Understanding High-Wage and Low-Wage Firms*

Horng Chern Wong†

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 covering 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: wage inequality, firm heterogeneity, market power, production technology

JEL codes: D24, D33, E2, J3, J42

*First version: August 2019. Special thanks to Thijs van Rens, Roland Rathelot, and Dennis Novy for their invaluable guidance, support, and patience. I am very grateful to Jan Eeckhout, Francis Kramarz, Tim Sang Yoon Lee, Jay Euijung Lee, Isabelle Mejean, Andrew Oswald, and Vincent Sterk for exceptionally helpful comments. For insightful discussions, I thank Treb Allen, Jesper Bagger, Pablo Beker, Dan Bernhardt, Paula Bustos, Christine Braun, Luis Candelaria, Danilo Cascaldi-Garcia, Emma Duchini, Dita Eckardt, James Fenske, Joaquin García-Cabo, Monika Gehrig-Merz, Basile Grassi, Daniel Habermacher, Simon Hong, Omiros Kouvavas, Alice Kuegler, Thomas Le Barbanchon, Timo Leidecker, Norman Loayza, Bogdan Marcu, 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. 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. 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, 2021) and the rise of "superstar" firms (Song, Price, Guvenen, Bloom, and Von Wachter, 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 firm wage premia? To make sense of them, I build a structural framework to interpret standard regression-based estimates of firm wage premia. In my framework, labor market frictions prevent firm wage premia from being competed away. Firm heterogeneity then determines how much a firm is willing to pay to hire a given worker, compared to other firms.

My main contribution is to show that firm wage premia can be structurally decomposed into two sets of firm characteristics. The role of the first set – *labor productivity* and *wage-setting power* – is relatively well-understood in research on firm wage premia.[1] On the other hand, the role of the second set – *product market power* and *labor share of production* – has received little attention so far.[2] Yet, recent research on the macroeconomics of market power and labor's share of income show that firms' *product market power* and *labor share of production* are key determinants of their labor demand.[3] Using rich administrative datasets on the universe of employers and employees in France, I estimate these firm characteristics and combine them with the model 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

[1]See Card, Cardoso, Heining, and Kline (2018) for a survey of the literature.

[2]"Labor share of production" refers to the elasticity of a firm's output with respect to labor inputs.

[3]Since research on misallocation and the labor share often work with perfectly competitive labor markets, firm wage premia cannot survive in those settings.

wage premium distribution. 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. 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.

Among these firm characteristics, my estimates uncover a negative correlation between labor productivity and the labor share of production. This finding: (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 dispersion in wage-setting power; (ii) provides a new explanation for why more productive firms have lower labor shares of revenue besides market power (De Loecker, Eeckhout, and Unger, 2020).

I begin by building a structural model to interpret statistical estimates of firm wage premia. In the model, labor market frictions sustain firm wage premia and firms endogenously differ from each other along multiple characteristics.[4] As in workhorse frictional labor market models in the Burdett and Mortensen (1998) tradition, firms differ in labor productivity and wage-setting power. 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*. 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 elasticity with respect to labor inputs*. 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.

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

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. 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 widely used in labor economics (Abowd et al., 1999; Bonhomme, Lamadon, and Manresa, 2019). 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.

Estimating firm wage premia and the four firm characteristics requires detailed information about workers and firms. I use large administrative datasets from France that covers the population of employers and employees between 1995 and 2014. I estimate the firm wage premia 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 relevant firm characteristics using firm balance sheet panel data, which contains information such as gross production, employment, and capital for over 2 million firms per year. The main advantages of these distinct datasets are that they are jointly available, can be linked, and are not limited to manufacturing or large firms.[5]

Understanding the extent of firms' wage-setting power is first-order for assessing its implications. Even with a high 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, price-cost markup,

[5]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.

and labor productivity, with most of these differences occuring within narrowly defined sectors (Syverson, 2011; De Loecker et al., 2020).

To gauge the importance of these firm characteristics for firm wage premia, 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. 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. The reason is that these firm characteristics are correlated with each other.

Among these correlations, my estimates reveal a strong 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. 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. It is an accurate proxy 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 more productive firms that are more constrained by labor market frictions can circumvent these frictions by substituting labor with other inputs. This mismeasurement overstates the aggregate productivity and output gains of removing labor market power distortions.

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 elasticities of output 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 extension with high and low-skilled labor, I show that the wage markdowns and labor elasticities of output among low-skilled, rather than high-skilled labor explain firm level labor shares.

Contributions to related literature. A large literature in labor economics estimates the 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 across countries, such as Brazil (Alvarez, Benguria, Engbom, and Moser, 2018), Denmark (Bagger, Christensen, and Mortensen, 2014), Portugal (Card et al., 2018), USA (Song et al., 2019), and Sweden (Bonhomme et al., 2019). A few recent papers provide structural interpretations of firm wage premia (Bagger et al., 2014; Lamadon, Mogstad, and Setzler, 2019). These studies provide fully microfounded models to study counterfactual scenarios. My paper differs by imposing just enough structure on the data to unpack firm wage premia, allowing the inclusion of a richer variety of firm characteristics.

By embedding a richer notion of a firm, the structural firm wage premium decomposition in my paper also speaks to broader recent work on the impact of productivity dispersion (Berlingieri, Blanchenay, and Criscuolo, 2017), labor market power (Azar, Marinescu, Steinbaum, and Taska, 2020), product market power (De Loecker et al., 2020), and the aggregate production technology (Karabarbounis and Neiman, 2014) on wages and the 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 different channels of firm heterogeneity in determining firm wage premia and highlights the importance of the cross-sectional relationships between each channel.

If firms share rents with their employees, then this opens the door for the productivity of individual firms to influence the wages they pay. An exciting body of work provides evidence for rent-sharing and explores their implications for wage determination (Barth, Bryson, Davis, and Freeman, 2016; Kline, Petkova, Williams, and Zidar, 2019; Garin and Silverio, 2019). My paper contributes by quantifying the extent to which productivity differences explain 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) provide evidence for monopsonistic labor markets. Recent papers by Azar et al. (2020), Gouin-Bonenfant (2020), Berger et al. (2020), Jarosch, Nimczik, and Sorkin (2021), and Brooks, Kaboski, Li, and Qian (2021) study the effects of labor market power on wages and the labor share. Caldwell and Danieli (2019) and Caldwell and Harmon (2019) study the effects of outside options 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, and (ii) quantifying the importance of wage markdowns for firm wage premia.

This paper is also inspired by recent work on the labor share of national income, which focuses on the role of product market power and the aggregate production technology. 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. 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 are also important in explaining firm wage premia and their labor shares.

2 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. Wage-setting throughout this paper is contemporaneous. I impose just enough structure on this framework to allow a number of channels of firm heterogeneity, while attempting to remain agnostic about 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.

2.1 Departures from standard labor market search 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 depend on 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 (Peters, 2020; De Loecker et al., 2020).

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 to each input to differ across firms. 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 Burdett and Mortensen (1998). In addition, firms can also increase recruitment effort, as in job search models such as the Diamond-Mortensen-Pissarides model. Each firm j posts piece-rate wages per efficiency unit of labor (Barlevy, 2008), 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. 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:

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

where $s_{jt} = s(\Phi_{jt}, A_{jt})$ 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 ($R_{jt} = R(\Phi_{jt}, A_{jt}, V_{jt})$) 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. All else equal, firms that offer higher wages and better non-wage amenities have a higher recruitment rate and lower separation rate. Equation (3) is therefore the firm-specific labor supply function.

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 Mortensen (2010). I also allow recruitment and separation to depend on non-wage amenities, as there is evidence that non-wage amenities are important determinants of worker flows between firms (Sorkin, 2018). 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 maximizes profits subject to (1), (2), and (3):

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

where $S_{jt} = \{X_{jt}, D_{jt}, A_{jt}, K_{jt-1}, H_{jt-1}\}$. 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.

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

$$\Phi_{jt} = WM_{jt} \cdot ARPH_{jt} \cdot PM_{jt}^{-1} \cdot LEO_{jt} = WM_{jt} \cdot MRPH_{jt} \tag{4}$$

which is log-linear in four channels of firm heterogeneity. The last three components form the marginal revenue product of labor ($MRPH$).

Wage markdown (WM). This component is the fraction of marginal revenue product of labor paid as wages. It can be written as:

$$WM_{jt} = \frac{\epsilon_{jt}^H}{1 + \epsilon_{jt}^H - \beta E_t \left(\frac{(1-s_{jt+1})J_{jt+1}}{c_{V,jt}V_{jt} + c(V_{jt})} \right) R_{V,jt}} \tag{5}$$

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 (5) shows that firms facing lower labor supply elasticities possess stronger wage-setting power: they mark down wages more. The firm-specific labor supply elasticity (ϵ^H_{jt}) depends on:

$$\epsilon^H_{jt} = \frac{R_{jt}}{H_{jt}}\epsilon^R_{\Phi,jt} - \frac{s_{jt}H_{jt-1}}{H_{jt}}\epsilon^s_{\Phi,jt} > 0$$

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 of having 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 (5) nests static monopsony models, in which firms use wages as the sole instrument for hiring workers. 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 depends on the microfoundation for the labor supply curve. In the Supplementary Material, I present a few possible microfoundations.

While I do not take a stance on the joint distribution of heterogeneity in primitives (idiosyncratic productivity (X), demand (D), and non-wage amenities(A)), it is worth discussing how they 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 on 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 same is true with differences in idiosyncratic demand (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.

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: (i) The labor market is characterized by search frictions and workers search on-the-job; (ii) the goods market is perfectly competitive and the production function is linear in labor inputs; (iii) firms attract new workers by posting wages only; (iv) firms are in their steady state.

The first assumption takes a stance 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. 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. 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:

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

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 $WM_j = \frac{\epsilon^H(\Phi_j)}{1+\epsilon^H(\Phi_j)}$.

Average revenue product of labor (ARPH). This is the theory-consistent measure of productivity for the firm wage premium.[6] 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 (4) 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.

The average revenue product of labor is increasing in firms' idiosyncratic productivity (X) and demand (D), and decreasing in the value of non-wage amenities (A). This latter 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 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, Midrigan, and Xu, 2015). Equation (4) 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). This is because more productive firms are able to produce with lower marginal costs and charge lower prices, thereby facing lower price elasticities of demand – they face less product market competition locally.

If two firms are identical except for the value of their non-wage amenities (A), then

[6] This component is also commonly referred to as "labor productivity".

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 (4) 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. The labor elasticity of output in this case is then sector-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:

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

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 productivity (X). Then the more productive firm has a lower labor elasticity of output if the elasticity of substitution between labor and other inputs is greater than one (substitutes). This is because the more productive firm, who wants to hire more workers, faces a higher relative cost of labor. The presence of labor market frictions imply that firms must pay higher wages to hire more workers. Since labor and other inputs are substitutes, the more productive firm substitutes labor with other inputs, reducing its labor intensity, and hence its labor elasticity of output.

Analogously, if two firms have different values of non-wage amenities (A), but are otherwise identical, then the firm with less desirable amenities (lower A) will have a lower labor elasticity of output if labor and other inputs are substitutes.

2.3 Discussion

The firm wage premium equation (4) has a few advantages. First, it features channels of firm heterogeneity related to the broader literature on between-firm wage inequality (Carlsson, Messina, and Skans, 2014; Song et al., 2019; Bell, Bukowski, and Machin, 2018) and the labor share of income (Autor et al., 2020; De Loecker et al., 2020) in a simple and transparent way. Second, the log-linear structure substantially simplifies a decomposition 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 (Postel-Vinay and Robin, 2002). On the theoretical front, introducing of diminishing returns to labor in a frictional labor market model comes with additional modelling challenges. 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 (Stole and Zwiebel, 1996). On the empirical front, the sequential auctions mechanism implies that worker mobility depends on the previous employer, a channel ruled out by standard AKM identifying restrictions.[7] 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.

Second, implicit in the efficiency units specification of the production function, I assume that worker types are perfect substitutes, although average worker efficiency and

[7] Di Addario, Kline, Saggio, and Solvsten (2020) find little evidence for a past-employer effect on worker mobility.

firm productivity are complements. This restrictive assumption implies that the model abstracts from worker-firm sorting based on production complementarities (Eeckhout and Kircher, 2011). In return, this assumption (i) delivers a mapping between AKM regressions and the structural firm wage premium equation; and (ii) keeps the production function estimation procedure computationally 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. In Online Appendix E, I extend the analysis to include high and low-skilled occupations.

3 Estimating the Firm Wage Premium Equation

3.1 Empirical approach

Before implementing the structural decomposition, 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. One approach to estimate these firm characteristics is 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. This approach measures firm-specific markups using 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 knowledge of the firm wage premium.[8] 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

[8] Morlacco (2019) exploits a similar idea to estimate firms' market power in foreign intermediate input markets.

wage markdowns distort only labor demand. 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 technology to be labor-augmenting; (ii) allowing labor elasticities of output to vary across firms; (iii) using a flexible control function estimation approach; and (iv) showing the firm wage premia are required in the control function when firms have labor market power.

3.2 Estimating firm wage premia

A common way of estimating firm wage premia is to estimate firm effects from an AKM regression (Abowd et al., 1999). 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 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). 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 to 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 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 similar average wages, but the shape of their internal wage distributions differ substantially, they are clustered into different 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 stable across estimation samples. Online Appendix B provides more detail on how I cluster firms and addresses the main restrictions underlying the AKM regression.

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 $ARPH = \frac{PY}{\bar{E}L}$. 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:

$$\begin{aligned} y_{jt} = {} & \beta_{h,s} h_{jt} + \beta_{k,s} k_{jt} + \beta_{m,s} m_{jt} + \beta_{o,s} o_{jt} + \beta_{hh,s} h_{jt}^2 + \beta_{kk,s} k_{jt}^2 \\ & + \beta_{mm,s} m_{jt}^2 + \beta_{oo,s} o_{jt}^2 + \beta_{hk,s} h_{jt} k_{jt} + \beta_{hm,s} h_{jt} m_{jt} + \beta_{ho,s} h_{jt} o_{jt} \\ & + \beta_{km,s} k_{jt} m_{jt} + \beta_{ko,s} k_{jt} o_{jt} + \beta_{mo,s} m_{jt} o_{jt} + x_{jt} + \epsilon_{jt} \end{aligned} \tag{6}$$

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 Online Appendix A, French administrative datasets measure these intermediate inputs separately. Define ϵ_{jt} as the error term orthogonal to firms' input choices. In Online Appendix C, I address some well-known challenges in production function estimation (De Loecker and Goldberg, 2014; Gandhi, Navarro, and Rivers, 2020).

Firms' input demand is an endogenous choice of the firm and depends on the firm's productivity realization x_{jt}. 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 researcher to "observe" the firms' idiosyncratic productivity by inverting their optimal input demand function for a fully flexible variable input. My variable input of choice 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 6 in Online Appendix A.

Using the first-order conditions for each factor input, I obtain the 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 variables that can affect firms' input demand, which includes location, sector, 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.[9] However, because I observe hourly wages at the 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 firms' wage-setting power, which distorts relative input prices and input demand.

To obtain the control function, I invert the service input demand function under the assumption that, conditional on the variables in the control function, service input demand is monotonically increasing in idiosyncratic productivity x_{jt}. The control function expresses firm productivity as a function of observed variables:

$$x_{jt} = x(h_{jt}, k_{jt}, m_{jt}, o_{jt}, \mathbf{Z}_{jt}, \phi_{jt}) \tag{7}$$

The production function can then be estimated following the two-step GMM approach described in Ackerberg et al. (2015). In step one, I combine (6) and (7) and estimate the

[9]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 that come from differences in worker composition and market power.

following by OLS:

$$y_{jt} = \Psi(h_{jt}, k_{jt}, m_{jt}, o_{jt}, \mathbf{Z}_{jt}, \phi_{jt}) + \epsilon_{jt} \tag{8}$$

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:

$$x_{jt} = g(x_{jt-1}) + \zeta_{jt} \tag{9}$$

where $g(.)$ is a flexible function and ζ_{jt} is a productivity shock. In step two, I estimate the production function parameters. Combining the control function (7), the predicted output from (8), and the law of motion for productivity (9), 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. I combine the two steps into one and implement Wooldridge (2009). I estimate production functions by 2-digit sectors within three time intervals: 1995-2000, 2001-2007, 2008-2014. I then compute the labor (LEO) and service input (OEO) elasticities of output as follows:

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

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

In the third step of the estimation of firm characteristics, 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, markups represent the only distortion to service intermediate input demand (De Loecker and Warzynski, 2012). One can then express price-cost markups as a function of the service input share and service input elasticity of output:

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

where $\exp(\hat{\epsilon}_{jt})$ removes measurement error. I then 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 is the wage markdown.

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. Firm wage premia are estimated using matched employer-employee data, which follow workers over time and employment spells at different firms. The four channels of firm heterogeneity are estimated using firm balance sheet panel data. While both types of datasets have become increasingly accessible, they are typically not jointly available. To the extent that balance sheet datasets are available, they often cover only a set of large firms or manufacturing firms, or do not contain a panel structure. I therefore use matched employer-employee and balance sheet panel data from France. These datasets cover the population of firms and workers 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 information. These datasets contain balance sheet information for all firms in France without restriction on the size of firms. From these datasets, I obtain information on variables such as sales, nominal value of production, employment, intermediate input and capital expenditure. I provide details on measurement in Online 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.[10] Because workers are followed over time and their employer identifiers are observed, I use this dataset to estimate the AKM-BLM regressions.

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. 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 following the procedure described in the previous section. 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 Sample selection

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 sector is excluded. I include only firms with at least 5 employees. I harmonize all industry codes to the latest available version (Nomenclature d'activités Française – NAF rév. 2). I drop 2-digit sectors with less than 500 observations for each estimation time interval. This is important when estimating flexible production function 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

[10] Only October-workers born in even years are observed prior to 2002.

each time interval.

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 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 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 to allocate each firm-year observation 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 the four firm characteristics. 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 7 in Online Appendix D.

5 What Explains Firm Wage Premia?

5.1 Product market power and labor elasticities of output matter for firm wage premia

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.[11]

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

$$\phi_{jt} = wm_{jt} + arph_{jt} - pm_{jt} + leo_{jt} \tag{10}$$

where lowercase letters are variables in logs.

To maximize interpretability, my preferred variance decomposition method is a Shapley decomposition (Shorrocks, 2013).[12] I implement this variance decomposition by running equation (10) as a linear regression and then decomposing the R^2 into four components. Because each channel of firm heterogeneity is exactly identified in my estimation approach, my decomposition is also exact. Each component represents the marginal contribution of a given channel 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 values between 0 and 1, and they sum up to 1. Table 1 presents the Shapley decomposition results. Figure 7 in Online Appendix D displays the Shapley decomposition year by year.

The first row of Table 1 shows that wage markdowns account for 25% of the firm wage premium distribution. As discussed in Section 2, wage markdowns can differ across firms for a number of reasons. In static oligopsonistic wage-posting models, wage markdowns depend on the firm-specific labor market shares of employment or wage bill (Jarosch et al., 2021; Berger et al., 2020). The Burdett and Mortensen (1998) model without heterogeneity generates wage dispersion purely by search frictions, which shows up in the form of heterogeneous wage markdowns. Alternatively, dispersion in wage markdowns can reflect heterogeneity in outside options (Caldwell and Danieli, 2019; Schubert, Stansbury, and Taska, 2019).

Widely used models of frictional labor markets often feature heterogeneous firm productivity in the form of the average revenue product of labor (Postel-Vinay and Robin,

[11] Online Appendix E introduces an extension that includes skill intensive and less skill intensive occupations.

[12] The Supplementary Material discusses the Shapley decomposition in detail and reports the results from two alternative decomposition methods: a standard variance decomposition and an ensemble decomposition (Sorkin, 2018).

2002; Bagger et al., 2014). In these models, firm productivity determines the wage premium a firm pays relative to other firms for identical workers. The third row of Table 1 shows that heterogeneous average revenue productivity of labor accounts for 35% of the firm wage premium distribution.

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.

Firm-level differences in price-cost markups often do not feature in frictional labor market models. As discussed in Section 2, price-cost markups affect firm wage premia through the firm-specific labor demand in my structural framework. In the macroeconomics of input misallocation, this is a theoretically and quantitatively important determinant of labor demand and firm size (Edmond et al., 2015). 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, accounting for 29% of the firm wage premium distribution, as the fifth row of Table 1 shows. This component often does not feature in frictional labor market models, but it is a key determinant of firm-specific labor demand in my framework. It is also an important component of aggregate labor demand (Karabarbounis and Neiman, 2014).

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. This 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).

To get a sense of how much the role of differences in labor productivity and wage markdowns could be overestimated, I first re-estimate 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 approximates the estimation of 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.

5.2 Large differences in the estimated firm characteristics

I start by reporting new estimates of wage markdowns and labor elasticities of output, which display large cross-sectional variation. I also confirm the well-documented existence of large productivity dispersion and the more recently documented price-cost markup dispersion across firms. Table 2 summarizes the empirical moments of each estimated channel of firm heterogeneity in 2014.[13]

The first row of Table 2 describes the distribution of wage markdowns (WM). France has one of the highest minimum wages worldwide and an almost universal coverage of collective bargaining agreements. Nevertheless, I find that most French firms possess significant wage-setting power – half of the firms in my sample pay less than 0.80 of the marginal revenue product of labor as wages. I also find substantial dispersion of wage markdowns across firms. 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,

[13] The Supplementary Material (Tables 2-9) reports these statistics by sector.

workers obtain approximately two-thirds of their marginal revenue productivity.

	Mean	Median	Var	25^{th} Pct	75^{th} 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
Ave. rev. product of labor (log)	4.52	4.44	0.37	4.08	4.90
Marg. rev. 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.

The set of direct estimates of wage markdowns is small. I start by comparing my estimates to those of Hershbein et al. (2020) and Mertens (2019), whose estimates are methodologically the closest to mine. They 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, $\frac{\epsilon^H}{1+\epsilon^H}$. 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 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.

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, Fernald, Fisher, and Kimball, 2013; Oberfield and Raval, 2021). 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.

Consistent with existing findings, I find that 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 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. Edmond, Midrigan, and Xu (2018) report an interquartile range for markups of 1.31-0.97=0.34. My estimates for the interquartile range is 0.30.

The second-to-last row of Table 2 reports the summary statistics for the average revenue product of labor ($ARPH$) in logs. The dispersion of firm productivity is well-documented (Syverson, 2011). I find that the average revenue product of labor 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. The average interquartile range within two-digit sectors is 1.98.

5.3 Firm wage premium dispersion is moderate

The firm wage premium distribution masks large underlying differences in firm characteristics. Although the average French firm possesses significant wage-setting power and the channels of firm heterogeneity in equation (10) are highly dispersed, they do not translate into similarly dispersed firm wage premia.

Table 3 reports statistics about firm wage premia. The variance of firm wage premia (ϕ) is modest (0.009), accounting for 4.5% of wage dispersion, similar to the numbers for the United States, Sweden, Austria, Norway, and Italy from Bonhomme, Holzheu, Lamadon, Manresa, Mogstad, and Setzler (2020). At the same time, the relevant firm characteristics in equation (10) are orders of magnitude more dispersed than firm wage premia, as Table 2 shows.

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. This gap is almost twice as large as the average gender wage gap among OECD countries.

	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	7,908,159	5,497,277

Table 3: Dispersion of firm wage premia in 2014.

5.4 High labor productivity firms have lower labor elasticities of output

The main reason why firm wage premium dispersion is moderate relative to the large differences in estimated firm characteristics is that firms' labor productivity and labor elasticities of output are strongly negatively correlated. The correlation matrix is reported in Table 4.

The third row, second column of Table 4 presents this main finding. This negative correlation is the most quantitatively important among the set of cross-terms that offset the effects of firm heterogeneity on the firm wage premium distribution. At the same time, the correlation between labor productivity and the material elasticity of output is positive (0.51), while the correlation between labor productivity and the service elasticity of output is negative (0.28). The correlation between labor productivity and the capital elasticity of output is also negative (-0.17).[14] I discuss the other pairwise correlations in Online Appendix D.

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

[14]The Supplementary Material (Tables 19-23) reports the correlations by sector.

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 deviations from constant returns to scale – 98% of firms fall within 5 percentage points of constant returns to scale. Recall that my production function estimation approach only imposes constant returns to scale within sectors *on average*, but does not otherwise preclude firms from departing from constant returns to scale.

	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 4: Correlation matrix for channels of firm heterogeneity in 2014.

The labor input substitution channel works as follows. Since firms face upward-sloping labor supply curves due to labor market frictions, firms must offer higher wages to hire more workers. Because more productive firms wish to grow larger than less productive firms, the former face higher relative costs of labor. 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.

6 Implications for Labor Misallocation and Labor Shares Across Firms

Role of input substitution in labor misallocation across firms. The cross-sectional dispersion of the marginal revenue product of labor signals misallocation of labor inputs (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). As shown in equation (10), the marginal revenue product of labor (*mrph*) consists of the average revenue product of labor (*arph*), price-cost markups (*pm*), and labor elasticity of output (*leo*). Conventional measures of labor misallocation approximate the *mrph* with the *arph*. This is a precise approximation if *pm* and *leo* are constant across firms. However, my finding that *arph* and *leo* are strongly negatively correlated implies that the *mrph* is considerably less dispersed than the *arph*. Consequently, conventional measures of labor misallocation overstate the variance of *mrph* and, hence, the degree of labor misallocation.

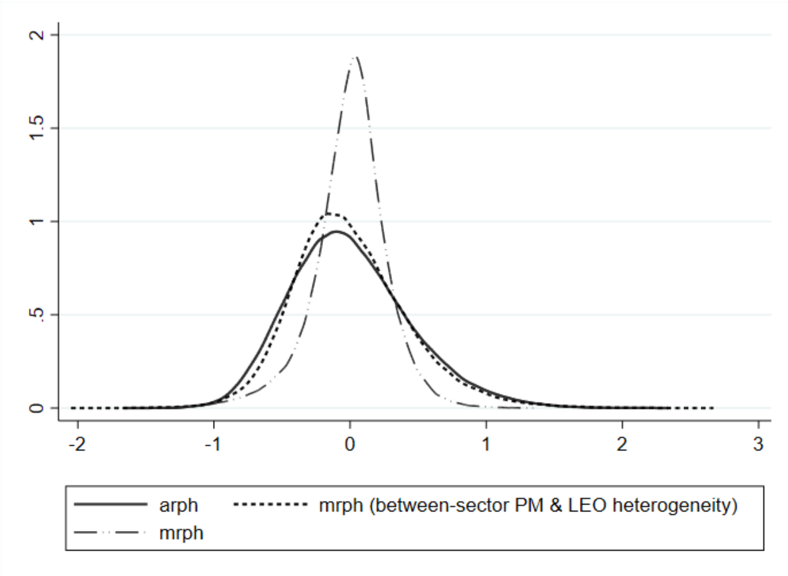

Figure 1: Distribution of the average and marginal revenue product of labor.
Year: 2014. The solid line denotes the average revenue product of labor, the short dashed line denotes the marginal revenue product of labor when price-cost markups and labor elasticities of output are sector-specific but not firm-specific, and the long dashed line denotes the marginal revenue product of labor when all dimensions are firm-specific. Each variable is de-meaned.

Table 2 shows that the variance of *arph* is over three times larger than the variance of *mrph*, $\frac{V(arph)}{V(mrph)} = 3.7$. This is graphically depicted in Figure 1, which plots the de-meaned *mrph* and *arph*. Indeed, Table 4 shows that the correlation coefficient between the two is only 0.30. Intuitively, this mismeasurement stems from the fact that 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. In Online Appendix D, I perform a back-of-the-envelope exercise, based on Hsieh and Klenow (2009), to get a sense of the extent to which measured labor misallocation is overstated when one uses the *arph* to approximate the *mrph*. I find that labor misallocation is overstated by close to 3 times.

New explanation for lower labor shares of revenue among more productive firms. 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, for example through capital-labor substitution (Karabarbounis and Neiman, 2014). However, recent research shows that the reallocation of sales towards highly productive "superstar" firms with low labor shares is the main driver (Autor et al., 2020; Kehrig and Vincent, 2020).

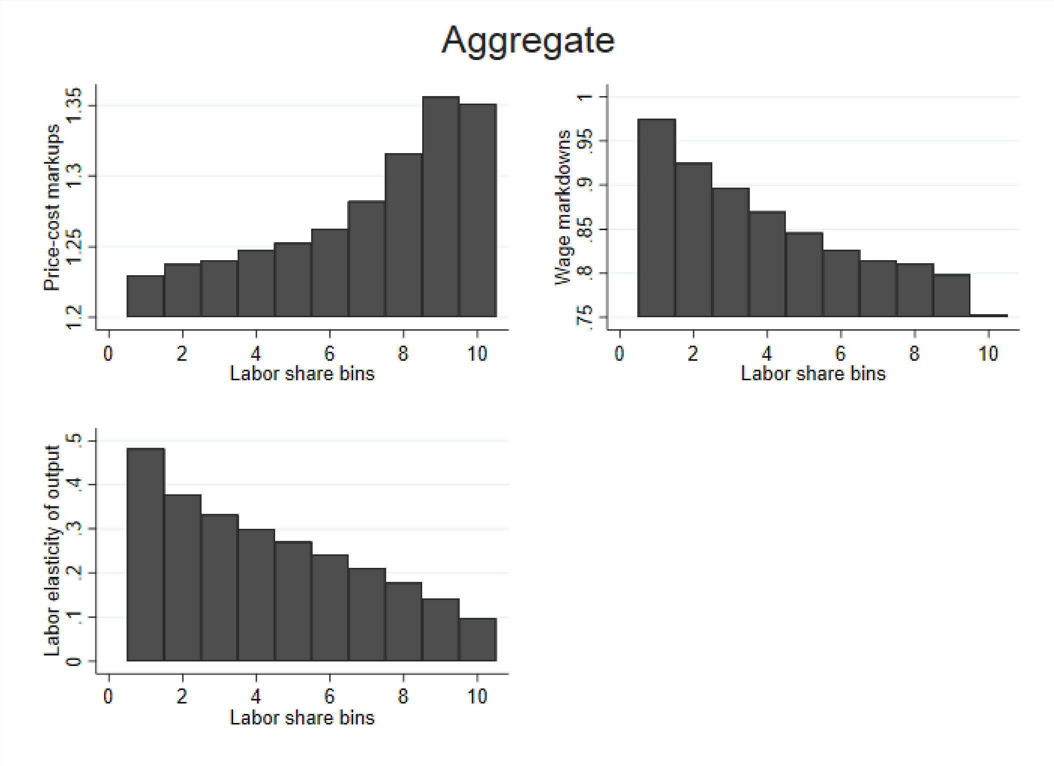

Figure 2: Labor shares and firm characteristics among French firms in 2014.

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 2 shows that 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 have more labor market power (Gouin-Bonenfant, 2020). Figure 2 also provides a new explanation for superstar firms' low labor shares: high productivity firms have 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 has the potential to play an important role.

	Agg	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 5: Labor share decomposition in 2014.

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

To get a sense of which channel of heterogeneity matters most, I decompose firm-level differences in labor shares. Write the labor revenue share in logs: $ls_j = wm_j - pm_j + leo_j$. Then, labor shares can be decomposed as such:

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

Table 5 reports the labor share decompositions. 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. I discuss the sector-specific differences in Online Appendix D.

7 Concluding Remarks

This paper investigates the firm characteristics underlying a well-known fact: some firms pay higher wages than other firms for identical workers. To do so, I develop and implement a new structural decomposition of firm wage premia. While it is well-understood that firms' labor productivity and wage-setting power matter for firm wage premia, this paper highlights the role of firm characteristics that are largely overlooked in research on firm wage premia: product market power and labor elasticities of output. My decomposition suggests that, without considering the role of firms' product market power and labor elasticity of output, workhorse models that generate firm wage premia overestimate the role of labor productivity and wage-setting power. By embedding a richer notion of a firm, the decomposition also uncovers a negative correlation between labor productivity and the labor elasticity of output underlying the firm wage premium distribution. This generates new implications for the measurement of labor misallocation and for the understanding of low labor shares among high productivity firms.

Bibliography

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.

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, M. Steinbaum, and B. Taska (2020). Concentration in US Labor Markets: Evidence from Online Vacancy Data. *Labour Economics 66*, 1018–86.

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*.

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*.

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. *Working paper*.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler (2020). How Much Should 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 (2021). Exploitation of Labor? Classical Monopsony and Labor's Share. *Journal of Development Economics (forthcoming)*.

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.

Carlsson, M., J. Messina, and O. Skans (2014). Wage Adjustment and Productivity Shocks. *Economic Journal 126*(595), 1739–1773.

Dauth, W., S. Findeisen, and J. Suedekum (2021). Adjusting to Globalization in Germany. *Journal of Labor Economics 39*(1), 492–533.

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 F. Warzynski (2012). Markups and Firm-Level Export Status. *American Economic Review 102*(6), 2437–2471.

Di Addario, S., P. Kline, R. Saggio, and M. Solvsten (2020). It Ain't Where You're From, It's Where You're At: Hiring Origins, Firm Heterogeneity, and Wages. *Working paper*.

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.

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.

Gandhi, A., S. Navarro, and D. Rivers (2020). On the Identification of Gross Output Production Functions. *Journal of Political Economy 128*(8), 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*.

Gouin-Bonenfant, E. (2020). Productivity Dispersion, Between-Firm Competition, and the Labor Share. *Working paper*.

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.

Hershbein, B., C. Macaluso, and C. Yeh (2020). Monopsony in the US Labor Market. *Working paper*.

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 (2021). Granular Search, Market Structure, and Wages. *Working paper*.

Karabarbounis, L. and B. Neiman (2014). The Global Decline of the Labor Share. *Quarterly Journal of Economics 129*(1), 61–103.

Kehrig, M. and V. Vincent (2020). The Micro-Level Anatomy of the Labor Share Decline. *Quarterly Journal of Economics (forthcoming)*.

Kline, P., N. Petkova, H. Williams, and O. Zidar (2019). Who Profits from Patents? Evidence from Innovative Firms. *Quarterly Journal of Economics 134*(3), 1343–1404.

Lamadon, T., M. Mogstad, and B. Setzler (2019). Imperfect Competition, Compensating Differentials, and Rent Sharing in the U.S. Labor Market. *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.

Manning, A. (2006). A Generalised Model of Monopsony. *Economic Journal 116*(508), 84–100.

Mertens, M. (2019). Micro-Mechanisms Behind Declining Labor Shares: Market Power, Production Processes, and Global Competition. *Working paper*.

Morlacco, M. (2019). Market Power in Input Markets: Theory and Evidence from French Manufacturing. *Working paper*.

Mortensen, D. (2010). Wage Dispersion in the Search and Matching Model. *American Economic Review: Papers and Proceedings 100*(2), 338–342.

Oberfield, E. and M. Raval (2021). Micro Data and Macro Technology. *Econometrica 89*(2), 703–732.

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 88*(5), 2037–2073.

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.

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. (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. (2011). What Determines Productivity? *Journal of Economic Literature 49*(2), 326–365.

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.

Online Appendix

Understanding High-Wage and Low-Wage Firms

Horng Chern Wong
CREST - École Polytechnique
Email: horngchernwong@gmail.com

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.

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

Table 6: Contemporaneous correlations between factor input growth rates and sales revenue growth rates, 1995-2014.

B K-Means Clustering and Addressing AKM Regression Restrictions

B.1 K-means clustering of firms into 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. 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.

B.2 AKM restrictions: conditional exogenous mobility and log-additivity

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 this section, 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.

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), 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 3, 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. (2018). Figure 3 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.

An alternative way to assess the AKM regression specification is to compare the changes in residual wages to changes in firm effects, following Sorkin (2018). This is similar to the above method. I run the following regression among all employer-to-employer transitions:

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

where $w^r_{it} = w_{it} - x'_{it}\hat{\beta}$ 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 4 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 (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 4 do not resemble a V-shape around zero.

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 5 and 6 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.

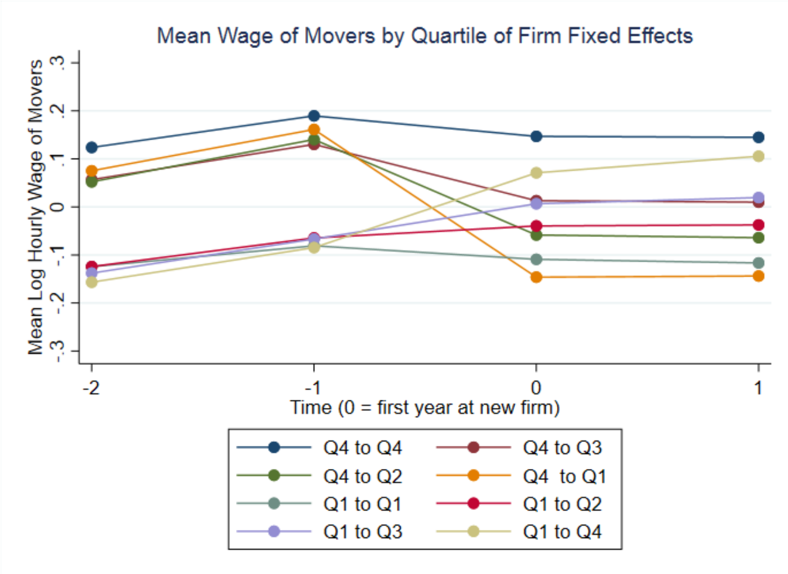

Figure 3: Worker mobility and wage changes by quartiles of firm effects (2009-2014).

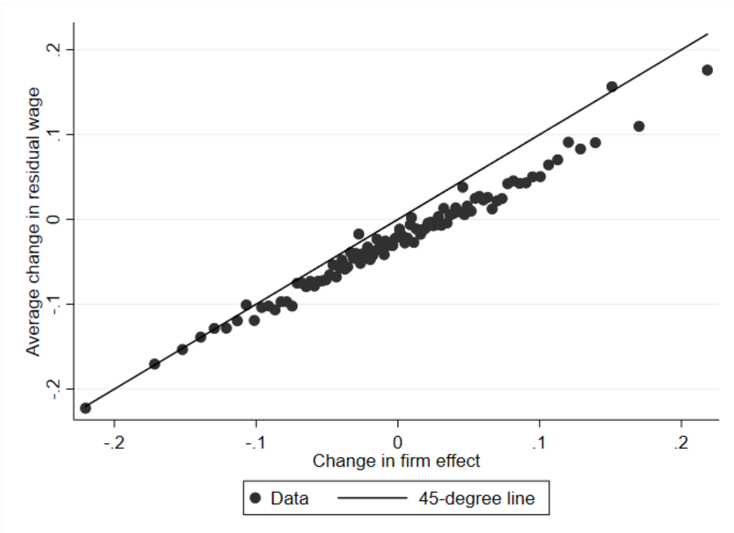

Figure 4: Average wage changes from worker mobility by declines of changes in firm premia (2009-2014).

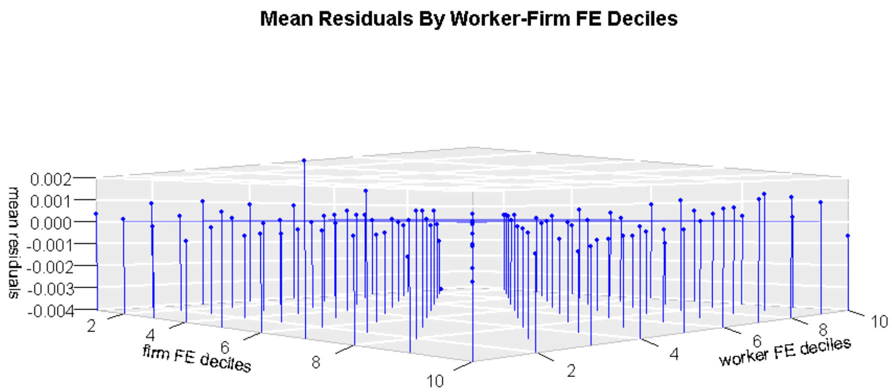

Figure 5: Mean estimated residuals by worker-firm deciles (2014)

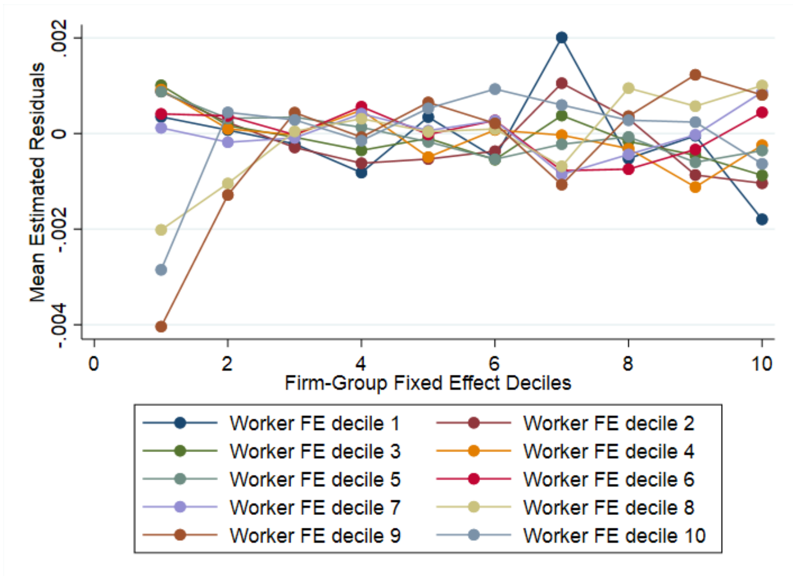

Figure 6: Mean estimated residuals by worker-firm deciles (2014)

C Additional Details on Production Function Estimation

C.1 Additional restrictions for gross output production functions

As Gandhi et al. (2020) 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. 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.

C.2 Unobserved output prices

A common challenge in the production function estimation literature is that output prices are rarely observed (De Loecker and Goldberg, 2014).[15] 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. 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 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

[15]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, e.g. US Manufacturing Census.

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 trade models such as 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 accumulate 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.[16]

[16] 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.

D Additional Figures, Tables, and Results

D.1 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 7: Summary statistics: employers and employees (1995-2014).

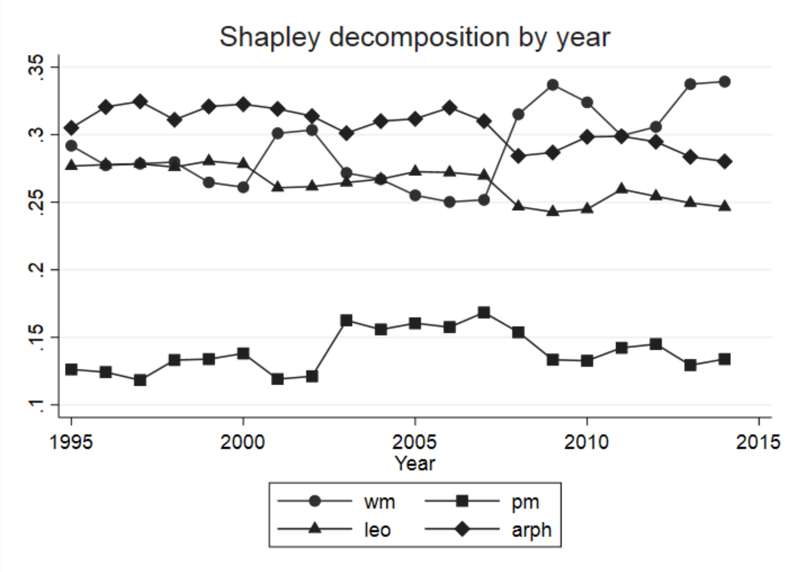

Figure 7: Shapley decomposition of the firm wage premium over time

D.2 Other correlations between the estimated 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., 2021). This prediction finds support in the last row of the first column in Table 4. The correlation between wm and $mrph$ is large (-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

($mrph = arph$).[17] This implies that the correlation and covariance between wm and $arph$ is the same as that between wm and $mrph$. Table 4 shows that these correlations are far from being identical. In particular, the correlation between wm and $arph$ (-0.26) is considerably weaker than the correlation 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 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 row, second column of Table 4 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, as shown in the Supplementary Material (Tables 19-23), 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 Table 4 shows 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.

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

D.3 The misallocation effects of market power on sectoral TFP

Section 6 shows that a key sufficient statistic for labor misallocation, the variance of the marginal revenue product of labor, is overstated when approximated by the average revenue product of labor, as is conventionally done. 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) type 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. Below, 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.

Derivations. 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:

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

Note that when markups are constant, this can be written as: $TFPR_{sj} \propto MRPH_{sj}^{\alpha_s^H}$. 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:

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

D.4 Why do labor shares of revenue vary across firms?

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 5 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.

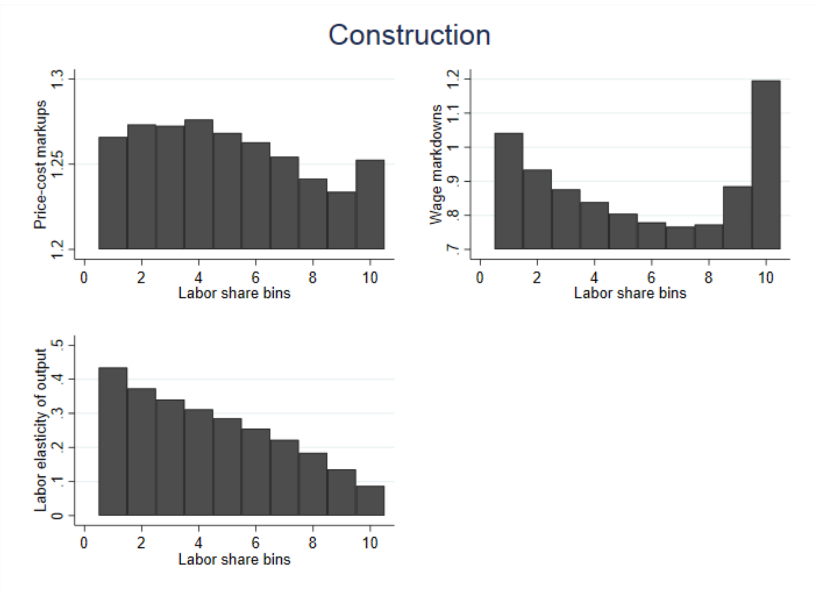

Figure 8: Labor shares and firm characteristics among construction firms in 2014.

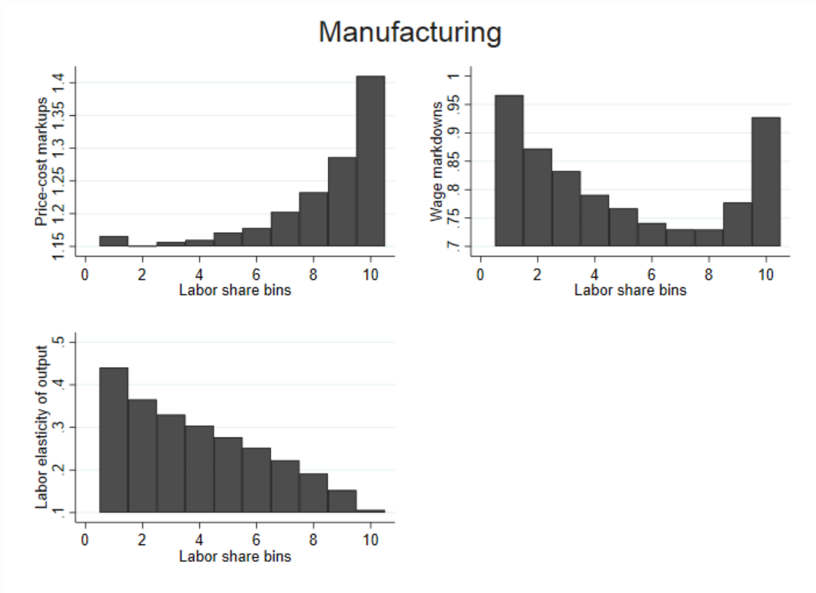

Figure 9: Labor shares and firm characteristics among manufacturing firms in 2014.

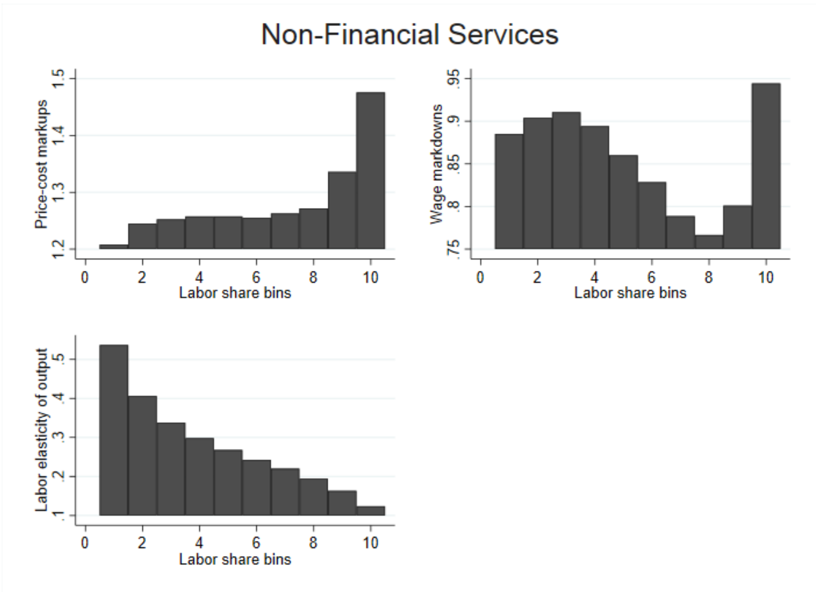

Figure 10: Labor shares and firm characteristics among non-financial services firms in 2014.

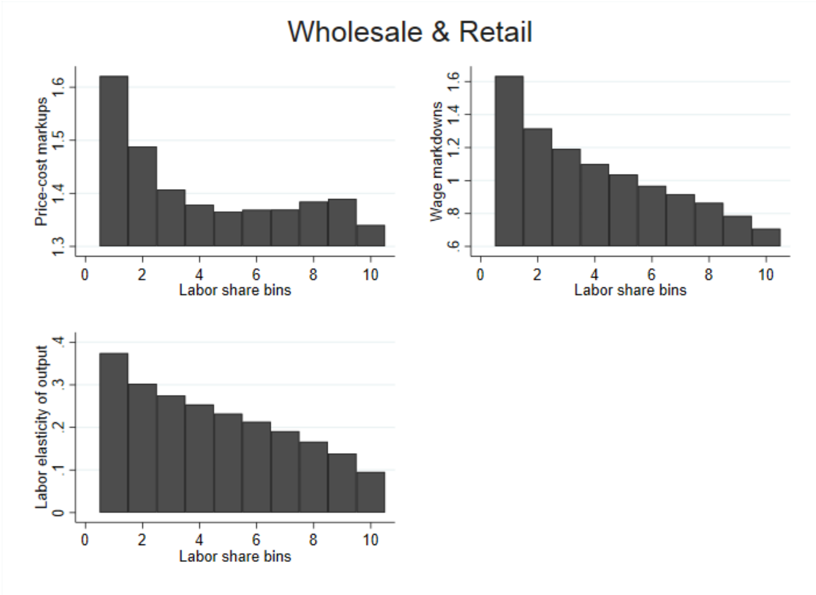

Figure 11: Labor shares and firm characteristics among wholesale and retail firms in 2014.

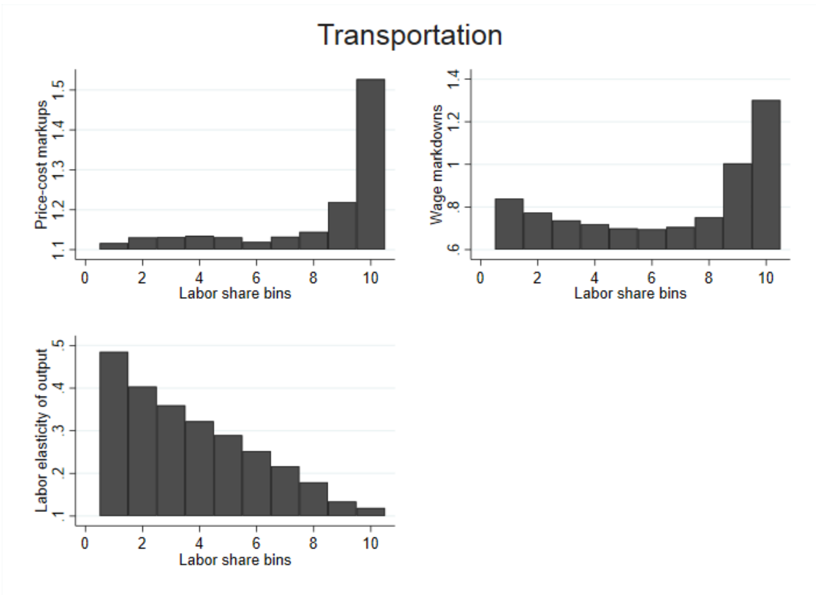

Figure 12: Labor shares and firm characteristics among transportation firms in 2014.

D.5 Other implications of the main findings

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 2, 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 above 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 $PM_j \cdot WM_j^{-\gamma_s^c} =$ 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 4 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.

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-skilled and a low-skilled occupation. The main findings of this extension can be summarized as follows: (i) Workers in high-skilled occupations are paid a greater share of their marginal revenue product compared to workers in low-skilled occupations; (ii) price-cost markups and labor elasticities of output matter for firm wage premia in both high-skilled and low-skilled occupations; (iii) the negative correlation between labor productivity and labor elasticities of output is large for both skill groups, but larger for low-skilled occupations; (iv) the variance of the marginal revenue product of labor is larger among high-skilled occupations than low-skilled 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-skilled occupations more, while the opposite is true for high-skilled occupations.

Defining high-skilled and low-skilled occupations. I use the one-digit occupation classifications in the DADS matched employer-employee data set to define low-skilled occupations as blue-collar occupations (e.g. maintenance workers and welders) and administrative support occupations (e.g. clerical workers and secretaries); I define high-skilled 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 subscript $s = \{h, l\}$ denote high and low-skilled 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 approximate $f(.)$ 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-skilled and low-skilled 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.

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

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 8 reports the summary statistics for wage markdowns among high-skilled and low-skilled workers. high-skilled workers typically incur a smaller markdown of wages below marginal revenue products compared to low-skilled workers.

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 9 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 9: 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 10 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 10: Shapley decomposition of firm wage premia, averaged over skill groups, in 2014.

Cross-sectional correlations of firm characteristics. Table 11 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-skilled occupations than high-skilled ones. One difference between high-skilled and low-skilled occupations is that wage markdowns increase more strongly with labor productivity among low-skilled occupations than among high-skilled 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 11: 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 12 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-skilled occupations than among high-skilled 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 12: 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-skilled and low-skilled labor, Figure 13 shows that the wage markdowns and labor elasticities of output among low-skilled, rather than high-skilled, labor are the main contributors to the lower labor shares of highly productive firms. Firms with lower labor shares have higher high-skilled labor elasticities of output and markdown high-skilled wages by less.

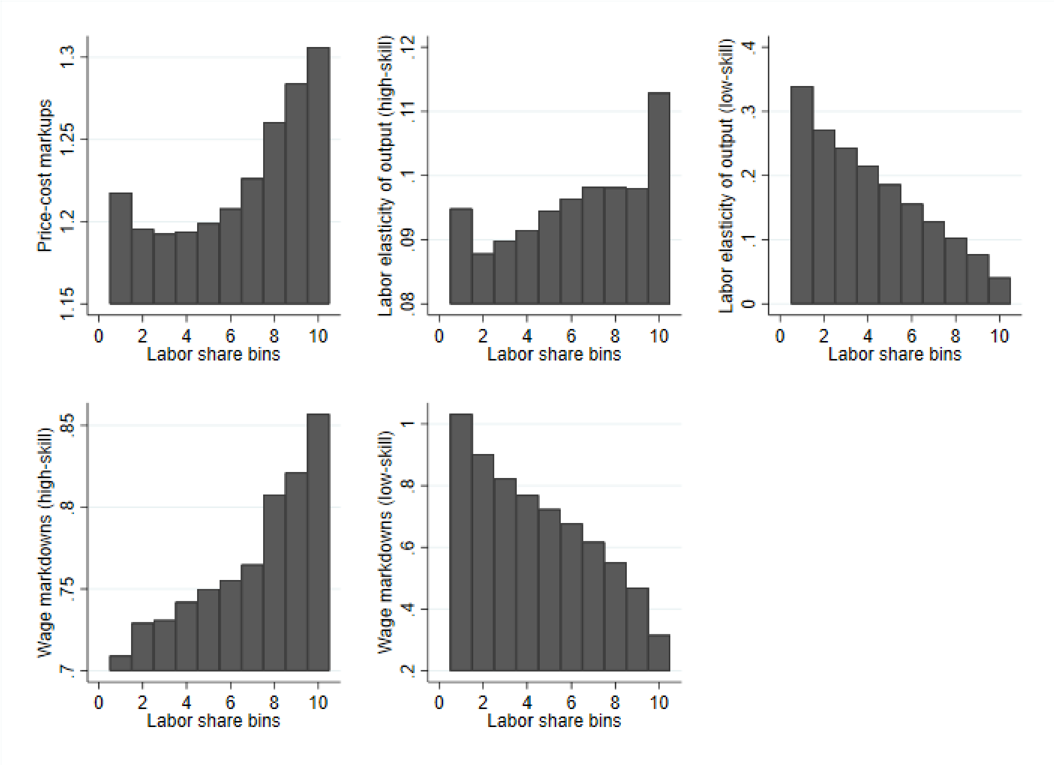

Figure 13: Firm characteristics across the labor share distribution by skill group in 2014.

Bibliography

Arnoud, A. (2018). Automation Threat and Wage Bargaining. *Working paper.*

De Loecker, J. and P. Scott (2016). Estimating Market Power: Evidence from the US Brewery Industry. *Working paper.*

Flynn, Z., A. Gandhi, and J. Traina (2019). Measuring Markups with Production Data. *Working paper.*

Gourio, F. and L. Rudanko (2014). Customer Capital. *Review of Economic Studies*

81 (3), 1102–1136.

Lopes de Melo, R. (2018). Firm Wage Differentials and Labor Market Sorting: Reconciling Theory and Evidence. *Journal of Political Economy 126* (1), 313–346.