firm-group effects following Bonhomme et al. (2019) between 2009 and 2014, and compare the changes in mean log wages for workers who move between quartiles of firm fixed effects, following Card et al. (2013). Figure 2 shows that workers who move up firm quartiles experience a wage gain similar in magnitude to the wage loss of workers who move down firm quartiles.

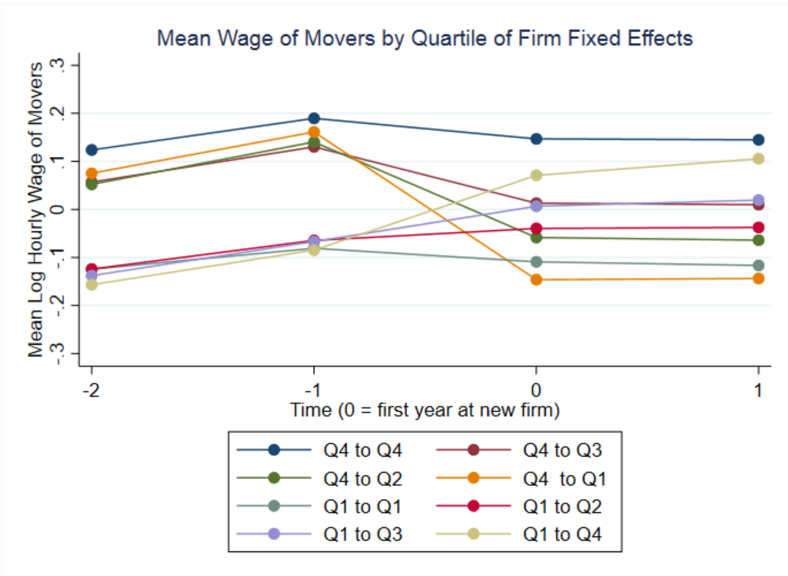

Figure 2: Worker mobility and wage changes by quartiles of firm effects (2009-2014).

An alternative way to assess the AKM regression specification is to compare the changes in residual wages to changes in firm effects, following Chetty, Friedman, and Rockoff (2014) and

Sorkin (2018). This is similar to the above method (Card et al., 2013). I run the following regression among all employer-to-employer transitions:

$$w_{it}^r - w_{it-1}^r = \alpha_0 + \alpha_1 \left(\phi_{g(j(i,t))} - \phi_{g(j(i,t-1))}\right) + \epsilon_{it} \;\; \forall \; (i,t), \; g(j(i,t)) \neq g(j(i,t-1))$$

where $w_{it}^r = w_{it} - x_{it}'\hat{\beta}$ denotes residualized wages and $\phi_{g(j(i,t))}$ are the firm-group fixed effects. If the AKM regression is not mis-specified, the estimated coefficient $\hat{\alpha_1}$ will equal 1. I find $\hat{\alpha}_1 = 0.857$, with a standard error of 0.007. To see this visually, figure 3 plots the changes in residual wages and the changes in firm fixed effects in 100 bins of changes in firm fixed effects. In models of assortative matching based on comparative advantage (Eeckhout and Kircher, 2011; Lopes de Melo, 2018), worker mobility is strongly driven the residual component of the AKM regression, implying that AKM regressions are mis-specified. As Sorkin (2018) shows, these models predict that worker mobility entails a wage gain, regardless of the direction of worker mobility in terms of the estimated firm effects, as workers move to firms at which they have a comparative advantage: there is a V-shape around zero changes in firm effects. The patterns of wage changes upon changes in firm fixed effects shown in figure 3 do not resemble a V-shape around zero.

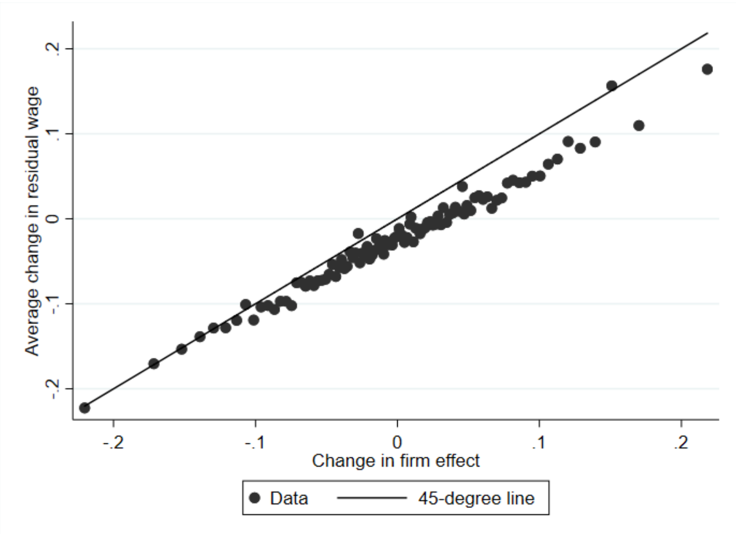

Figure 3: Average wage changes from worker mobility by declines of changes in firm premia (2009-2014).

Another way to assess the log additivity of the worker and firm components of wages is to group worker and firm fixed effects into 10 deciles each, generating 100 worker-firm fixed effect deciles, then plot the mean estimated residuals within each worker-firm fixed effect decile. If the firm wage premium depends strongly on the worker's skill type, log additivity would be severely violated, and one should observe that the estimated residuals systematically varies across worker-firm fixed effect deciles. Figures 4 and 5 show that the mean estimated residuals are approximately zero across worker-firm fixed effect deciles, with the exception of the very top deciles of high-wage workers who are employed at low-wage firms at the very bottom deciles. As a further robustness check, I follow Bonhomme et al. (2019) and run the BLM regression with worker-firm interactions, but with only 20 firm groups and 6 worker groups to maintain computational tractability. Moving from an additive to an interacted regression model gives a gain in R^2 of 0.01.

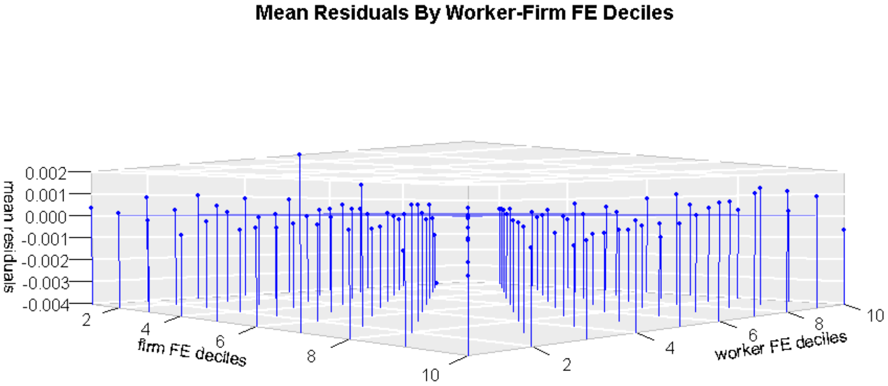

Figure 4: Mean estimated residuals by worker-firm deciles (2014)

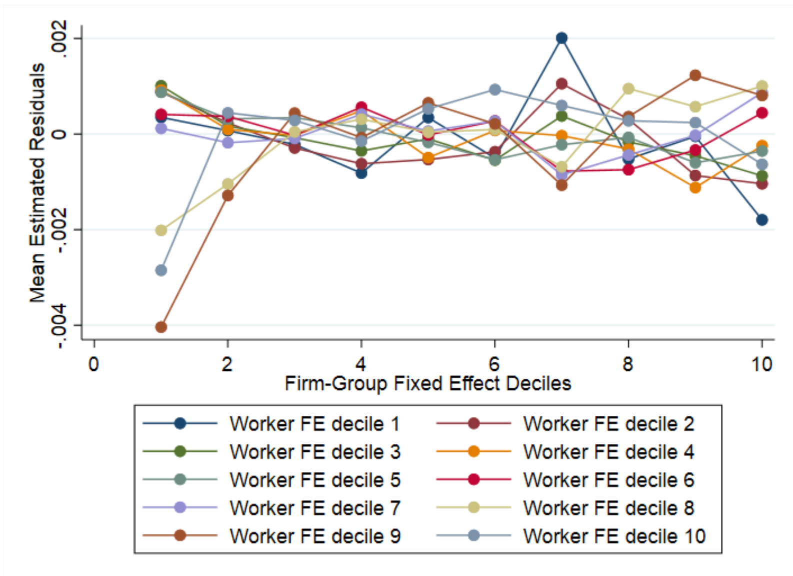

Figure 5: Mean estimated residuals by worker-firm deciles (2014)

Appendix D: Additional Figures and Tables

Summary Statistics

	Summary Statistics: Employees	
Sample size		
People-years	158,163,180	
Firm-years	4,907,010	
Average number of workers per year	7,908,159	
Average number of firms per year	245,351	
Wage distribution		
Mean log Wage	2.53	
Variance log wage	0.19	
Fraction between-firms	0.44	
Efficiency Units & Firm Premium		
Variance $\bar{e}$	0.05	
Variance ϕ	0.009	
Correlation $(\bar{e}, \phi)$	0.42	
	Summary Statistics: Employers	
	Mean	Variance
Log production value	13.71	1.30
Log employment	2.53	0.80
Log capital stock	12.13	2.67
Log intermediate inputs	12.82	2.02

Table 8: Summary statistics: employers and employees (1995-2014).

Distributions of Estimated Firm Characteristics

	Mean	Median	Variance	25th Pct	75th Pct
Wage markdown	0.85	0.82	0.14	0.69	0.96
Price-cost markup	1.27	1.26	0.12	1.14	1.39
Labor elasticity of output	0.41	0.41	0.02	0.57	0.24
Material elasticity of output	0.53	0.53	0.02	0.68	0.37
Service elasticity of output	0.53	0.53	0.02	0.68	0.37
Capital elasticity of output	0.06	0.06	0.00	0.09	0.03
log Average revenue product of labor	4.52	4.49	0.25	4.20	4.81
log Marginal revenue product of labor	2.83	2.82	0.09	2.67	2.98
Number of firms	243,453				

Table 9: Summary statistics of firm characteristics within 2-digit sectors in 2014. The overall mean of each dimension is kept constant.

Wage markdowns	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.85	0.80	1.17	0.97	0.67	0.55	0.15
Construction	0.87	0.84	1.11	0.97	0.72	0.62	0.11
Manufacturing	0.82	0.78	1.04	0.90	0.67	0.58	0.13
Non-Financial	0.87	0.81	1.17	0.98	0.69	0.58	0.15
Transportation	0.77	0.73	0.91	0.81	0.66	0.58	0.14
Wholesale-Retail	0.88	0.81	1.32	1.04	0.62	0.48	0.20

Table 10: Distribution of estimated wage markdowns by sector in 2014.

Price-cost markups	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	1.28	1.24	1.62	1.41	1.11	1.00	0.13
Construction	1.26	1.26	1.52	1.39	1.12	1.00	0.06
Manufacturing	1.19	1.17	1.39	1.27	1.07	0.98	0.08
Non-Financial	1.26	1.21	1.53	1.33	1.12	1.03	0.08
Transportation	1.14	1.12	1.33	1.22	1.02	0.95	0.04
Wholesale-Retail	1.38	1.27	1.74	1.58	1.17	1.00	0.24

Table 11: Distribution of estimated price-cost markups by sector in 2014.

Labor elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.26	0.24	0.43	0.33	0.17	0.11	0.02
Construction	0.30	0.30	0.42	0.37	0.24	0.17	0.01
Manufacturing	0.29	0.29	0.41	0.35	0.23	0.16	0.01
Non-Financial	0.32	0.27	0.56	0.42	0.21	0.17	0.02
Transportation	0.32	0.32	0.45	0.39	0.25	0.18	0.01
Wholesale-Retail	0.16	0.15	0.25	0.20	0.12	0.09	0.01

Table 12: Distribution of estimated labor elasticities of output by sector in 2014.

Material elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.41	0.41	0.72	0.55	0.26	0.13	0.05
Construction	0.31	0.32	0.39	0.36	0.28	0.22	0.01
Manufacturing	0.37	0.40	0.52	0.46	0.30	0.21	0.02
Non-Financial	0.28	0.25	0.53	0.48	0.12	0.03	0.04
Transportation	0.12	0.16	0.22	0.20	0.06	0.00	0.01
Wholesale-Retail	0.63	0.64	0.81	0.75	0.54	0.45	0.03

Table 13: Distribution of estimated intermediate input elasticities of output by sector in 2014.

Service elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.28	0.26	0.48	0.36	0.19	0.13	0.02
Construction	0.33	0.31	0.48	0.39	0.25	0.20	0.01
Manufacturing	0.28	0.27	0.42	0.34	0.21	0.17	0.01
Non-Financial	0.35	0.32	0.55	0.44	0.24	0.19	0.02
Transportation	0.49	0.47	0.70	0.59	0.38	0.30	0.02
Wholesale-Retail	0.19	0.18	0.28	0.24	0.12	0.08	0.01

Table 14: Distribution of estimated intermediate input elasticities of output by sector in 2014.

Capital elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.05	0.05	0.09	0.07	0.02	0.01	0.00
Construction	0.07	0.07	0.09	0.08	0.05	0.03	0.00
Manufacturing	0.05	0.05	0.08	0.07	0.04	0.02	0.00
Non-Financial	0.05	0.03	0.12	0.08	0.01	0.01	0.00
Transportation	0.07	0.08	0.12	0.10	0.05	0.02	0.00
Wholesale-Retail	0.02	0.02	0.04	0.03	0.01	0.00	0.00

Table 15: Distribution of estimated capital elasticities of output by sector in 2014.

Average revenue product of labor	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	4.52	4.44	5.39	4.90	4.08	3.79	0.37
Construction	4.32	4.28	4.90	4.58	4.02	3.79	0.20
Manufacturing	4.33	4.27	4.99	4.62	3.99	3.77	0.25
Non-Financial	4.25	4.24	4.89	4.57	3.89	3.59	0.27
Transportation	4.24	4.20	4.78	4.47	3.97	3.78	0.18
Wholesale-Retail	5.03	5.02	5.79	5.43	4.61	4.26	0.37

Table 16: Distribution of log average revenue product of labor in efficiency units by sector in 2014.

Marginal revenue product of labor	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	2.83	2.82	3.20	2.99	2.64	2.46	0.10
Construction	2.82	2.81	3.11	2.96	2.69	2.56	0.06
Manufacturing	2.86	2.86	3.17	3.00	2.71	2.56	0.09
Non-Financial	2.79	2.79	3.15	2.96	2.61	2.43	0.11
Transportation	2.90	2.89	3.17	2.99	2.80	2.69	0.06
Wholesale-Retail	2.82	2.81	3.33	3.07	2.57	2.33	0.18

Table 17: Distribution of log marginal revenue product of labor in efficiency units by sector in 2014.

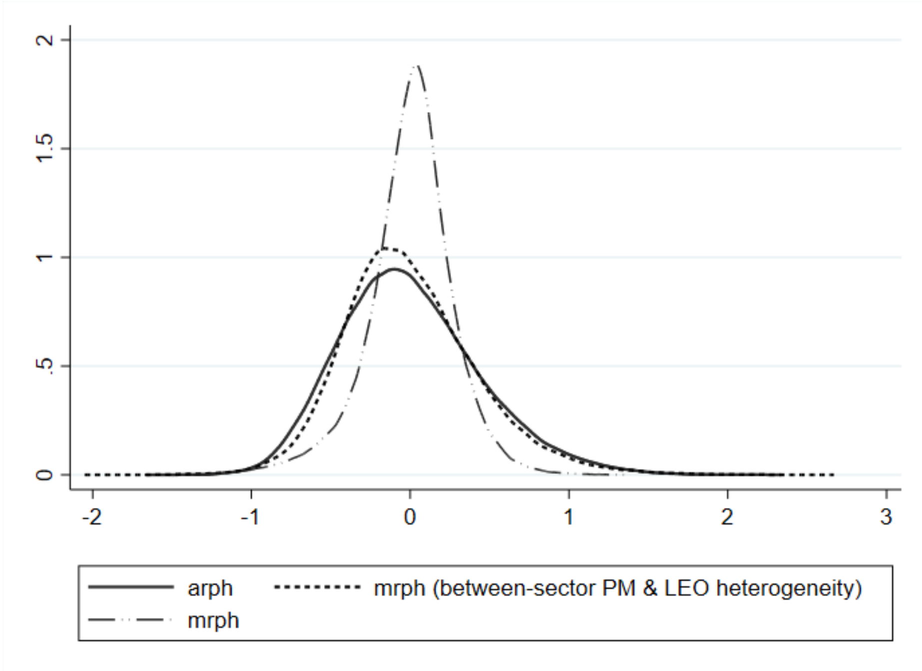

Figure 6: Distribution of the average and marginal revenue product of labor in 2014.
The solid line denotes the average revenue product of labor, the short dashed line denotes the marginal revenue product of labor when price-cost markups and labor elasticities of output are sector-specific but not firm-specific, and the long dashed line denotes the margianl revenue product of labor when all dimensions are firm-specific. Each variable is de-meaned.

Shapley Decomposition Results

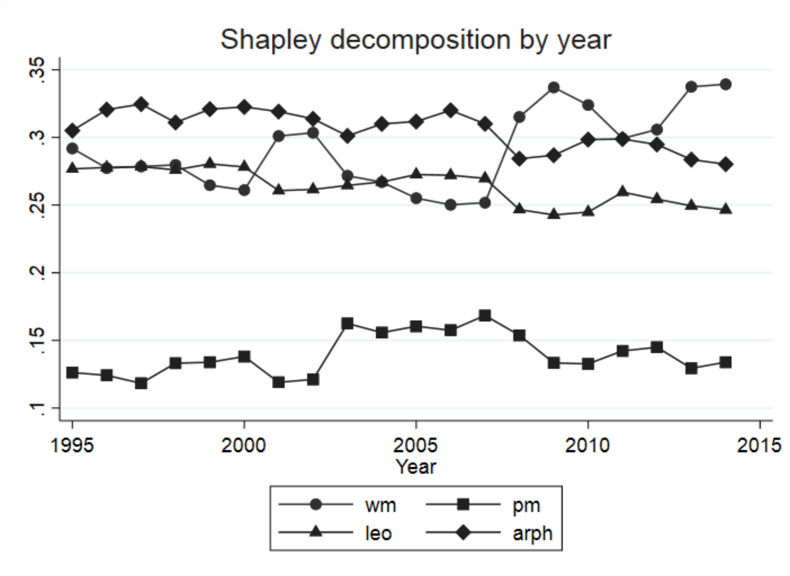

Figure 7: Shapley decomposition of the firm wage premium over time
Included sectors: Construction, Manufacturing, Non-Financial Services, Transportation, and Wholesale and Retail.

	Aggregate	Construction	Manufacturing
wm	0.25	0.26	0.24
arph	0.35	0.47	0.43
pm	0.11	0.09	0.08
leo	0.29	0.17	0.24
R^2	1	1	1
Number of firms	4,907,010	1,005,476	1,126,657

	Non-Financial	Transportation	Wholesale & Retail
wm	0.20	0.28	0.32
arph	0.39	0.41	0.36
pm	0.14	0.09	0.09
leo	0.29	0.22	0.23
R^2	1	1	1
Number of firms	1,126,785	178,487	1,469,287

Table 19: Shapley decomposition of the firm wage premium distribution by sectors, 1995-2014.

Ensemble Decomposition Results

$$V(\phi) = CV(\ln WM, \phi) + CV(-\ln PM, \phi) + CV(\ln LEO, \phi) + CV(\ln ARPH, \phi)$$

	Aggregate	Construction	Manufacturing	Non-Financial	Transportation	Wholesale & Retail
$CV(\ln WM, \phi)$	0.006	0.009	0.003	0.003	0.004	0.007
$CV(-\ln PM, \phi)$	-0.001	-0.002	-0.001	-0.002	-0.001	-0.002
$CV(\ln LEO, \phi)$	-0.002	-0.007	-0.005	-0.003	-0.004	-0.003
$CV(\ln ARPH, \phi)$	0.006	0.011	0.009	0.008	0.009	0.008
Number of firms	243,453	49,175	45,558	52,818	7,039	68,661

Table 20: Ensemble decomposition of the firm wage premium distribution by sectors (2014).

Standard Variance Decomposition Results

$$
\begin{aligned}
V(\phi) = {} & V(\ln WM) + V(\ln PM) + V(\ln LEO) + V(\ln ARPH) \\
& + 2CV(\ln WM, -\ln PM) + 2CV(\ln WM, \ln LEO) + 2CV(\ln WM, \ln ARPH) \\
& + 2CV(-\ln PM, \ln LEO) + 2CV(\ln PM, \ln ARPH) + 2CV(\ln LEO, \ln ARPH)
\end{aligned}
$$

	wm	*arph*	$-pm$	*leo*	*mrph*
wm	0.062				
arph	0.021	0.190			
$-pm$	-0.018	0.003	0.028		
leo	-0.014	-0.160	-0.014	0.181	
mrph	-0.053	0.032	0.016	0.007	0.055

Table 21: Firm heterogeneity variance-covariance matrix in 2014 (construction).

	wm	*arph*	$-pm$	*leo*	*mrph*
wm	0.078				
arph	0.034	0.232			
$-pm$	-0.021	-0.019	0.024		
leo	-0.021	-0.170	0.015	0.171	
mrph	-0.076	0.043	0.020	0.016	0.079

Table 22: Firm heterogeneity variance-covariance matrix in 2014 (manufacturing).

	wm	*arph*	$-pm$	*leo*	*mrph*
wm	0.095				
arph	-0.016	0.249			
$-pm$	-0.025	-0.019	0.033		
leo	0.050	-0.207	-0.010	0.244	
mrph	-0.092	0.024	0.023	0.047	0.095

Table 23: Firm heterogeneity variance-covariance matrix in 2014 (non-financial services).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.050				
$arph$	-0.011	0.156			
$-pm$	-0.009	-0.011	0.023		
leo	-0.048	-0.148	-0.004	0.195	
$mrph$	-0.046	-0.002	0.008	0.043	0.049

Table 24: Firm heterogeneity variance-covariance matrix in 2014 (transportation).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.169				
$arph$	-0.106	0.347			
$-pm$	-0.071	0.007	0.052		
leo	0.015	-0.241	-0.009	0.213	
$mrph$	-0.161	0.114	0.068	-0.019	0.163

Table 25: Firm heterogeneity variance-covariance matrix in 2014 (wholesale and retail).

Cross-sectional Correlations

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	0.194	1			
$-pm$	-0.432	0.034	1		
leo	-0.132	-0.864	0.199	1	
$mrph$	-0.904	-0.313	0.416	0.070	1

Table 26: Firm heterogeneity correlation matrix in 2014 (construction).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	0.251	1			
$-pm$	-0.479	-0.256	1		
leo	-0.182	-0.854	-0.239	1	
$mrph$	-0.960	0.315	0.464	0.141	1

Table 27: Firm heterogeneity correlation matrix in 2014 (manufacturing).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.103	1			
$-pm$	-0.459	-0.209	1		
leo	0.331	-0.837	-0.107	1	
$mrph$	-0.969	0.156	0.420	-0.311	1

Table 28: Firm heterogeneity correlation matrix in 2014 (non-financial services).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.123	1			
$-pm$	-0.260	-0.180	1		
leo	-0.484	-0.845	-0.063	1	
$mrph$	-0.923	-0.024	0.241	0.441	1

Table 29: Firm heterogeneity correlation matrix in 2014 (transportation).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.437	1			
$-pm$	-0.751	0.053	1		
leo	0.080	-0.884	0.084	1	
$mrph$	-0.971	0.477	0.739	-0.099	1

Table 30: Firm heterogeneity correlation matrix in 2014 (wholesale and retail).

Deriving the Effects of Market Power on Sectoral TFP Through Misallocation

Let s be a sector identifier. The sector-specific CES aggregator over firm-level output is $Y_s = \left(Y_{sj}^{\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}}$. To derive closed-form solutions for sectoral TFP, I impose the assumption that firms operate sector-specific Cobb-Douglas constant returns-to-scale production functions $Y_{sj} = X_{sj} K_{sj}^{\alpha_s^K} H_{sj}^{\alpha_s^H} M_{sj}^{\alpha_s^M}$. Firms face firm-specific product demand and labor supply curves, as in Section 2. The firm-specific price is then a markup (PM_{sj}) over marginal costs:

$$P_{sj} = PM_{sj} \frac{1}{X_{sj}} \left(\frac{R^K}{\alpha_s^K}\right)^{\alpha_s^K} \left(\frac{WM_{sj}^{-1}\Phi_{sj}}{\alpha_s^H}\right)^{\alpha_s^H} \left(\frac{P^M}{\alpha_s^M}\right)^{\alpha_s^M}$$

where WM_{sj}^{-1} denotes the inverted wage markdowns. The firm-specific revenue TFP can then be written as:

$$TFPR_{sj} = P_{sj} X_{sj} \propto PM_{sj} \cdot (WM_{sj}^{-1}\Phi_{sj})^{\alpha_s^H}$$

Following Hsieh and Klenow (2009), the expression sectoral TFP can be derived as:

$$TFP_s = \left[\sum_{j \in s} \left(X_{sj} \frac{\overline{TFPR_s}}{TFPR_s}\right)^{\rho-1}\right]^{\frac{1}{\rho-1}}$$

where $\overline{TFPR}_s$ denotes the mean revenue TFP within sector s. Finally, as shown in Hsieh and Klenow (2009), under the assumption that quantity TFP (X_{sj}) and revenue TFP ($TFPR_{sj}$) are jointly log-normally distributed, I obtain an analytical expression for sector-specific TFP:

$$\ln TFP_s = \frac{1}{\rho-1} \log\left(\sum_{j \in s} X_{sj}^{\rho-1}\right) - \frac{\rho}{2} V_s \left(\ln(PM_{sj} \cdot (WM_{sj}^{-1}\Phi_{sj})^{\alpha_s^H})\right)$$

As section 5.3 shows, the variance of firm wage premia is modest. I therefore assume that $\Phi_j \approx \Phi \ \forall j$. Therefore, approximately,

$$\ln TFP_s \approx \frac{1}{\rho-1} \log\left(\sum_{j \in s} X_{sj}^{\rho-1}\right) - \frac{\rho}{2} V_s \left(\ln(PM_{sj} \cdot WM_{sj}^{-\alpha_s^H})\right)$$

Denote TFP_s^* as aggregate sectoral TFP when there are no labor market frictions. Then, the potential gains to aggregate sectoral productivity from removing labor market frictions is:

$$\ln TFP_s^* - \ln TFP_s \approx \frac{\rho}{2} V_s \left(\ln(PM_{sj} \cdot WM_{sj}^{-\alpha_s^H})\right) - \frac{\rho}{2} V_s \left(\ln(PM_{sj})\right)$$

Why do labor shares of revenue vary across firms?

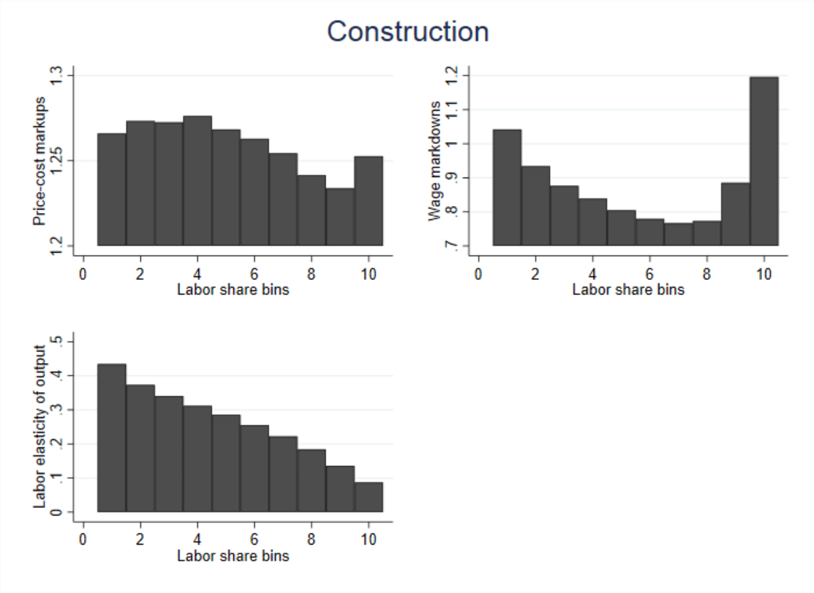

Figure 8: Labor shares and firm characteristics among construction firms in 2014.

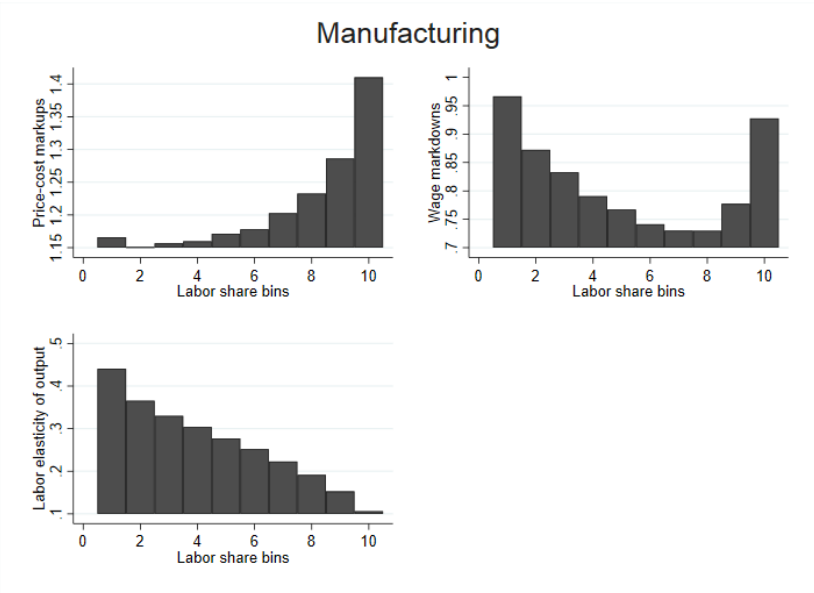

Figure 9: Labor shares and firm characteristics among manufacturing firms in 2014.

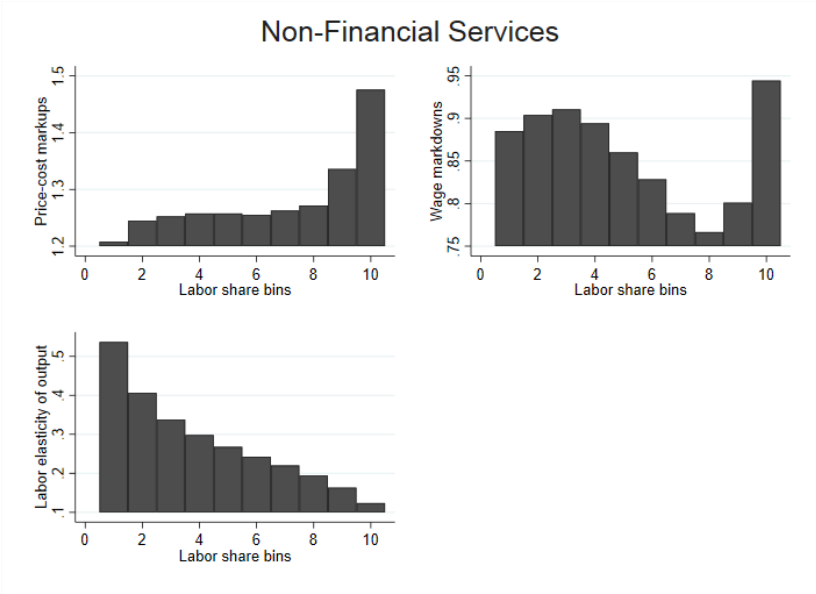

Figure 10: Labor shares and firm characteristics among non-financial services firms in 2014.

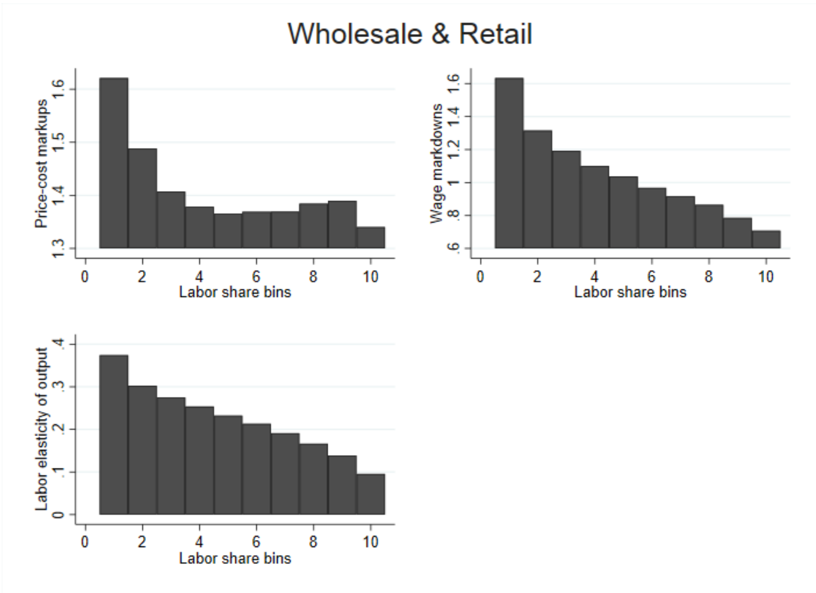

Figure 11: Labor shares and firm characteristics among wholesale and retail firms in 2014.

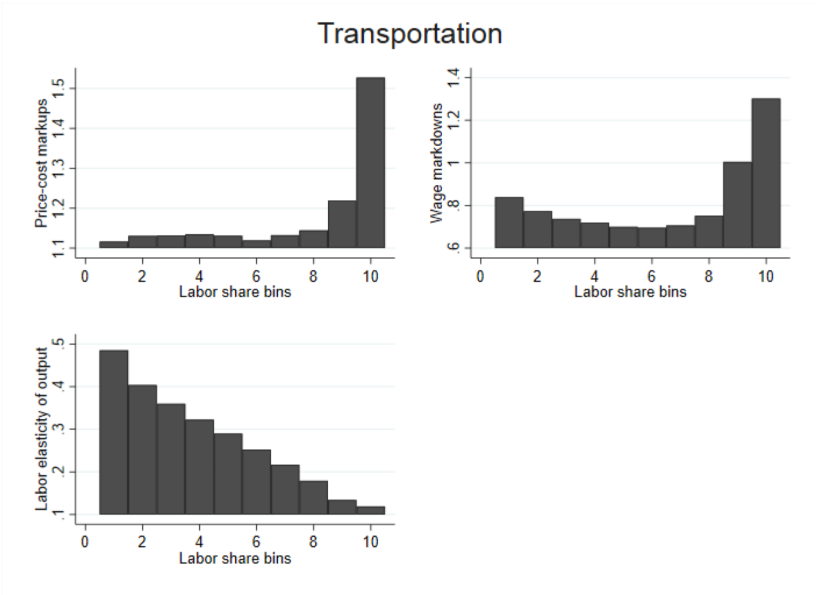

Figure 12: Labor shares and firm characteristics among transportation firms in 2014.

Appendix E: Extension with Differentially Skilled Occupations

As discussed in Section 2.3, one limitation of the current analysis is that workers enter the production function in efficiency units. To relax this assumption, I extend the analysis in sections 5 and 6 to a setting with a high-skill and a low-skill occupation. The main findings of this extension can be summarized as follows: (i) Workers in high-skill occupations are paid a greater share of their marginal revenue product compared to workers in low-skill occupations; (ii) price-cost markups and labor elasticities of output matter for firm wage premia in both high-skill and low-skill occupations; (iii) the negative correlation between labor productivity and labor elasticities of output is large for both skill groups, but larger for low-skill occupations; (iv) the variance of the marginal revenue product of labor is larger among high-skill occupations than low-skill occupations, using the average revenue product of labor as a proxy would have indicated the opposite; (v) firms with lower labor shares of revenue mark down wages of low-skill occupations more, while the opposite is true for high-skill occupations.

Defining high-skill and low-skill occupations. I use the one-digit occupation classifications in the DADS matched employer-employee data set to define low-skill occupations as blue-collar occupations (e.g. maintenance workers and welders) and administrative support occupations (e.g. clerical workers and secretaries); I define high-skill occupations as senior staff in top management positions (e.g. head of logistics or human resources), employees in supervisory roles (e.g. accounting and sales managers), and technical workers (e.g. IT and quality control technicians).

Structural firm wage premium equation by skill-group. Let the subcript $s = \{h, l\}$ denote high and low-skill labor. The firm wage premium equation is:

$$\phi_{j,s} = wm_{j,s} - pm_j + leo_{j,s} + arph_{j,s}$$

where the wage markdown, labor elasticity of output, and average revenue product of labor are now firm-skill-group-specific. The average revenue product of labor for a given skill group in this case is (log) total revenue divided by the total efficiency units of that skill group.

Estimating firm wage premia. The estimation procedure is as described in Section 3.2. However, firm-group effects are now occupation-specific:

$$\ln W_{it} = a_i + \phi_{g(j(i,t))} + \text{Occ}_{o(i,t)} \times \phi_{g(j(i,t))} + X'_{it}\beta + \nu_{it}$$

where i denotes the individual, j denotes the firm, $g(j)$ denotes the group of firm j at time t, $o(i,t)$ denotes worker i's occupational group at time t, a_i are worker fixed effects, $\phi_{g(j(i,t))}$ are firm-group fixed effects, and X_{it} is a vector of time-varying worker characteristics.

Estimating firm characteristics. The production function now looks as follows:

$$y = f(h, l, k, m, o) + x$$

I approxmimate $f(.)$ with a translog functional form, where k denotes capital, m denotes materials, and o denotes services. With a slight abuse of notation, $h = \bar{e}_h + n_h$ and $l = \bar{e}_l + n_l$ now denote high-skill and low-skill labor in efficiency units, where $\bar{e}_h$ and $\bar{e}_l$ denote the average ability of each skill group, and n_h and n_l denote total hours in each skill group. Skill-group-specific average ability can be measured as the difference between the skill-group-specific average wage at a firm j and the corresponding firm wage premium, $\bar{w}_{j,s} = \bar{e}_{j,s} + \phi_{g(j),s}$ where $s = \{h, l\}$. All lowercase letters represent variables in logs.

The procedure to back out price-cost markups and skill-group-specific wage markdowns are identical to the one described in Section 3.3. Table 31 reports the summary statistics for wage markdowns among high-skill and low-skill workers. High-skill workers typically incur a smaller markdown of wages below marginal revenue products compared to low-skill workers.

	Mean	Median	90th	75th	25th	10th	Variance
High-skill wage markdown	0.99	0.81	1.66	1.11	0.55	0.36	0.76
Low-skill wage markdown	0.68	0.65	1.11	0.90	0.43	0.22	0.14
Number of firms				147,347			

Table 31: Employment-weighted distribution of wage markdowns by skill group in 2014.

Shapley decomposition of firm wage premia. The overarching message of the Shapley decomposition by skill groups remains the same – price-cost markups and labor elasticities of output matter for firm wage premia. Table 32 presents the decomposition by skill groups.

	Marginal contribution to the R^2	
	Low-skill	High-skill
Wage markdown	0.51	0.38
Average revenue product of labor	0.19	0.24
Price-cost markup	0.05	0.09
Labor elasticity of output	0.25	0.29
Number of firms	147,347	

Table 32: Shapley decomposition of firm wage premia by skill groups in 2014.

To compare this decomposition with the one in Table 1, I average the skill-group-specific firm wage premium components at the firm level, weighted by their total shares of hours within the firm, and implement the decomposition. Table 33 shows that the decomposition results are similar to those in Table 1.

	Marginal contribution to the R^2
Wage markdown	0.33
Average revenue product of labor	0.32
Price-cost markup	0.12
Labor elasticity of output	0.23
Number of firms	147,347

Table 33: Shapley decomposition of firm wage premia, averaged over skill groups, in 2014.

Cross-sectional correlations of firm characteristics. Table 34 shows that the negative correlation between labor productivity and the labor elasticity of output documented in the main body of the paper also holds when disaggregating by skill group, with a somewhat larger coefficient among low-skill occupations than high-skill ones. One difference between high-skill and low-skill occupations is that wage markdowns increase more strongly with labor productivity among low-skill occupations than among high-skill ones.

	Low-skill				High-skill			
	wm_s	$arph$	$-pm$	leo_s	wm_s	$arph_s$	$-pm$	leo_s
wm_s	1				1			
$arph_s$	-0.68	1			-0.39	1		
$-pm$	-0.28	-0.08	1		-0.39	-0.03	1	
leo_s	0.51	-0.92	0.10	1	0.00	-0.87	0.03	1

Table 34: Firm heterogeneity correlation matrix by skill group in 2014.

Labor misallocation by skill groups. One key implication of the negative correlation between labor productivity and labor elasticities of output explored in the main body of the paper is that conventional approaches to measure labor misallocation – using average, rather than marginal revenue products of labor – overstate the degree of misallocation. The intuition is that firms can sidestep labor market frictions that force firms to pay higher wages to attract workers by substituting workers with other inputs. Disaggregating this result by skill group, Table 35 shows a similar finding. Further, a skill-group-specific measure of labor misallocation using average, instead of marginal, revenue products of labor would lead to the conclusion that labor misallocation is worse among low-skill occupations than among high-skill ones ($var(arph_l) = 0.78 > 0.65 = var(arph_h)$), when estimated marginal revenue products of labor would suggest otherwise ($var(mrph_l) = 0.16 < 0.21 = var(mrph_h)$).

	Mean	Median	90th	75th	25th	10th	Variance
Low-skill *arph*	4.70	4.55	5.94	5.23	4.04	3.71	0.78
Low-skill *mrph*	2.55	2.54	3.00	2.76	2.33	2.10	0.16
High-skill *arph*	5.75	5.73	6.80	6.29	5.22	4.74	0.65
High-skill *mrph*	2.97	2.99	3.49	3.22	2.72	2.19	0.21
Number of firms				147,347			

Table 35: Distribution of estimated average and marginal revenue products of labor by skill group in 2014.

Labor revenue share. Section 5 shows that wage markdowns and labor elasticities of output account for considerable shares of the low labor shares of revenue among high productivity firms. Comparing high-skill and low-skill labor, Figure 13 shows that the wage markdowns and labor elasticities of output among low-skill, rather than high-skill, labor are the main contributors to the lower labor shares of highly productive firms. Firms with lower labor shares have higher high-skill labor elasticities of output and markdown high-skill wages by less.

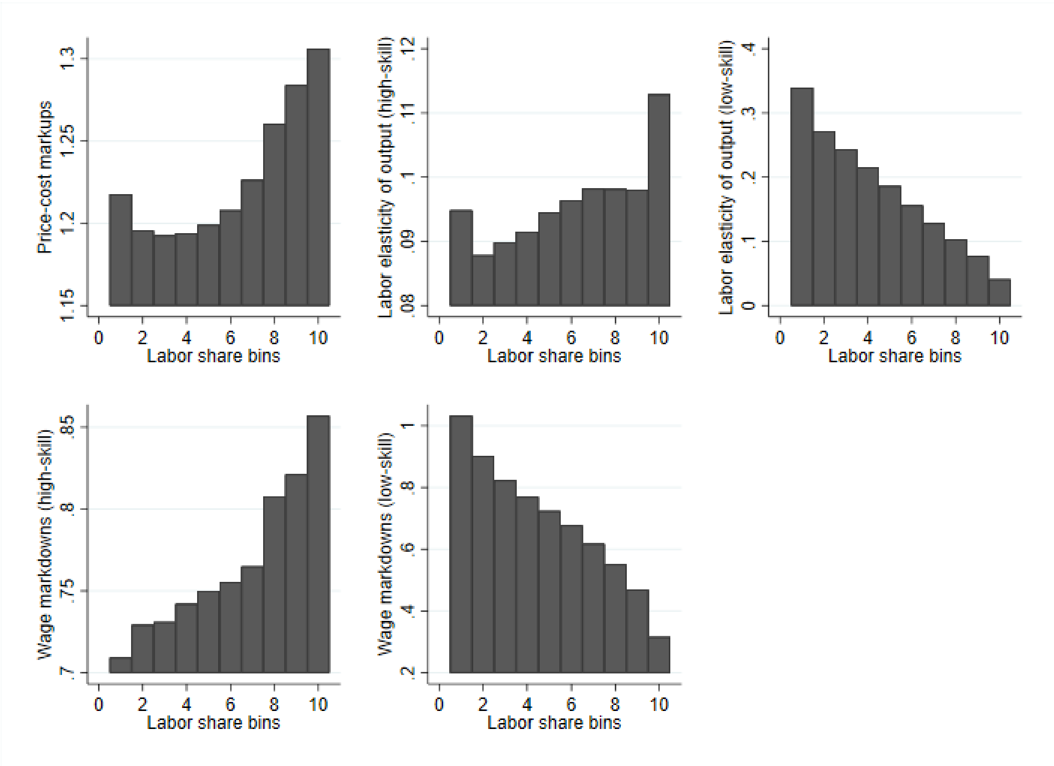

Figure 13: Firm characteristics across the labor share distribution by skill group in 2014.

Appendix F: Wage-Posting and Wage-Bargaining Frameworks

Random Search Wage-Bargaining Framework

The structural framework presented in section 2 does not take a stance on the specific frictions generating upward-sloping labor supply curves. I present here a model in which labor markets are characterized by search frictions and wages are set via bargaining over the match surplus. I derive the firm wage premium equation (1) from this model and discuss the interpretation of the wage markdown in this model. I draw from the multiworker-firm random search models of Mortensen (2010) and Elsby, Michaels, and Ratner (2018), in which workers are allowed to search on-the-job. I assume that there are no aggregate shocks.

Matching in the labor market is governed by a matching function $\Lambda_t = \Lambda(\bar{U}_t + \xi(\bar{H}_t - \bar{U}_t), \bar{V}_t)$, where $\bar{H}$ and $\bar{U}$ denote total skill-adjusted population of workers and unemployed workers, and $\bar{V}$ denotes aggregate vacancies. The search intensity of employed workers is ξ. Labor market tightness is the ratio of vacancies to jobseekers $\theta_t \equiv \frac{\bar{V}_t}{\bar{U}_t + \xi(\bar{H}_t - \bar{U}_t)}$. The vacancy contact rate is then $q(\theta_t) = \Lambda(\theta_t^{-1})$, and the unemployed and employed worker job finding rates are $f(\theta_t)$ and $\xi f(\theta_t)$.

On the firm side, the hiring rate for a firm providing a value V_{jt}^e to its workers is:

$$\lambda(V_{jt}^e) = q(\theta_t)\left[\frac{\bar{U}_t}{\bar{U}_t + \xi(\bar{H}_t - \bar{U}_t)} + \frac{\xi(\bar{H}_t - \bar{U}_t)}{\bar{U}_t + \xi(\bar{H}_t - \bar{U}_t)} G_E(V_{jt}^e)\right]$$

where $G_E(.)$ denotes the cumulative distribution function of the realized value of employment to workers across employed workers. Similarly, the separation rate of this firm is:

$$s(V_{jt}^e) = \delta_s + (1 - \delta_s)\xi f(\theta_t)\left(1 - F_V(V_{jt}^e)\right)$$

where δ_s is an exogenous separation rate, and $F_V(.)$ is the cumulative distribution function of the offered value of employment to workers among vacancies.

The unemployed worker's value function is:

$$U_t = b + \beta[(1 - f(\theta_{t+1}))U_{t+1} + f(\theta_{t+1})E_t(V_{t+1}^e)]$$

which is a function of the flow value of unemployment b and the expected utility next period.

Since there are no aggregate shocks, $U_t = U_{t+1}$. The employed worker's value function is:

$$\begin{aligned} V_{jt}^e = & \, u(\Phi_{jt}, A_{jt}) \\ & + \beta\{\delta_s U_{t+1} + (1-\delta_s)E_t[(1-\xi f(\theta_{t+1}))V_{jt+1}^e \\ & + \xi f(\theta_{t+1})F(V_{jt+1}^e)V_{jt+1}^e \\ & + \xi f(\theta_{t+1})(1-F(V_{jt+1}^e))E_t(V_{t+1}^e | V_{t+1}^e \geq V_{jt+1}^e)]\} \end{aligned}$$

which depends on the wage Φ_{jt} and non-wage amenities A_{jt} this period through a constant returns to scale utility function $u(.)$, the expected utility next period if the worker is exogenously separated from the firm, and the expected utility if the worker is not exogenously separated. The last component depends on the expected utility of being employed at the same firm, and the expected utility of moving to a new employer conditional on the new employer offering a higher utility. I assume that: (i) the flow utility function $u(.,.)$ is homogenous of degree one in its inputs, (ii) there is no savings mechanism, (iii) the value of non-wage amenities is proportional to worker efficiency, $A_{ijt} = E_{it}A_{jt}$, and (iv) worker efficiency is allowed vary over time due to random shocks: $E_{it+1} = E_{it} + \zeta_{it+1}$, where ζ_{it+1} is a mean-zero random shock.[46] Therefore, the value of unemployment and employment is proportional to worker efficiency. A worker with efficiency E_{it} obtains a value of $E_{it}U_t$ while unemployed and $E_{it}V_{jt}^e$ while employed.

The firm's profit maximization problem can be written as:

$$\Pi_{jt} = \max_{K_{jt}, M_{jt}, V_{jt}} P_{jt}Y_{jt} - R_t^K K_{jt} - P_t^m M_{jt} - \Phi(H_{jt})H_{jt} - c_t(V_{jt})V_{jt} + \beta E_t[\Pi_{jt+1}]$$

subject to the law of motion for employment:

$$H_{jt} = (1 - s(V_{jt}^e))H_{jt-1} + \lambda(V_{jt}^e)V_{jt} \tag{13}$$

and (2) and (3). The average skill of workers at firm j is denoted as $\bar{E}_j$. The vacancy posting cost function $c_t(V_{jt})$ is assumed to be twice differentiable, monotonically increasing in vacancies $c_t'(V_{jt}) > 0$, and the marginal cost of vacancies is increasing $c_t''(V_{jt}) > 0$.

Wages are determined via Stole and Zwiebel (1996) bargaining between the firm and the marginal worker over the marginal match surplus. This generalizes the Nash bargaining protocol in models with constant marginal returns to labor to the case of diminishing marginal returns to labor. Employers do not make counteroffers. The bargained wage $\Phi(H_{jt})$ is a function of the firm's size, since diminishing marginal returns to labor implies that, all else equal, the marginal revenue product of labor, and hence total match surplus, is decreasing in firm size. The maginal

[46] Alternatively, for a slightly more realistic human capital accumulation process, one can also envision a model in which each worker i's efficiency grows at a deterministic rate, and the worker may receive an exogenous death shock, in which case the worker is replaced by a newly-born worker in the model. See, for example, Bagger, Fontaine, Postel-Vinay, and Robin (2014).

surplus to be bargained over is:

$$\kappa_{jt}J_{jt} = (1-\kappa_{jt})(V_{jt}^e - U_t)$$

where κ_{jt} is the worker's relative bargaining weight, which is allowed to differ across firms, and $J_{jt} \equiv \frac{\partial \Pi_{jt}}{\partial H_{jt}}$ is the firm's marginal surplus from an additional skill-adjusted worker. I obtain the following familiar equation for the firm's wage (premium):

$$\Phi_{jt} = \kappa_{jt}(MRPH_{jt} - \frac{\partial \Phi_{jt}}{\partial H_{jt}}H_{jt} + \beta E_t[(1-s(\Phi_{jt+1}, A_{jt+1}))J_{jt+1}]) + (1-\kappa_{jt})W_{jt}^r$$

This equation shows that the firm's wage is a weighted average of the value of the worker to the firm and the worker's reservation wage.

Combining the wage bargaining protocol with the first-order condition with respect to vacancies, I rearrange the above firm wage equation to obtain the firm wage premium equation (1), in which the firm's wage markdown component can be written as:

$$WM_{jt} = \frac{\left(\frac{\kappa_{jt}}{1-\kappa_{jt}}\right)\left(1+\frac{W_{jt}^r}{\Phi_{jt}-W_{jt}^r}\right)}{1+(1-|\frac{\partial \Phi_{jt}}{\partial H_{jt}}\frac{H_{jt}}{\Phi_{jt}}|)\left(\frac{\kappa_{jt}}{1-\kappa_{jt}}\right)\left(1+\frac{W_{jt}^r}{\Phi_{jt}-W_{jt}^r}\right) - \beta E_t\left(\frac{(1-s(\Phi_{jt+1},A_{jt+1}))J_{jt+1}}{c_{V,jt}V_{jt}+c(V_{jt})}\right)\lambda(\Phi_{jt}, A_{jt})} \tag{14}$$

Note that $\frac{\partial \Phi_{jt}}{\partial H_{jt}}\frac{H_{jt}}{W_{jt}} < 0$ is no longer the inverse labor supply elasticity. It takes a negative value. This is because, with multilateral bargaining, the firm bargains with all of its worker over the marginal surplus of a match. With diminishing marginal returns to labor, if the firm and worker do not agree on a wage, the match is not formed, and the marginal revenue product of labor is higher for the remaining workers. This is an additional channel on top of workers' bargaining weight from which workers extract rents from the match.

The numerator of the wage markdown shows that the firm's wage markdown depends on its workers' relative bargaining power (κ_{jt}) and the reservation wage (W_{jt}^r). The higher the workers' bargaining power or reservation wage, the higher the fraction of marginal revenue product of labor workers obtain (higher wage markdown). The denominator shows that the wage markdown is also increasing in the expected future value of the worker to the firm.

Random Search Wage-Posting Framework

I now replace the wage-setting protocol of the random search framework above with wage-posting and discuss the determinants of the wage markdown. This model generates equation (1) and provides one microfoundation for the wage markdown derived from the structural framework presented in section 2.

The firm's profit maximization problem is:

$$\Pi_{jt} = \max_{K_{jt}, M_{jt}, V_{jt}, \Phi_{jt}} P_{jt}Y_{jt} - R_t^K K_{jt} - P_t^m M_{jt} - \Phi_{jt}H_{jt} - c_t(V_{jt})V_{jt} + \beta E_t[\Pi_{jt+1}]$$

subject to the law of motion for employment:

$$H_{jt} = (1 - s(\Phi_{jt}, A_{jt}))H_{jt-1} + \lambda(\Phi_{jt}, a_{jt})V_{jt} \tag{15}$$

and (2) and (3). The wage markdown in this model is as follows:

$$WM_{jt} = \frac{\epsilon_{jt}^H}{1 + \epsilon_{jt}^H - \beta E_t\left(\frac{(1-s(\Phi_{jt+1}, A_{jt+1}))J_{jt+1}}{c_{V,jt}V_{jt} + c(V_{jt})}\right)\lambda(\Phi_{jt}, A_{jt})}$$

where the firm-specific labor supply elasticity (ϵ_{jt}^H) can be written as:

$$\epsilon_{jt}^H = \frac{\lambda(\Phi_{jt}, A_{jt})V_{jt}}{H_{jt}}\epsilon_{\Phi,jt}^{\lambda} - \frac{s(\Phi_{jt}, A_{jt})H_{jt-1}}{H_{jt}}\epsilon_{\Phi,jt}^{s} > 0$$

which depends on the elasticity of the firm's hiring rate with respect to the firm's wage ($\epsilon_{\Phi,jt}^{\lambda} > 0$) weighted by the share of new hires among its workforce, minus the elasticity of the firm's separation rate with respect to the firm's wage ($\epsilon_{\Phi,jt}^{s} < 0$) weighted by the share of workers who separate from the firm among its workforce.

Directed Search Wage-Posting Framework

The random search model assumes that workers have no information about wages when they search for a job. An alternative assumption is that workers observe the full menu of wages in the economy when searching for jobs – directed or competitive search (Moen, 1997). I now replace random search with directed search in the otherwise identical wage-posting model. I show in this environment that the firm wage premium equation (1) can be obtained and the wage markdown is identical as the model with random search.[47] The following timing assumption applies. First, idiosyncratic firm productivity and worker efficiency shocks are realized. Next, firms post wages and workers decide on where to search. Then, matching and separations take place. Finally, production begins.

In this model, workers can choose the firm or market at which she searches for a job by trading off offered utility and job-finding probability. The worker observes the offered utility V_{jt}^e at each firm j, before matching takes place. The pre-matching value to the unemployed

[47] For a comprehensive discussion of the theory and applications of directed search, see Wright, Kircher, Julien, and Guerrieri (2018).

worker is:

$$U_t^{bm} = \max_{V_{jt}^e} (1 - f(\theta(V_{jt}^e)))U_t + f(\theta(V_{jt}^e))V_{jt}^e$$

and the pre-matching value to a worker employed at firm j is:

$$V_{jt}^{e,bm} = \max_{V_{kt}^e} \delta_s U_t + (1-\delta_s)\left[(1 - sf(\theta(V_{kt}^e)))\, V_{jt}^e + sf(\theta(V_{kt}^e))V_{kt}^e\right]$$

where the offered utility $V(W_{jt}, A_{jt})$ at any firm j depends on both the offered wages and non-wage amenities. As I show in the next subsection, no two workers with different utility V^e will search for employment at the same firm. Relative to a worker with lower utility, the worker with a higher utility will search for employment at a firm that offers an even higher utility, at the cost of a lower probability of this employment relationship materializing.

Firms post wages taking into account its effect on both recruitment and retention. Each firm recruits from other firms who offer a lower utility to their employees. From the employed worker's value function above, given the value of employment at a firm that offers $\underline{V}_t^e$, this worker optimally searches for employment at firm j, where $V_{jt}^e > \underline{V}_t^e$. Denote this unique solution as $V_{jt}^e = v(\underline{V}_t^e)$. Therefore, firm j recruits workers from this market. Similarly, firm j loses workers due to quits to a higher utility firm who pays $\overline{V}_t^e$. The optimal search strategy of a worker employed at firm j is then $\overline{V}_t^e = v(V_{jt}^e)$. Next, note that the firm-specific separation rate is now $s_{jt} = \delta_s + (1-\delta_s)sf(\theta(\overline{V}_t^e))$. Using the law of motion for employment, the firm-specific "labor supply" curve is then:

$$\begin{aligned} H_{jt} &= (1 - s(\overline{V}_t^e))H_{jt-1} + q(\underline{V}_t^e)V_{jt} \\ &= (1 - s(V_{jt}^e))H_{jt-1} + q(V_{jt}^e)V_{jt} \end{aligned}$$

The second line obtains by inverting the employed worker's optimal search function $v(.)$, which is monotonically increasing in its argument. Solving for the firm wage premium equation (1) gives the same wage markdown expression as the random search wage-posting model above.

Workers' search behavior in a Directed Search Model

I now show that in the directed search model above, relative to workers employed at lower offered utility firms, workers employed at a higher offered utility firm will choose to search for employment at a firm that offers even higher utility, at the cost of a lower probability of finding employment there (see Wright et al. (2018)). Consider worker 1 employed at firm 1, searching optimally for employment at firm j; and worker 2 employed at firm 2, searching optimally for employment at firm k. Suppose that firm 2 offers a strictly higher utility than firm 1, $V_{2t}^e > V_{1t}^e$.

The utility of either workers can be written as:

$$V_{2t}^e = U_t + \left(\frac{1-\delta_s}{\delta_s}\right) sf(V_{kt}^e)[V_{kt}^e - V_{2t}^e]$$

$$V_{1t}^e = U_t + \left(\frac{1-\delta_s}{\delta_s}\right) sf(V_{jt}^e)[V_{jt}^e - V_{1t}^e]$$

Under utility maximization:

$$f(V_{kt}^e)[V_{kt}^e - V_{2t}^e] \geq f(V_{jt}^e)[V_{jt}^e - V_{2t}^e]$$

$$f(V_{kt}^e)[V_{kt}^e - V_{1t}^e] \leq f(V_{jt}^e)[V_{jt}^e - V_{1t}^e]$$

which implies that:

$$f(V_{kt}^e)[V_{1t}^e - V_{2t}^e] > f(V_{jt}^e)[V_{1t}^e - V_{2t}^e]$$

Since the utility of worker 2 is strictly larger than that of worker 1

$$f(V_{kt}^e) < f(V_{jt}^e)$$

Therefore, relative to worker 1, worker 2 who is employed at a higher utility firm searches for a job at a firm which has an even higher offered utility, at the cost of a lower probability of matching.

Workplace Differentiation Monopsonistic Wage-Posting Framework

This section presents a static monopsonistic model based on the imperfect substitutability of firm-specific non-wage amenities (Card et al., 2018). Worker i's indirect utility when employed at firm j is:

$$u_{ijt} = \gamma \ln(W_{ijt}) + a_{jt} + \eta_{ijt}$$

where $W_{ijt} = E_{it}\Phi_{jt}$ is the wage obtained by worker i with efficiency E_{it} earning a wage premium Φ_{jt}. The common value of the firm-specific non-wage amenity a_{jt}. Worker's preferences over non-wage amenities are subject to idiosyncratic shocks η_{ijt}, which is identically and independently drawn from a type I extreme value distribution.

Each worker i maximizes utility by choosing where to work:

$$j = \arg\max_j u_{ijt}$$

The firm-specific labor supply curve is then:

$$\frac{H_{jt}}{\bar{H}_t} = \frac{exp(\gamma \ln(\Phi_{jt}) + a_{jt})}{\sum_{k=1}^{J} exp(\gamma \ln(\Phi_{kt}) + a_{kt})}$$

Firm j's profit-maximization problem is:

$$\Pi_{jt} = \max_{K_{jt}, M_{jt}, \Phi_{jt}} P_{jt} Y_{jt} - R_t^K K_{jt} - P_t^m M_{jt} - \Phi(H_{jt}) H_{jt} + \beta E_t[\Pi_{jt+1}]$$

subject to the firm-specific labor supply curve and equations (2) and (3). This model gives the firm wage premium equation (1) and the following expression for the wage markdown:

$$WM_{jt} = \frac{\epsilon_{jt}^h}{1 + \epsilon_{jt}^h}$$

where the labor supply elasticity $\epsilon_{jt}^h = \gamma(1 - \frac{H_{jt}}{\bar{H}_t})$ depends on the labor market share of firm j. This equation shows that the wage markdown is decreasing in the firm's labor market share, as firm's with a high market share face a low labor supply elasticity. This expression provides a mapping between labor market shares, labor market concentration, and wages (Azar, Marinescu, and Steinbaum, 2017; Benmelech et al., 2018).