Understanding High-Wage and Low-Wage Firms*

Horng Chern Wong†

JOB MARKET PAPER

Latest version here.

October 2019

Abstract

I study why some firms pay a wage premium relative to other firms for identical workers. To unpack the firm wage premium distribution, I develop and implement a novel structural decomposition using datasets on the universe of employers and employees in France. Existing research shows that firm wage premia depend on firms' *productivity* and *wage-setting power*. This paper shows that they also depend on firms' *product market power* and *labor share of production*. My decomposition reveals strong correlations between these firm characteristics. First, the negative relationship between productivity and labor share of production (i) provides a new explanation for exceptionally productive superstar firms' low labor shares of revenue, and (ii) implies that conventional measures of labor misallocation overstate the degree of misallocation. Second, firms' product and labor market power are in general negatively correlated, but superstar firms have more market power in both markets than other firms.

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

*Special thanks to my supervisors, Thijs van Rens and Roland Rathelot, for their invaluable guidance, support, and patience. I am very grateful to Dennis Novy and Vincent Sterk for numerous constructive feedback and generosity with their time. For very helpful comments, I thank Jesper Bagger, Pablo Beker, Dan Bernhardt, Christine Braun, Luis Candelaria, Danilo Cascaldi-Garcia, Emma Duchini, James Fenske, Joaquin García-Cabo, Monika Gehrig-Merz, Daniel Habermacher, Simon Hong, Omiros Kouvavas, Alice Kuegler, Jay Lee, Tim Lee, Timo Leidecker, Roberto Pancrazi, Ayush Pant, Carlo Perroni, Franck Portier, Fabien Postel-Vinay, Camilla Roncoroni, Federico Rossi, Alessandro Ruggieri, Ayşegül Şahin, Sylvia Sarpietro, Ben Smith, Isaac Sorkin, Liliana Varela, Marija Vukotic, Jonathan Yeo, and many seminar and conference participants. I gratefully acknowledge financial support from the ESRC and data access from DARES, DGFiP, INSEE, and CASD.

†Department of Economics, University of Warwick. Email: Horng.Wong@warwick.ac.uk

1 Introduction

Some firms pay higher wages than others for identically skilled workers. This is known as the *firm wage premium*. Following the pioneering work of Abowd et al. (1999), a large body of empirical research confirms this finding in a number of countries.[1] The firm wage premium distribution plays an important role in explaining a range of labor market phenomena, from classic questions such as the long-term wage loss of displaced workers (Lachowska et al., 2018) and the gender wage gap (Card et al., 2015), to recent questions about how globalization (Dauth et al., 2018) and the rise of "superstar" firms (Song et al., 2019) impact the wage distribution.

What determines the firm wage premium distribution? Existing research shows that (i) labor market frictions prevent firm wage premia from being competed away, and (ii) firm heterogeneity in *productivity* and *wage-setting power* determine the wage premium a firm pays relative to other firms.[2] On the other hand, recent discussions surrounding the drivers of the labor share of national income (i) emphasize the importance of firms' *product market power* and *labor share of production*, but (ii) largely abstract from labor market frictions.[3,4] In the presence of labor market frictions, each of these firm characteristics are likely to determine firm wage premia, as they are important determinants of labor demand. Yet, because they are often studied separately, less is known about their interrelationships and their relative importance for the firm wage premium distribution.

This paper quantifies the relative importance of firm-level differences in productivity, wage-setting power, product market power, labor share of production, and their interrelationships in accounting for the firm wage premium distribution. I build a frictional labor market framework in which firms are heterogeneous along these dimensions and use it to interpret common regression estimates of firm wage premia.[5] Using rich administrative datasets on the universe of employers and employees in France, I estimate each of these dimensions. I then combine the model with my estimates to unpack the firm wage premium distribution.

My central finding is that differences in product market power and the labor share of production are quantitatively important, accounting for 13% and 24% of the firm wage premium distribution. These dimensions have received little attention in the firm wage premium literature so far.[6] My decompositions also uncover important relationships between pairs of firm characteristics. First, the strong negative correlation between firm productivity and the labor share of production (i) provides a new explanation for exceptionally productive superstar firms'

[1] See, for example, Card et al. (2013), Song et al. (2019), and Alvarez et al. (2018).

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

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

[4] For example, Karabarbounis and Neiman (2014) argues that aggregate capital-labor substitution drives the U.S. labor share decline, while De Loecker and Eeckhout (2018) argue that product market power is key.

[5] These are the firm fixed effects in what is commonly known as an AKM regression (Abowd et al., 1999).

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

low labor shares of revenue besides market power (De Loecker and Eeckhout, 2018), and (ii) implies that conventional measures of labor misallocation (Hsieh and Klenow, 2009) based on revenue per hour overstate the extent of productivity gains from removing labor market frictions, because these measures do not account for firms' ability to sidestep these frictions by substituting labor with other inputs. Second, I find that in general firms' product and labor market power are negatively correlated. However, superstar firms have more market power than others in both markets. The effects of firms' product and labor market power on labor misallocation are often studied separately (Edmond et al., 2018; Berger et al., 2018). Yet, I show that whether these dimensions amplify or dampen each others' effects on labor misallocation depend on their correlation.

To establish these findings, I develop and implement a novel structural decomposition of firm wage premia to quantify the contributions of each firm characteristic. I begin by building a structural model to interpret regression estimates of firm wage premia. In the model, labor market frictions sustain firm wage premia and firms are heterogeneous along multiple dimensions. As in workhorse frictional labor market models, firms differ in productivity and wage-setting power (Burdett and Mortensen, 1998). Wage-setting power is defined as the fraction of marginal revenue product of labor paid as wages. I refer to this fraction as the *wage markdown*.[7] Compared to these models, the novel features of my framework are firm heterogeneity 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*.[8]

I estimate the dimensions of firm heterogeneity using empirical methods from industrial organization. In particular, I provide novel estimates of the distribution of wage markdowns and output elasticities across firms in a range of industries. These dimensions have received increasing attention but their cross-sectional properties are not yet well-documented.[9,10,11] I do so by extending the production-based approach of De Loecker and Warzynski (2012) and Flynn et al. (2019) to accommodate imperfectly competitive labor markets. 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 et al., 2015), in

[7]In a perfectly competitive labor market, wages equal the marginal revenue product of labor, therefore there are no wage markdowns. In a frictional labor market, the wage markdown can vary across firms due to differences in labor supply elasticities, outside options, or relative bargaining positions (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Berger et al., 2018).

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

[9]A growing literature documents monoposonistic labor market competition, for example, Azar et al. (2018), Rinz (2018), and Lamadon et al. (2019)

[10]For sectoral estimates of output elasticities, see Hubmer (2019), Oberfield and Raval (2014), and Basu et al. (2013). For theoretical work on sectoral output elasticities, see Acemoglu and Restrepo (2017).

[11]For recent model-based estimates of wage markdowns, see Berger et al. (2018), Webber (2015), and Tortarolo and Zarate (2018). For manufacturing sector estimates, see Mertens (2019) and Brooks et al. (2019).

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 realizations. 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.[12]

I use large administrative datasets from France, covering the population of employers and employees between 1995 and 2014. Estimating firm wage premia and firm heterogeneity require detailed information about workers and firms. I estimate the former using matched employer-employee panel data, which includes key information on hourly wages and employer identifiers for over 25 million workers per year. I estimate the latter using firm balance sheet panel data, which contains information on gross production, employment, capital, and intermediate inputs for over 2 million firms per year. The main advantages of these distinct datasets from France are that they are jointly available and are not limited to manufacturing or large firms.[13,14]

With the firm heterogeneity estimates in hand, I then use the structural firm wage premium equation to decompose its empirical distribution. To maximize interpretability, my decomposition allocates each dimension its marginal contribution. I find that wage markdowns, productivity, price-cost markups, and labor elasticities of output contribute 25%, 38%, 13%, and 24%.[15] The last three components form the marginal revenue product of labor, accounting for 75% of the variation. These results indicate that firm characteristics at the center of the labor share debate – price-cost markups and labor elasticities of output – are quantitatively important drivers of the firm wage premium distribution.

Next, I document the dispersion of each dimension of firm heterogeneity. My estimates show that firms differ enormously in their wage markdowns. In France, the firm at the 90^{th} percentile of the wage markdown distribution pays 99% of the marginal revenue product of labor as wages, while the 10^{th} percentile firm pays 52%. Moreover, most firms have considerable influence over the wages they pay – half of the firms pay a wage markdown of less than 70%.

I also find substantial heterogeneity across firms in the labor elasticity of output. The raw 90^{th} percentile labor elasticity of output is 0.61 while the 10^{th} percentile counterpart is 0.20, a difference of 0.41. Consistent with existing sector-level estimates, I find moderate dispersion in sectoral output elasticities (Basu et al., 2013; Oberfield and Raval, 2014). Most of the dispersion occurs within sectors: the average within-sector 90-10 difference is 0.33. Finally, my estimates of large productivity and price-cost markup dispersion across firms are also in line with existing

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

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

[14] 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.

[15] This decomposition is implemented on the R^2.

literature (Syverson, 2011; De Loecker and Eeckhout, 2018).

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

At first glance, this suggests that labor markets are highly competitive. However, my estimates do not support this interpretation for two reasons. First, the previous finding shows that the median firm is able to pay a wage markdown considerably below one. Second, I compute the skill-adjusted marginal revenue product of labor and find sizable dispersion: a 90-10 ratio of 1.89. Instead, my findings suggest that the negative relationships between different firm characteristics offset each other's effect on firm wage premia.

Quantitatively, I find that the negative correlation between firm productivity and the labor elasticity of output is the main explanation for the coexistence of substantial firm heterogeneity and relatively moderate firm wage premium dispersion. At the same time, I find that more productive firms have a higher intermediate input and capital elasticity of output than less productive firms. I interpret this empirical pattern as suggesting that more productive firms are more likely to substitute labor with other factor inputs than less productive firms, perhaps through automation or outsourcing.[16]

Through the lens of the structural framework, one possible reason for why more productive firms are more likely to substitute labor with other inputs is that they 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. Under this interpretation, the ability of more productive firms to substitute labor with other inputs partially offsets their higher labor demand relative to less productive firms and reduces their willingness to pay higher wage premia to compete in hiring workers.

Exceptionally productive "superstar" firms (Autor et al., 2017) play an outsized role in driving between-firm wage inequality (Song et al., 2019) and the aggregate labor share (Kehrig

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

and Vincent, 2018). My firm heterogeneity estimates indicate important differences between superstar firms and other firms. I find that superstar firms tend to operate production technologies that are substantially less labor-intensive (low labor elasticity of output), consistent with the previous finding. Superstar firms also pay significantly lower wage markdowns, and charge disproportionately higher price-cost markups compared to other firms. Therefore, while my estimates are consistent with product market power (De Loecker and Eeckhout, 2018) as a determinant of superstar firms' low labor shares of revenue, they also (i) provide empirical support for the hypothesis that superstar firms have considerable labor market power (Gouin-Bonenfant, 2018), and (ii) offer a new explanation that point to the low labor elasticity of superstar firms' production technologies.

The negative correlation between firm productivity and labor elasticity of output also 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). I find that revenue per worker (or per hour), a common proxy for the marginal revenue product of labor, overstates the variance of the latter by about three times. This mismeasurement overstates the productivity and output gains of removing labor market frictions. The reason is that revenue per worker can be an accurate proxy of the marginal revenue product of labor only under the assumption of a Cobb-Douglas production technology within sectors. This assumption implies that output elasticities are constant across firms and are uncorrelated with firm productivity within a given sector.

Finally, my decomposition also shows that, apart from superstar firms, firms with more product market power are in general not firms with more labor market power: they are negatively correlated. Firms' product and labor market power as distortions to the allocation of factor inputs across firms are often studied separately (Edmond et al., 2018; Berger et al., 2018). However, their cross-sectional relationship has important implications for allocative efficiency. I show that when product and labor market power are positively correlated across firms, they tend to distort the labor demand of the same firms, thereby amplifying each other's negative effects on allocative efficiency. When they are negatively correlated, the opposite is true.

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

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

A large strand of work examines firm productivity and rent-sharing as a driver of wage inequality between firms. There is ample evidence that firms share rents with employees, therefore firm productivity determines the wages they pay (Carlsson et al., 2014; Kline et al., 2017; Bell et al., 2018; Garin and Silverio, 2019). Recent papers study the link between widening firm productivity distribution and rising between-firm wage inequality (Faggio et al., 2007; Berlingieri et al., 2017). My paper contributes by quantifying the extent to which rent-sharing explains firm wage premia.

A growing number of researchers study the effect of labor market power on wages. Dube et al. (2018), Naidu et al. (2016) provide evidence for labor market monopsonistic competition. Recent work by Azar et al. (2018), Benmelech et al. (2018), and Abel et al. (2018) studies the effects of labor market concentration on wages. Berger et al. (2018), Jarosch et al. (2019), Brooks et al. (2019) each provide a structural framework that maps labor market concentration into firm-specific wage markdowns. Gouin-Bonenfant (2018) studies how the firm productivity distribution affects the aggregate labor share through labor market imperfect competition in a wage-posting model. Caldwell and Danieli (2019) and Caldwell and Harmon (2019) study the effects of relative bargaining positions on wages. My paper adds to this literature by (i) documenting the economy-wide distribution of firm-specific wage markdowns (labor market power) that apply to a subset of wage-posting and wage-bargaining models, (ii) quantifying the importance of wage markdowns for firm wage premia, and (iii) showing that superstar firms pay significantly lower wage markdowns than other firms.

Recent work on the labor share of national income focuses on the role of product market power (price-cost markups) and the labor intensity of the aggregate production technology (labor elasticity of output). Elsby et al. (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 (2017) makes the case for growing product market power as a key driver. Autor et al. (2017) and Kehrig and Vincent (2018) show that the falling US labor share is due to the rising market share of "superstar" firms with low labor shares of revenue. De Loecker and Eeckhout (2018) show that superstar firms charge high markups. Much of the debate in this literature abstracts from labor market frictions and thus does not speak to their effects on firm wage premia. By incorporating these margins into a frictional labor market framework, my contribution is to (i) document the distribution of output elasticities with respect to labor across firms within sectors, (ii) quantify their relative importance for the firm wage premium distribution, and (iii) show that superstar firms have low labor shares of revenue not only because they have considerable market power, but also because they use production technologies that are significantly less labor-intensive.

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

2 A Framework to Decompose Firm Wage Premia

I present a wage-posting framework with frictional labor markets from which I derive an equation for the firm-specific wage premium. The framework is a dynamic version of the Manning (2006) generalized model of monopsony augmented with imperfectly competitive product markets and a general production function with capital, labor, and intermediate inputs. The model is set in partial equilibrium. I impose just enough structure on this framework to allow a number of endogeneously determined dimensions firm heterogeneity, but leave unspecified much 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 premia to survive. Firm heterogeneity then determines the wage premium a firm is willing to pay to hire workers of a given skill. This framework features firm heterogeneity in (average revenue labor) productivity ($ARPH$) and wage markdowns (WM), as in standard frictional labor market models such as Burdett and Mortensen (1998). On top of that, firms also differ in price-cost markups (PM) and the labor elasticity of output (LEO). Each concept is made clear later in this section. I show that the framework produces the following equation for the firm-specific wage premium (Φ):

$$\Phi = WM \cdot ARPH \cdot PM^{-1} \cdot LEO \tag{1}$$

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

2.1 Departures from standard frictional labor market models

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

$$P_{jt} = D_s(Y_{jt}, d_{jt}) \tag{2}$$

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. The firm's idiosyncratic demand d_{jt} can be a function of aggregate, sectoral, or firm-specific demand shifters. The assumption of imperfectly competitive goods markets generates a distribution of firm-specific price-cost markups, an important determinant of firms' labor demand in the macroeconomics of resource allocation and the labor share of income (Edmond et al., 2015; Peters, 2017; De Loecker and Eeckhout, 2018).

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

$$Y_{jt} = X_{jt} F_{st}(K_{jt}, H_{jt}, M_{jt}) \tag{3}$$

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.[18] I discuss what these output elasticities and price-cost

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

[18] Moreover, diminishing marginal returns implies that the notion of firms in this framework is based on optimal

markups depend on in the next subsection.

2.2 Deriving the firm-specific wage premium equation

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

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

$$H_{jt} = (1 - s_{jt})H_{jt-1} + R_{jt} \tag{4}$$

with:

$$s_{jt} = s(\Phi_{jt}, a_{jt}) \tag{5}$$

$$R_{jt} = R(\Phi_{jt}, a_{jt}, e_{jt}) \tag{6}$$

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

The assumption that firm-specific separation and recruitment rates depend on the wages offered is informed by models of on-the-job search such as Burdett and Mortensen (1998) and Mortensen (2010), or directed search models such as Kaas and Kircher (2015). I also allow recruitment and separation to depend on non-wage amenities, as there is evidence that non-wage amenities are important determinants of worker flows between firms (Sorkin, 2018). Together, equations (4), (5), and (6) form the firm-specific upward-sloping labor supply curve.

Firms' recruitment efforts are subject to recruitment costs $c(e_{jt})$. I assume that the recruit-

firm sizes. In contrast, firms are a collection of jobs with the same productivity in standard frictional labor market models with linear production functions in labor.

ment cost function is twice differentiable, and that $c_e(.) > 0$ and $c_{ee}(.) > 0$, so that the marginal cost of recruitment effort is increasing in recruitment.

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

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

subject to (2), (3), (4), (5), and (6). Let R_t^K and P_t^m denote the competitive price of capital and intermediate inputs. The timing of events is as follows. First, firms obtain an idiosyncratic draw of productivity and demand. Then firms post wages, exert recruitment effort, and employ workers and other inputs. Finally, firms produce. This timing assumption is consistent with the recent class of multiworker firm models (for example, Kaas and Kircher (2015), Elsby and Michaels (2013), and Schaal (2017)).

Solving for the first-order condition with respect to Φ gives the firm-specific wage premium, equation (1):

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

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

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

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

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

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

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

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

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

which is simply a function of labor supply elasticities.

The specific functional form for labor supply elasticities (ϵ_{jt}^H) depends on the microfoundation for firm-specific labor supply curves (formed by equations (4), (5), and (6)) pursued by the researcher. In Appendix C, I show in a random search and a directed search wage-posting model with on-the-job search that this elasticity depends on the elasticity of the job-filling and separation rates with respect to wages. In a monopsonistic model in which upward-sloping labor supply curves are microfounded by workplace differentiation, I show that the firm-specific labor supply elasticity depends on the firm's labor market share. Finally, in a random search model with wage-bargaining, I show that the labor supply elasticity in the wage markdown replaced by a function of relative bargaining power and workers' value of outside options.

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

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

The first assumption takes a stand on the source of firms' monopsony power in the labor market. As Burdett and Mortensen (1998) show, the combination of search frictions and on-the-job search implies a non-degenerate wage distribution, even when workers and firms are homogenous.[20] The second assumption ensures that the firms' revenue functions exhibit constant marginal returns to labor. This assumption implies that the output elasticity with respect

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

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

to labor is equal to 1 across all firms. The third assumption is standard in traditional monopsony models.[21] The fourth assumption implies that the wage markdown is only a function of the firm-specific labor supply elasticities. Under these assumptions, the firm's profit-maximization problem reduces to:

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

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

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

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

which is the ratio of sales revenue over efficiency units of labor. The firm wage premium equation (1) 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.

Price-cost markup (PM). This component captures firms' price-setting power. It is the ratio of prices over marginal costs. It can be written as:

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

where which ϵ^G_{jt} is the firm-specific price elasticity of demand. The specific functional form for the price elasticity of demand depends on the researcher's microfoundation for the product demand curve (2). For example, with an oligopolistic competition market structure and a nested constant elasticity of substitution (CES) demand system, it depends on the firm's market share of sales (Edmond et al., 2015).[23] Equation (1) shows that, all else equal, firms with higher markups pay a lower wage premium. The intuition is that firms that are able to charge posi-

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

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

[23] Since the demand function in equation (2) is static, the price-setting problem is also static. This is the most common formulation of product demand. However, there are increasingly used dynamic formulations, in which firms' price-setting decisions affect future demand, for example, due to customer accummulation (Gourio and Rudanko, 2014).

tive 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.

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

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

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

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

$$LEO_s = \alpha_s^H$$

The CES production function is:

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

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

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

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

For comparison with the Cobb-Douglas case, when I estimate firm-specific output elasticities in the next section, I assume a translog production function, which can be written as:

$$Y_j = X_j H_j^{\alpha_s^H(H_j,K_j,M_j)} K_j^{\alpha_s^K(H_j,K_j,M_j)} M_j^{\alpha_s^M(H_j,K_j,M_j)}$$

I leave the elasticity of substitution between each pair of inputs unrestricted.

Marginal revenue product of labor (MRPH). The last three components of the firm wage premium equation (1) form the marginal revenue product of labor. This component has two interpretations. In wage-posting models, such as the one presented here, wages are determined *ex ante* forming an employment relationship. Therefore, the firm wage premium reflects a firm's willingness to pay for a worker of a given efficiency and the marginal revenue product of labor reflects the firm's labor demand. In wage-bargaining models, such as the one presented in Appendix C, wages are determined through bargaining over the total match surplus *ex post* matching. Since, all else equal, the total match surplus is larger for high marginal revenue product firms, bargained wages are also higher. Therefore, dispersion of the firm wage premium in a bargaining model due to differences in the marginal revenue product of labor reflects surplus sharing, holding wage markdowns constant across firms.

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

2.3 Discussion

The firm wage premium equation (1) has a few advantages. First, it features firm heterogeneity in dimensions related to the broader literature on between-firm wage inequality (Barth et al., 2016; Song et al., 2019; Faggio et al., 2007; Bell et al., 2018; Berlingieri et al., 2017) and the labor share of income (Autor et al., 2017; Kehrig and Vincent, 2018; De Loecker and Eeckhout, 2018; Berger et al., 2018; Gouin-Bonenfant, 2018; Karabarbounis and Neiman, 2014; Hubmer, 2019) in a transparent way. Second, the log-linear structure substantially simplifies a decomposition of the distribution of firm wage premia without requiring the researcher to fully specify and

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

estimate the underlying primitives of the model, such as the joint distribution of firms' intrinsic productivity and non-wage amenities.

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

Second, implicit in the efficiency units specification of the production function, I assume that worker types are perfect substitutes (within sectors), although the average worker efficiency and firm productivity are complements. This assumption implies that the production function is not log supermodular or submodular in worker and firm productivity, and thus abstracts from worker-firm sorting based on production complementarities (Eeckhout and Kircher, 2011; Bagger and Lentz, 2019).[26] In return, this assumption (i) provides a mapping between the widely-estimated two-way fixed effect regressions (Abowd et al., 1999; Bonhomme et al., 2019) and the structural firm wage premium equation; and (ii) keeps the firm heterogeneity estimation procedure computationally affordable and data requirements feasible. This is because the estimation strategy involves estimating flexible production functions without restrictions on the elasticity of substitution between pairs of factor inputs. Relaxing this assumption by introducing multiple worker types exponentially increases the number of parameters to be estimated and quickly renders the estimation procedure infeasible. To relax this assumption, future work can use the random coefficients approach to estimate firm-specific production functions (Kasahara et al., 2017; Li and Sasaki, 2017) combined with the interacted two-way fixed effects model of Bonhomme et al. (2019) to reduce the dimensionality problem.

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

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

3 Estimating the Structural Firm Wage Premium Equation

3.1 Empirical approach

To use the structural firm wage premium equation to decompose the empirical distribution, I require firm-specific measures of the wage markdown, average revenue product of labor, price-cost markup, and labor elasticity of output, which are unobserved variables.

One approach to estimate each dimension of heterogeneity would be to estimate a fully-specified structural framework. However, this requires the researcher to specify the market structure in each product and labor market. Alternatively, a common approach to measure firm-specific price-cost markups is the cost share approach (Foster et al., 2008). This approach measures firms-specific markups using the firm-specific sales-to-total-cost ratio. 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 and Eeckhout (2018) 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-specific wage premium. Then, once output elasticities obtained, I show that price-cost markups and wage markdowns can be obtained disentangled by exploiting the fact that price-cost markups distort each input demand, while wage markdowns distort only labor demand.

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

The second and third steps extend the production-based approach of De Loecker and Warzynski (2012) and De Loecker and Eeckhout (2018). In the second step, I estimate a production function to obtain firm-specific output elasticities. I posit the following sector-specific translog production function, which is a second-order approximation of any well-behaved production function:[27]

$$
\begin{aligned}
y_{jt} = {} & \beta_{h,s}h_{jt} + \beta_{m,s}m_{jt} + \beta_{k,s}k_{jt} + \beta_{hh,s}h_{jt}^2 + \beta_{mm,s}m_{jt}^2 + \beta_{kk,s}k_{jt}^2 \\
& + \beta_{hm,s}h_{jt}m_{jt} + \beta_{hk,s}h_{jt}k_{jt} + \beta_{mk,s}m_{jt}k_{jt} + \beta_{hmk,s}h_{jt}m_{jt}k_{jt} + x_{jt} + \epsilon_{jt}
\end{aligned}
\tag{8}
$$

where lowercase letters represent the ln counterparts of variables written in uppercase letters. Define ϵ_{jt} as the error term orthogonal to firms' input choice, which can be measurement error.

As Gandhi et al. (2019) show, the control function approach does not generally identify the production function parameters when considering a gross output production function. Therefore, output elasticities and markups cannot generally be separately identified. To address this issue, I follow Flynn et al. (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.[28] This entails the following parameter restrictions:

$$\beta_{hmk,s} = 0$$

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

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

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

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

Firms' input demand is an endogenous choice of the firm and depends on the firm's productivity realization x_{jt}. This is likely to bias the production function parameters upwards. To address this endogeneity issue, I follow a control function approach (Olley and Pakes, 1996). This approach allows the reseacher to "observe" the firms' idiosyncratic productivity by inverting their optimal input demand function for a variable input. The control function is then a function of the variable input and other state variables that I observe in the data. I assume that intermediate inputs are fully flexible and use this to obtain the control function (Levinsohn and Petrin, 2003; Ackerberg et al., 2015).

Using the first-order conditions for intermediate inputs, labor, and capital, I obtain the

[27]This can be thought of as a Cobb-Douglas-like production function in which output elasticities are firm-specific and depend on firm-specific factor intensities and sector-specific pairwise elasticities of substitution between any pair of inputs. In logs: $y_{jt} = x_{jt} + \theta_{h,s}(h_{jt}, m_{jt}, k_{jt})h_{jt} + \theta_{m,s}(h_{jt}, m_{jt}, k_{jt})m_{jt} + \theta_{k,s}(h_{jt}, m_{jt}, k_{jt})k_{jt}$.

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

following optimal intermediate input demand function:

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

where $\mathbf{Z}_{jt}$ is a vector of exogenous firm characteristics that can affect its input demand, which includes location fixed effects, sector fixed effects, and year fixed effects. Since firm-specific input unit prices, especially for intermediate and capital inputs, are unobserved in most existing datasets, my estimation operates under the assumptions that firms are price-takers in intermediate and capital input markets, and firms within a given sector and location face the same input prices.[29] However, because I observe hourly wages at the individual worker level, my datasets enable me to extend the estimation procedure to allow imperfectly competitive labor markets. This extension entails augmenting the control function to include firm-specific wage premia ϕ. This inclusion controls for the fact that in this environment firms have some degree of wage-setting power, which distorts relative input prices, hence, relative input demand.

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

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

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

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

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

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

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

where $g(.)$ is a flexible function and ζ_{jt} is a productivity shock. In step two, I estimate the production function parameters. Combining the control function (9), the predicted output

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

from (10), and the law of motion for productivity (11), I form the following moment conditions:

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

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

$$\begin{aligned}\mathbf{X}_{jt} =[&m_{jt-1} \quad m_{jt-1}^2 \quad h_{jt-1} \quad h_{jt-1}^2 \quad k_{jt} \quad k_{jt}^2 \\ &k_{jt-1}h_{jt-1} \quad k_{jt-1}m_{jt-1} \quad h_{jt-1}m_{jt-1} \quad k_{jt-1}h_{jt-1}m_{jt-1} \quad \phi_{jt-1} \\ &\phi_{jt-1}\mathbf{F}_{jt-1} \quad \mathbf{Z}_{jt}]'\end{aligned}$$

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

A common challenge in the production function estimation literature is that output prices are rarely observed (De Loecker and Goldberg, 2014).[30] 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:

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

where $p_{jt} + x_{jt}$ is the revenue TFP (Foster et al., 2008). The control function is therefore for revenue-TFP rather than quantity-TFP. The potential negative correlation between output prices and input demand could lead to a downward output price bias. The intuition is that, all else equal, firms that set higher prices tend to sell less output, which in turn requires less inputs to produce. It is therefore important to discuss the conditions under which unobserved output prices do not bias estimates of output elasticities.

If firm heterogeneity in prices (markups over marginal costs) is driven by differences in production costs due to productivity x, the firm wage premium ϕ, or regional or sectoral differences in capital or intermediate input prices, these are controlled for in the control function. However, differences in idiosyncratic demand uncorrelated with TFP could still drive markup (hence, price) variation *beyond* what is controlled for by arguments in the control function. Therefore, I additionally include controls for markup heterogeneity. Informed by oligopolistic competition and trade models such as Atkeson and Burstein (2008) and Edmond et al. (2015),

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

I include export status and market shares as additional controls. Informed by models of customer capital (Gourio and Rudanko, 2014), which predict that firms accummulate customers over time, I also include firm age. The lags of these additional controls therefore also appear in the vector $\mathbf{X}$ in the moment conditions of the estimation procedure $E[\zeta_{jt}(\beta)\mathbf{X}_{jt}] = \mathbf{0}$.

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

$$LEO_{jt} = \beta_h + 2\beta_{hh}h_{jt} + \beta_{hm}m_{jt} + \beta_{hk}k_{jt} + \beta_{hmk}m_{jt}k_{jt}$$

$$MEO_{jt} = \beta_m + 2\beta_{mm}m_{jt} + \beta_{hm}h_{jt} + \beta_{mk}k_{jt} + \beta_{hmk}h_{jt}k_{jt}$$

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

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

$$PM_{jt} = MEO_{jt}\frac{P_{jt}Y_{jt}/\exp(\hat{\epsilon}_{jt})}{P_{M,jt}M_{jt}}$$

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

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

$$WM_{jt} = \frac{\Phi_{jt}H_{jt}}{P_t^M M_{jt}} \cdot \frac{MEO_{jt}}{LEO_{jt}} = \frac{\bar{W}_{jt}L_{jt}}{P_t^M M_{jt}} \cdot \frac{MEO_{jt}}{LEO_{jt}}$$

Since the price-cost markup is a common input distortion, it cancels out and therefore does not feature in this equation.

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

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

3.2 Estimating firm wage premia

I estimate the firm-specific wage premium by estimating firm(-group) fixed effects following Bonhomme et al. (2019) (BLM henceforth):[31]

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

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

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 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.

Before implementing the regression, firms are grouped into clusters using a weighted k-means clustering algorithm. 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.[32] This choice is motivated by the following tradeoff: having a coarse classification of firms into fewer groups leads to many more workers who switch between firms per firm-

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

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

group, which substantially improves the precision of firm wage premium estimates. However, this comes at the cost of potentially averaging away considerable amounts of multidimensional firm heterogeneity within firm-groups. In practice, I apply the clustering algorithm by 2-digit sectors for the following intervals of time: 1995-1998, 1999-2002, 2003-2005, 2006-2008, 2009-2011, 2012-2014. The time intervals are chosen to keep the number of observations similar across estimation samples. This produces an average of 4,035 firm-groups for an average of 273,031 firms per year.

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. If the conditional exogenous mobility assumption is a reasonable approximation, then one should observe systematic worker mobility up and down the firm effect quartiles. Moreover, workers should experience approximately symmetric wage changes as they move along the firm effect quartiles, given the log additive regression specification. On the other hand, in structural models of worker-firm sorting based on comparative advantage (Eeckhout and Kircher, 2011; Lopes de Melo, 2018), worker mobility is based on the match-specific component of wages, which is captured by the residual component of wages in the AKM regression. In this class of models the AKM regression is misspecified in the sense that the wage gains depend on value of the particular worker-firm match, for example, if highly skilled workers have a comparative advantage in high productivity firms. In Appendix B, the event study Figure 1 compares the changes in mean log wages for workers who move between quartiles of firm fixed effects, following Card et al. (2013). Figure 1 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.

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 on the worker's skill type, log additivity would be violated, and one should observe that the estimated residuals systematically varies across worker-firm fixed effect deciles. In Appendix B, Figure 3 and 4 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.

4 Data Description

4.1 Administrative datasets from France

Estimating the structural firm wage premium equation using the approach described above requires two types of datasets. The empirical distribution of the firm wage premium is estimated with matched employer-employee datasets, which follow workers over time and employment spells at different firms. The four dimensions of firm heterogeneity in the model are estimated with firm balance sheet panel datasets. While both types of datasets have become increasingly accessible, they are typically not jointly available.[33] To the extent that firm balance sheet datasets are available, most cover only a set of large firms or manufacturing firms, or do not contain a panel structure.[34] I therefore use matched employer-employee and firm balance sheet panel data from France covering the population of firms and workers in the private sector between 1995 and 2014.

My sources for firm balance sheet information are the *Fichier de comptabilité unifié dans SUSE* (FICUS) and *Fichier approché des résultats d'Esane* (FARE) datasets, jointly available from 1995 to 2014. FICUS and FARE are compiled by the fiscal authority of France, *Direction Générale des Finances Publiques* (DGFiP), from compulsory filings of firms' annual accounting information. These datasets contain balance sheet information for all firms in France without restriction on the size of firms. There are over 2 million firms per year. From these datasets, I obtain information on variables such as sales, nominal value of production, employment, intermediate input and capital expenditure.

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

The first DADS dataset is the DADS-Panel, which provides information on all employed workers in the private sector born in October in a panel structure (only October-workers born

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

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

in even years are observed prior to 2002). Because workers are followed over time and their employer identifiers are observed, I use this dataset to run the AKM-BLM regressions described in the previous section to estimate firm wage premia.

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

4.2 Sample selection

I restrict firm level observations from the FICUS-FARE balance sheet data to several broad industries: agriculture, construction, manufacturing, financial services, non-financial services, transportation, and wholesale and retail. Education and utilities are excluded. I include only firms with at least 5 employees. I harmonize all 2-digit and 4-digit industry codes to the latest available version (Nomenclature d'activités Française – NAF rév. 2). I drop 2-digit sectors with less than 500 observations within each 7-year interval (1995-2000, 2001-2007, 2008-2014). This is important when estimating production functions, especially flexible specifications such as the translog, as this procedure would be demanding on small sample sizes, and could lead to imprecise estimates of the production function parameters. In practice, few two-digit sectors have less than 500 observations in this time interval. I also drop firms within each 7-year interval that only appear once since estimating production functions requires at least two consecutive years of data.

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

Firm wage premia (firm fixed effects) in the AKM-BLM regression are only identified for the sets of firms connected by worker mobility. I therefore focus on the largest connected set of firms. In practice, due to the clustering of firms into firm-groups using the DADS-Postes, my analysis pertains to the largest connected set of firm-groups, of which very few firms are not a part. This group consists of 174,305,521 people-year observations, an average of 8,715,276 per year. After clustering firms into groups, I link the DADS-Postes and DADS-Panel via the firm identifier (SIREN) to allocate each firm-year observation in the panel data a firm-group identifier and construct the estimation sample for firm wage premia. I implement the AKM-BLM regression on this sample.

After estimating firm wage premia, I collapse the dataset to the firm level and link it to the FICUS-FARE firm balance sheet data to construct the estimation sample for each dimension of firm heterogeneity. I implement the production function estimation routine on this sample. There are 5,884,663 firm-year observations in total and an average of 294,233 firms per year in this sample. Summary statistics for worker and firm characteristics are reported in Table 6.

5 Firm Heterogeneity and the Firm-Specific Wage Premium

5.1 Substantial firm heterogeneity in each dimension

This section documents the empirical moments of each of the four dimensions of firm heterogeneity. I start by discussing novel estimates of the dispersed wage markdown and labor elasticity of output distributions. I then confirm the well-documented existence of large productivity dispersion and the more recently documented price-cost markup dispersion across firms (Syverson, 2004; De Loecker and Eeckhout, 2018). Table 1 summarizes the empirical moments of each estimated dimension of firm heterogeneity in 2014.[35] Table 7 in Appendix B shows moments related to the within-sector distribution of each dimension. Tables 8, 9, 10, and 13 in Appendix B report the variances of each dimension of firm heterogeneity by broad industries.

The wage markdown has received increasing attention as a potentially important driver of wage inequality (Azar et al., 2018; Schubert et al., 2019; Caldwell and Danieli, 2019) and the distribution of labor shares (Berger et al., 2018; Gouin-Bonenfant, 2018; Jarosch et al., 2019; Brooks et al., 2019). Despite its theoretical relevance, its empirical distribution is not well-documented. The first row of Table 1 describes the distribution of wage markdowns (WM). I find substantial dispersion of wage markdowns across firms. My estimates show that firms at the 90^{th} percentile of the wage markdown distribution pay a wage markdown of 0.99, ap-

[35] In Tables 14, 15, 16, 17, and 18 in Appendix B, I show that each dimension of firm heterogeneity is correlated with observed firm characteristics, including firm size, age, productivity, and market concentration.

proximately the level of marginal revenue product of labor that firms would pay in a perfectly competitive labor market. At the 10^{th} percentile, workers obtain approximately half of their marginal revenue productivity (0.52). Figure 6 plots the kernel density of wage markdowns. As discussed in section 2, this large dispersion of the wage markdown reflects differences in labor supply elasticities in a wage-posting model, or bargaining power and outside options in a wage-bargaining model. In either model, it could also reflect differences across firms in the future value of a worker, if the employment relationship remains intact.

Firm Heterogeneity	Mean	Median	Variance	90th Pct	10th Pct
Wage markdown	0.74	0.69	0.05	0.99	0.52
Inverted wage markdown	1.47	1.45	0.16	1.94	1.01
Price markup	1.27	1.18	0.12	1.60	1.02
Inverted price markup	0.82	0.85	0.02	0.98	0.51
Labor elasticity of output	0.41	0.41	0.02	0.61	0.20
Intermediate elasticity of output	0.53	0.53	0.02	0.72	0.33
Capital elasticity of output	0.06	0.06	0.00	0.10	0.02
Average revenue product of labor (log)	4.17	4.13	0.20	4.76	3.63
Marginal revenue product of labor (log)	2.96	2.98	0.07	3.28	2.64
Number of firms	294,233				

Table 1: Summary statistics of firm heterogeneity in 2014.

I also find that most firms possess significant wage-setting ability – half of the firms in my sample pay less than 0.70 of the marginal revenue product of labor as wages. This suggests that there is ample room for wage increases at the typical firm. One way to assess firms' wage-setting ability is to compare it to firms' price-setting ability. I do so by inverting the wage markdown and then comparing it to the price-cost markup. Figure 7 plots the distribution of inverted wage markdowns against price-cost markups. Table 1 shows that the inverted wage markdown is 27 percentage points higher than price-cost markups at the median firm (1.45 compared to 1.18). Inverted wage markdowns are also more dispersed than price-cost markups.

Since there is little systematic documentation of the distribution of wage markdowns, it is not straightforward to compare my estimates with existing work. One way to do so is to assume that my wage markdown estimates are generated by a static wage-posting model. As discussed in section 2, wage markdowns in this case are entirely determined by labor supply elasticities. I consider a Burdett-Mortensen model, in which the wage markdown is $\frac{\epsilon^H}{1+\epsilon^H}$, and back out the implied labor supply elasticities.[36] This gives firm-specific labor supply elasticities of 0.44, 1.89, 5.98 at the 10^{th}, 50^{th}, and 90^{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 0.26, 0.85, 2.13, at the 10^{th}, 50^{th}, and 90^{th} percentiles. Berger et al. (2018) find firm-specific labor supply elasticities driven by differences in market shares in an oligopsonistic model between 0.76 and 3.74 in the US. Relative to Webber (2015) and Berger et al. (2018), my wage markdown estimates for France imply, on average, a higher labor supply elasticity and more dispersion than the US.

The labor elasticity of output is a central part of the debate about the causes of the U.S. aggregate labor share decline. Existing estimates for labor elasticities of output are usually at an aggregated level, for example, at the sector level or for the entire macroeconomy (Basu et al., 2013; Karabarbounis and Neiman, 2014; Oberfield and Raval, 2014). My estimates for firm-specific labor elasticities of output reported in the fifth row of Table 1 display substantial heterogeneity across firms, particularly within sectors. The 90^{th} percentile labor elasticity of output is 0.61, while it is 0.20 at the 10^{th} percentile, a 90-10 difference of 0.41. 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 90-10 difference slightly to 0.33.[37,38] These findings suggest that labor elasticities of output are potentially important determinants of the distribution of firm wage premia and labor shares within sectors. This is explored further in section 5.5.[39]

I now confirm that, consistent with existing findings, price-cost markups and firm productivity are highly dispersed across firms. The third row of Table 1 reports the summary statistics for price-cost markups. The median markup is 1.18. This is in the ballpark of existing estimates. De Loecker and Warzynski (2012) estimate markups using Slovenian manufacturing firm data and find median markups between 1.10 and 1.28. De Loecker and Eeckhout (2018) use Compustat data and find median markups in the US in 2014 of about 1.20. Edmond et al. (2018) use Compustat data and find a median markup of 1.4 in 2012 following the methodology of De Loecker and Eeckhout (2018), and a median markup of 1.12 using a calibrated structural

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

[37] The reason that the average labor elasticity of output is lower than commonly used calibration targets of around 0.7 is because I estimate these from a gross output production function rather than value-added production function. The implicit assumption underlying the latter is that the production function is Leontief in intermediate inputs.

[38] As detailed in section 3, production functions are estimated at the 2-digit level. Removing differences across 4-digit sectors reduces the 90-10 difference to 0.31.

[39] As discussed in section 2, the labor elasticity of output depends on the elasticity of substitution between each pair of inputs and the factor intensities (input ratios). Elsby et al. (2013) proposes offshoring as an important driver of the US labor share decline (intermediate-input-labor substitution), while Karabarbounis and Neiman (2014) argue for the role of capital-labor substitution in aggregate. My estimates of significant dispersion of labor elasticities of output point to differential offshoring, outsourcing, or automation patterns across firms as potentially important for explaining firm level labor shares.

model with heterogeneous markups. De Loecker and Eeckhout (2018) find markups at the 90^{th} percentile between 1.9 and 2.3 in 2014 in the US, which is higher than my estimates for France of 1.60 in 2014. Edmond et al. (2018) report markups at the 90^{th} percentile between 1.24-1.69 in 2012. While the markups at the 10^{th} percentile are not reported, Edmond et al. (2018) report an interquartile range for markups of $1.69 - 0.97 = 0.73$ using the the methodology of De Loecker and Eeckhout (2018), and an interquartile range of $1.19 - 1.10 = 0.09$ using their structural model. My estimates for the interquartile range is signficantly smaller, $1.32 - 1.09 = 0.23$.

The second-to-last row of Table 1 reports the distributional statistics for the average revenue product of labor ($ARPH$) in logs. The dispersion of firm productivity is well-documented (Foster et al., 2008; Syverson, 2011) and a key feature of models of heterogeneous firms (Melitz, 2003). I find that the average revenue product of labor (in efficiency units) has a 90-10 ratio of $\exp(4.76 - 3.63) = 3.09$. Most of the dispersion in productivity occurs within sectors, consistent with existing work (Syverson, 2011). Table 7 shows that the average 90-10 ratio within two-digit sectors is $\exp(4.69 - 3.70) = 2.70$.[40]

5.2 Each dimension matters for the firm wage premium distribution

Having shown that each dimension of firm heterogeneity in equation (1) is dispersed, this section quantifies their relative importance for the empirical firm wage premium distribution. I show that the additional dimensions of heterogeneity introduced into an otherwise standard frictional labor market framework – price-cost markups (PM) and labor elasticities of output (LEO) – account for sizable shares of the firm wage premium distribution.

Recall that the structural firm wage premium equation is a log-linear function of the four dimensions of heterogeneity. Taking logs on equation (1) gives:

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

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

To maximize interpretability, my preferred decomposition method is a Shapley decomposition (Shorrocks, 2013).[41,42] I implement this decomposition by running equation (12) as a linear

[40] Syverson (2004) shows that the average 90-10 ratio of TFP within four-digit manufacturing sectors is 1.92. I find an average 90-10 ratio for the average revenue product of labor in efficiency units across four-digit manufacturing sectors in France of 2.57.

[41] Appendix A discusses the Shapley decomposition in detail.

[42] An alternative decomposition method is the ensemble decomposition (Sorkin, 2018). It can be written as: $1 = \frac{CV(\ln WM,\phi)}{V(\phi)} + \frac{CV(-\ln PM,\phi)}{V(\phi)} + \frac{CV(\ln LEO,\phi)}{V(\phi)} + \frac{CV(\ln ARPH,\phi)}{V(\phi)}$. I show this in Table 22 in Appendix B.

regression and then decomposing the R^2 into four components. Each component represents the marginal contribution of a dimension of firm heterogeneity to the cross-sectional firm wage premium variation. Relative to a standard variance decomposition, the Shapley decomposition is easier to interpret because (i) it is more parsimonious; (ii) the marginal contributions take values between 0 and 1, and they sum up to the R^2, which is equal to 1. In section 5.4, where I study the importance of the relationships between firm characteristics, I present results from the standard variance decomposition. Table 2 presents the Shapley decomposition results.

Firm heterogeneity	Marginal contribution to the R^2
Wage markdown (wm)	0.25
Marginal revenue product of labor ($mrph$)	0.75
Average revenue product of labor ($arph$)	0.38
Price markup (pm)	0.13
Labor elasticity of output (leo)	0.24
R^2	1
Number of firms	294,233

Table 2: Shapley decomposition of the firm wage premium distribution in 2014.

The first row of Table 2 shows that wage markdowns account for 25% of the firm wage premium distribution. Theoretically, wage markdowns can differ across firms for a number of reasons. As discussed in section 2, wage markdowns in the dynamic frictional wage-posting framework are a function of the firm-specific labor supply elasticity and the discounted expected marginal profits of keeping the worker next period. In static monopsonistic and oligopsonistic wage-posting models, wage markdowns depend on the firm-specific labor market shares of employment or wage bill (Boal and Ransom, 1997; Berger et al., 2018; Azar et al., 2019).[43] The structural framework in section 2 also nests the workhorse Burdett and Mortensen (1998) model of frictional labor markets. In its simplest form, in which workers and firms are homogenous, a well-known prediction of the Burdett-Mortensen model is that the wage distribution is non-degenerate. This is also known as "frictional wage dispersion" (Hornstein et al., 2011) and it shows up in the form of heterogeneous wage markdowns.[44]

Alternatively, as shown in Appendix C, wage markdown dispersion in wage-bargaining models of the labor market reflect heterogeneity in workers' share of the match surplus (relative bargaining power), outside options (captured by reservation wages), and the discounted expected

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

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

marginal profits of retaining the worker. Estimating heterogeneous outside options and their effect on wages and the labor share of income is a growing literature (Caldwell and Danieli, 2019; Caldwell and Harmon, 2019; Schubert et al., 2019; Jarosch et al., 2019).

Commonly used models of frictional labor markets often feature heterogeneous firm productivity in the form of the average revenue product of labor (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Mortensen, 2010; Bagger et al., 2014; Elsby and Michaels, 2013; Kaas and Kircher, 2015; Engbom and Moser, 2018; Gouin-Bonenfant, 2018; Lamadon et al., 2019). In these models, firm productivity determines the extent of wage premium a firm pays relative to other firms for an identically skilled worker, with more productive firms paying workers of a given skill higher wages. It is well-known that firm productivity is highly dispersed (Syverson, 2004), that changes in productivity dispersion are correlated with changes in between-firm wage dispersion (Faggio et al., 2007; Berlingieri et al., 2017), and that firm level productivity shocks pass through to wages even conditional on worker ability (Card et al., 2018; Kline et al., 2017). The third row of Table 2 shows that heterogeneous average revenue productivity of labor accounts for 38% of the firm wage premium distribution. This result implies that a substantial share of the variation is due to the two additional dimensions of heterogeneity – price-cost markups and labor elasticities of output.

Heterogeneity in price-cost markups often does not feature in frictional labor market models. However, this is a theoretically and quantitatively important determinant of labor demand and firm size in macroeconomic models of the labor share and resource misallocation (Edmond et al., 2015; Peters, 2017). There is increasing evidence that markups vary significantly across firms (De Loecker and Eeckhout, 2018; Edmond et al., 2018). At the same time , these models do not speak to the distribution of firm wage premia due to the assumption of perfectly competitive labor markets. As discussed in section 2, price-cost markups affect firm wage premia through firms' labor demand in my structural framework. The fourth row of Table 2 shows that price-cost markups account for 13% of the firm wage premium distribution.

The fifth row of Table 2 quantifies the importance of heterogeneous labor elasticities of output for the firm wage premium distribution. The labor elasticity of output is a key component of labor demand in macroeconomic models of the labor share (Karabarbounis and Neiman, 2014; Oberfield and Raval, 2014; Hubmer, 2019), but often does not feature in frictional labor market models. I find that heterogeneous labor elasticities of output account for 24% of the firm wage premium distribution.

Finally, the second row of Table 2 shows that the marginal revenue product of labor in efficiency units (MRPH), which is equal to the sum of the last three components of equation (12), accounts for three quarters of the cross-sectional variation of firm wage premia. Since firm wage premia are often estimated regression objects, this result provides a quantitative structural interpretation. In wage-posting models, the contributions of this component can be thought of as reflecting differences in firms' labor demand, and hence, firms' willingness to pay for a given worker. This is because in wage-posting models wages are determined before commencing an employment relationship. Alternatively, in wage-bargaining models, wages are decided ex-ante through a bargaining process. In this case, the contribution of the MRPH can be interpreted as arising from surplus sharing.

5.3 Each dimension is significantly more dispersed than firm wage premia

While each dimension of firm heterogeneity in equation (1) is highly dispersed and accounts for important shares of the firm wage premium distribution, this section shows that they do not translate into a highly *dispersed* firm wage premium distribution. This is despite the finding that the typical firm has significant wage-setting ability in section 5.1, with half of the firms paying less than 70% of the marginal revenue product of labor as wages. The main message of this section is that the modest firm wage premium distribution masks substantial underlying firm heterogeneity in each dimension, and their interactions offset their effects on the firm wage premium distribution. The next section explores each pair of interactions in detail.

Table 3 reports statistics about the dispersion of firm wage premia in 2014. The variance of firm wage premia (ϕ) is modest (0.008), accounting for 4.5% of the wage distribution. At the same time, the dimensions of firm heterogeneity are orders of magnitude more dispersed than firm wage premia. As the diagonals of table 4 show, the variances of the logs of the wage markdown (wm), average revenue product of labor ($arph$), price-cost markups (pm), and labor elasticity of output (leo) are 0.073, 0.242, 0.039, 0.259.

	Firm-Specific Wage Premium (ϕ)
Variance	0.008
Fraction of Total Variance	4.5%
90-10 ratio	1.25
90-50 ratio	1.13
50-10 ratio	1.11
Number of firms	294,233
Number of firm-groups	4,017
Number of workers	8,715,276

Table 3: Dispersion of firm wage premia in 2014.

Nevertheless, the dispersion of firm wage premia is a quantitatively important deviation from the law of one wage. Table 3 shows that a firm at the 90th percentile of the firm wage premium distribution pays a given worker a wage that is on average 25% more than a firm at the 10th percentile of the distribution. This gap is almost twice as large as the average gender wage gap in OECD countries and it is comparable to the gender wage gap in Japan, among the highest in OECD countries.[45]

At first glance, these results suggest that the labor market is highly competitive. In a perfectly competitive labor market, the marginal revenue product of labor is constant across firms, and workers obtain the full amount of the latter. However, my findings in the previous sub-section do not support this interpretation. First, as table 1 shows, the marginal revenue product of labor in efficiency units is highly dispersed, with a 90-10 ratio of 1.89. Second, the median firm is able to pay a wage markdown of less than 0.70.

Rather, these results suggest that the relationships between dimensions of firm heterogeneity in equation (12) offset each other's impact on the firm wage premium distribution. To show how important these negative relationships are, I run the following exercise. I start by restricting the structural framework to the workhorse Burdett-Mortensen model, which features only wage markdown and productivity heterogeneity. I do so by re-imposing the assumptions that product markets are perfectly competitive and production functions are linear in labor inputs. The firm wage premium equation then becomes:

$$\phi_{jt} = wm_{jt} + arph_{jt} \tag{13}$$

This equation shows that under these assumptions, the marginal revenue product of labor ($mrph$) is equal to, and can be measured as, the average revenue product of labor ($arph$). Next, I take my estimates for wage markdowns and productivity as given and feed these data into equation (13). I find an implied variance of firm wage premia (0.233) that is much larger than observed in the data (0.008).[46] This implies that correlations between dimensions of firm heterogeneity in workhorse frictional labor market models and those emphasized in the labor share literature compress the firm wage premium distribution.

5.4 Negative correlations between dimensions of firm heterogeneity compress the firm wage premium distribution

I now discuss which pair of interactions between the dimensions of firm heterogeneity compress the firm wage premium distribution. The main finding in this section is the following: firm

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

[46] An alternative exercise allows price-cost markups and labor elasticities of output to vary across sectors but not within sectors. Doing so, I find a predicted firm wage premium variance of 0.186.

productivity and labor elasticity of output are strongly negatively correlated. More productive firms tend to use less labor intensive production technologies by substituting labor with other inputs. This negative relationship implies that more productive firms generally have a less-than-proportionately higher labor demand compared to a less productive firms, therefore they pay a less-than-proportionately higher wage premium.

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

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

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

I start by examining the cross-sectional relationship 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, 2018; Berger et al., 2018). The intuition is that since firms with high $mrph$ (labor demand) pay higher wages, they face a locally less elastic labor supply curve, reflecting less labor market competition locally. Therefore, high $mrph$ firms have less incentives to pay a high fraction of $mrph$ as wages. Wage-bargaining models that allow outside options or bargaining power to vary by $mrph$ also share this prediction (Postel-Vinay and Robin, 2002; Jarosch et al., 2019). This prediction finds support in the last row of the first column in Tables 4 and 5. The covariance between wm and $mrph$ of -0.07 is large relative to most other covariance terms, and the correlation is -0.95. Therefore, given the distribution of marginal revenue productivity of labor across firms, wage markdowns are quantitatively important mechanisms that compress the distribution of firm wage premia.

However, under the assumptions that product markets are perfectly competitive and production technologies are linear in labor in standard frictional labor market models, the marginal revenue product of labor is equal to the average revenue product of labor ($mrph = arph$).[47] This implies that the correlation and covariance between wm and $arph$ is the same as that between wm and $mrph$. Table 5 shows that these correlations are far from the same. In particular, the covariance between wm and $arph$ (-0.02) is considerably weaker than the covariance between wm and $mrph$ (-0.07). Unpacking the latter, the first column of tables 4 and 5 shows that

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

the covariance between wm and inverted price-cost markups $-pm$ (-0.02), and the covariance between wm and labor elasticities of output leo (-0.03) matter. The negative correlation between wage markdowns and inverted price-cost markups 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. The negative correlation between wage markdowns and labor elasticities of output suggests that firms that use labor intensive production technologies tend to have stronger wage-setting power. One rationale for this result 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 (Arnoud, 2018).

The fourth row in the third column of Tables 4 and 5 show that firms that charge higher markups (pm) tend to have higher labor elasticities of output (leo). While a higher leo raises the firm's labor demand, higher pm offsets it. A potential explanation for this correlation is that higher quality goods fetch higher markups due to a lower price elasticity of demand (Coibion et al., 2007; Manova and Zhang, 2012; Atkin et al., 2015) and are more labor intensive to produce (Jaimovich et al., 2019).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.073				
$arph$	-0.022	0.242			
$-pm$	-0.019	0.004	0.039		
leo	-0.030	-0.211	-0.024	0.259	
$mrph$	-0.071	0.035	0.018	0.024	0.077

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

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.165	1			
$-pm$	-0.354	0.036	1		
leo	-0.221	-0.841	-0.243	1	
$mrph$	-0.951	0.254	0.332	0.171	1

Table 5: Firm heterogeneity correlation matrix in 2014.

The second rows, second columns of Tables 4 and 5 show that the relationship between firm productivity and price-cost markups is relatively weak and somewhat negative. This is consistent with the estimates of De Loecker and Eeckhout (2018), who find a negative correlation between firm size (sales) and markups in the cross-section of firms and sectors. However, within sectors there is evidence that price-cost markups and firm productivity are positively correlated. Consistent with theory (Edmond et al., 2015), Tables 23 and 30 in Appendix B show that more productive firms ($arph$) tend to charge higher markups (pm) in both manufacturing and non-financial services. Overall, these results imply that price-cost markups are not the main compressors of the firm wage premium distribution.

The third rows of the second column of Tables 4 and 5 present the main finding of this section – more productive firms ($arph$) have a lower labor elasticity of output (leo). This negative relationship is the most quantitatively important among the set of interaction terms that offset the effects of firm heterogeneity on the firm wage premium distribution ($CV(leo, arph) = -0.21$, $\text{corr}(leo, arph) = -0.84$). At the same time, the correlation between firm productivity and the intermediate input elasticity of output is large and positive (0.69), and the correlation between firm productivity and the capital elasticity of output is also positive but weaker (0.19). This result shows that more productive firms substitute away from labor towards capital and in particular, intermediate inputs. In this process, they use production technologies that are less elastic with respect to labor inputs.

While higher productivity raises the firm's labor demand, this input substitution mechanism works in the opposite direction. The presence of this mechanism implies that firm productivity ($arph$) is more dispersed than labor demand ($mrph$), therefore, more productive firms do not pay proportionately higher wage premia. The intuition, through the lens of the structural framework of section 2, is the following. Since firms face upward-sloping labor supply curves due to labor market frictions, firms that want to hire more workers must offer higher wages. Because more productive firms wish to grow larger than less productive firms, the former face a higher cost of labor relative to other inputs. If labor and other inputs are imperfect substitutes, more productive firms substitute labor with other inputs to avoid higher relative costs of employing labor. In this case, the labor elasticity of output is decreasing in the firm's input intensity of other inputs, reducing the firm's labor demand and offered wage premium.

This finding speaks to the role of production technologies in the determination of the labor share of national income. Earlier studies focus on the role of aggregate changes in production technology through capital-labor substitution (Karabarbounis and Neiman, 2014) or intermediate-input-labor substitution (Elsby et al., 2013). On the other hand, recent studies emphasize the importance of firm level labor shares, showing that the US labor share is entirely driven by a reallocation of sales from high to low labor share firms (Autor et al., 2017; Kehrig and Vincent, 2018). My estimates show that production technologies are important determinants not only of firm wage premia, but also firm-level labor shares. In the next section, I discuss how superstar firms, which are large and have low labor shares, differ from the rest.

5.5 Implications

Superstar firms versus the rest. The decline of the U.S. aggregate labor share of income has attracted significant academic attention. Earlier studies make the case for changes in the aggregate production technology, either through capital-labor substitution (Karabarbounis and Neiman, 2014; Oberfield and Raval, 2014) or intermediate-input-labor substitution (Elsby et al., 2013). However, recent research shows that the U.S. labor share decline is explained by the reallocation of sales towards highly productive "superstar" firms, which have low labor shares (Autor et al., 2017; Kehrig and Vincent, 2018). Hypotheses based on changes in the aggregate production technology do not account for this pattern.

What are the key differences between superstar firms and other firms? My estimates of firm heterogeneity allow me to assess these differences. I categorize firms by size (sales revenue) into ten equal-sized groups within each two-digit sector. Then, following Autor et al. (2017), I define superstar firms as the four largest firms in terms of sales in each sector.[48]

Figures 13, 14, 15, and 16 confirm that large superstar firms are more productive, pay higher average wages, pay a higher firm wage premium, and have a lower labor share of revenue than other firms. Consistent with De Loecker and Eeckhout (2018), Figure 17 shows that superstar firms charge significantly higher price-cost markups. However, Figure 18 shows that superstar firms also tend to pay lower wage markdowns, which provides empirical support for the hypothesis that superstar firms reduce aggregate labor shares because they have more labor market power (Gouin-Bonenfant, 2018). On top of that, Figure 19 provides a novel explanation for superstar firms' low labor shares. It shows that they operate production technologies with low labor elasticity of output. Therefore, while the aggregate production technology cannot account for the fact that superstar firms' low labor shares are the main drivers of the U.S. labor share decline, my findings point to superstar firms' production technologies as potentially important drivers.

Measuring labor misallocation. The marginal revenue product of labor ($MRPH$) is a key component of wage determination in structural models of imperfectly competitive labor markets (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Card et al., 2018; Berger et al., 2018). Its variance is also a key statistic for the misallocation of labor inputs across firms (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). The firm wage premium equation (1) shows that the marginal revenue product of labor consists of the average revenue product of labor ($ARPH$), price-cost markups (PM), and labor elasticity of output (LEO). If PM and LEO are constant across firms (a common assumption), then the $MRPH$ is proportional

[48] Autor et al. (2017) also define superstar firms as the top 20 firm in terms of sales by sector. This alternative definition does not affect my results.

to, and can be measured by, the $ARPH$. However, the finding in section 5.4 that $ARPH$ and LEO are strongly negatively correlated implies that the $MRPH$ is considerably less dispersed than the $ARPH$. Indeed, the variance of (log) $ARPH$ is almost three times larger than the variance of (log) $MRPH$, $\frac{CV(arph)}{CV(mrph)} = 2.86$. Figure 11 plots the de-meaned $mrph$ and $arph$. This mismeasurement stems from the fact that the when the elasticities of substitution between labor and other inputs are different from one, the $arph$ and leo are correlated.[49] When this elasticity is greater than one, leo declines as more productive firms substitute labor with other inputs to circumvent labor market frictions, which require firms to pay higher wages to hire more workers. This result suggests that counterfactual output and productivity gains from eliminating labor market frictions are overstated when the $mrph$ is measured as the $arph$.

Role of market power in labor misallocation. The relationship between wage markdowns and price-cost markups have important implications for the allocative efficiency of labor inputs. Market power in the product (Edmond et al., 2015; Peters, 2017) and labor markets (Berger et al., 2018) have been separately shown to distort the allocation of labor across firms. However, whether they amplify or dampen each other's effects on labor misallocation depends on their cross-sectional correlation. Since both product and labor market power reduce firm size below the perfect competition benchmark, these distortions amplify each other's effects when they are positively correlated, as they tend to distort labor demand of the same firms. When they are negatively correlated, the opposite is true.

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

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

where $\gamma_s^a > 0$, $\gamma_s^b > 0$, and $\gamma_s^c \in [0, 1]$ are constants. The higher the inverted wage markdown (WM^{-1}), the stronger the firm's labor market power. This equation shows that if product and labor market power are perfectly negatively correlated ($PM_j \cdot WM_j^{-\gamma_s^c} =$ constant $\forall j \in s$), not only are there are no TFP gains to equalizing market power distortions within sector s, but policies that generate dispersion in the joint market power component $PM_j \cdot WM_j^{-\gamma_s^c}$ lead to input misallocation and TFP losses. As Table 5 shows, the correlation between price-cost markups and inverted wage markdowns is -0.35, 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).

[49] Section 2 provides more detail.

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.

Income inequality. There is a negative relationship between two types of income inequality – between-firm wage inequality and the aggregate labor share. The former is a measure of income inequality between employees working at different firms, while the latter is a measure of the income distribution between workers and firm owners. There is evidence that more productive firms pay higher wages, contributing to between-firm wage inequality (Berlingieri et al., 2017). At the same time, evidence suggests that more productive firms have lower labor shares, and that the reallocation of sales to low labor share firms drives the decline of the US aggregate labor share (Autor et al., 2017; Kehrig and Vincent, 2018). These two measures of income inequality are linked in the following sense:

$$\text{Labor share}_j = \frac{\Phi_j H_j}{P_j Y_j} = \frac{\Phi_j}{ARPH_j} = WM_j \cdot PM_j^{-1} \cdot LEO_j$$

where j denotes the firm. This equation shows that (i) the relationship between the firm wage premium and the average revenue product of labor (productivity) determines firms' labor shares; (ii) the more strongly positively related firm wage premia and firm productivity are, the higher the labor share of more productive firms (hence a higher aggregate labor share), but all else equal this also comes with higher between-firm wage dispersion; (iii) the strength of this relationship is determined by the wage markdown, price-cost markup, and labor elasticity of output. This implies that policies that aim to reduce the labor market power of low labor share firms, such as codetermination, could potentially raise the aggregate labor share while also leading to higher between-firm wage inequality.[50] To assess the impact of this policy and others, such as minimum wages and size-dependent taxes, on these two measures of income inequality, future work would require a fully-specified model suitable for policy simulations.

Inferring bargaining power. The covariance of wages and the marginal revenue product of labor at the firm-level helps identify workers' relative bargaining power in the structural estimation of frictional labor market models (Cahuc et al., 2006; Bagger et al., 2014). This bargaining parameter is important, for example, because it determines the relative importance

[50] Codetermination refers to the participation of employees in corporate management decisions.

of firm productivity and outside options in wage-setting (Caldwell and Harmon, 2019). As discussed above, while the conceptually informative moment is $CV(\phi, mrph)$, the $mrph$ is often measured as the $arph$ under the assumptions stated above. From equation (12), the slope between the firm wage premium ϕ and the $arph$ can be written as:

$$\frac{CV(\phi, arph)}{V(arph)} = 1 + \frac{CV(wm, arph)}{V(arph)} + \frac{CV(-pm, arph)}{V(arph)} + \frac{CV(leo, arph)}{V(arph)}$$

Because the average revenue product of labor and labor elasticities of output are negatively correlated ($CV(leo, arph) < 0$), using $CV(\phi, arph)$ instead of $CV(\phi, mrph)$ as a moment to match potentially understates workers' relative bargaining power.

6 Concluding Remarks

I investigate how firm characteristics determine the wage premium a firm pays relative to other firms for identically skilled workers. To do so, I develop and implement a novel structural decomposition of the firm wage premium distribution. While a large literature emphasizes the importance of firm productivity and wage-setting power in a frictional labor market, this paper highlights the role of firms' product market power and the labor share of production. My decomposition also uncovers important correlations between these firm characteristics. First, there is a negative relationship between firm productivity and the labor share of production. Second, product and labor market power are negatively correlated in the cross-section.

These findings have important implications. First, my findings show that exceptionally productive superstar firms are different from other firms in several ways (Autor et al., 2017). I confirm that superstar firms charge disproportionately higher price-cost markups (De Loecker and Eeckhout, 2018), but also provide empirical support for the hypothesis that these firms pay markedly stronger labor market power (Gouin-Bonenfant, 2018), and offer a new explanation for their low labor shares of revenue: low labor share of production. Second, the negative relationship between firm productivity and the labor share of production implies that conventional measures of the variance of the marginal revenue product of labor, a sufficient statistic for labor misallocation (Hsieh and Klenow, 2009), overstates the degree of labor misallocation across firms. Third, while the effects of product and labor market power on input misallocation are often studied separately (Edmond et al., 2018; Berger et al., 2018), their cross-sectional relationship decides whether they amplify or dampen each other's effects on misallocation.

The structural decomposition framework also has a number of potential applications. One application could be to study the extent to which the long-term wage loss from losing the firm wage premium for outsourced (Goldschmidt and Schmieder, 2017) or displaced workers (Heining et al., 2018; Lachowska et al., 2018) is due to the loss of bargaining power or to moving to a less productive firm. Other applications could be to use this framework to understand the rising dispersion of the firm wage premium in countries such as Germany (Card et al., 2013), or to decompose the fall in the US aggregate labor share into contributions of each dimension of firm heterogeneity (Autor et al., 2017; Kehrig and Vincent, 2018).

Bibliography

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

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

Acemoglu, D. and P. Restrepo (2017). Robots and jobs. *working paper*.

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

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

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

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

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

Atkin, D., A. Chaudhry, S. Chaudry, A. Khandelwal, and E. Verhoogen (2015). Markup and cost dispersion across firms: Direct evidence from producer surveys in pakistan. *American Economic Review: Papers and Proceedings 105*(5), 537–544.

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2017). The fall of the labor share and the rise of superstar firms. *working paper No. 23396, National Bureau of Economic Research 98*(1), 394–425.

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

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

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

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

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

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

Barkai, S. (2017). Declining labor and capital shares. *Stigler Center working paper*.

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. *working paper*.

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

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

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

Berger, D., K. Herkenhoff, and S. Mongey (2018). Labor market concentration and welfare. slideshow.

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

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

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

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

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

Cahuc, P., F. Postel-Vinay, and J. Robin (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica 74*(2), 323–364.

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

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

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

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

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

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

Coibion, O., L. Einav, and J. Hallak (2007). Equilibrium demand elasticities across quality segments. *International Journal of Industrial Organization 25*(1), 13–30.

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

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

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

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

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

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

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

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

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

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

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

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

Elsby, M., R. Michaels, and M. Ratner (2018). Vacancy chains. *working paper 1*.

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

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

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

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

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

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

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

Gouin-Bonenfant, E. (2018). Productivity dispersion, between-firm competition, and the labor share. *working paper*.

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

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

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

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

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

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

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

Jaimovich, N., S. Rebelo, and A. Wong (2019). Trading down and the business cycle. *Journal of Monetary Economics 102*, 96–121.

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

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

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

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

Kasahara, H., P. Schrimpf, and M. Suzuki (2017). Identification and estimation of production function with unobserved heterogeneity. *working paper*.

Kehrig, M. and V. Vincent (2018). The micro-level anatomy of the labor share decline. *working paper*.

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

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

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

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

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

Li, T. and Y. Sasaki (2017). Constructive identification of heterogeneous elasticities in the cobb-douglas production function. *working paper*.

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

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

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

Manova, K. and Z. Zhang (2012). Export prices across firms and destinations. *Quarterly Journal of Economics 127*(1), 379–436.

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

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

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

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

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

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

Oberfield, E. and M. Raval (2014). Micro data and macro technology. *working paper*.

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

Peters, M. (2017). Heterogeneous markups, growth and endogenous misallocation. *working paper*.

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

Wright, R., P. Kircher, B. Julien, and V. Guerrieri (2018). Directed search: A guided tour. *working paper 1*.

Appendix A:

6.1 Decomposition Method: Shapley

Recall the firm wage premium equation (1), written in natural logs:

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

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

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

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

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

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

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

Appendix B: Figures and Tables

Summary Statistics

	Summary Statistics: Employees	
Sample size		
People-years	174,305,521	
Firm-years	5,884,663	
Average number of workers per year	8,715,276	
Average number of firms per year	294,233	
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 6: Summary statistics: Employees and employers (1995-2014).

Conditional Exogenous Mobility & Symmetry

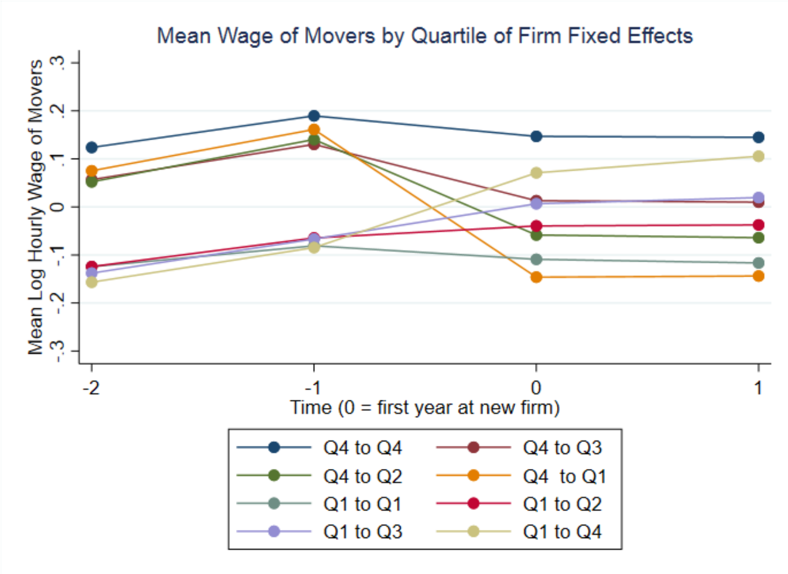

Figure 1: Worker mobility and wage changes by quartiles of firm effects (2009-2014).

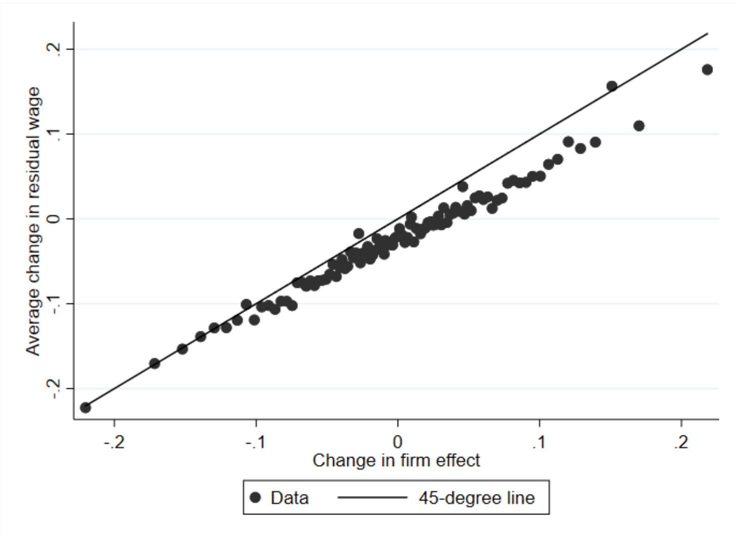

Figure 2: Average wage changes from worker mobility by declines of changes in firm premia (2009-2014).

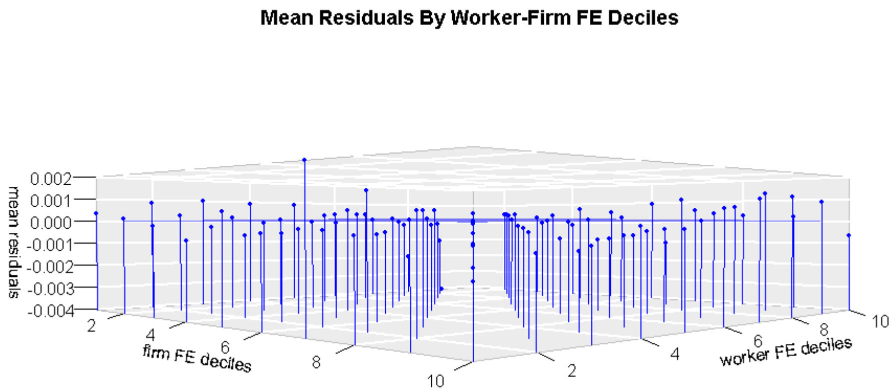

Figure 3: Mean estimated residuals by worker-firm deciles (2014)

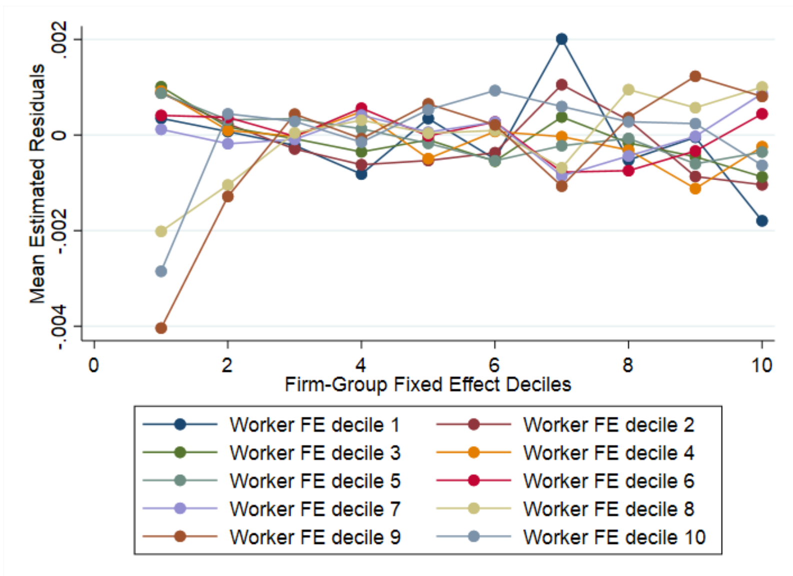

Figure 4: Mean estimated residuals by worker-firm deciles (2014)

Distributions of Firm Heterogeneity

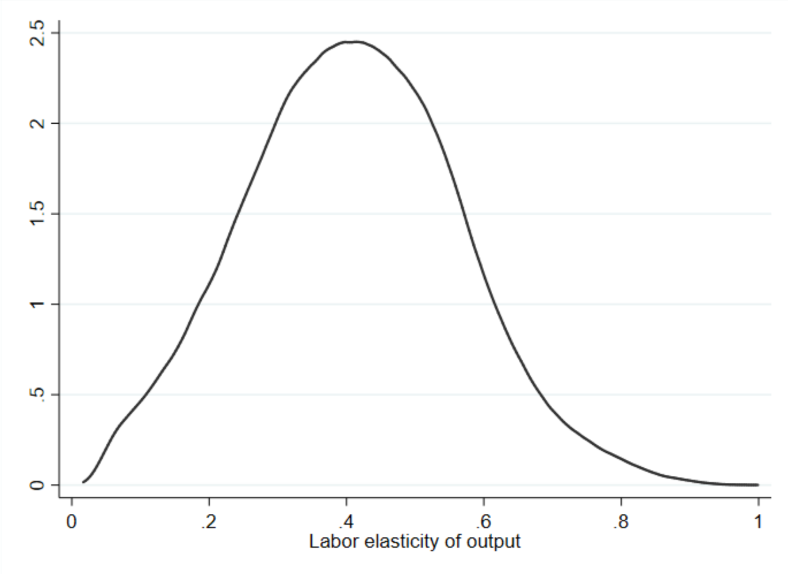

Figure 5: Distribution of labor elasticities of output, 2014.

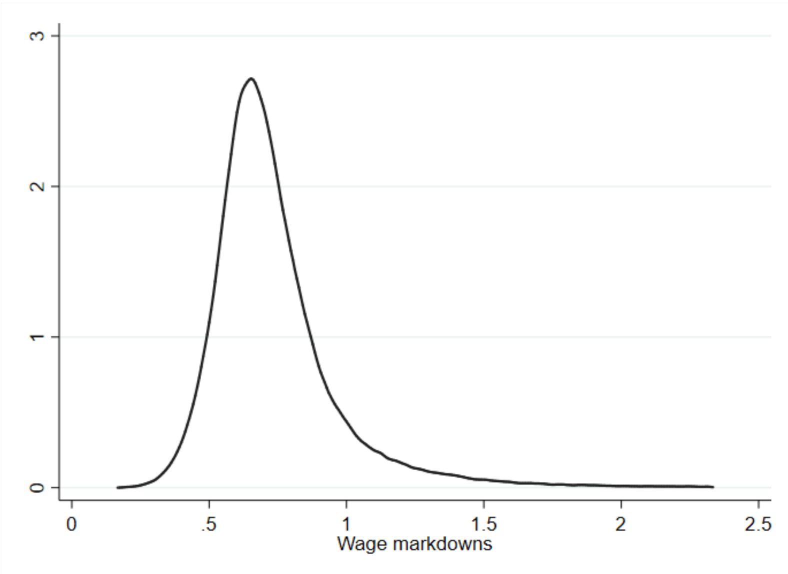

Figure 6: Distribution of wage markdowns, 2014.

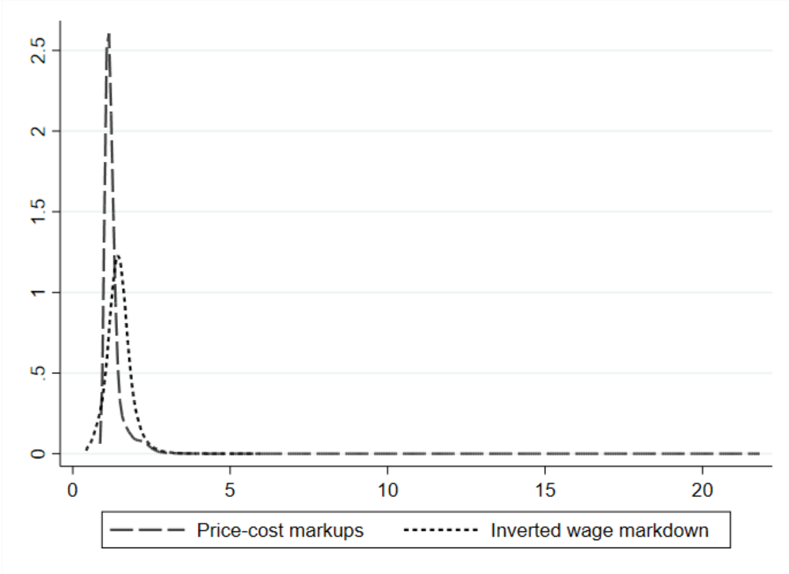

Figure 7: Distribution of price-cost markups and (inverse) wage markdowns, 2014.

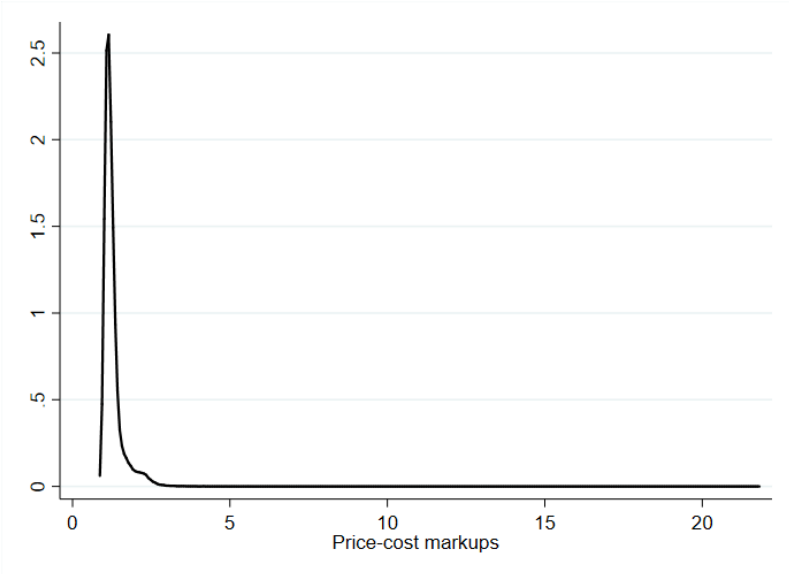

Figure 8: Distribution of price-cost markups, 2014.

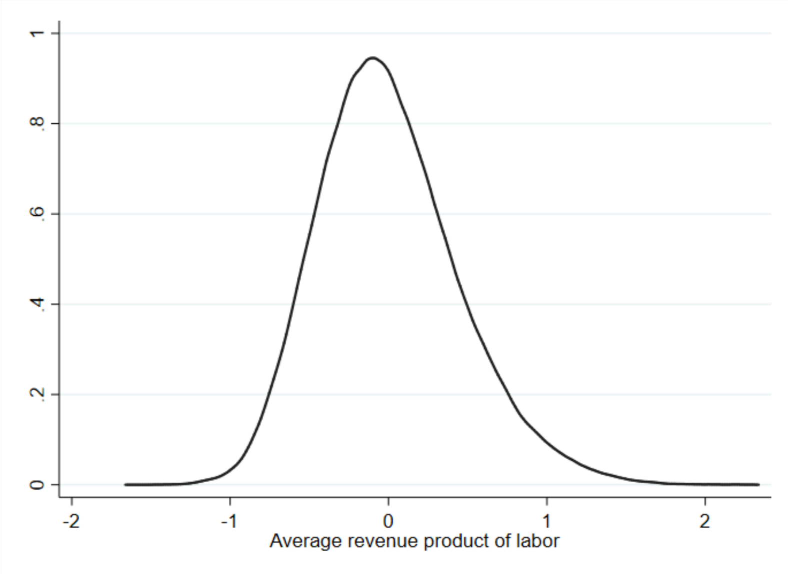

Figure 9: Distribution of the average revenue product of labor, 2014.

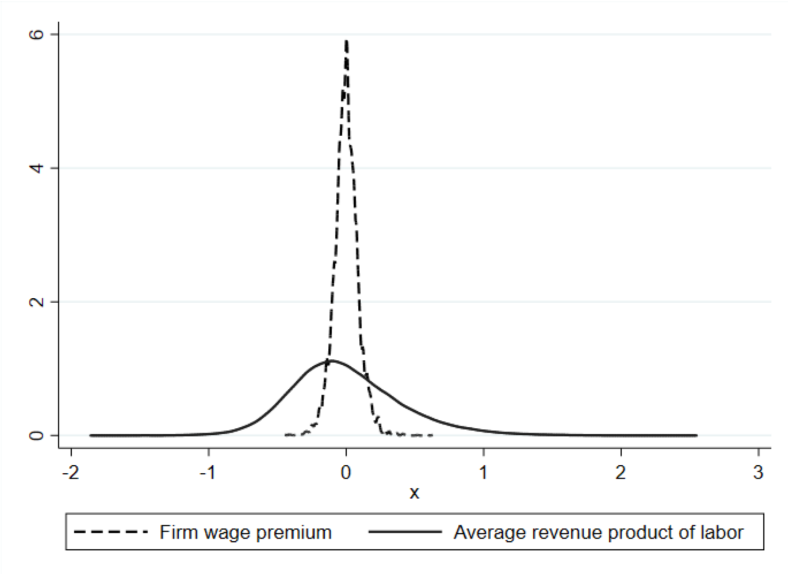

Figure 10: Distribution of firm wage premia and the average revenue product of labor, 2014.

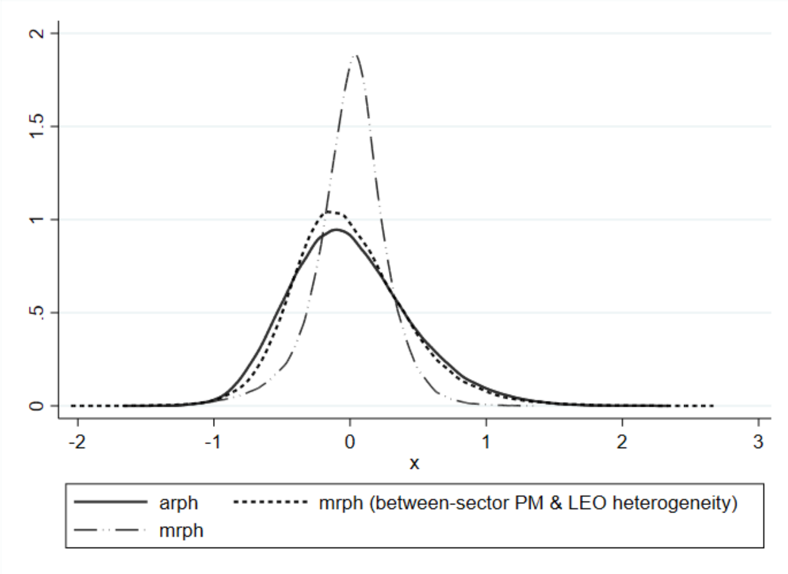

Figure 11: Distribution of the average and marginal revenue product of labor, 2014.
The solid line denotes the average revenue product of labor, the short dashed line denotes the marginal revenue product of labor when price-cost markups and labor elasticities of output are sector-specific but not firm-specific, and the long dashed line denotes the margianl revenue product of labor when all dimensions are firm-specific. Each variable is de-meaned.

Firm Heterogeneity	Mean	Median	Variance	90th Pct	10th Pct
Wage markdown	0.74	0.71	0.04	0.95	0.53
Inverted wage markdown	1.47	1.45	0.13	1.88	1.06
Price markup	1.27	1.23	0.10	1.53	1.01
Inverted price markup	0.82	0.83	0.02	0.97	0.67
Labor elasticity of output	0.41	0.41	0.02	0.57	0.24
Intermediate elasticity of output	0.53	0.53	0.02	0.68	0.37
Capital elasticity of output	0.06	0.06	0.00	0.09	0.03
ln Average revenue product of labor	4.17	4.13	0.16	4.69	3.70
Number of firms	292,157				

Table 7: Summary statistics of firm heterogeneity within 2-digit sectors in 2014. The overall mean of each dimension is kept constant.

Wage markdowns	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.74	0.69	0.99	0.82	0.60	0.52	0.05
Agriculture	0.77	0.75	0.97	0.84	0.66	0.57	0.04
Construction	0.83	0.76	1.16	0.92	0.65	0.57	0.08
Manufacturing	0.76	0.71	0.99	0.82	0.63	0.56	0.05
Financial	0.72	0.69	1.06	0.87	0.53	0.38	0.08
Non-Financial	0.68	0.66	0.87	0.75	0.58	0.51	0.03
Transportation	0.70	0.64	0.98	0.73	0.58	0.52	0.06
Wholesale-Retail	0.74	0.69	1.06	0.84	0.57	0.48	0.06

Table 8: Distribution of estimated wage markdowns by sector in 2014.

Price-cost markups	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	1.27	1.18	1.60	1.32	1.09	1.02	0.12
Agriculture	1.32	1.23	1.70	1.34	1.15	1.06	0.15
Construction	1.11	1.10	1.22	1.16	1.05	1.00	0.01
Manufacturing	1.21	1.13	1.45	1.23	1.06	1.01	0.10
Financial	1.50	1.46	1.84	1.66	1.28	1.10	0.27
Non-Financial	1.33	1.25	1.70	1.40	1.14	1.05	0.13
Transportation	1.12	1.06	1.31	1.14	1.00	0.97	0.12
Wholesale-Retail	1.33	1.22	2.01	1.35	1.12	1.14	0.14

Table 9: Distribution of estimated price-cost markups by sector in 2014.

Labor elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.41	0.41	0.61	0.51	0.30	0.20	0.02
Agriculture	0.34	0.34	0.52	0.43	0.26	0.17	0.02
Construction	0.33	0.33	0.52	0.43	0.25	0.17	0.02
Manufacturing	0.34	0.33	0.52	0.43	0.24	0.15	0.02
Financial	0.47	0.48	0.56	0.52	0.44	0.40	0.00
Non-Financial	0.45	0.46	0.68	0.59	0.34	0.24	0.03
Transportation	0.35	0.35	0.50	0.41	0.26	0.18	0.02
Wholesale-Retail	0.43	0.44	0.58	0.51	0.36	0.28	0.01

Table 10: Distribution of estimated labor elasticities of output by sector in 2014.

Intermediate input elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.52	0.51	0.70	0.61	0.42	0.34	0.02
Agriculture	0.59	0.59	0.72	0.62	0.45	0.36	0.02
Construction	0.58	0.59	0.73	0.66	0.50	0.41	0.02
Manufacturing	0.59	0.60	0.79	0.71	0.52	0.43	0.02
Financial	0.49	0.47	0.53	0.49	0.42	0.39	0.00
Non-Financial	0.49	0.48	0.67	0.58	0.35	0.26	0.02
Transportation	0.58	0.58	0.74	0.66	0.50	0.42	0.02
Wholesale-Retail	0.52	0.51	0.63	0.55	0.41	0.35	0.01

Table 11: Distribution of estimated intermediate input elasticities of output by sector in 2014.

Capital elasticity of output	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	0.06	0.06	0.10	0.08	0.04	0.02	0.00
Agriculture	0.07	0.07	0.09	0.08	0.05	0.03	0.00
Construction	0.09	0.08	0.10	0.08	0.05	0.04	0.00
Manufacturing	0.07	0.07	0.11	0.08	0.04	0.03	0.00
Financial	0.04	0.05	0.09	0.07	0.02	0.01	0.00
Non-Financial	0.06	0.06	0.08	0.07	0.04	0.02	0.00
Transportation	0.07	0.07	0.11	0.09	0.05	0.03	0.00
Wholesale-Retail	0.05	0.05	0.08	0.07	0.03	0.01	0.00

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

Average revenue product of labor	Mean	Median	90th	75th	25th	10th	Variance
Aggregate	4.17	4.13	4.76	4.44	3.85	3.63	0.20
Agriculture	4.27	4.19	4.90	4.50	3.97	3.80	0.19
Construction	4.26	4.24	4.76	4.51	4.00	3.78	0.14
Manufacturing	4.30	4.27	4.89	4.59	3.99	3.76	0.20
Financial	4.20	4.15	4.77	4.44	3.92	3.72	0.18
Non-Financial	4.14	4.11	4.78	4.45	3.78	3.55	0.23
Transportation	4.20	4.15	4.80	4.43	3.90	3.65	0.22
Wholesale-Retail	4.05	4.00	4.57	4.27	3.78	3.60	0.16

Table 13: Distribution of log average revenue product of labor in efficiency units by sector in 2014.

Correlations Between Dimensions of Firm Heterogeneity and Observed Firm Characteristics

This section verifies the correlation between measured price-cost markups, wage markdowns, and labor elasticities of output, and observed firm characteristics. To do so, I run the following regression:

$$Y_{jt} = \gamma X_{jt} + Sector_{s(j,t)} \times Location_{l(j,t)} \times Year_t + \epsilon_{jt}$$

where $Y_{jt} = \{pm_{jt}, wm_{jt}, leo_{jt}\}$, X_{jt} is an observed firm characteristics, which is either log employment size, log sales, log revenue total factor productivity, or firm age (relative to firms below age 10). The regression includes interacted fixed effects by 2-digit sector, location (French *département*), and year.

		pm	
ln Employment	-0.054***		
	(0.006)		
ln TFP		0.169***	
		(0.079)	
Firm age			0.010***
			(0.001)
Sector-*département*-year F.E.	√	√	√
Adj. R^2	0.27	0.32	0.31
Total # of firm-years	5,884,663	5,884,663	5,884,663

Table 14: Correlation of goods markups with observed firm characteristics. The fourth to last row displays the average number of observations per year, and the third to last row shows the total number of observations over 1994-2014. The second to last row displays the average number of firms per year, and the last row shows the total number of distinct firms over 1994-2014. Standard errors are clustered at the sector-region-year level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

	wm		
ln Employment	-0.058***		
	(0.015)		
ln TFP		-0.152***	
		(0.058)	
Firm age			0.009
			(0.008)
Sector-*département*-year F.E.	√	√	√
Adj. R^2	0.25	0.24	0.24
Total # of firm-years	5,884,663	5,884,663	5,884,663

Table 15: Correlation of wage markdowns with observed firm characteristics. The fourth to last row displays the average number of observations per year, and the third to last row shows the total number of observations over 1995-2014. The second to last row displays the average number of firms per year, and the last row shows the total number of distinct firms over 1995-2014. Standard errors are clustered at the sector-region-year level.***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

	leo		
ln Employment	0.019*		
	(0.010)		
ln TFP		-0.127**	
		(0.055)	
Firm age			0.021***
			(0.006)
Sector-*département*-year F.E.	√	√	√
Adj. R^2	0.26	0.28	0.25
Total # of firm-years	5,884,663	5,884,663	5,884,663

Table 16: Correlation of labor intensity with observed firm characteristics. The fourth to last row displays the average number of observations per year, and the third to last row shows the total number of observations over 1994-2014. The second to last row displays the average number of firms per year, and the last row shows the total number of distinct firms over 1994-2014. Standard errors are clustered at the sector-region-year level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

I now show that a measure of market concentration, the Herfindahl-Hirschman index (HHI), is correlated with market-level averages of price-cost markups and wage markdowns. To do so, I run the following regression:

$$Y_{s,l,t} = \gamma \ln \mathrm{HHI}_{s,l,t} + Sector_s \times Year_t + Location_l \times Year_t + \epsilon_{jt}$$

where $Y_{s,l,t} = \{pm_{s,l,t}, wm_{s,l,t}\}$ are the natural logs of the average price-cost markups and wage markdowns with a given sector-location-year cell, and $\ln HHI_{,l,t}$ is the HHI in natural logs. The HHI indices are computed as follows: $HHI_{s,l,t} = \sum_{jt} \mathrm{SHARE}^2_{j,s(j,t),l(j,t),t}$, where $\mathrm{SHARE}_{j,s(j,t),l(j,t),t}$ denotes firm j's market share at time t. In the case of product markets, the HHIs are computed using firms' market shares of sales. In the case of labor markets, the HHIs are computed using firms' market shares of employment.

I define a product market as a sector-*département*, assuming that firms within a given 4-digit sector compete directly within a given *département*, but not across. I chose this definition to allow for imperfect tradability of goods and services across locations. There are 99 *département*s and 579 sectors in my sample. An alternative is to view a sector as a product market, this assumes that firms within a sector competes directly with each other, regardless of their location. Another alternative is to view a sector-*commune* pair as a product market. There are about 36,000 communes in France. Therefore, doing so would lead to a high measure of concentration within product markets. I define a labor market as a sector-commuting-zone pair. This is consistent with the definition of labor markets in some recent work, such as Benmelech et al. (2018) and Rinz (2018). There are 510 commuting zones in my sample.

	pm
ln HHI	0.006***
	(0.000)
Sector-year FE	√
Département-year FE	√
Adj. R^2	0.782
Total # of observations	555,927

Table 17: Correlation of goods markups with the Herfindahl-Hirschman index. Robust standard errors are reported in parentheses. Sample from 1995 to 2014. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

	wm
ln HHI	-0.024***
	(0.002)
Sector-year FE	$\surd$
Commuting-zone-year FE	$\surd$
Adj. R^2	0.451
Total # of market-years	1,308,343

Table 18: Correlation of wage markdowns with the Herfindahl-Hirschman index. Robust standard errors are reported in parentheses. Sample from 1995 to 2014. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

Shapley Decomposition Results

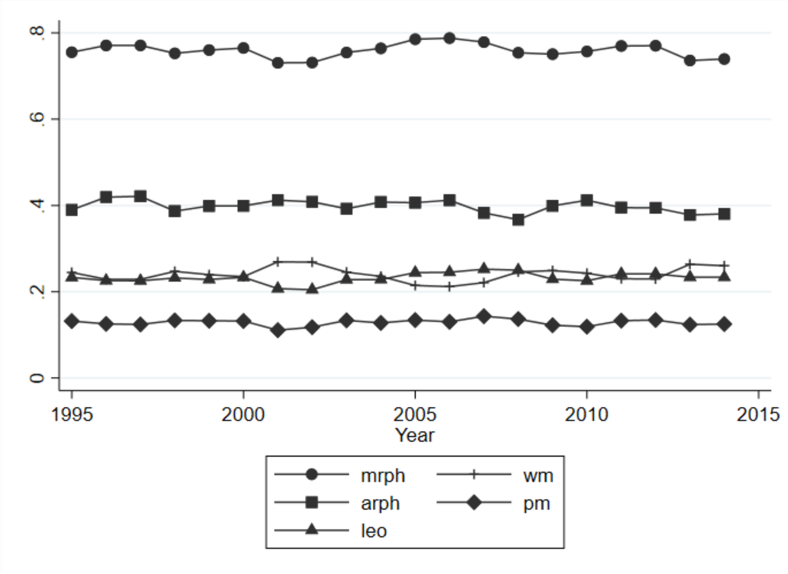

Figure 12: Shapley decomposition of the firm wage premium over time
Included sectors: Agriculture, Construction, Manufacturing, Financial Services, Non-Financial Services, Transportation, and Wholesale and Retail.

	Aggregate	Agriculture	Construction	Manufacturing
WM	0.25	0.26	0.14	0.25
MRPH	0.75	0.74	0.86	0.75
PM	0.13	0.08	0.09	0.10
LEO	0.24	0.26	0.17	0.22
ARPH	0.38	0.40	0.60	0.43
R^2	1	1	1	1
Number of firms	292,157	14,836	47,965	34,422

	Financial	Non-Financial	Transportation	Wholesale & Retail
WM	0.33	0.19	0.17	0.25
MRPH	0.67	0.81	0.83	0.75
PM	0.12	0.13	0.08	0.14
LEO	0.14	0.31	0.23	0.18
ARPH	0.41	0.37	0.52	0.43
R^2	1	1	1	1
Number of firms	2,030	96,726	15,366	80,812

Table 20: Shapley decomposition of the firm wage premium distribution by sectors in 2014.

Ensemble Decomposition Results

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

	Aggregate	Agriculture	Construction	Manufacturing
$CV(\ln WM, \phi)/V(\phi)$	0.24	-0.07	0.31	0.14
$CV(\ln MRPH, \phi)/V(\phi)$	0.76	1.07	0.69	0.86
$CV(-\ln PM, \phi)/V(\phi)$	-0.09	0.11	-0.05	-0.25
$CV(\ln LEO, \phi)/V(\phi)$	-0.84	-2.56	-0.40	0.26
$CV(\ln ARPH, \phi)/V(\phi)$	1.69	3.51	1.14	1.37
R^2	1	1	1	1
Number of firms	292,157	14,836	47,965	34,422

	Financial	Non-Financial	Transportation	Wholesale & Retail
$CV(\ln WM, \phi)/V(\phi)$	0.05	-0.17	0.23	0.83
$CV(\ln MRPH, \phi)/V(\phi)$	0.95	1.17	0.77	0.17
$CV(-\ln PM, \phi)/V(\phi)$	-0.08	0.05	-0.56	-0.33
$CV(\ln LEO, \phi)/V(\phi)$	-0.26	-0.49	-0.51	-0.57
$CV(\ln ARPH, \phi)/V(\phi)$	1.29	1.60	1.84	1.07
R^2	1	1	1	1
Number of firms	2,030	96,726	15,366	80,812

Table 22: Ensemble decomposition of the firm wage premium distribution by sectors in 2014.

Standard Variance Decomposition Results

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

	wm	*arph*	$-pm$	*leo*	*mrph*
wm	0.101				
arph	0.008	0.319			
$-pm$	0.001	-0.002	0.040		
leo	-0.110	-0.307	-0.038	0.442	
mrph	-0.101	0.010	-0.000	0.097	0.107

Table 23: Firm heterogeneity variance-covariance matrix in 2014 (agriculture).

	wm	*arph*	$-pm$	*leo*	*mrph*
wm	0.094				
arph	0.010	0.135			
$-pm$	-0.019	0.002	0.009		
leo	-0.083	-0.137	0.007	0.209	
mrph	-0.091	-0.001	0.019	0.079	0.097

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

	wm	*arph*	$-pm$	*leo*	*mrph*
wm	0.077				
arph	0.001	0.178			
$-pm$	-0.004	-0.017	0.036		
leo	-0.074	-0.153	-0.017	0.241	
mrph	-0.076	0.008	0.002	0.072	0.082

Table 25: Firm heterogeneity variance-covariance matrix in 2014 (manufacturing).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.195				
$arph$	-0.188	0.247			
$-pm$	-0.078	0.026	0.081		
leo	0.071	-0.074	-0.030	0.031	
$mrph$	-0.194	0.199	0.077	-0.073	0.203

Table 26: Firm heterogeneity variance-covariance matrix in 2014 (financial services).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.057				
$arph$	-0.037	0.260			
$-pm$	-0.024	-0.006	0.045		
leo	0.003	-0.205	-0.014	0.213	
$mrph$	-0.058	0.048	0.025	-0.006	0.067

Table 27: Firm heterogeneity variance-covariance matrix in 2014 (non-financial services).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.065				
$arph$	-0.033	0.226			
$-pm$	-0.018	-0.015	0.039		
leo	-0.012	-0.163	-0.011	0.181	
$mrph$	-0.063	0.049	0.014	0.007	0.069

Table 28: Firm heterogeneity variance-covariance matrix in 2014 (transportation).

	wm	$arph$	$-pm$	leo	$mrph$
wm	0.075				
$arph$	-0.047	0.163			
$-pm$	-0.028	0.020	0.031		
leo	0.005	-0.128	-0.024	0.143	
$mrph$	-0.069	0.054	0.026	-0.009	0.071

Table 29: Firm heterogeneity variance-covariance matrix in 2014 (wholesale and retail).

Cross-Sectional Correlations

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	0.046	1			
$-pm$	0.016	-0.015	1		
leo	-0.522	-0.816	-0.290	1	
$mrph$	-0.975	0.055	-0.001	0.446	1

Table 30: Firm heterogeneity correlation matrix in 2014 (agriculture).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	0.092	1			
$-pm$	-0.642	0.053	1		
leo	-0.590	-0.818	0.169	1	
$mrph$	-0.956	-0.008	0.618	0.557	1

Table 31: Firm heterogeneity correlation matrix in 2014 (construction).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	0.011	1			
$-pm$	-0.077	-0.214	1		
leo	-0.539	-0.738	-0.179	1	
$mrph$	-0.958	0.066	0.044	0.510	1

Table 32: Firm heterogeneity correlation matrix in 2014 (manufacturing).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.857	1			
$-pm$	-0.618	0.185	1		
leo	0.919	-0.848	-0.608	1	
$mrph$	-0.978	0.890	0.601	-0.931	1

Table 33: Firm heterogeneity correlation matrix in 2014 (financial services).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.304	1			
$-pm$	-0.475	-0.059	1		
leo	0.026	-0.871	-0.147	1	
$mrph$	-0.946	0.368	0.447	-0.053	1

Table 34: Firm heterogeneity correlation matrix in 2014 (non-financial services).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.274	1			
$-pm$	-0.365	-0.155	1		
leo	-0.107	-0.805	-0.128	1	
$mrph$	-0.939	0.388	0.262	0.066	1

Table 35: Firm heterogeneity correlation matrix in 2014 (transportation).

	wm	$arph$	$-pm$	leo	$mrph$
wm	1				
$arph$	-0.422	1			
$-pm$	-0.585	0.276	1		
leo	0.053	-0.840	-0.369	1	
$mrph$	-0.953	0.506	0.554	-0.094	1

Table 36: Firm heterogeneity correlation matrix in 2014 (wholesale and retail).

Deriving Effects of Market Power and Misallocation on Sectoral TFP

The representative final goods firm aggregates sector-specific output using a Cobb-Douglas aggregator $Y = \Pi_s Y_s^{\theta_s}$, where s denotes the sector, and θ_s denotes the sector-specific shares which sum to 1. The sector-specific CES aggregator 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 (μ_{sj}) over marginal costs:

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

where λ_{sj} denotes the inverted wage markdowns. As section 5.3 shows, the variance of firm wage premia is modest. I therefore assume that $\Phi_j \approx \Phi \; \forall j$. The firm-specific revenue TFP can then be written as:

$$TFPR_{sj} = P_{sj} X_{sj} \propto \mu_{sj} \cdot \lambda_{sj}^{\alpha_s^H}$$

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:

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

Superstar Firms vs Other Firms

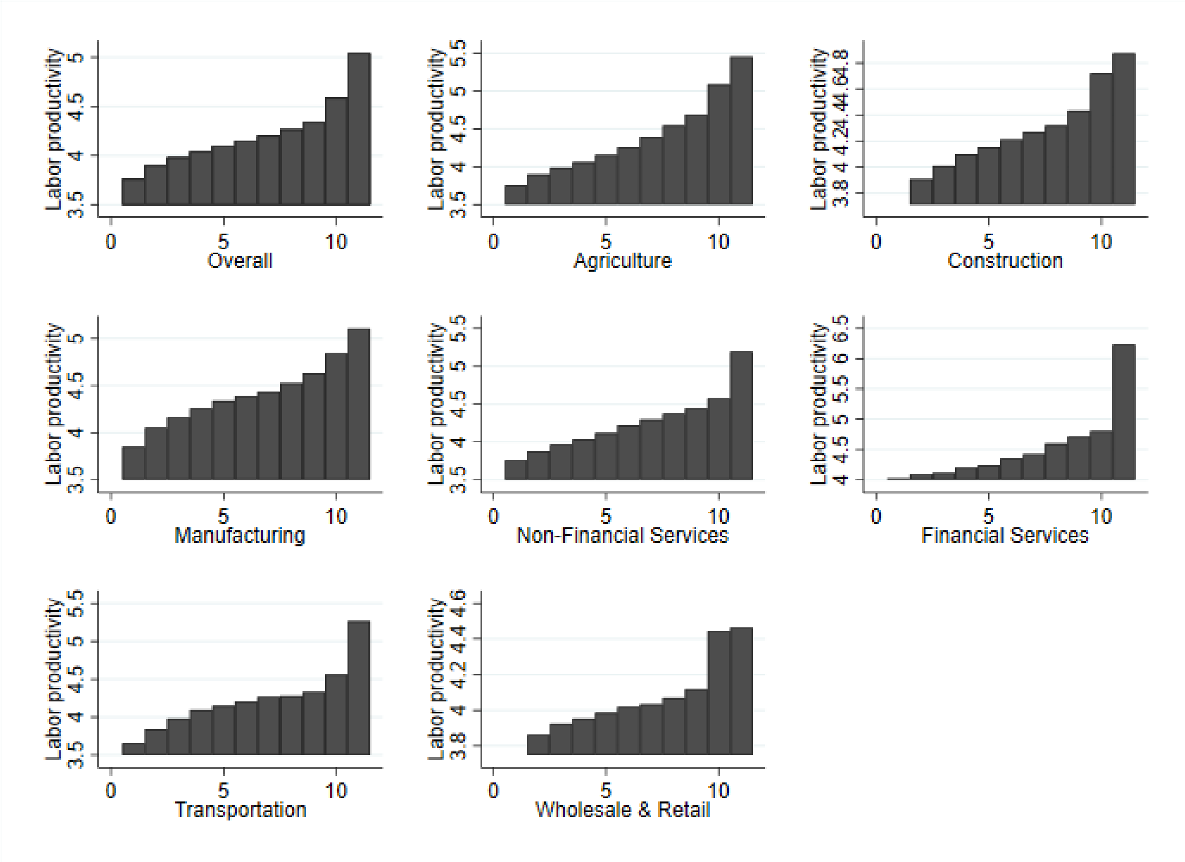

Figure 13: Superstar firms and other firms: average revenue product of labor (2014)
Firm-size (sales) groups 1 to 10 denote “other” firms of increasing sizes, while group 11 denotes “superstar” firms.

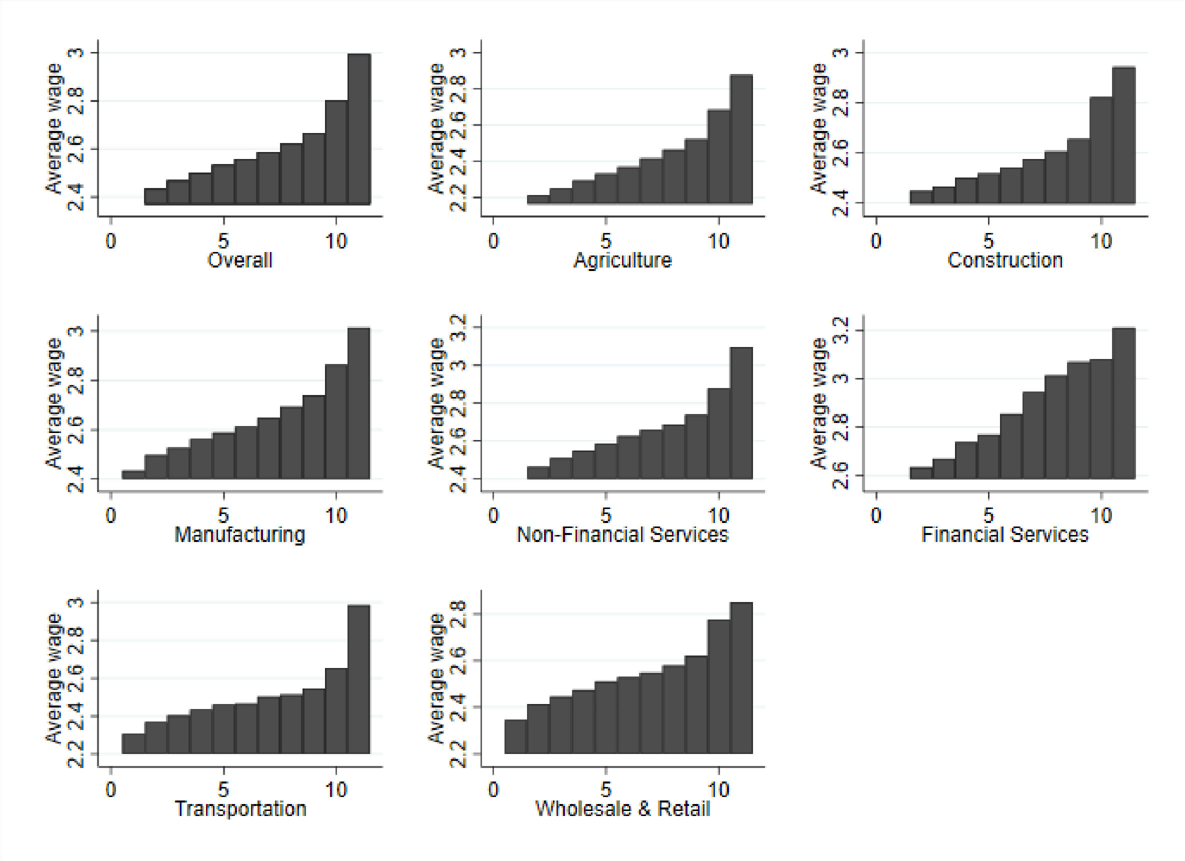

Figure 14: Superstar firms and other firms: average wage (2014)
Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.

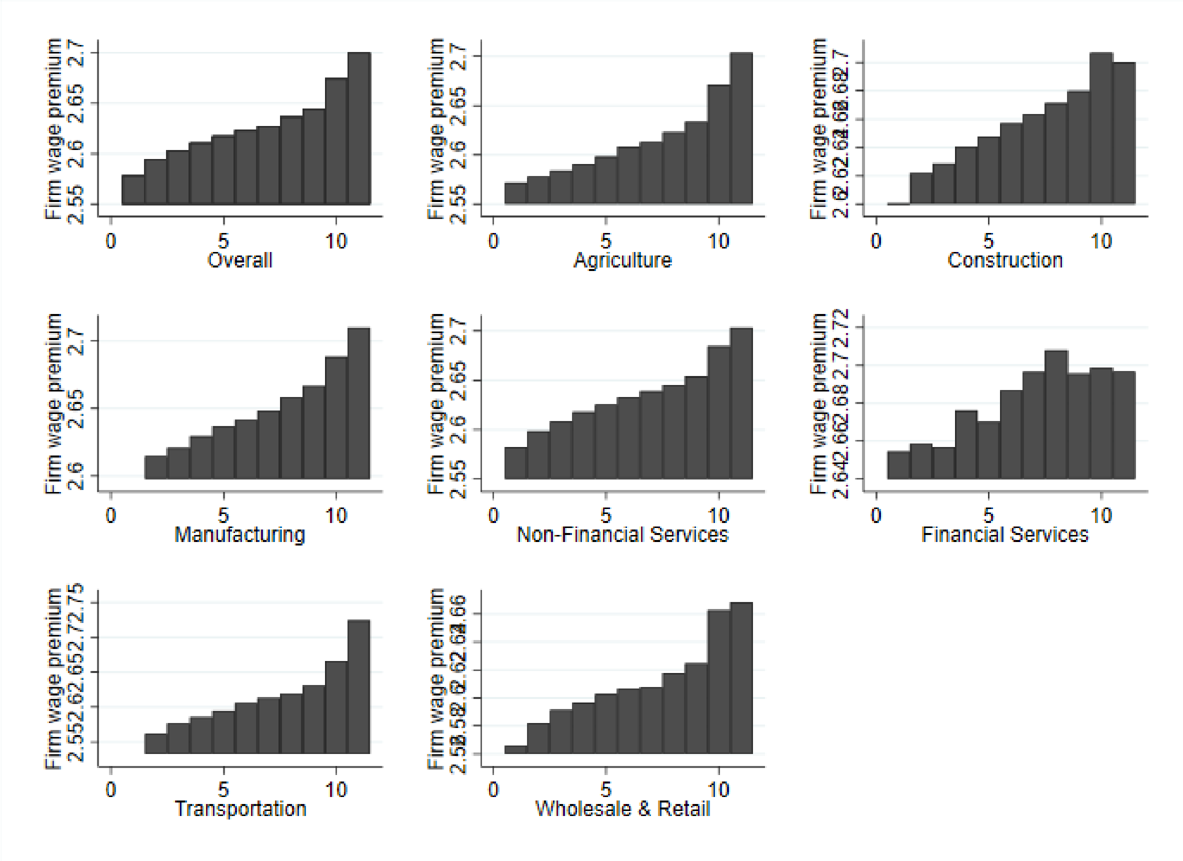

Figure 15: Superstar firms and other firms: firm wage premium (2014)
Firm-size (sales) groups 1 to 10 denote “other” firms of increasing sizes, while group 11 denotes “superstar” firms.

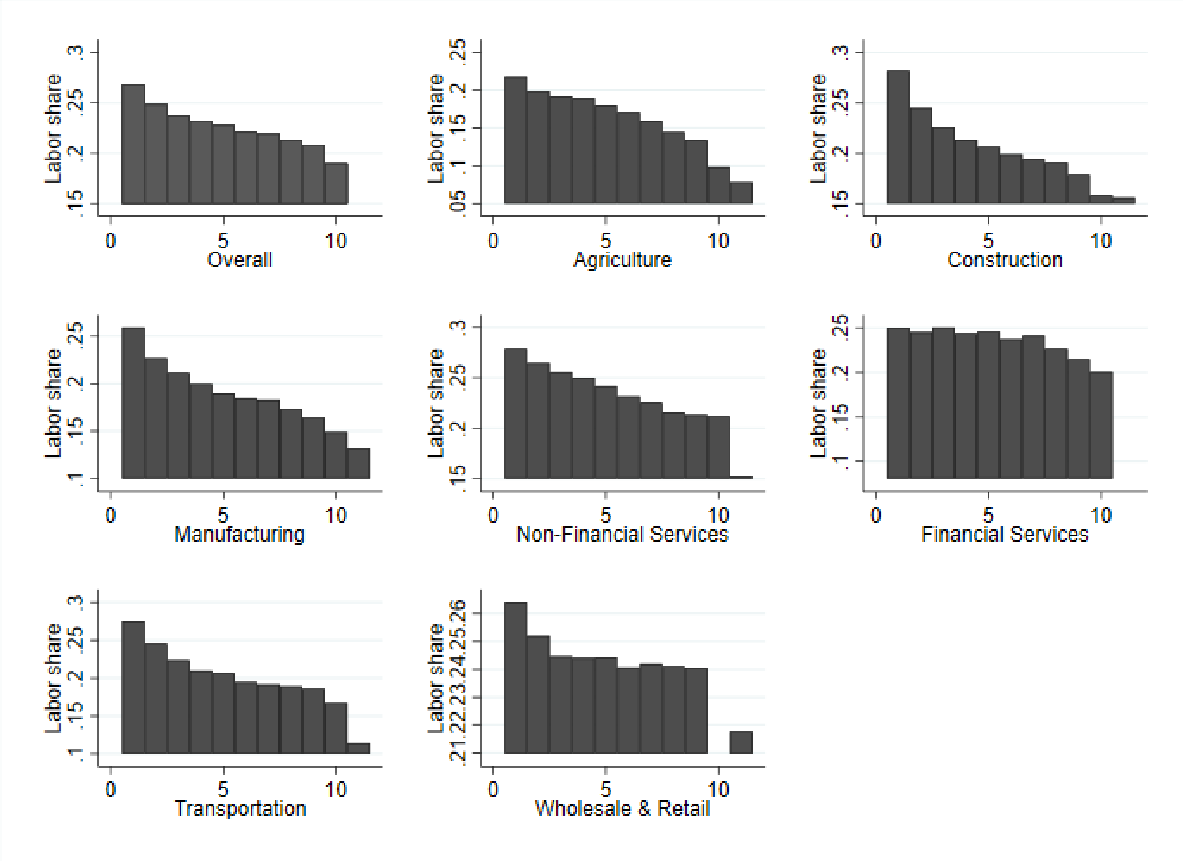

Figure 16: Superstar firms and other firms: labor share (2014)
Firm-size (sales) groups 1 to 10 denote “other” firms of increasing sizes, while group 11 denotes “superstar” firms.

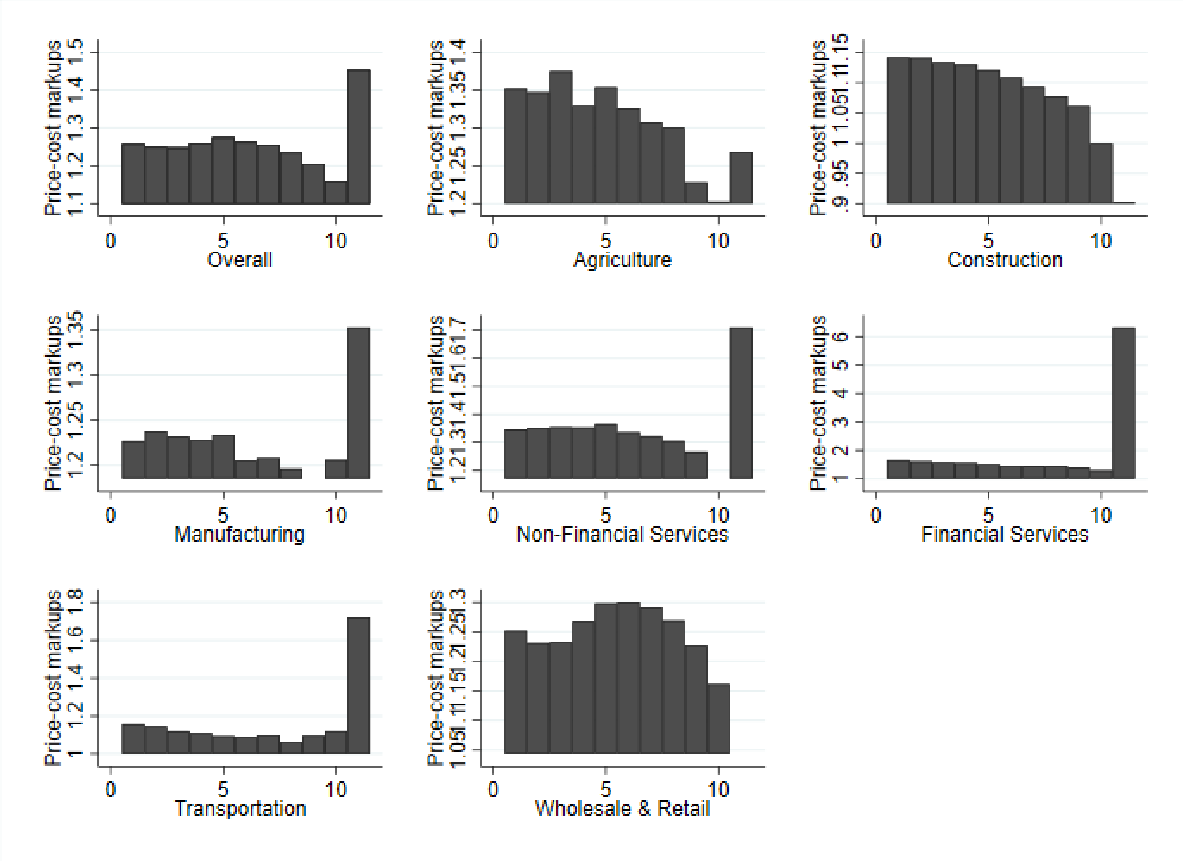

Figure 17: Superstar firms and other firms: price-cost markups (2014)
Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.

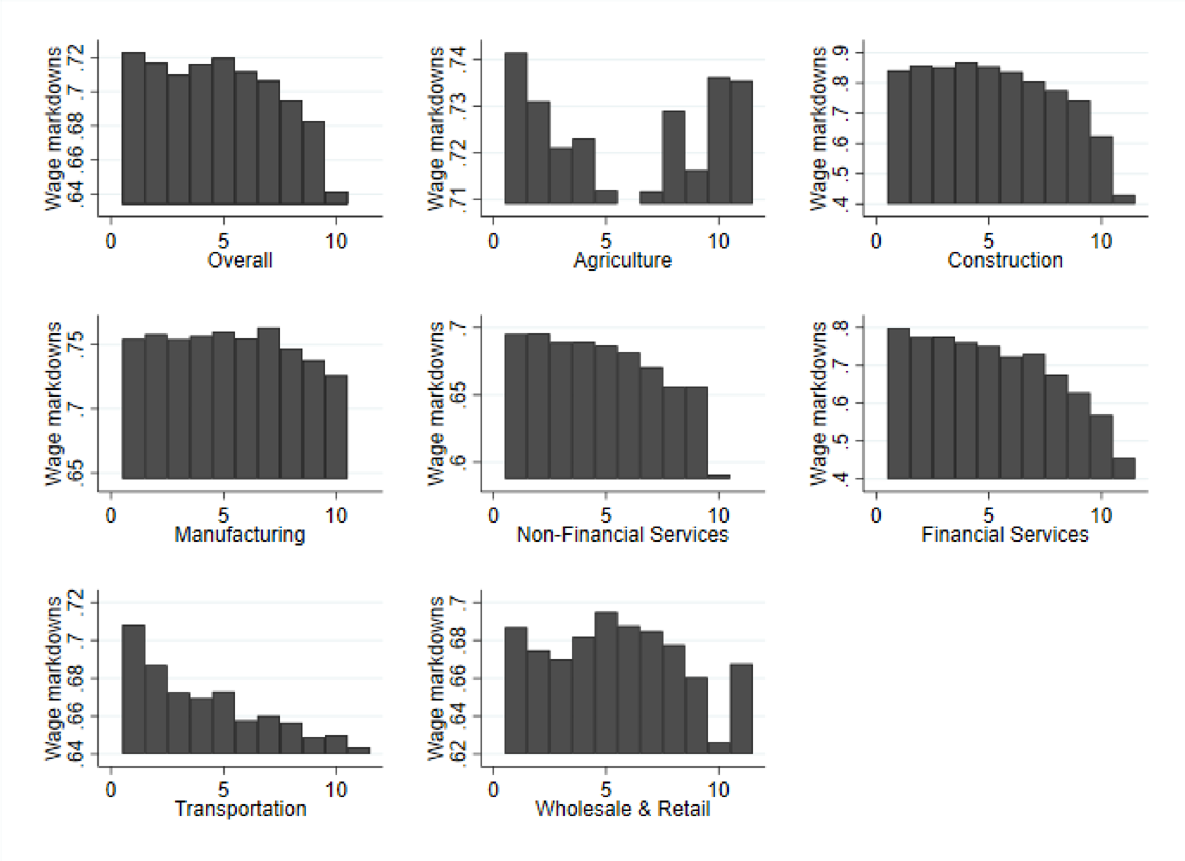

Figure 18: Superstar firms and other firms: wage markdowns (2014)
Firm-size (sales) groups 1 to 10 denote “other” firms of increasing sizes, while group 11 denotes “superstar” firms.

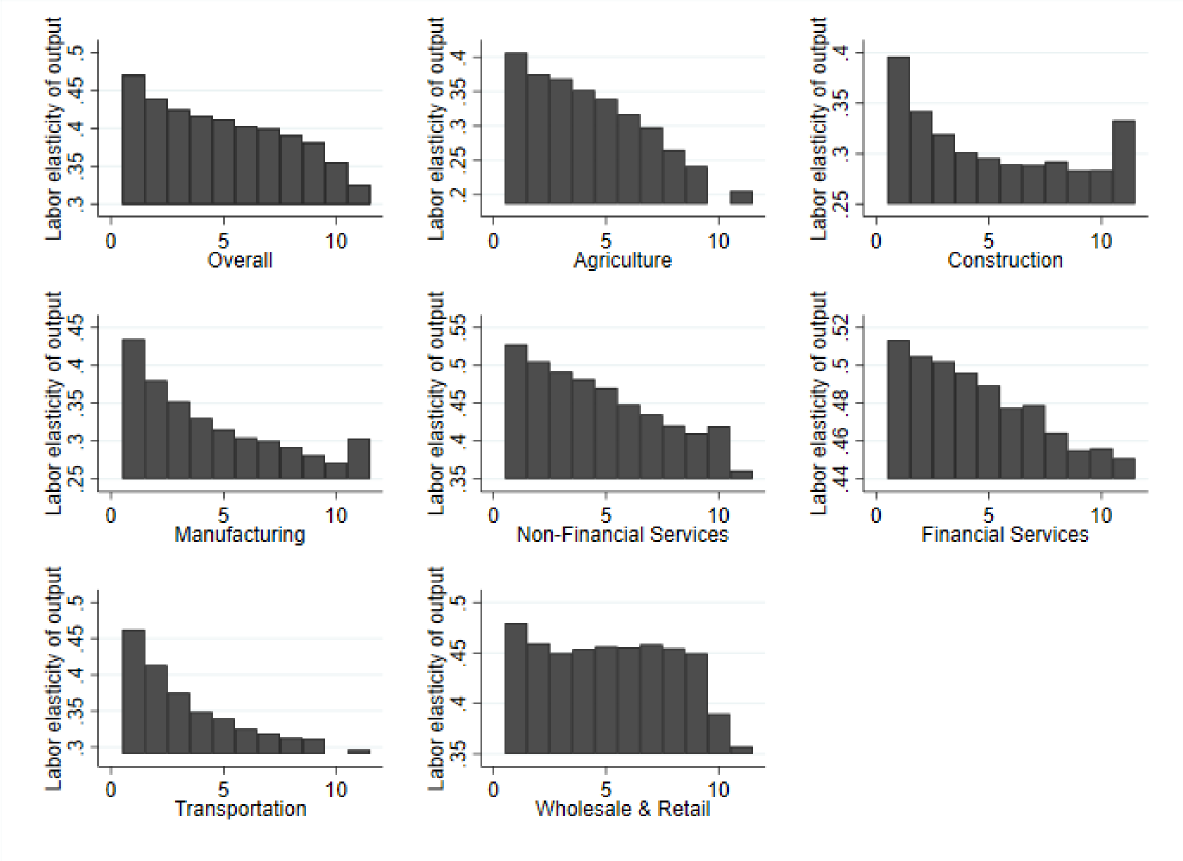

Figure 19: Superstar firms and other firms: labor elasticity of output (2014)
Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.

Appendix C: Wage-Posting and Wage-Bargaining Frameworks

Random Search Wage-Bargaining Framework

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

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

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

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

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

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

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

The unemployed worker's value function is:

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

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

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

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

which depends on the wage Φ_{jt} and non-wage amenities a_{jt} this period through a constant returns to scale utility function $u(.)$, the expected utility next period if the worker is exogenously separated from the firm, and the expected utility if the worker is not exogenously separated. The last component depends on the expected utility of being employed at the same firm, and the expected utility of moving to a new employer conditional on the new employer offering a higher utility. I assume that: (i) the flow utility function $u(.,.)$ is homogenous of degree one in its inputs, (ii) there is no savings mechanism, (iii) the value of non-wage amenities is proportional to worker efficiency, $a_{ijt} = E_{it}a_{jt}$, and (iv) worker efficiency is allowed vary over time due to random shocks: $E_{it+1} = E_{it} + \zeta_{it+1}$, where ζ_{it+1} is a mean-zero random shock.[51] Therefore, the value of unemployment and employment is proportional to worker efficiency. A worker with efficiency E_{it} obtains a value of $E_{it}U_t$ while unemployed and $E_{it}V^e_{jt}$ while employed.

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

$$\Pi_{jt} = \max_{K_{jt}, M_{jt}, V_{jt}} P_{jt}Y_{jt} - R^K_t K_{jt} - P^m_t M_{jt} - \Phi(H_{jt})H_{jt} - c_t(V_{jt})V_{jt} + \beta E_t[\Pi_{jt+1}]$$

subject to the law of motion for employment:

$$H_{jt} = (1 - s(V^e_{jt}))H_{jt-1} + \lambda(V^e_{jt})V_{jt} \tag{14}$$

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

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

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

surplus to be bargained over is:

$$\kappa_{jt} J_{jt} = (1-\kappa_{jt})(V^e_{jt} - U_t)$$

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

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

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

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

$$WM_{jt} = \frac{\left(\frac{\kappa_{jt}}{1-\kappa_{jt}}\right)\left(1 + \frac{W^r_{jt}}{\Phi_{jt} - W^r_{jt}}\right)}{1 + (1 - |\frac{\partial \Phi_{jt}}{\partial H_{jt}} \frac{H_{jt}}{\Phi_{jt}}|)\left(\frac{\kappa_{jt}}{1-\kappa_{jt}}\right)\left(1 + \frac{W^r_{jt}}{\Phi_{jt} - W^r_{jt}}\right) - \beta E_t \left(\frac{(1-s(\Phi_{jt+1}, a_{jt+1}))J_{jt+1}}{c_{V,jt}V_{jt} + c(V_{jt})}\right) \lambda(\Phi_{jt}, a_{jt})} \tag{15}$$

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

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

Random Search Wage-Posting Framework

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

The firm's profit maximization problem is:

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

subject to the law of motion for employment:

$$H_{jt} = (1 - s(\Phi_{jt}, a_{jt}))H_{jt-1} + \lambda(\Phi_{jt}, a_{jt})V_{jt} \tag{16}$$

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

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

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

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

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

Directed Search Wage-Posting Framework

The random search model assumes that workers have no information about wages when they search for a job. An alternative assumption is that workers observe the full menu of wages in the economy when searching for jobs – directed or competitive search (Moen, 1997). I now replace random search with directed search in the otherwise identical wage-posting model. I show in this environment that the firm wage premium equation (1) can be obtained and the wage markdown is identical as the model with random search.[52] The following timing assumption applies. First, idiosyncratic firm productivity and worker efficiency shocks are realized. Next, firms post wages and workers decide on where to search. Then, matching and separations take place. Finally, production begins.

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

[52] For a comprehensive discussion of the theory and applications of directed search, see Wright et al. (2018).

worker is:

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

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

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

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

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

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

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

Workers' search behavior in a Directed Search Model

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

The utility of either workers can be written as:

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

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

Under utility maximization:

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

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

which implies that:

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

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

$$f(V^e_{kt}) < f(V^e_{jt})$$

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

Workplace Differentiation Monopsonistic Wage-Posting Framework

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

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

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

Each worker i maximizes utility by choosing where to work:

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

The firm-specific labor supply curve is then:

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

Firm j's profit-maximization problem is:

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

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

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

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