A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/160254

Essays on Firms' Roles in Labor Markets and International Trade

by

Horng Chern Wong

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Economics

July 2020

THE UNIVERSITY OF

WARWICK

Contents

List of Tables

List of Figures

vi

Acknowledgments

This Ph.D. dissertation represents the culmination of years of support, guidance, faith, and patience from my supervisors, coauthors, friends, family, and the many kind souls I have had the privilege to meet. All of them have contributed in significant ways to making my time in graduate school immensely enjoyable and to making my exciting journey ahead possible.

I am deeply grateful to my supervisors, Thijs van Rens and Roland Rathelot, whose apparently limitless supply of guidance, support, generosity, and encouragement have contributed so much to my growth, both as a researcher and personally. They have constantly pushed me hard to be ambitious and to think hard about how to learn something about the world. From them, I have learned so much, from how to think critically about research to how to communicate effectively. Needless to say, I will continue to draw inspiration from them as I navigate through academia and life in general.

I am also deeply indebted to Dennis Novy and Vincent Sterk (University College London) whose generosity with their time and constructive comments pushed me to think about the bigger picture and profoundly shaped my job market paper. A substantial part of my dream to pursue an academic career was made possible because of their contributions.

It is hard to overstate how much I benefited, through countless interactions, from a great number of faculty members at the University of Warwick and outside of Warwick. So, my gratitude goes out to Warwick Economics faculty: Robby Akerlof, Manuel Bagues, Pablo Beker, Dan Bernhardt, Christine Braun, Luis Candelaria, Natalie Chen, Emma Duchini, Roger Farmer, James Fenske, Omer Moav, Andrew Oswald, Roberto Pancrazi, Carlo Perroni, Claudia Rei, Eric Renault, Giovanni Ricco, Camilla Roncoroni, Federico Rossi, Christopher Roth, Marta Santamaria, and Marija Vukotic. Outside of Warwick: Thomas Le Barbanchon (Bocconi), Dita Eckardt (LSE), Alice Kuegler (UCL), Tim Lee (Queen Mary), Isabelle Mejean (CREST), Fabien Postel-Vinay (UCL), Alessandro Ruggieri (Nottingham), Jesse Shapiro (Brown), Ben Smith (FTC), and Isaac Sorkin (Stanford). Each of them have been so kind and generous with their time and comments, which substantially improved my work.

Declarations

This thesis is submitted to the University of Warwick in accordance with the requirements of the degree of Doctor of Philosophy. I declare that the thesis is my own and original work. Chapters 1 and 2 are sole-authored work. Chapter 3 is joint work with Natalie Chen, Dennis Novy, and Carlo Perroni, all of whom are at the University of Warwick. My coauthors can attest to the significance of my contributions to the project. I also declare that any material contained in this thesis has not been submitted for a degree to any other university.

Horng Chern Wong

July 31, 2020

Abstract

This thesis contains three essays on how firms shape wage differences in the labor market and spatial trade patterns in international goods markets. Chapter 1 asks: why do 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 covering the universe of employers and employees in France. Existing research shows that firm wage premia depend on firms' *labor productivity* and *wage-setting power*. This paper shows that they also depend on firms' *product market power* and *labor share of production*. My decomposition reveals important relationships between these firm characteristics. First, the negative relationship between labor productivity and the 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 with more product market power generally do not have more labor market power, but superstar firms have more market power in both markets than other firms.

Chapter 2 asks: why do larger firms pay higher wages? I revisit the well-known positive relationship between employer size and wages using administrative micro-data from France. I find that this empirical relationship is substantially weaker in services and retail than in manufacturing. To explore the channels through which this relationship can arise, I estimate a structural decomposition framework building on chapter 1. I find that differences in the wage-size relationship between manufacturing and non-manufacturing sectors arise because in the latter (i) the relationship between firm size and average worker productivity, and (ii) the relationship between firm size and the firm-specific component of wages, are comparatively weaker. Worker composition explains approximately three-quarters of the wage-size relationship, while firm-specific heterogeneity explains one-quarter. Overall, my decompositions suggest rent-sharing as the main driver behind the positive relationship between the firm-specific component of wages and firm size, as opposed to efficiency wage considerations.

Chapter 3 asks: why are there large differences in export performance across firms? We study the importance of firms' geographical location within a country relative to well-known differences in firm productivity and how they are related. Using French administrative datasets on the universe of workers, firms, and trade flows, we find that exporter characteristics differ systematically by their proximity to foreign markets. To interpret these new empirical patterns, we build a trade model in which firms differ in terms of productivity and location and derive testable predictions. Our theory shows that the relative importance of domestic and international trade opportunities can shape the spatial distribution of non-exporters and exporters, and in turn, determine the size and productivity of cities.

Chapter 1

Understanding High-Wage and Low-Wage Firms

1.1 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 (Schmieder 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? On one hand, existing research shows that firm heterogeneity in *labor productivity* and *wage-setting power* determine the wage premium a firm pays relative to other firms.[2] On the other hand, recent discussions surrounding a closely related object, the labor share of income, emphasize the importance of firms' *product market power* and *labor share of production*.[3,4] Since these firm characteristics are often studied separately, little is known about their interrelationships or their relative importance for the firm wage premium distribution. Yet, in the presence of labor market frictions, all four firm characteristics jointly determine the firm wage premium distribution, because they are important determinants of labor demand.

This paper quantifies the relative importance of firm-level differences in labor 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

---

[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).

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.[6,7] Without taking them into account, workhorse models of frictional labor markets would overestimate the role that firms' labor productivity and wage-setting power differences play in explaining firm wage premia. To the extent that the role of wage-setting power is overestimated, its negative impact on the allocative efficiency of workers would be overstated, and by extension, room for labor market policy interventions. To the extent that the relative importance of labor productivity is overestimated, distortions to the allocation of workers due to differences in product market power would be missed.[8]

This finding has important implications for understanding the low labor shares of revenue among exceptionally productive superstar firms and for inferring the extent of labor misallocation across firms. The is because there are strong relationships between the estimated firm characteristics. First, the negative correlation between labor productivity and the labor share of production (i) provides a new and quantitatively significant explanation for superstar firms' low labor shares of revenue besides market power (De Loecker et al., 2020), and (ii) implies that conventional measures of labor misallocation (Hsieh and Klenow, 2009) based on revenue per hour overstate the extent of aggregate productivity gains from removing labor market frictions, because these measures do not account for firms' ability to sidestep these frictions by substituting labor with other inputs. Second, while superstar firms have more market power in both product and labor markets, I find that in general firms' product and labor market power are negatively correlated. The effects of firms' product and labor market power on labor misallocation are often studied in isolation (Edmond et al., 2018; Berger et al., 2018). Yet, with the negative correlation, product and labor market power generally dampen each other's effect on labor misallocation.

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.[9] As in workhorse frictional labor market models, firms differ in labor productivity and wage-setting power (Burdett and Mortensen, 1998). Wage-setting power is defined as the fraction of marginal revenue product of labor paid as wages. I refer to this fraction as the *wage markdown*.[10] Compared to these models, the novel features of my framework are firm

---

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

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

[8]For example, in monopsonistic models with constant wage markdowns, labor market power does not lead to misallocation of workers across firms, since there is no dispersion in wage markdowns (Card et al., 2018; Lamadon et al., 2019). This is analogous to models with constant price-cost markups commonly used in macroeconomics and international trade.

[9]These dimensions of heterogeneity are equilibrium outcomes of the model, I therefore also refer to them as channels of heterogeneity.

[10]In a perfectly competitive labor market, wages equal the marginal revenue product of labor, therefore there are

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*.[11]

I estimate the dimensions of firm heterogeneity using a combination of empirical methods from industrial organization and labor economics. In particular, I provide new 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.[12,13,14] I do so by building on the production-based approach of De Loecker and Warzynski (2012) and Flynn et al. (2019) to accommodate imperfectly competitive labor markets and worker heterogeneity, which involves panel data methods commonly used in labor economics (Abowd et al., 1999; Bonhomme et al., 2020). This approach has the advantage that it does not require the researcher to specify particular market structures in a wide array of product and labor markets. To separately estimate output elasticities from productivity, I estimate production functions using a control function approach (Ackerberg et al., 2015), in which I use firms' past input choices to instrument for their current choices under the following timing assumption: firms' past input choices are orthogonal to current realized productivity shocks. To separately identify firms' wage markdowns from price-cost markups, I exploit the fact that labor market power is a distortion only on labor demand, while product market power is a common distortion on the demand for each input.[15]

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.[16,17]

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,

---

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).

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

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

[13]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).

[14]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).

[15]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.

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

[17]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.

price-cost markups, and labor elasticities of output contribute 25%, 38%, 13%, and 24%.[18] 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. Without taking them into account, the model suggests that the explanatory power of firm productivity and wage-setting power would be overstated, accounting for up to 54% and 46% 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 90th percentile of the wage markdown distribution pays 99% of the marginal revenue product of labor as wages, while the 10th percentile firm pays 52%. Moreover, most firms have considerable influence over the wages they pay – half of the firms markdown wages by more than 30%.

I also find substantial heterogeneity across firms in the labor elasticity of output. The raw 90th percentile labor elasticity of output is 0.61 while the 10th 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 literature (Syverson, 2011; De Loecker et al., 2020).

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 (0.69). 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

---

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

through automation or outsourcing.[19]

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 play an outsized role in driving the aggregate labor income share (Autor et al., 2017; Kehrig and Vincent, 2018). My previous finding offers a new explanation for superstar firms' low labor shares of revenue besides product and labor market power: superstar firms operate production processes that are substantially less labor-intensive than other firms (low labor elasticity of output). However, consistent with the existing literature, I also find that superstar firms charge disproportionately higher price-cost markups (De Loecker et al., 2020) and pay significantly larger wage markdowns compared to other firms. My estimates therefore also provide empirical support for the hypothesis that superstar firms have considerable labor market power (Gouin-Bonenfant, 2018). Quantitatively, a back-of-the-envelope exercise suggests that low labor intensity production processes explain the bulk of superstars' low labor shares of revenue, followed by wage markdowns and price-cost markups.

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.

---

[19]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.

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.

This 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 et al., 2020), and the aggregate production technology (Karabarbounis and Neiman, 2014) on wages and the aggregate labor share. I discuss each below. The most closely related paper is Mertens (2019), who studies how manufacturing firms' production technology and market power explain the decline of the German manufacturing labor share. In contrast, my paper studies the role of 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 (Katz and Summers, 1989; Blanchflower et al., 1996; Hildreth and Oswald, 1997; 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 et al. (2020) 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 processes that are significantly less labor-intensive.

Road map. In the next section, I present the structural framework for firm wage premia. In section 1.3, I describe how the structural firm wage premium equation is estimated. Section 1.4 provides information on the French administrative datasets. In section 1.5, I present the main findings. Section 1.6 discusses the implications. The conclusion is in section 1.7.

1.2   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 endogenously 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):

$$\Phi = WM \cdot ARPH \cdot PM^{-1} \cdot LEO \tag{1.1}$$

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.[20] 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.

1.2.1 Departures from standard frictional labor market models

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

$$P_{jt} = \tilde{D}_s(Y_{jt}, D_{jt}) \tag{1.2}$$

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, with $\tilde{D}_{s,y} < 0$ and $\tilde{D}_{s,yy} > 0$. The firm's idiosyncratic demand D_{jt} can be a function of aggregate, sectoral, or firm-specific demand shifters. The assumption of imperfectly competitive goods markets generates a distribution of firm-specific price-cost markups, an important determinant of firms' labor demand in the macroeconomics of resource allocation and the labor share of income (Edmond et al., 2015; Peters, 2017; De Loecker et al., 2020).

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

$$Y_{jt} = X_{jt}F_{st}(K_{jt}, H_{jt}, M_{jt}) \tag{1.3}$$

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 $\ln X_{jt} = G(\ln X_{jt-1}) + \epsilon_{jt}^x$ where ϵ_{jt}^x 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.[21] I discuss what these output elasticities and price-cost markups depend on in the next subsection.

---

[20]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).

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

1.2.2 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 1.3.2. 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:

$$H_{jt} = (1 - s_{jt})H_{jt-1} + R_{jt} \tag{1.4}$$

with:

$$s_{jt} = s(\Phi_{jt}, A_{jt}) \tag{1.5}$$

$$R_{jt} = R(\Phi_{jt}, A_{jt}, V_{jt}) \tag{1.6}$$

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

The assumption that firm-specific separation and recruitment rates depend on the wages offered is informed by models of on-the-job search such as Burdett and Mortensen (1998) and Mortensen (2010), or directed search models such as Kaas and Kircher (2015). I also allow recruitment and separation to depend on non-wage amenities, as there is evidence that non-wage amenities are important determinants of worker flows between firms (Sorkin, 2018). Together, equations (1.4), (1.5), and (1.6) form the firm-specific upward-sloping labor supply curve.

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

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

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

subject to (1.2), (1.3), (1.4), (1.5), and (1.6). 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 (1.1):

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

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

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

$$WM_{jt} = \frac{\epsilon_{jt}^H}{1 + \epsilon_{jt}^H - \beta E_t \left(\frac{(1-s_{jt+1})J_{jt+1}}{c_{V,jt}V_{jt}+c(V_{jt})} \right) R_{V,jt}} \tag{1.7}$$

where $\epsilon_{jt}^H = \epsilon^H(\Phi_{jt}, a_{jt}, V_{jt})$ is the firm-specific labor supply elasticity, $c_{V,jt}V_{jt} + c(V_{jt})$ is the marginal recruitment cost, and J_{jt+1} is the marginal profit to the firm of having an additional worker next period. Equation (1.7) 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 (1.7) 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.[22] 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 (1.4), (1.5), and (1.6)) 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.

Although I do not take a stand on the joint distribution of the heterogeneity in primitives (idiosyncratic productivity (X), demand (D), and non-wage amenities(A)), I now discuss how these primitives map into wage markdowns. Consider two firms that are identical along all dimensions, but one has higher productivity than the other. Then the firm with the higher productivity will have a higher labor demand and pay higher wages (Φ). Since the more productive firm pays higher wages, it locates itself at the part of the labor supply curve where the labor supply elasticity (ϵ^H) is lower: it faces less labor market competition compared to the less productive firm. The lower labor supply elasticity reflects the lower recruitment (ϵ_Φ^R) and separation elasticity with respect to wages (ϵ_Φ^s): the high-wage firm cannot raise the recruitment rate and reduce the separation rate by much if it offers yet higher wages, since it already pays the highest wages. The wage markdown is therefore lower. The same is true in a comparison of two firms which are identical along every dimension except idiosyncratic demand (D). The prediction that more productive firms have lower wage markdowns is standard in monopsonistic or oligopsonistic models, such as Burdett and Mortensen (1998).

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

It is worth noting that this structural framework nests a workhorse model of frictional labor markets - the Burdett and Mortensen (1998) model. This model will be a useful benchmark for interpreting some of the decomposition results in section 1.5. To obtain the Burdett-Mortensen

---

[22]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.

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.[23] The second assumption ensures that the firms' revenue functions exhibit constant marginal returns to labor. This assumption implies that the output elasticity with respect to labor is equal to 1 across all firms. The third assumption is standard in traditional monopsony models.[24] 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.[25] 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 (1.1) 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.

Since total revenue is increasing in firms' idiosyncratic productivity (X) and demand (D), all else equal, the average revenue product of labor is increasing in these underlying firm primitives. Moreover, the average revenue product of labor is decreasing in the value of non-wage amenities (A), all else equal. This is because the firm with less desirable amenities

---

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

[24]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.

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

will have to pay higher wages to hire a given number of workers, reducing its total number of recruits. Since the production function satisfies diminishing marginal returns to labor, the average revenue product is higher for firms with less desirable amenities.

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

$$PM_{jt} = \frac{\epsilon_{jt}^{G}}{\epsilon_{jt}^{G} - 1}$$

where which ϵ_{jt}^{G} is the firm-specific price elasticity of demand. The specific functional form for the price elasticity of demand depends on the researcher's microfoundation for the product demand curve (1.2). 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).[26] Equation (1.1) shows that, all else equal, firms with higher markups pay a lower wage premium. The intuition is that firms that are able to charge positive markups maximize profits by producing less than they would in the perfectly competitive benchmark, which reduces their labor demand and the wage premium they are willing to pay.

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

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

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

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

---

[26]Since the demand function in equation (1.2) 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).

Equation (1.1) 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.

Consider two firms that have different idiosyncratic productivities (X). Then the more productive firm has a higher labor elasticity of output if the elasticity of substitution between labor and other inputs is less than one (complements), while the opposite is true if the elasticity of substitution is greater than one (substitutes). As shown in section 1.5, the empirically relevant case is the latter. The more productive firm wants to hire more workers to produce higher output. However, because of labor market frictions, firms must pay higher wages to hire more workers. Therefore, the more productive firm faces a higher relative cost of labor compared to the less productive firm. Since labor and other inputs are substitutes, the more productive firm substitutes labor with other inputs, increasing the capital-labor ratio and intermediate-input-labor ratio, reducing the labor elasticity of output. The same is true when two firms have different idiosyncratic demand (D), but are otherwise identical.

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

Marginal revenue product of labor (MRPH). The last three components of the firm wage premium equation (1.1) 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.[27] I explore these implications further in section 1.5.

1.2.3   Discussion

The firm wage premium equation (1.1) 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 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 only consider wage-setting protocols of a static nature: contemporaneous wage-posting and wage-bargaining. In doing so, I abstract

---

[27]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.

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.[28] 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).[29] 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.

1.3   Estimating the Structural Firm Wage Premium Equation

1.3.1   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 effective 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

---

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

[29]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.

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 et al. (2020) to accommodate imperfectly competitive labor markets. In the original approach, one first estimates the output elasticities, then computes price-cost markups from a variable input's expenditure share of revenue. I show that when labor markets are imperfectly competitive, estimating output elasticities requires knowledge of the firm-specific wage premium.[30] 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:[31]

$$
\begin{aligned}
y_{jt} = {} & \beta_{h,s}h_{jt} + \beta_{m,s}m_{jt} + \beta_{k,s}k_{jt} + \beta_{hh,s}h_{jt}^2 + \beta_{mm,s}m_{jt}^2 + \beta_{kk,s}k_{jt}^2 \\
& + \beta_{hm,s}h_{jt}m_{jt} + \beta_{hk,s}h_{jt}k_{jt} + \beta_{mk,s}m_{jt}k_{jt} + \beta_{hmk,s}h_{jt}m_{jt}k_{jt} + x_{jt} + \epsilon_{jt}
\end{aligned}
\tag{1.8}
$$

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. In essence, returns-to-scale 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

---

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

[31]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: $y_{jt} = x_{jt} + \theta_{h,s}(h_{jt}, m_{jt}, k_{jt})h_{jt} + \theta_{m,s}(h_{jt}, m_{jt}, k_{jt})m_{jt} + \theta_{k,s}(h_{jt}, m_{jt}, k_{jt})k_{jt}$.

allowing returns-to-scale to depend on firms' input choices besides the proxy variable input.[32] 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 researcher 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 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.[33] 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

---

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

[33]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.

express idiosyncratic total factor productivity as a function of observed variables:

$$x_{jt} = x(m_{jt}, k_{jt}, h_{jt}, \mathbf{Z}_{jt}, \phi_{jt}) \tag{1.9}$$

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 (1.8) and (1.9) and estimate the following by OLS:

$$y_{jt} = \Psi(k_{jt}, h_{jt}, m_{jt}, \mathbf{Z}_{jt}, \phi_{jt}) + \epsilon_{jt} \tag{1.10}$$

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:

$$x_{jt} = g(x_{jt-1}) + \zeta_{jt} \tag{1.11}$$

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 (1.9), the predicted output from (1.10), and the law of motion for productivity (1.11), 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:

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

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

A common challenge in the production function estimation literature is that output prices are rarely observed (De Loecker and Goldberg, 2014).[34] In typical firm-level balance sheet data, output is usually measured in terms of sales revenue or the nominal value of production. Under

---

[34]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.

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), I include export status and market shares as additional controls. Informed by models of customer capital (Gourio and Rudanko, 2014), which predict that firms accummulate customers over time, I also include firm age. The lags of these additional controls therefore also appear in the vector $\mathbf{X}$ in the moment conditions of the estimation procedure $E[\zeta_{jt}(\beta)\mathbf{X}_{jt}] = \mathbf{0}$. The key assumption here is that these additional controls sufficiently capture variation in markups uncorrelated with TFP. This assumption rules out a role for differences price elasticities of demand due, for example, to product quality differences, conditional on firms' TFP.[35]

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

---

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

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:

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

1.3.2 Estimating firm wage premia

I estimate the firm-specific wage premium by estimating firm(-group) fixed effects following Bonhomme et al. (2019) (BLM henceforth):[36]

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

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

---

[36]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.

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.[37] 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-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 1.1 compares the changes in mean log wages for workers who move between quartiles of firm fixed effects, following Card et al. (2013). Figure 1.1 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. Figure 1.2 depicts a similar exercise, following Sorkin (2018).

Another way to assess the log additivity of the worker and firm components of wages is

---

[37]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.

to group worker and firm fixed effects into 10 deciles each, generating 100 worker-firm fixed effect deciles, then plot the mean estimated residuals within each worker-firm fixed effect decile. If the firm wage premium depends strongly on the worker's skill type, log additivity would be severely violated, and one should observe that the estimated residuals systematically varies across worker-firm fixed effect deciles. In Appendix B, Figure 1.3 and 1.4 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.

1.4   Data Description

1.4.1   Administrative datasets from France

Estimating the structural firm wage premium equation using the approach described above requires two types of datasets. The empirical distribution of the firm wage premium is estimated with matched employer-employee datasets, which follow workers over time and employment spells at different firms. The four channels of firm heterogeneity in the model are estimated with firm balance sheet panel datasets. While both types of datasets have become increasingly accessible, they are typically not jointly available.[38] 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.[39] 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

---

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

[39]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.

observed, allowing researchers to construct and study variation in hourly wages. This addresses concerns that variation in earnings simply reflect variation in hours worked. They also include employer identifiers, called SIREN, which enables linking with firm balance sheet data. One disadvantage is that information about workers' education is not available.

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

1.4.2   Sample cleaning

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 1.6.

1.5 Firm Heterogeneity and the Firm-Specific Wage Premium

1.5.1 Firms have very different product market power and labor elasticities of output

This section documents the empirical moments of each of the four channels 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 1.1 summarizes the empirical moments of each estimated dimension of firm heterogeneity in 2014.[40] Table 1.7 in Appendix B shows moments related to the within-sector distribution of each dimension. Tables 1.8, 1.9, 1.10, and 1.13 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 1.1 describes the distribution of wage markdowns (WM). I find substantial dispersion of wage markdowns across firms. My estimates show that firms at the 90th percentile of the wage markdown distribution pay a wage markdown of 0.99, approximately the level of marginal revenue product of labor that firms would pay in a perfectly competitive labor market. At the 10th percentile, workers obtain approximately half of their marginal revenue productivity (0.52). Figure 1.6 plots the kernel density of wage markdowns.

---

[40]In Tables 1.14, 1.15, 1.16, 1.17, and 1.18 in Appendix B, I show that each dimension of firm heterogeneity is correlated with observed firm characteristics, including firm size, age, productivity, market concentration, and the job-to-job transition rate.

As discussed in section 1.2, 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 rev. product of labor ($\log$)	4.17	4.13	0.20	4.76	3.63
Marginal rev. product of labor ($\log$)	2.96	2.98	0.07	3.28	2.64
Number of firms	294,233				

Table 1.1: 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 1.7 plots the distribution of inverted wage markdowns against price-cost markups. Table 1.1 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 1.2, 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.[41] This gives firm-specific labor supply elasticities of 0.44, 1.89, 5.98 at the 10th, 50th, and 90th 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 10th, 50th, and 90th 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

---

[41]In section 1.2, I show how the Burdett-Mortensen model can be obtained from my structural framework.

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 1.1 display substantial heterogeneity across firms, particularly within sectors. The 90th percentile labor elasticity of output is 0.61, while it is 0.20 at the 10th 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.[42,43] These findings suggest that labor elasticities of output are potentially important determinants of the distribution of firm wage premia and labor shares within sectors. This is explored further in Section 1.6.[44]

I now confirm that, consistent with existing findings, price-cost markups and firm productivity are highly dispersed across firms. The third row of Table 1.1 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 model with heterogeneous markups. De Loecker and Eeckhout (2018) find markups at the 90th 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 90th percentile between 1.24-1.69 in 2012. While the markups at the 10th 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 1.1 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, 2003a). 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 1.7 shows that the average 90-10 ratio

---

[42]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.

[43]As detailed in section 1.3, production functions are estimated at the 2-digit level. Removing differences across 4-digit sectors reduces the 90-10 difference to 0.31.

[44]As discussed in section 1.2, 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.

within two-digit sectors is $\exp(4.69 - 3.70) = 2.70$.[45]

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

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 channels 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. Without taking them into account, standard models of frictional labor markets risk overstating the explanatory power of other firm characteristics – firm productivity ($ARPH$) and wage markdowns (WM).

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

$$\phi_{jt} = wm_{jt} + arph_{jt} - pm_{jt} + leo_{jt} \tag{1.12}$$

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).[46,47] I implement this decomposition by running equation (1.12) as a linear 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 1.5.4, where I study the importance of the relationships between firm characteristics, I present results from the standard variance decomposition. Table 1.2 presents the Shapley decomposition results.

The first row of table 1.2 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 1.2, 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).[48] The structural framework in section 1.2 also nests the workhorse Burdett and Mortensen

---

[45]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.

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

[47]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 1.23 in Appendix B.

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

([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.[49]

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

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 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 [1.2](#) 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 channels of heterogeneity – price-cost markups and labor elasticities of output. Therefore, without taking them into account, workhorse models of frictional labor markets would overestimate the explanatory power of firm productivity and wage markdowns for firm wage premia.

---

[49]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.

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 1.2, price-cost markups affect firm wage premia through firms' labor demand in my structural framework. The fourth row of table 1.2 shows that price-cost markups account for 13% of the firm wage premium distribution.

The fifth row of table 1.2 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 1.2 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 (1.12), 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.

If one were to estimate a standard frictional labor market model without taking into account the two channels of firm heterogeneity – price-cost markups and labor elasticities of output – the model would overstate the explanatory power of firm productivity and wage markdowns for firm wage premia.[50] In this case, my decomposition suggests that firm productivity ($ARPH$) and wage markdowns (WM) account for 54% and 46% of the firm wage premium distribution. This is important for two reasons. First, by overstating the role of heterogeneous wage markdowns, the model overstates the extent to which labor market policies can address distortions due to labor market frictions. For example, when firms' wage markdowns are quantitatively important distortions to labor demand, the minimum wage can be an effective tool to correct such distortions and lead to welfare improvements (Berger et al., 2018). Second, by overstating the role of firm productivity, the model overestimates the extent to which firm wage premium dispersion is efficient, reallocating workers from less productive to more productive firms as workers search on-the-job for better-paying firms. This worker reallocation role of wage dispersion is a key driver of aggregate productivity and wage growth in workhorse models of frictional labor markets (Postel-Vinay and Robin, 2002).

---

[50]Section 1.2 shows that in this case, the model presented converges to a standard Burdett and Mortensen (1998) model.

1.5.3   Moderate firm wage premium dispersion despite substantial firm heterogeneity

While each dimension of firm heterogeneity in equation (1.1) 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 1.5.1, 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 1.3 reports statistics about the dispersion of firm wage premia in 2014. The variance of firm wage premia (ϕ) is modest (0.008), accounting for 4.5% of the wage distribution, similar to the numbers for the United States, Sweden, Austria, Norway, and Italy from Bonhomme et al. (2020). At the same time, the dimensions of firm heterogeneity are orders of magnitude more dispersed than firm wage premia. As the diagonals of table 1.4 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)	
Method	BLM	AKM
Variance	0.008	0.017
Fraction of Total Variance	4.5%	7.7%
90-10 ratio	1.25	1.45
90-50 ratio	1.13	1.19
50-10 ratio	1.11	1.22
Number of firms	294,233	78,268
Number of firm-groups	4,017	78,268
Number of workers	8,715,276	5,497,277

Table 1.3: 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 1.3 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.[51]

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

---

[51]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.

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 1.1 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 channels of firm heterogeneity in equation (1.12) 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:

$$\phi_{jt} = wm_{jt} + arph_{jt} \tag{1.13}$$

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 (1.13). I find an implied variance of firm wage premia (0.233) that is much larger than observed in the data (0.008).[52] This implies that correlations between channels of firm heterogeneity in workhorse frictional labor market models and those emphasized in the labor share literature compress the firm wage premium distribution.

1.5.4 More productive firms have lower labor elasticities of output

I now discuss which pair of correlations between the channels of firm heterogeneity compress the firm wage premium distribution. The main finding in this section is the following: firm productivity and labor elasticity of output are strongly negatively correlated. More productive firms tend to 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 (1.12), 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 1.5.3. The covariance terms show the importance of the relationships between each pair of firm characteristic. These are shown in tables 1.4 and 1.5.

---

[52]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.

More productive firms have a lower labor elasticity of output. The third rows of the second column of Tables 1.4 and 1.5 present this main finding. This negative relationship is the most quantitatively important among the set of cross-terms that offset the effects of firm heterogeneity on the firm wage premium distribution ($CV(leo, arph) = -0.21$, 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 is a prediction of the model in section 1.2, conditional on my production function estimates. This result shows that more productive firms substitute labor with capital and in particular, intermediate inputs. In the process, they use production technologies that are less elastic with respect to labor inputs.

The structural framework of Section 1.2 provides the following intuition. 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.

1.5.5 Other correlations between pairs of firm characteristics

Negative correlation between wage markdowns (wm) and the marginal revenue product of labor ($mrph$). Wage-posting models that allow wage markdowns to vary across firms predict that this pair of variables is negatively correlated across firms (Burdett and Mortensen, 1998; Gouin-Bonenfant, 2018; Berger et al., 2018), consistent with the model in Section 2. The intuition is that since firms with high $mrph$ (labor demand) pay higher wages, they face a locally less elastic labor supply curve, reflecting less labor market competition locally. Therefore, high $mrph$ firms have less incentives to pay a high fraction of $mrph$ as wages. Wage-bargaining models that allow outside options or bargaining power to vary by $mrph$ also share this prediction (Postel-Vinay and Robin, 2002; Jarosch et al., 2019). This prediction finds support in the last row of the first column in tables 1.4 and 1.5. 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 ($mrph = arph$).[53] This implies that the correlation and covariance between wm and $arph$ is the same as that between wm and $mrph$. Table 1.5 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 1.4 and 1.5 shows that 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.

Negative correlation between wage markdowns (wm) and inverted price-cost markups (pm). This negative correlation suggests that firms with more market power in product markets are generally not the same firms as those with more market power in labor markets. This has important implications for the aggregate productivity gains of equalizing product and labor market power distortions across firms. I explore the implications further in the next section.

	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 1.4: 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 1.5: Firm heterogeneity correlation matrix in 2014.

Negative correlation between wage markdowns (wm) and labor elasticities of output (pm). This negative correlation suggests that firms that use labor intensive production technologies tend to have stronger wage-setting power. In a model with frictional labor markets and firms experience random opportunities to automate or outsource production processes such as Arnoud

---

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

(2018), one rationale for this correlation could be that more labor intensive firms have a stronger bargaining position relative to their employees as they can threaten to substitute capital or intermediate inputs for labor.

Little correlation between product market power (pm) and firm productivity ($arph$). The second rows, second columns of Tables 1.4 and 1.5 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 et al. (2020), 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 existing models such as (Edmond et al., 2015), tables 1.24 and 1.31 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.

Positive correlation between product market power (pm) and labor elasticities of output (leo). The fourth row in the third column of tables 1.4 and 1.5 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).

1.6 Implications

New explanation for superstar firms' low labor share of revenue. The literature proposes two main explanations for superstar firms' low labor revenue shares: product market power (De Loecker et al., 2020) and labor market power (Gouin-Bonenfant, 2018). My finding of a negative correlation between firm productivity and the labor elasticity of output adds a new explanation: low labor share of production.

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. To understand the decline of the aggregate labor income share, it is therefore important to understand the differences between superstar firms and other firms.

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.[54,55]

Consistent with De Loecker et al. (2020), figure 1.17 shows that superstar firms charge significantly higher price-cost markups. Figure 1.18 shows that superstar firms also tend have large 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 1.19 provides a novel explanation for superstar firms' low labor shares. It shows that they operate production processes 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 processes as potentially important drivers.

To obtain a sense of which channel of heterogeneity matters more for superstar firms' labor revenue shares, I conduct the following accounting exercise. Recall that the labor revenue share is:

$$\text{Labor revenue share}_j = WM_j \cdot PM_j^{-1} \cdot LEO_j$$

For each sector, I replace *one* channel of heterogeneity in superstar firms' labor revenue share with the average of that channel among non-superstar firms, then compute the mean labor revenue share among superstar firms. I do this for each channel and compare the implied labor revenue share of superstars with the data counterpart. Figure 1.20 shows that, overall, the ranking in decreasing order of quantitative importance is labor elasticity of output (LEO_j), wage markdowns (WM_j), and price-cost markups (PM_j^{-1}). All else equal, replacing superstars' labor elasticities of output with the non-superstar mean raises superstars' labor revenue share from 0.15 to 0.20. Doing the same for wage markdowns raises superstars' labor revenue share from 0.15 to 0.18. The equivalent exercise for price-cost markups raises superstars' labor revenue share from 0.15 to 0.155.

However, figures 1.21, 1.22, and 1.23 show that there is some heterogeneity between sectors in whether market power or production processes matters more for superstar firms' labor revenue shares. Consider three large sectors: manufacturing, non-financial services, and wholesale and retail. Figure 1.21 shows that the low labor revenue shares of superstar firms in the manufacturing sector is mainly driven by wage markdowns and labor elasticities of output. In the non-financial service sector, figure 1.22 shows that labor elasticities of output is the main driver, followed by price-cost markups and wage markdown (with similar quantitative importance). In the wholesale and retail sector, however, figure 1.23 shows that low labor elasticities of output and wage markdowns entirely explain superstars' low labor revenue shares.

Role of input substitution in labor misallocation. The cross-sectional dispersion of the marginal revenue product of labor is a well-known measure of misallocation of labor inputs (Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009). Conventional measures of labor

---

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

[55]In terms of observed characteristics, Figures 1.13, 1.14, 1.15, and 1.16 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.

misallocation overstate the variance of the marginal revenue product of labor and, hence, the degree of labor misallocation. This is because conventional measures of the variance of the marginal revenue product of labor do not account for firms' ability to substitute labor with other inputs in the presence of labor market frictions. However, my finding of a negative correlation between firm productivity and the labor elasticity of output shows that more productive firms tend to substitute labor with other inputs.

The firm wage premium equation (1.1) 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 to, and can be measured by, the $ARPH$. However, the finding in section 5.4 that $ARPH$ and LEO are strongly negatively correlated implies that the $MRPH$ is considerably less dispersed than the $ARPH$. Indeed, in 2014 the variance of (log) $ARPH$ is almost three times larger than the variance of (log) $MRPH$, $\frac{V(arph)}{V(mrph)} = 2.86$. The correlation coefficient between the two is 0.25. Figure 1.11 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.[56] 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. Recall that equation (1.1) can be written as:

$$WM_j^{-1} \cdot \Phi_j = MRPH_j$$

This equation shows that, given the observed firm wage premium (Φ_j), conventional measures of the $MRPH$ would overstate the degree of dispersion in labor market distortions (WM_j^{-1}). This result suggests that counterfactual output and productivity gains from eliminating labor market frictions are overstated when the $mrph$ is measured as the $arph$.

To get a sense of the potential extent to which aggregate efficiency gains from removing labor market frictions could be overstated, I perform a Hsieh and Klenow (2009) exercise to compare the implied efficiency gains from using the conventional measure and my estimated measure of $MRPH$ dispersion. Let s denote the sector. As in Hsieh and Klenow (2009), assume that the sector-specific CES aggregator over firm-level output is $Y_s = \left(Y_{sj}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}}$. To derive closed-form solutions for aggregate sectoral efficiency (TFP_s), I impose the assumption that firms operate sector-specific Cobb-Douglas constant returns-to-scale production functions $Y_{sj} = X_{sj} K_{sj}^{\alpha_s^K} H_{sj}^{\alpha_s^H} M_{sj}^{\alpha_s^M}$. As in section 1.2, firms face firm-specific labor supply curves. I assume that labor market frictions are the only distortions present. In Appendix B, I show that in this world the sectoral TFP gains from removing labor market frictions is given by:

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

where TFP^* denotes the sectoral efficiency (total factor productivity) in a world without labor market frictions. Let $M\tilde{R}PH$ denote the measure of the marginal revenue product of labor

---

[56]Section 1.2 provides more detail.

that does not account for the negative correlation between $ARPH$ and LEO, while $\hat{MRPH}$ denotes the measure that does. Then, the average relative sectoral efficiency gains from removing labor market frictions is:

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

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

Role of product and labor market power in labor misallocation. I now show that my finding of a negative cross-sectional correlation between product and labor market power implies that these channels of heterogeneity partially dampen each other's effect on labor input misallocation. However, among superstar firms, product and labor market power amplify each other's effect on labor misallocation, since superstar firms have more market power in both markets than other firms.

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 1.5 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). 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.

Relationship between aggregate labor income share and between-firm wage inequality. There is plenty of evidence that larger and more productive firms pay higher wages, contributing to between-firm wage inequality (Berlingieri et al., 2017). At the same time, evidence suggests that larger and more productive firms have lower labor shares of revenue, 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). Therefore, there is a negative relationship between two types of income inequality – between-firm wage inequality and the aggregate labor income share. These two measures of income inequality are linked in the following sense:

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

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 of revenue; (ii) the more strongly positively related firm wage premia and firms' labor productivity are, the higher labor shares are at 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 the positive relationship between firm wage premia and labor productivity 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.[57] To assess the impact of this policy and others, such as minimum wages and size-dependent taxes and subsidies, on these two measures of income inequality, future work would require a fully-specified model suitable for policy simulations, taking into account firm entry and exit.

Inferring workers' relative bargaining power. The negative correlation between firm productivity and labor elasticity of output implies that using the joint distribution of wages and firm productivity as a moment to help structurally estimated models infer workers' relative bargaining power would understate workers' 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 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

---

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

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.

1.7 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 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 (Schmieder 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).

1.8 Appendix A: Shapley Decomposition

Recall the firm wage premium equation (1.1), 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.

1.9 Appendix B: Figures and Tables

1.9.1 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 1.6: Summary statistics: Employees and employers (1995-2014).

1.9.2 Conditional Exogenous Mobility & Symmetry
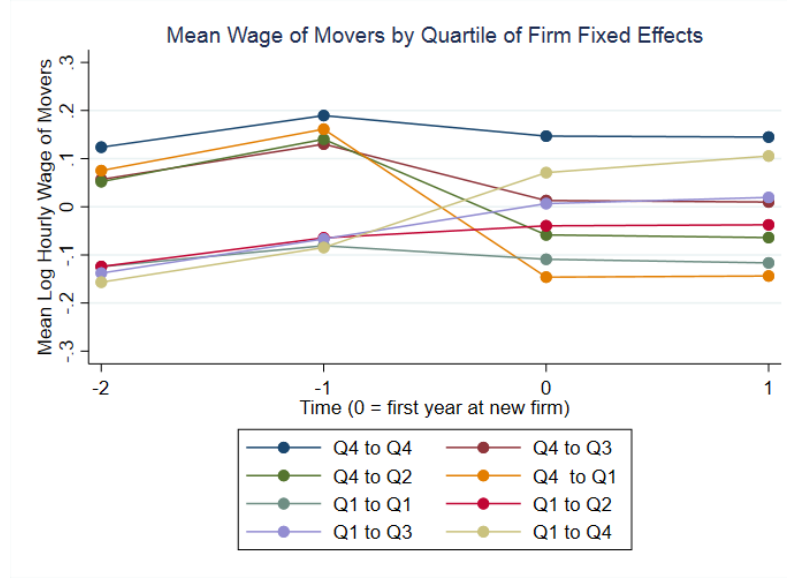
Figure 1.1: Worker mobility and wage changes by quartiles of firm effects (2009-2014).

An alternative way to assess the AKM regression specification is to compare the changes is residual wages to changes in firm effects, following Chetty et al. (2014) and Sorkin (2018). This is similar to the above method (Card et al., 2013). I run the following regression among all employer-to-employer transitions:

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

where $w_{it}^r = w_{it} - x_{it}'\hat{\beta}$ denotes residualized wages and $\phi_{g(j(i,t))}$ are the firm-group fixed effects. If the AKM regression is not mis-specified, the estimated coefficient $\hat{\alpha_1}$ will equal 1. I find $\hat{\alpha_1} = 0.857$, with a standard error of 0.007. To see this graphically, Figure 1.2 plots the changes in residual wages and the changes in firm fixed effects in 100 bins of changes in firm fixed effects. In models of assortative matching based on comparative advantage (Eeckhout and Kircher, 2011; Lopes de Melo, 2018), worker mobility is strongly driven the residual component of the AKM regression, implying that AKM regressions are mis-specified. As Sorkin (2018) shows, these models predict that worker mobility entails a wage gain, regardless of the direction of worker mobility in terms of the estimated firm effects, as workers move to firms at which they have a comparative advantage: there is a V-shape around zero changes in firm effects. The patterns of wage changes upon changes in firm fixed effects shown in Figure 1.2 do not resemble a V-shape around zero.
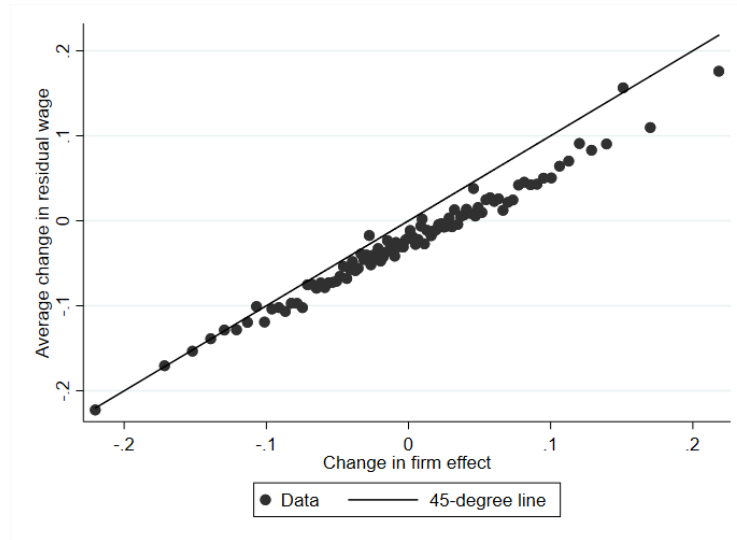
Figure 1.2: Average wage changes from worker mobility by declines of changes in firm premia (2009-2014).
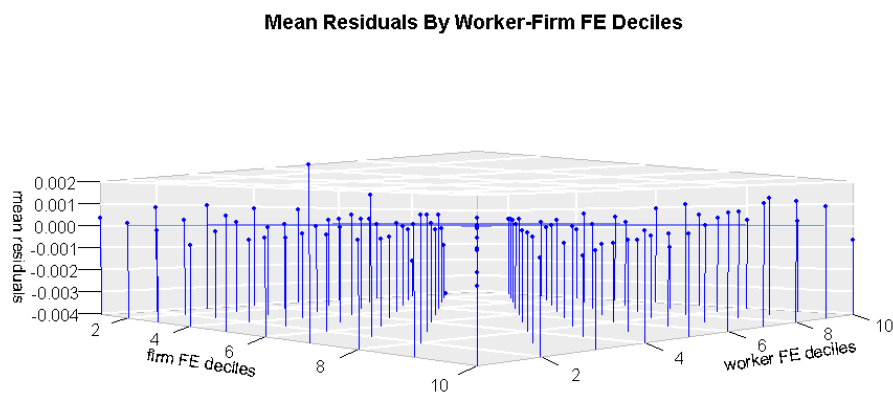
Figure 1.3: Mean estimated residuals by worker-firm deciles (2014)
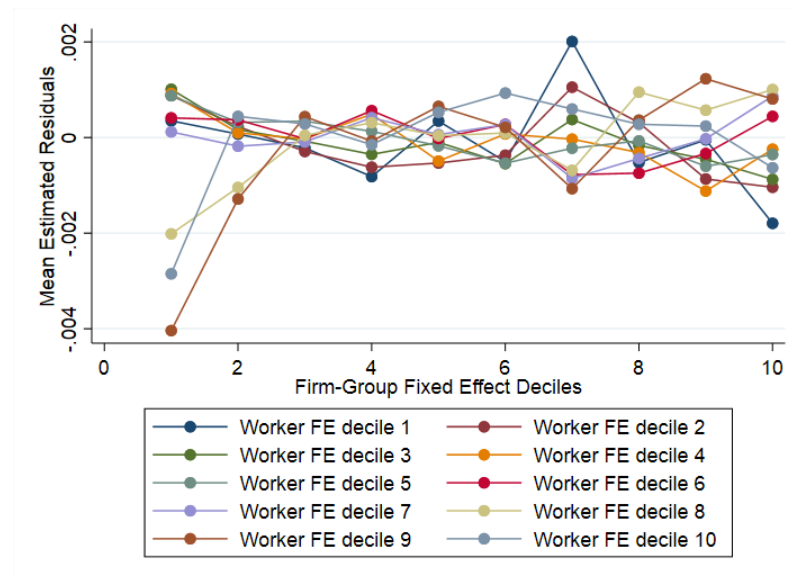
Figure 1.4: Mean estimated residuals by worker-firm deciles (2014)

1.9.3 Distributions of Firm Heterogeneity
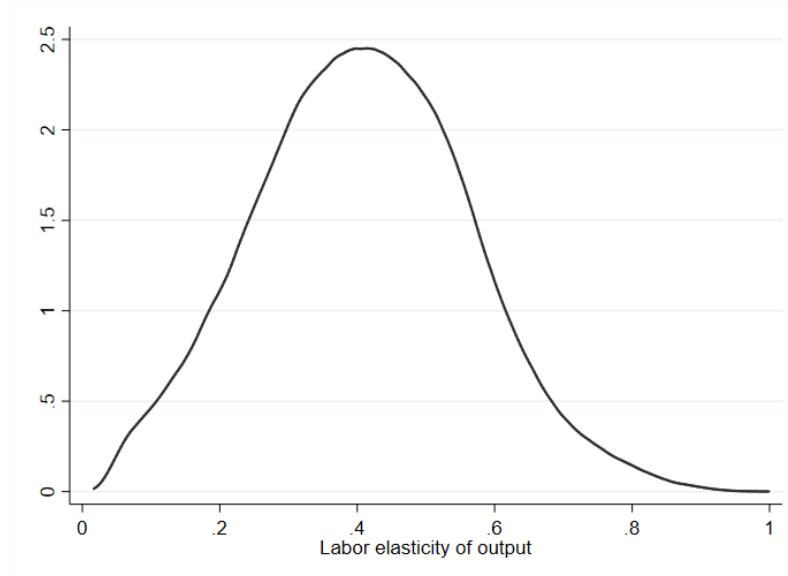
Figure 1.5: Distribution of labor elasticities of output, 2014.
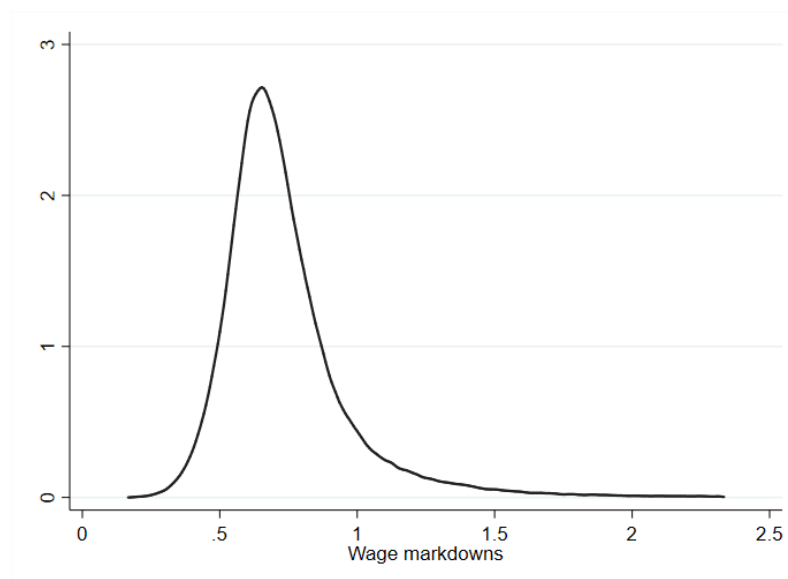
Figure 1.6: Distribution of wage markdowns, 2014.
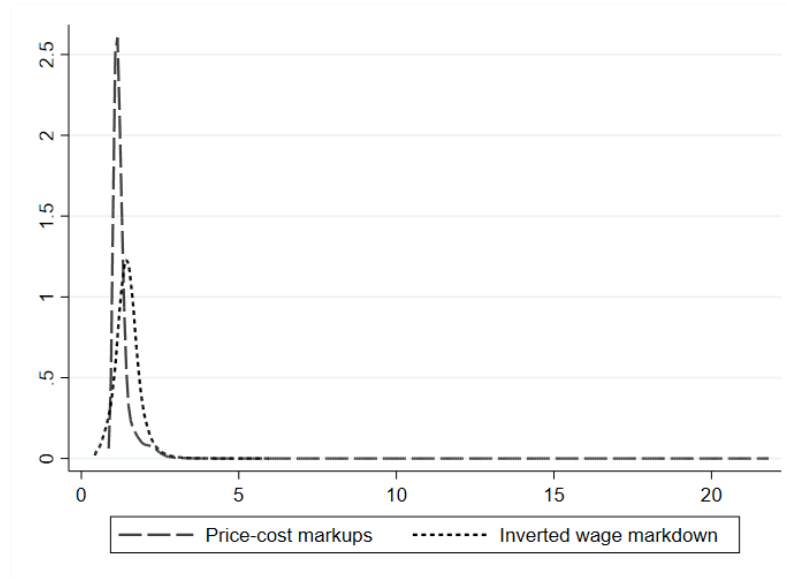
Figure 1.7: Distribution of price-cost markups and (inverse) wage markdowns, 2014.
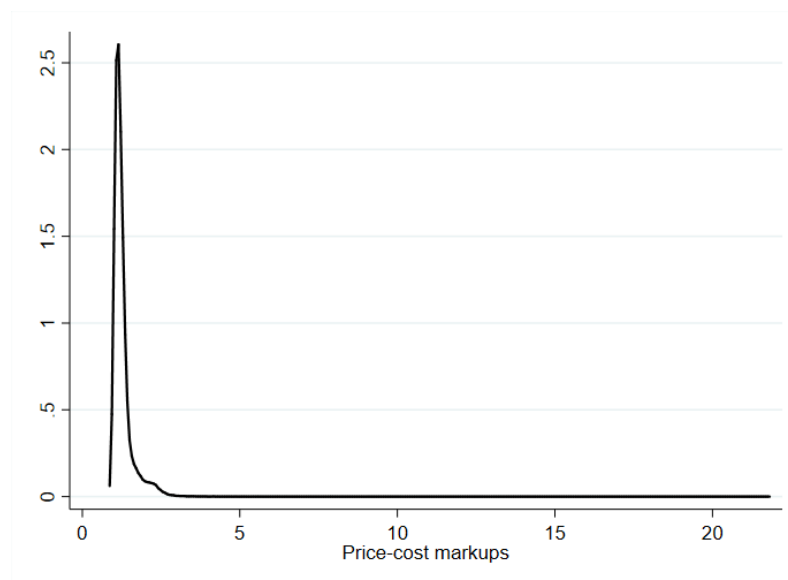
Figure 1.8: Distribution of price-cost markups, 2014.

47
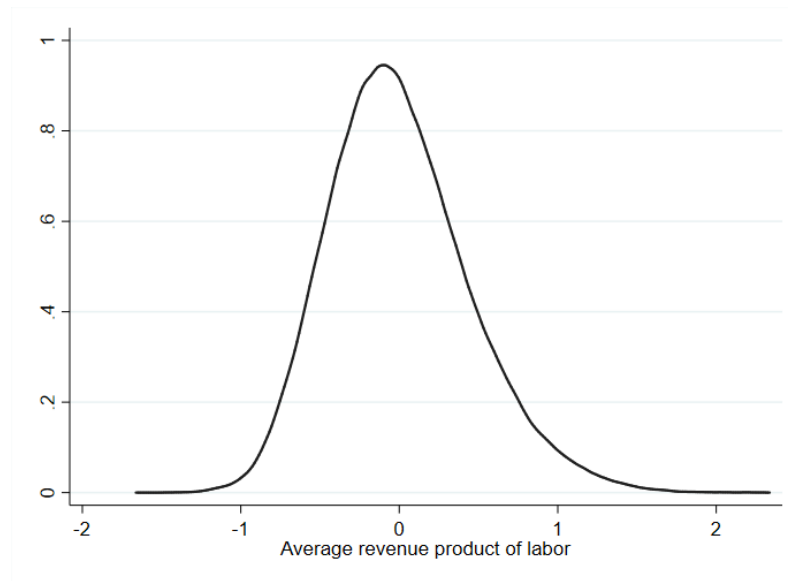
Figure 1.9: Distribution of the average revenue product of labor, 2014.
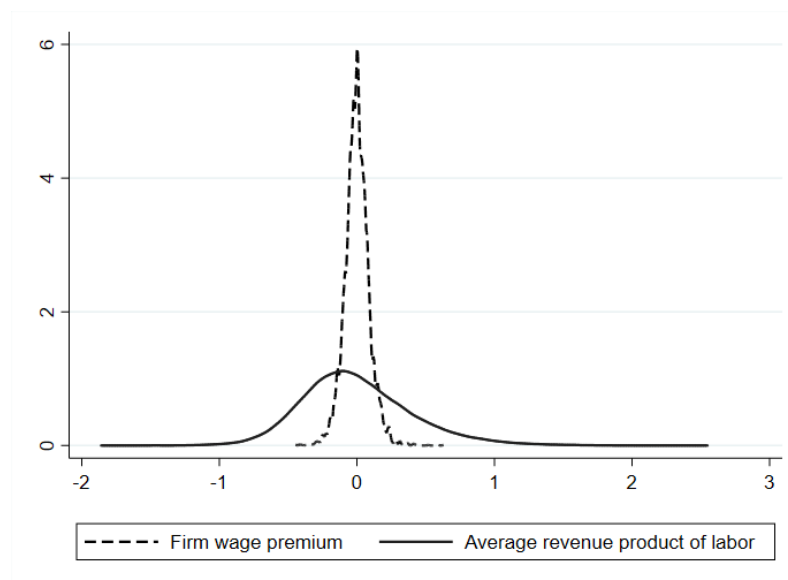
Figure 1.10: Distribution of firm wage premia and the average revenue product of labor, 2014.
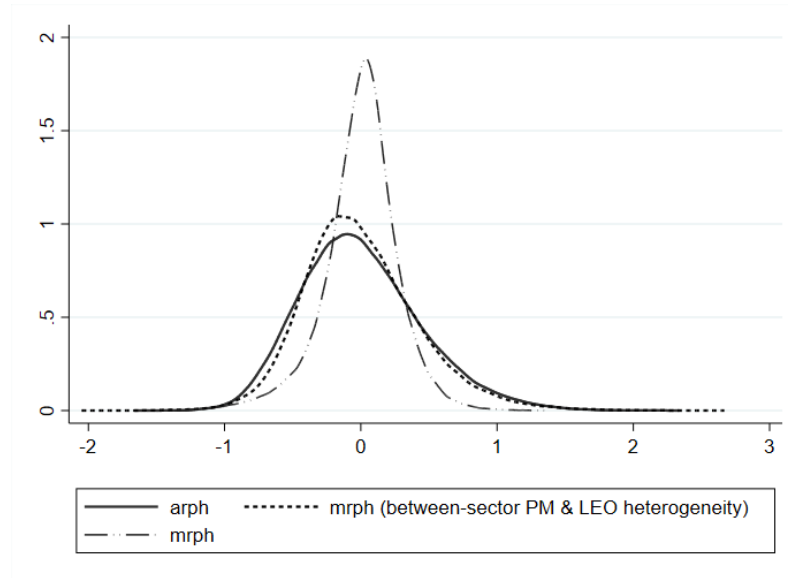
Figure 1.11: Distribution of the average and marginal revenue product of labor, 2014.
The solid line denotes the average revenue product of labor, the short dashed line denotes the marginal revenue product of labor when price-cost markups and labor elasticities of output are sector-specific but not firm-specific, 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 1.7: Summary statistics of firm heterogeneity within 2-digit sectors in 2014.

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 1.8: 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 1.9: 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 1.10: 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 1.11: 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 1.12: 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 1.13: Distribution of log average revenue product of labor in efficiency units by sector in 2014.

1.9.4 Correlations Between Dimensions of Firm Heterogeneity and Observed Firm Characteristics

This section shows that measured price-cost markups, wage markdowns, and labor elasticities of output are correlated with 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 TFPR		0.169***	
		(0.079)	
Firm age			0.010***
			(0.001)
Sector-*département*-year F.E.	$\checkmark$	$\checkmark$	$\checkmark$
Adj. R^2	0.27	0.32	0.31
Total # of firm-years	5,884,663	5,884,663	5,884,663

Table 1.14: 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 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 1994-2014. Standard errors are clustered at the sector-region-year level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

Table 1.14 shows that larger firms have lower price-cost markups, similar to the finding of De Loecker et al. (2020). The table also shows that more productive firms charge higher markups, consistent with models of variable markups such as Atkeson and Burstein (2008), Edmond et al. (2015), and Peters (2017). Moreover, older firms tend to charge higher markups than younger firms, consistent with models of customer accumulation, such as Gourio and Rudanko (2014) and Kaas and Kimasa (2019).

Table 1.15 shows that larger and more productive firms have lower wage markdowns, consistent with predictions of frictional models of wage determination (Burdett and Mortensen, 1998; Postel-Vinay and Robin, 2002; Jarosch et al., 2019) and oligopsonistic models of the labor market (Berger et al., 2018). I do not find any statistically significant difference in wage markdowns between younger and older firms.

Finally, Table 1.16 shows that larger firms (in terms of employment) are more labor

intensive. In addition, firms with higher revenue TFP and younger firms are associated with lower labor elasticities of output.

	wm		
ln Employment	-0.058***		
	(0.015)		
ln TFPR		-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 1.15: 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 TFPR		-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 1.16: 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 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 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:

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

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: $HHI_{s,l,t} = \sum_{jt} \text{SHARE}^2_{j,s(j,t),l(j,t),t}$, where $\text{SHARE}_{j,s(j,t),l(j,t),t}$ 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.

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

Table 1.17: 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	$\checkmark$
Commuting-zone-year FE	$\checkmark$
Adj. R^2	0.451
Total # of market-years	1,308,343

Table 1.18: 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.

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 choose this definition

to allow for some degree of imperfect tradability of goods and services across locations. There are 99 *départements* 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.

Table 1.17 shows that product market concentration is positively correlated with price-cost markups. A 1% higher HHI is associated with a 0.6 percentage point higher price-cost markup. Similarly, Table 1.18 shows that labor market concentration is negatively correlated with wage markdowns, consistent with an interpretation of higher labor market concentration as indicative of stronger labor market power (Azar et al., 2018; Jarosch et al., 2019; Berger et al., 2018). A 1% higher HHI is associated with a 2.4 percentage point lower wage markdown.

	wm
J2J	0.019***
	(0.002)
Sector-year FE	$\checkmark$
Département-zone-year FE	$\checkmark$
Adj. R^2	0.591
Total # of market-years	586,773

Table 1.19: Correlation of wage markdowns with the job-to-job transition rate.

Robust standard errors are reported in parentheses. Sample from 2003 to 2014. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels.

I now examine the relationship between wage markdowns and the job-to-job transition rate of a local labor market by running the following regression:

$$wm_{s,l,t} = \gamma \text{J2J}_{s,l,t} + Sector_s \times Year_t + Location_l \times Year_t + \epsilon_{jt}$$

where $\text{J2J}_{s,l,t}$ denotes the job-to-job transition rate, defined as the fraction of workers who switched employers within a market between year t and $t-1$. As above, the definition of a local labor market here is a 4-digit sector and commuting zone pair. I compute the job-to-job transition rate using the DADS Panel, which allows me to track workers over time as they switch employers. In the case in which workers held two jobs or more in a given calendar year, I choose the one with the highest wage. The DADS Panel does not contain information about the commuting zone in which a firm operates in. I obtain this information from the DADS Postes. Table 1.19 shows that wage markdowns are higher in markets where the job-to-job transition rate is higher, consistent with models where workers search on-the-job (Burdett and Mortensen,

1998; Postel-Vinay and Robin, 2002). A 1 percentage point higher job-to-job transition rate is associated with a 0.019 percentage point higher wage markdown.

1.9.5 Shapley Decomposition Results
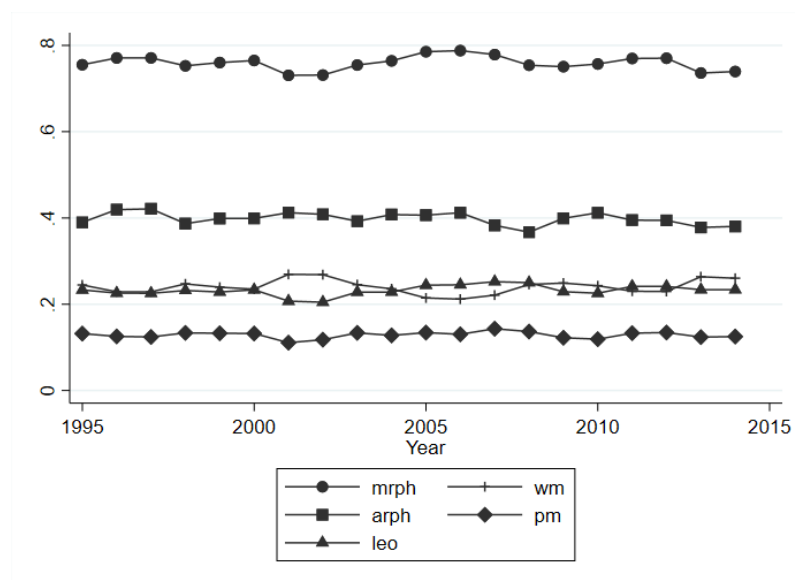
Figure 1.12: Shapley decomposition of the firm wage premium over time

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

Ensemble Decomposition Results

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

Standard Variance Decomposition Results

$$V(\phi) = V(\ln WM) + V(\ln PM) + V(\ln LEO) + V(\ln ARPH)$$
$$+ 2CV(\ln WM, -\ln PM) + 2CV(\ln WM, \ln LEO) + 2CV(\ln WM, \ln ARPH)$$
$$+ 2CV(-\ln PM, \ln LEO) + 2CV(\ln PM, \ln ARPH) + 2CV(\ln LEO, \ln ARPH)$$

	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 1.24: 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 1.25: 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 1.26: 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 1.27: 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 1.28: 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 1.29: 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 1.30: Firm heterogeneity variance-covariance matrix in 2014 (wholesale and retail).

1.9.6 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 1.31: 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 1.32: 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 1.33: 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 1.34: 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 1.35: 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 1.36: 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 1.37: Firm heterogeneity correlation matrix in 2014 (wholesale and retail).

1.9.7 Deriving the Effects of Market Power on Sectoral TFP Through Misallocation

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

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

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

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

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

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

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

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

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

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

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

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

1.9.8 Superstar Firms vs Other Firms
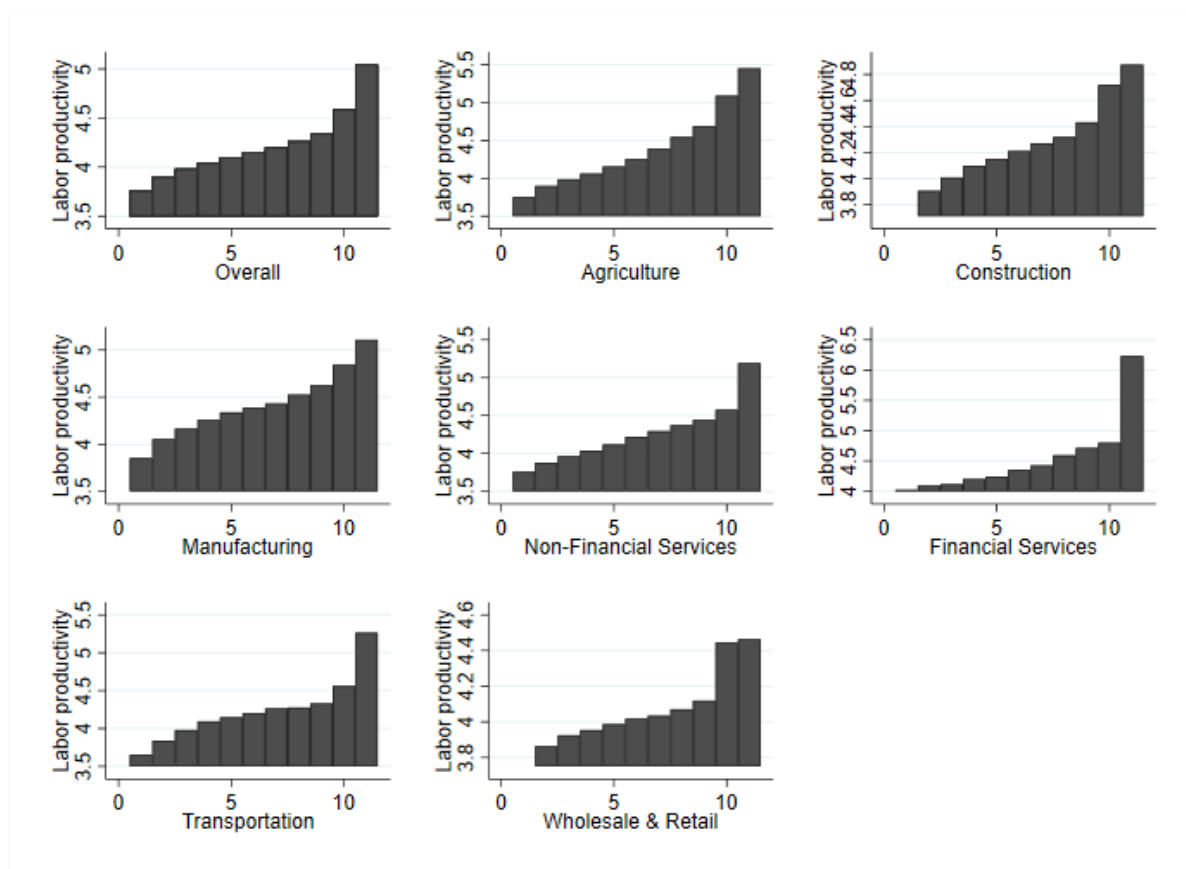
Figure 1.13: Superstar firms and other firms: average revenue product of labor (2014)

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.
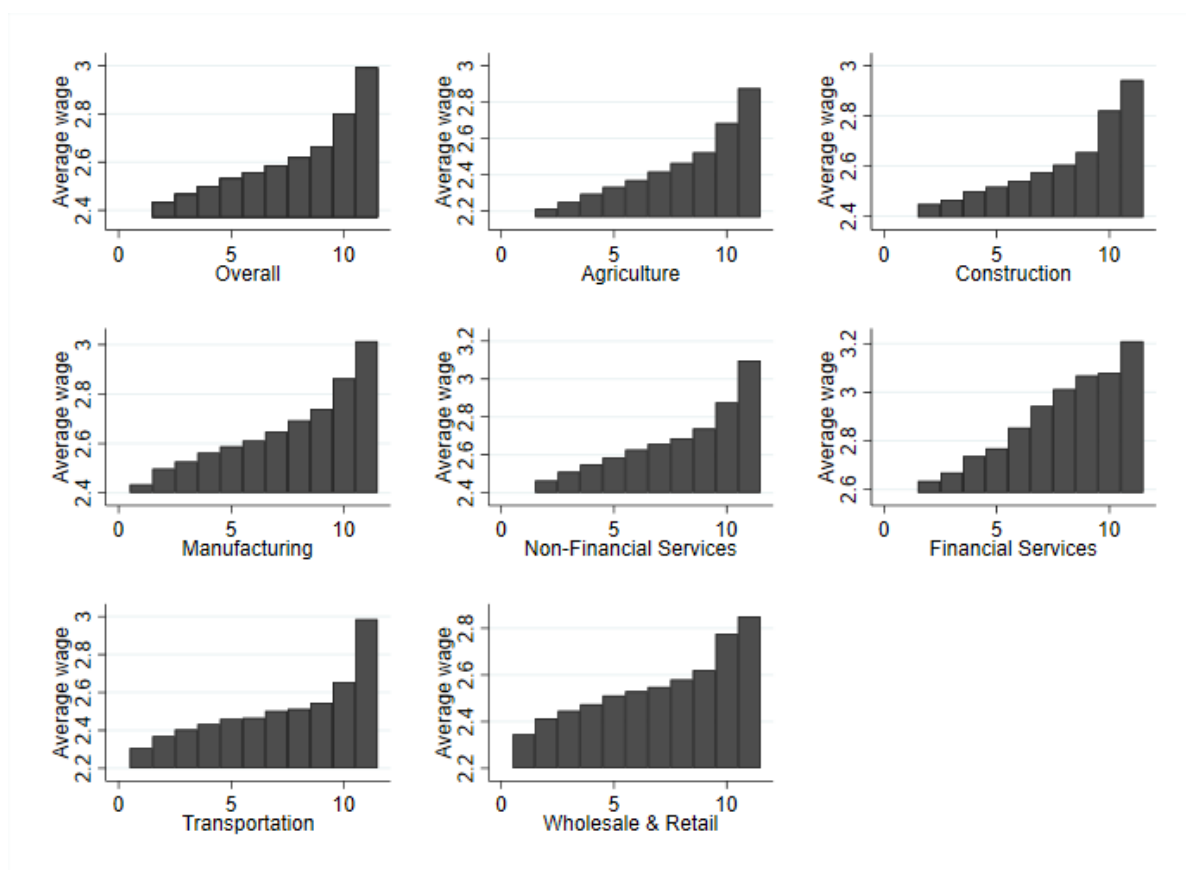
Figure 1.14: Superstar firms and other firms: average wage (2014)
Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.
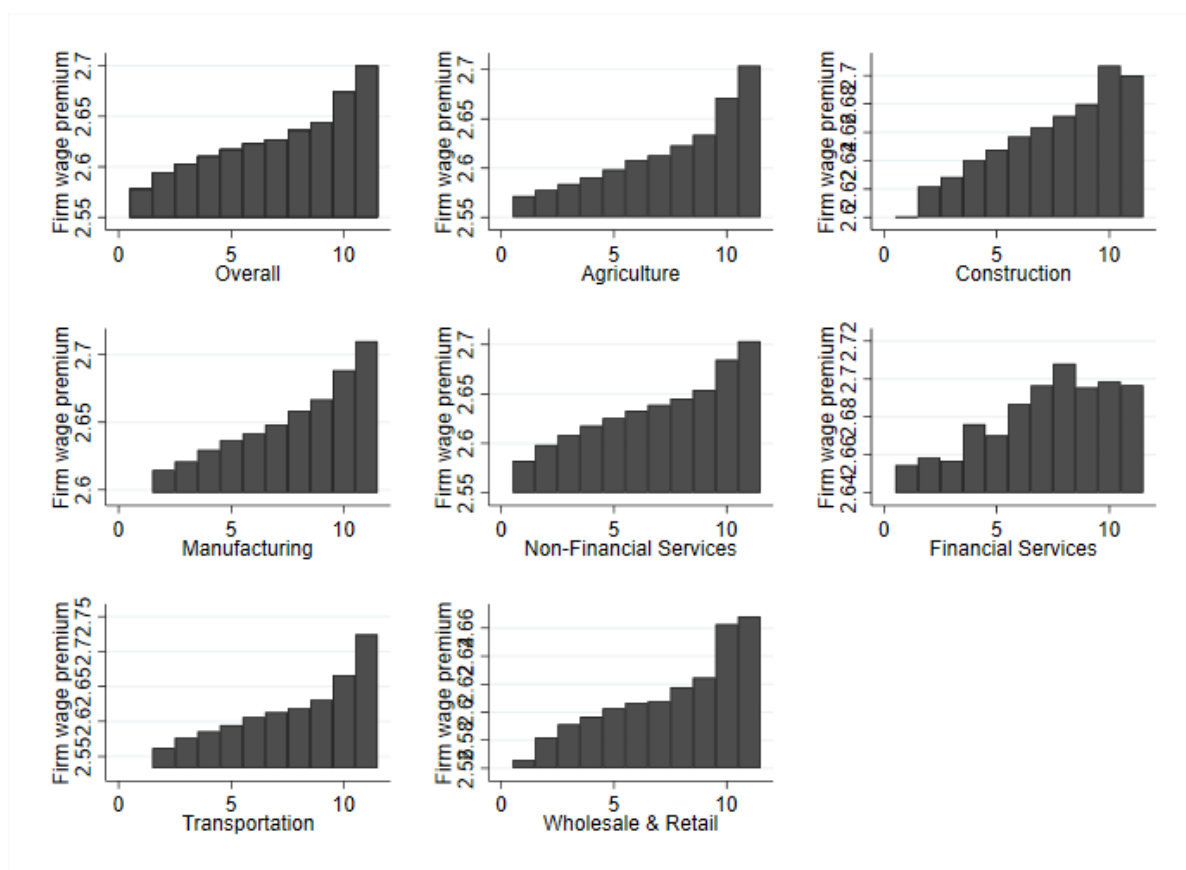
Figure 1.15: Superstar firms and other firms: firm wage premium (2014)

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.
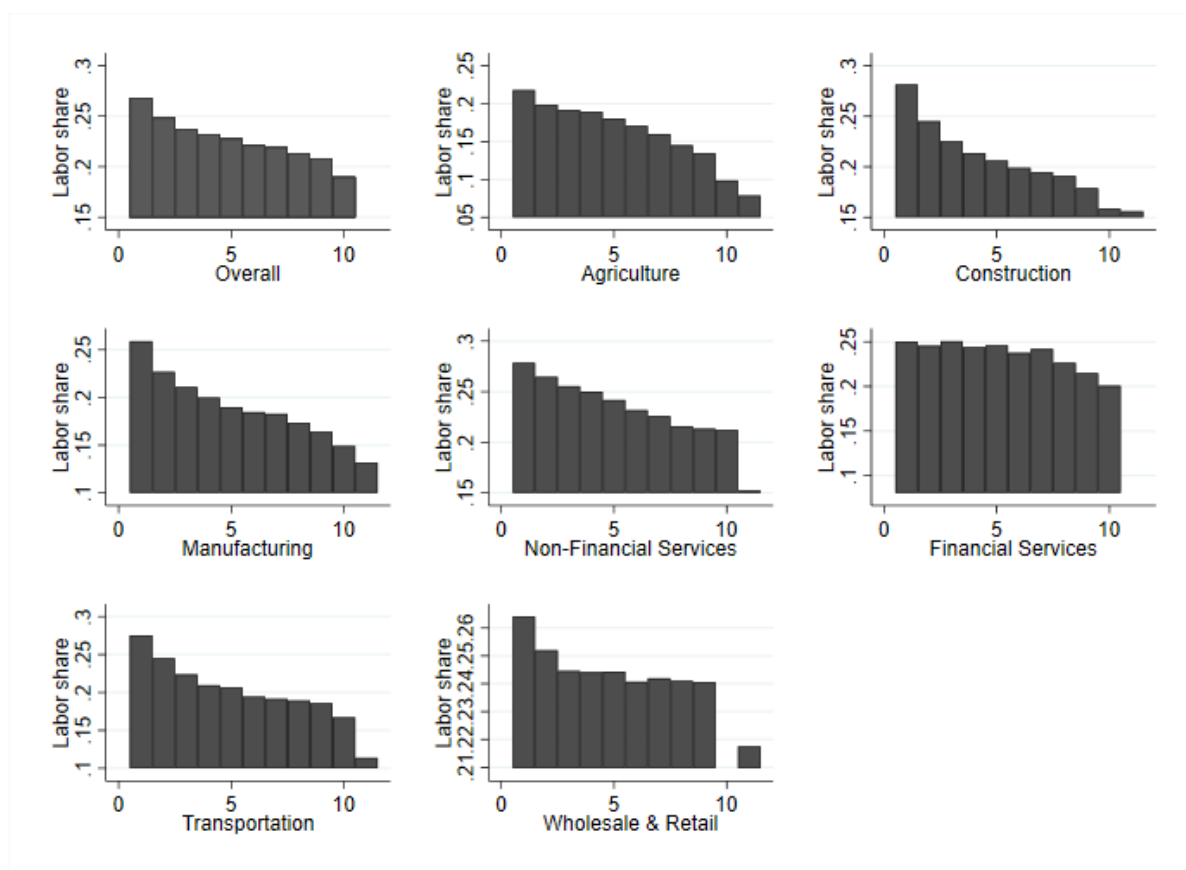
Figure 1.16: Superstar firms and other firms: labor share (2014)

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.
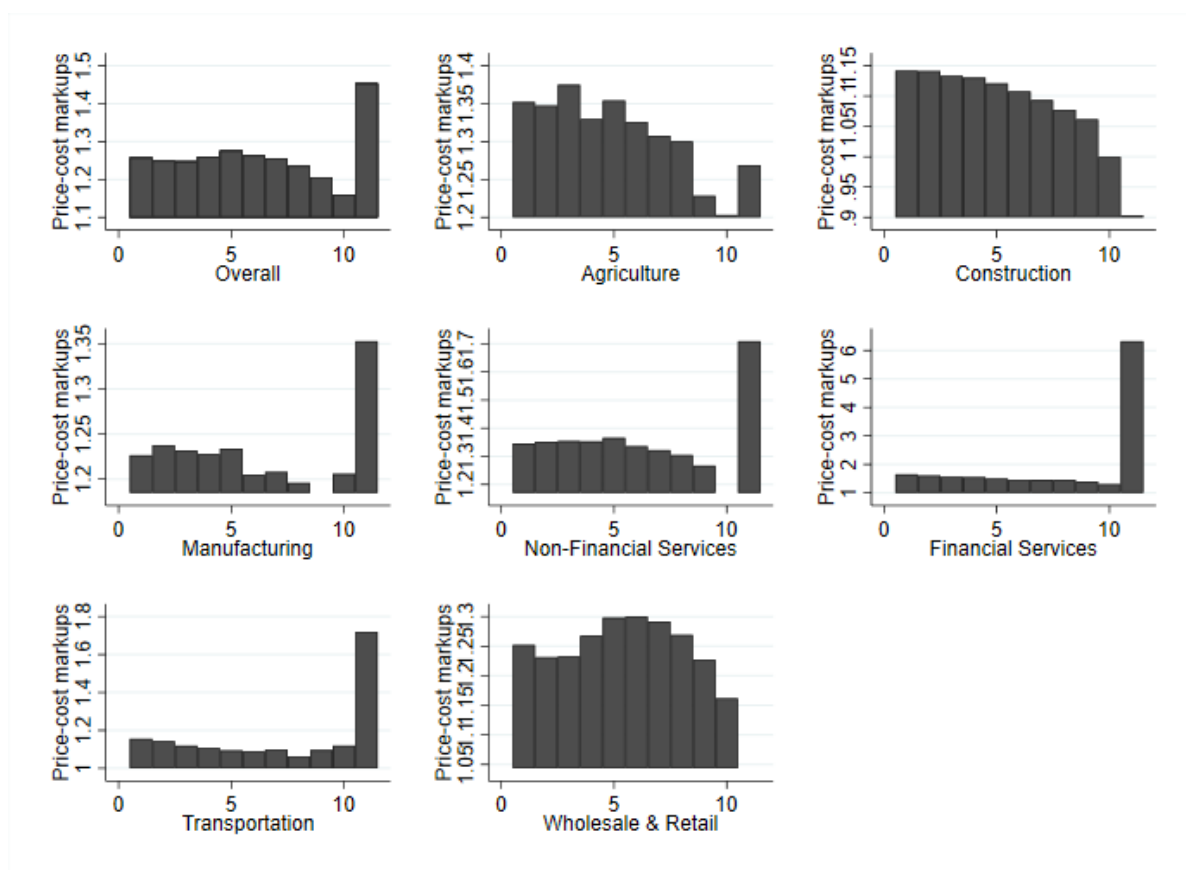
Figure 1.17: Superstar firms and other firms: price-cost markups (2014)

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.
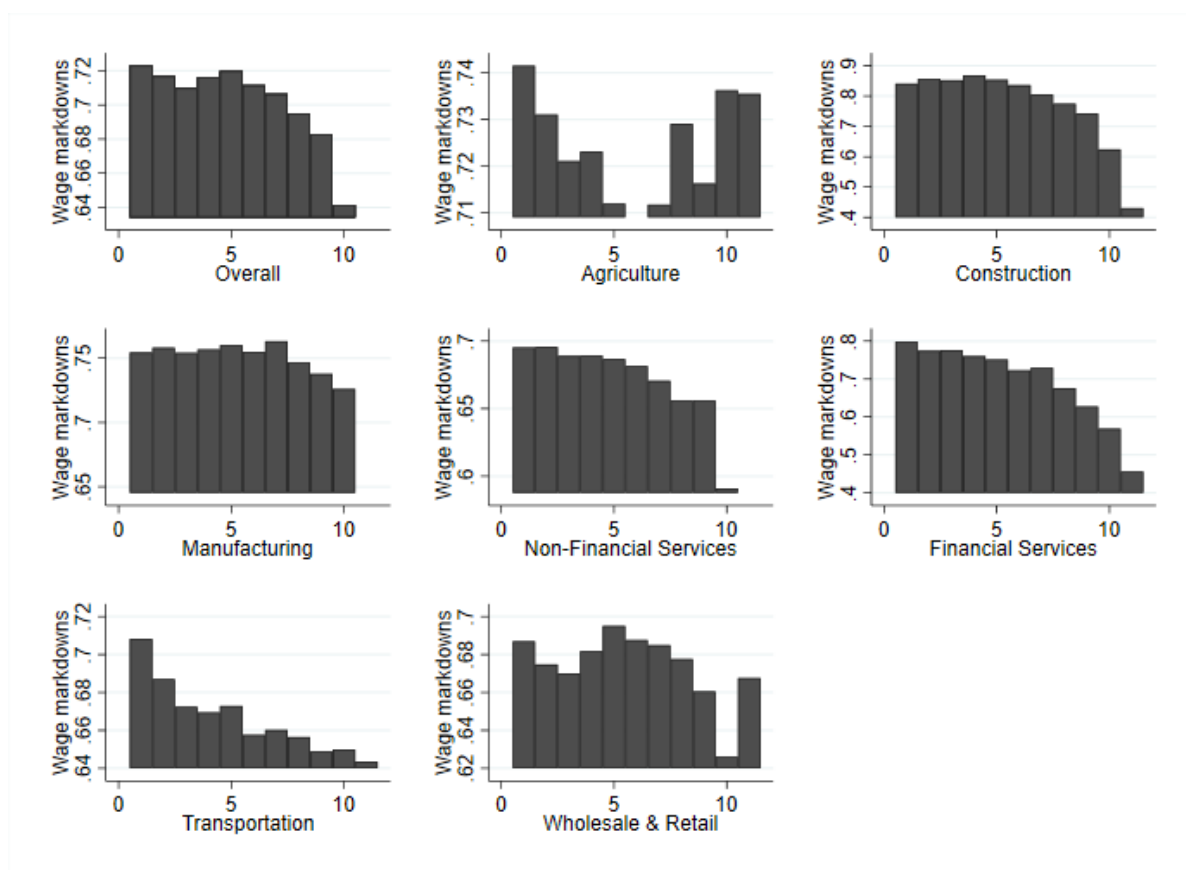
Figure 1.18: Superstar firms and other firms: wage markdowns (2014)

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.
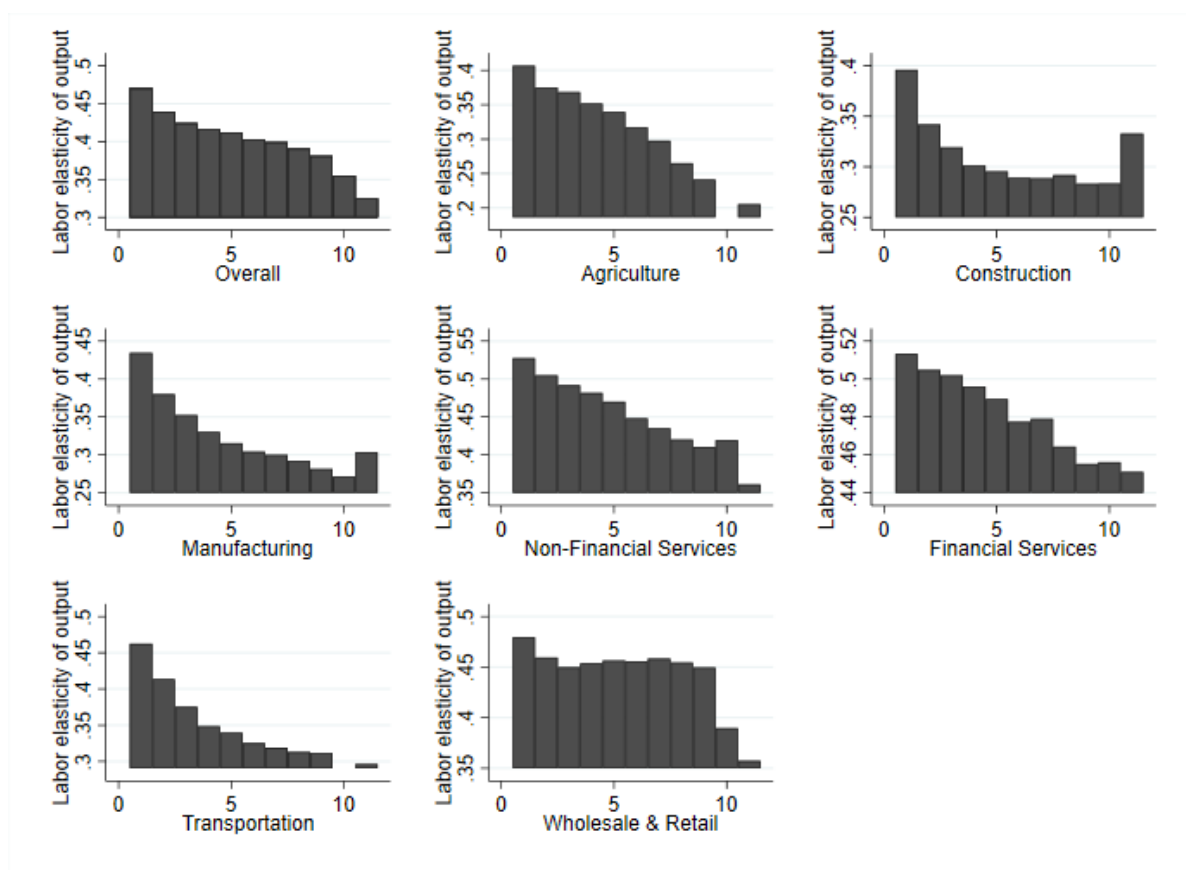
Figure 1.19: Superstar firms and other firms: labor elasticity of output (2014)

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms.

1.9.9 Superstar Firms' Low Labor Revenue Shares: Technology vs Market Power
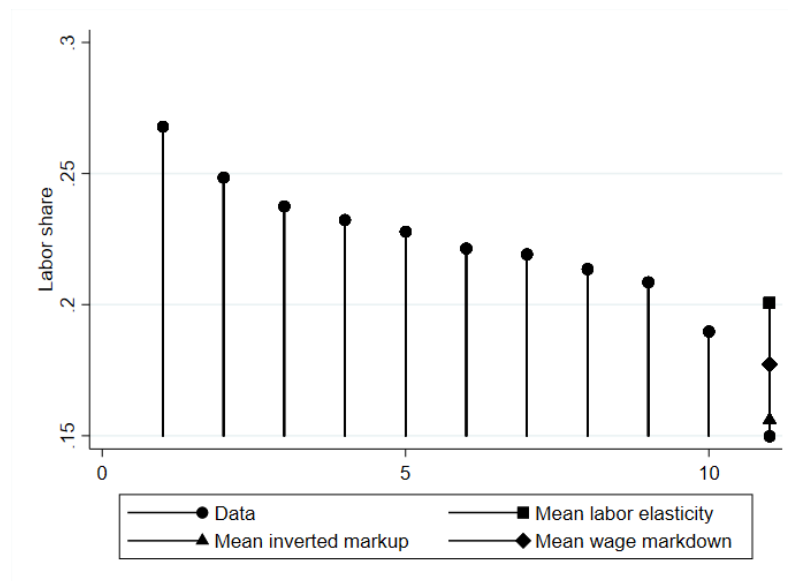
Figure 1.20: Superstar firms' labor share of revenue (2014): Aggregate

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms. The square market shows the implied labor revenue share when I replace superstars' labor elasticities of output with the mean labor elasticity of output among non-superstar firms. The equivalent applies to wage markdowns and price-cost markups.
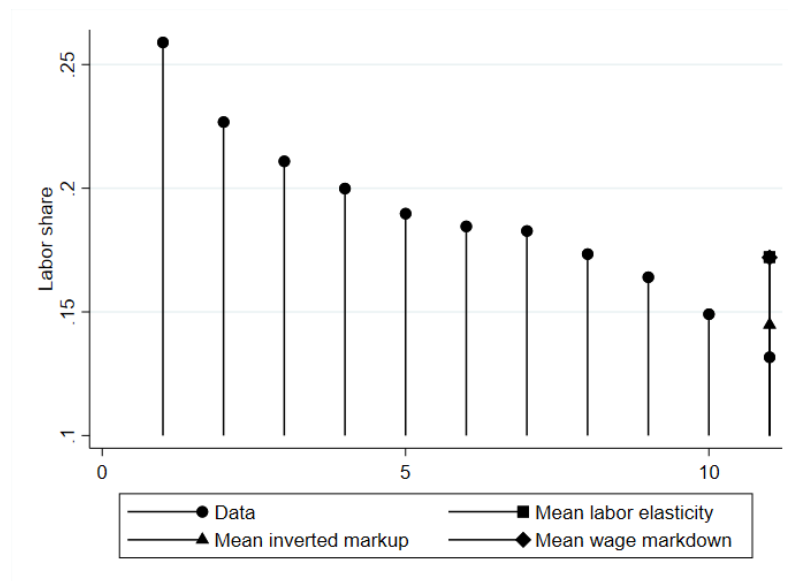
Figure 1.21: Superstar firms' labor share of revenue (2014): Manufacturing

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms. The square market shows the implied labor revenue share when I replace superstars' labor elasticities of output with the mean labor elasticity of output among non-superstar firms. The equivalent applies to wage markdowns and price-cost markups.
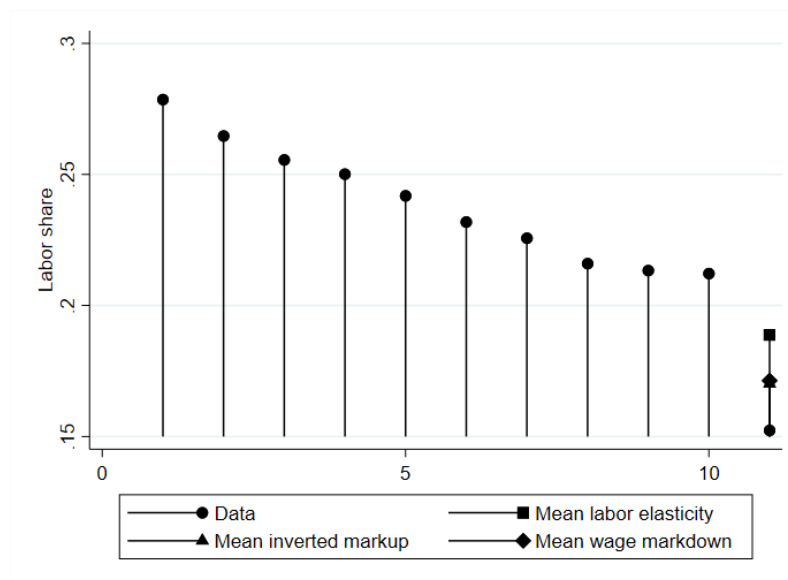
Figure 1.22: Superstar firms' labor share of revenue (2014): Non-Financial Services

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms. The square market shows the implied labor revenue share when I replace superstars' labor elasticities of output with the mean labor elasticity of output among non-superstar firms. The equivalent applies to wage markdowns and price-cost markups.
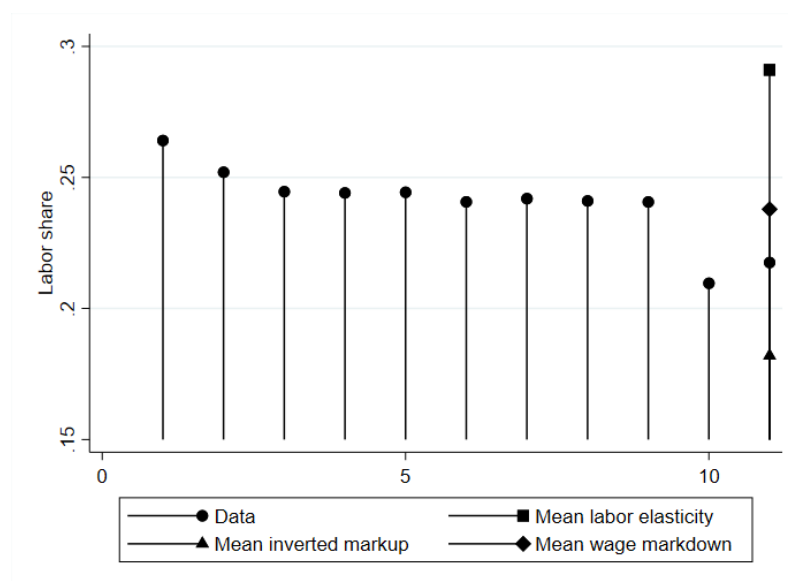
Figure 1.23: Superstar firms' labor share of revenue (2014): Wholesale and Retail

Firm-size (sales) groups 1 to 10 denote "other" firms of increasing sizes, while group 11 denotes "superstar" firms. The square market shows the implied labor revenue share when I replace superstars' labor elasticities of output with the mean labor elasticity of output among non-superstar firms. The equivalent applies to wage markdowns and price-cost markups.

1.10 Appendix C: Wage-Posting and Wage-Bargaining Frameworks

Random Search Wage-Bargaining Framework

The structural framework presented in section 1.2 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_{jt}^e to its workers is:

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

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

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

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

The unemployed worker's value function is:

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

which is a function of the flow value of unemployment b and the expected utility next period. Since there are no aggregate shocks, $U_t = U_{t+1}$. The employed worker's value function is:

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

which depends on the wage Φ_{jt} and non-wage amenities A_{jt} this period through a constant returns to scale utility function $u(.)$, the expected utility next period if the worker is exogenously separated from the firm, and the expected utility if the worker is not exogenously separated. The

last component depends on the expected utility of being employed at the same firm, and the expected utility of moving to a new employer conditional on the new employer offering a higher utility. I assume that: (i) the flow utility function $u(.,.)$ is homogenous of degree one in its inputs, (ii) there is no savings mechanism, (iii) the value of non-wage amenities is proportional to worker efficiency, $A_{ijt} = E_{it}A_{jt}$, and (iv) worker efficiency is allowed vary over time due to random shocks: $E_{it+1} = E_{it} + \zeta_{it+1}$, where ζ_{it+1} is a mean-zero random shock.[58] Therefore, the value of unemployment and employment is proportional to worker efficiency. A worker with efficiency E_{it} obtains a value of $E_{it}U_t$ while unemployed and $E_{it}V_{jt}^e$ while employed.

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

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

subject to the law of motion for employment:

$$H_{jt} = (1 - s(V_{jt}^e))H_{jt-1} + \lambda(V_{jt}^e)V_{jt} \tag{1.14}$$

and (1.2) and (1.3). 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 surplus to be bargained over is:

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

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

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

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

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

---

[58] 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).

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

(1.15)

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

Random Search Wage-Posting Framework

I now replace the wage-setting protocol of the random search framework above with wage-posting and discuss the determinants of the wage markdown. This model generates equation (1.1) and provides one microfoundation for the wage markdown derived from the structural framework presented in section 1.2.

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:

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

(1.16)

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 ($\epsilon^s_{\Phi,jt} < 0$) 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 (1.1) can be obtained and the wage markdown is identical as the model with random search.[59] 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^e_{jt} at each firm j, before matching takes place. The pre-matching value to the unemployed worker is:

$$U^{bm}_t = \max_{V^e_{jt}}(1 - f(\theta(V^e_{jt})))U_t + f(\theta(V^e_{jt}))V^e_{jt}$$

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

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

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 $\underline{V}^e_t$, this worker optimally searches for employment at firm j, where $V^e_{jt} > \underline{V}^e_t$. Denote this unique solution as $V^e_{jt} = v(\underline{V}^e_t)$. Therefore, firm j recruits workers from this market. Similarly, firm j loses workers due to quits to a higher utility firm who pays $\overline{V}^e_t$. The optimal search strategy of a worker employed at firm j is then $\overline{V}^e_t = v(V^e_{jt})$. Next, note that the firm-specific separation rate is now $s_{jt} = \delta_s + (1 - \delta_s)sf(\theta(\overline{V}^e_t))$. Using the law of motion for employment, the firm-specific "labor supply" curve is then:

$$H_{jt} = (1 - s(\overline{V}^e_t))H_{jt-1} + q(\underline{V}^e_t)V_{jt}$$
$$= (1 - s(V^e_{jt}))H_{jt-1} + q(V^e_{jt})V_{jt}$$

---

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

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 (1.1) gives the same wage markdown expression as the random search wage-posting model above.

Workers' search behavior in a Directed Search Model

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

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

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

Under utility maximization:

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

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

which implies that:

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

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

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

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

Workplace Differentiation Monopsonistic Wage-Posting Framework

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

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

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

Each worker i maximizes utility by choosing where to work:

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

The firm-specific labor supply curve is then:

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

Firm j's profit-maximization problem is:

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

subject to the firm-specific labor supply curve and equations (1.2) and (1.3). This model gives the firm wage premium equation (1.1) 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).

Chapter 2

Why Do Larger Firms Pay Higher Wages?

2.1 Introduction

Larger firms pay higher wages. Established at least since the work of Moore (1911), this well-known empirical relationship holds across a host of countries (Brown and Medoff, 1989; Schmidt and Zimmermann, 1991; Main and Reilly, 1993; Troske, 1999; Oi and Idson, 1999; Haltiwanger et al., 2017; Berlingieri et al., 2018; Bloom et al., 2018). Understanding the relationship between firm size and wages is important because large firms employ a disproportionately large share of workers (Autor et al., 2017), have a large impact on young workers' career progression (Arellano-Bover, 2020), drive fluctuations in employment (Moscarini and Postel-Vinay, 2012), and job losses at large firms typically come with sizable long-term wage losses (Schmieder et al., 2018; Helm et al., 2020).

What explains why larger firms pay higher wages? This paper first documents how the relationship between firm size and wages differs across sectors using detailed worker and firm microdata from France, and then disentangles the drivers behind this relationship using a structural model-based decomposition. The structural framework features heterogeneous firms and an imperfectly competitive labor market, allowing widely hypothesized channels – rent-sharing and efficiency wage considerations – to explain the wage-size relationship.

My main findings are twofold. First, while it is well-known that larger firms pay significantly higher wages in the manufacturing sector, this relationship is considerably weaker in the non-financial service and the wholesale and retail sectors. Both are larger sectors than manufacturing in terms of employment. Second, to account for the different wage-size patterns between manufacturing and non-manufacturing firms, my decompositions point to two reasons: (i) larger firms tend to employ more productive workers in the manufacturing sector, but the positive relationship between worker productivity and firm size is considerably weaker among non-manufacturing firms; (ii) rent-sharing appears to be the main driver behind why identical workers earn higher wages at larger firms in the manufacturing and non-manufacturing sectors. However, there are less rents to be shared at large firms in the latter than in the former.

To revisit the empirical relationship between firm size and wages, I use large and highly

detailed administrative micro-data from France. Existing work on this topic is largely based on data on manufacturing firms or worker survey data due to the lack of information on non-manufacturing firms (Troske, 1999; Main and Reilly, 1993).[1] The linked worker-firm administrative datasets from France provide information on hourly wages and employment beyond manufacturing firms, comprising of firms in all sectors of the French economy. This makes it possible to assess whether two of the largest French business sectors in terms of employment size – non-financial services and wholesale-retail – show similar patterns between firm size and wages. Relative to worker survey data, the linked worker-firm data contains a panel structure, which makes it possible to estimate worker and firm-specific contributions to the wage-size relationship (Abowd et al., 1999). The French administrative datasets confer a further advantage: the linked worker-firm data can be matched with firm balance sheet information for all active firms in France, offering a chance to investigate the mechanisms underlying the wage-size relationship.

My findings show that the positive relationship between firm size and wages in France is substantially stronger in the manufacturing sector than in the non-financial service sector and wholesale and retail sector. In the manufacturing sector, firms with more than 2000 employees pay hourly wages that are on average 40% to 50% higher than firms with less than 20 employees. However, in the non-financial service sector and the wholesale and retail sector, the corresponding number is between 10% and 20%. This is consistent with recent evidence that the long-term wage loss of job displacement is larger in manufacturing than in non-financial services (Helm et al., 2020).

I then decompose the wage-size relationship by estimating the contributions of worker and firm heterogeneity directly. I do this by estimating a twoway fixed effect regression, in which the firm-specific component of wages is separately identified from the worker-specific component by using information on firm-to-firm worker mobility (Abowd et al., 1999; Bonhomme et al., 2019).[2]

My results show that differences in worker composition account for the bulk of the wage-size relationship and how this relationship differs between sectors. However, the firm-specific component of wages remains important for understanding higher wages at larger firms. Across sectors, approximately 3/4 to 4/5 of the firm size-wage relationship is tied to larger firms employing more productive workers. Further, the profile of average worker productivity over firm size follows closely the profile of hourly wages over firm size in each sector, accounting for the different firm size-wage relationship in the manufacturing and non-manufacturing sectors.

To explore the mechanisms behind the relationship between firm size and the firm component of wages, I implement a decomposition of this relationship derived from a structural model. In the model, homogenous workers and heterogeneous firms meet in a search frictional labor market. Firms therefore pay workers a wage that is a markdown below their marginal revenue productivity. The relationship between the wage markdown, marginal revenue product of labor, and firm size contains information about widely hypothesized channels behind why larger firms pay higher wages – rent-sharing and efficiency wage considerations (Brown and

---

[1] The few recent exceptions are Bloom et al. (2018), Colonelli et al. (2018), and Berlingieri et al. (2018).

[2] I address the main concerns with twoway fixed effect regressions in Wong (2020) in the context of France.

Medoff, 1989; Manning, 2011). The rent-sharing explanation implies a positive relationship between firm size and the marginal revenue product of labor, controlling for worker composition differences. At the same time, the model predicts that larger firms pay a lower fraction of workers' marginal revenue product as wages (larger wage markdown), since firms paying higher wages face lower labor market competition locally, partially offsetting the higher wages at larger firms (Gouin-Bonenfant, 2018). In contrast, if larger firms pay higher wages because it is more costly for them to monitor workers' effort, then larger firms tend to pay a higher fraction of workers' marginal revenue productivity as wages to ensure that workers do not shirk.

My decomposition suggests rent-sharing as the main driver behind the relationship between firm size and the firm-component of wages in both the manufacturing and non-financial service sectors. In the wholesale and retail sector, efficiency wage considerations appear to be potentially more relevant drivers among larger firms.[3] In the manufacturing and non-financial service sectors, the marginal revenue product of labor (MRPH), net of worker composition differences, is monotonically increasing with firm size. In manufacturing, a firm with over 2000 employees has an MRPH that is on average 30% higher than firms with less than 20 employees. In non-financial services, this number is 20%. However, much of this is offset by larger firms having a larger wage markdown. In manufacturing, firms with more than 2000 employees pay on average approximately 65% of the MRPH as wages, while firms with less than 20 employees pay on average 75% as wages. In non-financial services, the corresponding numbers for firms with more than 2000 employees and those with less than 20 employees are 60% and 70%. Among wholesale and retail firms, these patterns differ – the profile of the MRPH is U-shaped over firm size, while the profile of wage markdowns is inverted U-shaped. This suggests that efficiency wage considerations and non-wage amenities could play a role in explaining the wage-size relationship among larger wholesale and retail firms.

The literature seeking to understand the relationship between firm size and wages largely documents the positive wage-size relationship using worker survey data or manufacturing sector data. Brown and Medoff (1989) and Oi and Idson (1999) use worker-level survey data to study a range of reasons for the size-wage relationship. Troske (1999), Schmidt and Zimmermann (1991), and Main and Reilly (1993) use manufacturing firm data to study the wage-size relationship. Bloom et al. (2018) show that the wage-size relationship is disappearing in the US. Colonelli et al. (2018) show that the size-wage relationship displays different trends across countries. This literature advances four main hypotheses for the wage-size relationship: worker composition, rent-sharing, higher productivity standards at larger firms, and costly monitoring of worker efforts (Manning, 2011).

This paper contributes in a few ways to this body of work. First, this paper contributes by documenting how this relationship differs across sectors using comprehensive matched employer-employee data with information on hourly wages and firm balance sheet data. This combination of data makes it possible to (i) decompose the wage-size relationship into worker and firm components, (ii) decompose the firm component into the contributions of various

---

[3]To disentangle the importance of non-wage amenities from efficiency wages, this project aims to infer the value of non-wage amenities from workers' firm-to-firm transitions that entail a wage loss not explained by mass layoffs (Sorkin, 2018). This work is left for the near future as recent university closures due to the COVID-19 pandemic has hindered access to data.

firm characteristics, and (iii) deal with differences in earnings due to differences in work hours. Second, this paper presents a new approach to understanding the drivers of the wage-size relationship – rent-sharing and efficiency wage considerations have different implications for how the wage markdown differs across firm size. Third, this paper shows that worker composition and rent-sharing appear to be the main channels for the wage-size relationship and how it differs across sectors. The key assumption underlying my approach is that labor markets are imperfectly competitive.

The most closely related paper is Berlingieri et al. (2018), who also find a flatter relationship between firm size and wages per worker in the service sector than the manufacturing sector using balance sheet panel data of firms in 24 OECD countries. Relative to their work, this paper (a) shows that the same is also true in the wholesale and retail sector, (b) uses matched employer-employee data, which makes it possible to assess the importance of worker composition differences and control for differences in hours worked, and (c) proposes a structural decomposition to learn about the potential drivers behind the wage-firm size relationship.

2.2 A Model to Interpret the Wage-Size Relationship

To interpret the empirical relationship between wages and firm size, and the empirical findings presented in Section 2.4, I start by setting up a frictional labor market model drawing on Kaas and Kircher (2015). In the model, the relationship between firm size and wages reflect rent-sharing and efficiency wages (Shapiro and Stiglitz, 1984; Piyapromdee, 2018). I use the model to study what rent-sharing and costly effort-monitoring imply for the relationship between the wage markdown, the marginal revenue product of labor, and firm size. I show that rent-sharing implies that the wage markdown is larger for larger firms, while costly monitoring implies the opposite.

Environment. Time is continuous and the economy is in the steady state. The final goods market is perfectly competitive. Let the final goods price be the numeraire. Workers engage in directed search in a labor market with search frictions. Workers do not search while on the job. The matching technology is Cobb-Douglas. Define the job-finding rate as $f_j = \lambda_j^{\kappa-1}$ and the job-filling rate as $m_j = \lambda_j^{\kappa}$, where λ_j is the queue length for a vacancy and $\kappa \in (0,1)$ the elasticity of the job-filling rate with respect to the queue length.

Workers. Workers are homogenous. They direct their search by choosing from a menu of wages, job-finding probability, and the effort-monitoring rate. Workers can choose to shirk or exert effort at a utility cost ϵ. A worker who exerts effort produces a unit of output. Workers who choose to shirk will be caught at a rate q_j, depending on firm j. Let Φ_j represent the wage paid by firm j, r the discount rate and δ the exogenous rate of job destruction. The value function of a shirker ($V_{E,j}^S$) and non-shirker ($V_{E,j}^N$) while being employed at firm j is:

$$rV_{E,j}^S = \Phi_j + (\delta + q_j)(V_{U,j} - V_{E,j}^S) \tag{2.1}$$

$$rV_{E,j}^N = \Phi_j - \epsilon + \delta(V_{U,j} - V_{E,j}^N) \tag{2.2}$$

The value of unemployment is:

$$rV_{U,j} = b + f_j(\max(V_{E,j}^S, V_{E,j}^N) - V_{U,j}) \tag{2.3}$$

where b represents the flow value of unemployment. In equilibrium, the value of unemployment ($V_{U,j} = V_U$) is equalized across firms and all workers exert effort ($V_{E,j}^N \geq V_{E,j}^S$). The former, a no-shirking condition, implies an inverse relationship between wages and the monitoring rate:

$$\Phi_j = b + \epsilon + \rho + \left(\frac{(r+\delta)\epsilon}{q_j}\right) \tag{2.4}$$

The latter, an indifference condition, implies an inverse relationship between wages and the job-finding rate:

$$\Phi_j = b + \epsilon + \rho + \left(\frac{(r+\delta)\rho}{f_j}\right) \tag{2.5}$$

where $\rho \equiv rV_U - b$.[4] Equations (2.4) and (2.5) show that higher wages serves two purposes: (i) it increases the job-filling rate by attracting more jobseekers, and (ii) it discourages shirking by increasing the penalty of getting caught.[5]

Firms. There is an exogenous mass J of firms. Firms produce output with labor inputs using a decreasing returns-to-scale production technology, $Y_j = X_j F(H_j)$, where X_j represents firm j's productivity. Firms maximize steady state profits by posting a job-filling rate (m_j), vacancies (V_j), and monitoring rates (q_j), taking into account the no-shirking condition (2.4) and indifference condition (2.5):

$$\Pi_j = \max_{m_j, V_j, q_j} X_j F(H_j) - \Phi_j H_j - c(V_j) - \Psi(q_j)H_j$$

where the vacancy posting cost $c(.)$ and monitoring cost $\Psi(.)$ are both twice differentiable and convex. This gives a wage equation of the form:

$$\Phi_j = WM_j \cdot MRPH_j \tag{2.6}$$

where the $MRPH$ denotes the marginal revenue product of labor and WM denotes the wage markdown – the fraction of $MRPH$ paid as wages. Since the labor market is frictional, firm heterogeneity in productivity X leads to dispersion in the $MRPH$ and wages. It is well-known that in such models, more productive firms are larger and pay higher wages. In this model, this is because: (a) more productive firms make larger rents in the labor market, and (b) more

---

[4]As Kaas and Kircher (2015) show, the equilibrium value of ρ is pinned down by a free entry condition $\bar{\Pi}(\rho) = f_e$, where $\bar{\Pi}(\rho)$ represents expected profits and f_e denotes the fixed entry cost.

[5]The role that wages play in attracting jobseekers in this directed search model is similar to classic wage posting models such as Burdett and Mortensen (1998) and Weiss (1980). However, since workers in my model do not search on the job, wages do not affect worker quits and retention (Salop, 1979). In a richer model with on the job search, wages also play the role of retention, as in Burdett and Mortensen (1998) and Piyapromdee (2018).

productive firms grow larger, implying a higher total effort-monitoring cost, and hence higher efficiency wages.

Firm size and wage markdowns with costly monitoring. How does the wage markdown differ across the productivity distribution, and hence firm size? The wage markdown can be written as:

$$WM_j = \left(1 + \frac{1}{\epsilon_{\Phi,j}^H} + \left[\frac{\epsilon_{\Phi,j}^\Psi}{\epsilon_{\Phi,j}^H} + \frac{\Psi_j}{\Phi_j}\right]\right)^{-1} \tag{2.7}$$

where $\epsilon_{\Phi,j}^H$ represents the elasticity of firm size with respect to wages and $\epsilon_{\Phi,j}^\Psi$ represents the elasticity of monitoring costs with respect to wages. In standard wage-posting models, the wage markdown is a function of only labor supply elasticities ($\epsilon_{\Phi,j}^H$), since effort-monitoring is costless (Burdett and Mortensen, 1998). That is, the terms in the square brackets, which capture changes in the costs of monitoring workers, would be absent. In these models, it is well-known that labor supply elasticities are decreasing in firm productivity and firm size. In my model, the intuition is that since the largest and most productive firms pay the highest wages, their vacancies attract the most number of jobseekers, hence they face less labor market competition locally. Therefore, when rent-sharing is the only driver behind the wage-size relationship, larger firms markdown wages more (lower WM).

With costly effort-monitoring, the wage markdown also depends on how monitoring costs change in response to changes in wages (in square brackets). To see how this component changes across over firm size, we can rewrite equation (2.7) as:

$$WM_j = \left(1 + \frac{\left(\frac{1-\kappa}{\kappa}\right)(r+\delta)\rho m_j^{\frac{1-\kappa}{\kappa}}}{\Phi_R + (r+\delta)\rho m_j^{\frac{1-\kappa}{\kappa}}} + \left[\frac{\theta\left(\frac{\epsilon}{\rho}m_j^{\frac{\kappa-1}{\kappa}}\right)^\eta}{\Phi_R + \left(\frac{1-\kappa}{\kappa}\right)(r+\delta)\rho m_j^{\frac{1-\kappa}{\kappa}}} - \eta\frac{1-\kappa}{\kappa}\right]\right)^{-1}$$

where $\Phi_R \equiv b + \epsilon + \rho$ denotes the reservation wage and I assume the monitoring cost takes the form $\Psi(q_j) = \theta q_j^\eta$, with $\eta > 1$. Since more productive firms wish to grow larger, they choose a higher job-filling rate (m_j), which entails a higher wage. The above equation shows that the first term, which represents the inverse of the labor supply elasticity ($\frac{1}{\epsilon_\Phi^H}$), is increasing in m_j, implying that the fraction of $MRPH$ paid as wages is decreasing in m_j. The second term in square brackets, due to costly monitoring, is decreasing in m_j, which implies that the fraction of $MRPH$ paid as wages in increasing in m_j. The intuition for this second component is that a more productive firm wishes to grow larger, which leads to higher monitoring costs. By paying a higher fraction of $MRPH$ as wages, the firm increases the cost to workers of shirking, thereby reducing firms' monitoring costs.

How do wage markdowns and the marginal revenue product of labor ($MRPH$) differ over firm size in the presence of costly effort monitoring? Figures 2.2 and 2.1 illustrate these relationships when monitoring costs are high versus when they are low.[6] When monitoring costs

---

[6]The calibrated parameter values for this illustration are in table 2.1.
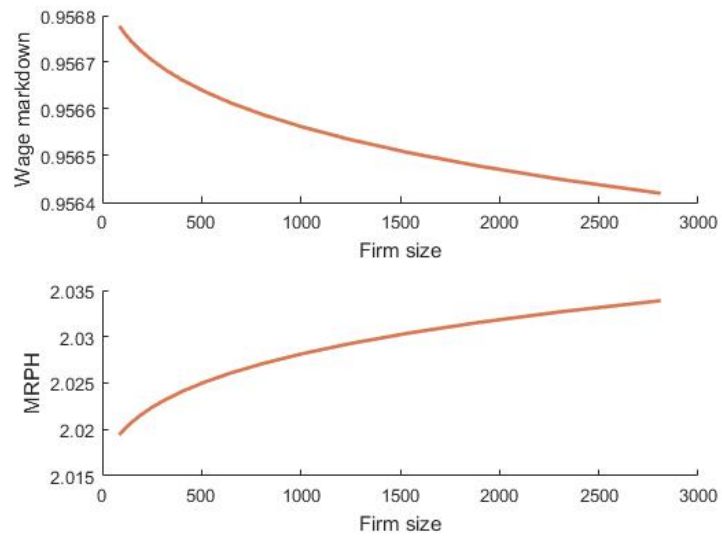
Figure 2.1: Firm size, wage markdowns, and MRPH under low monitoring costs ($\theta = 0.1$, $\eta = 0.1$)
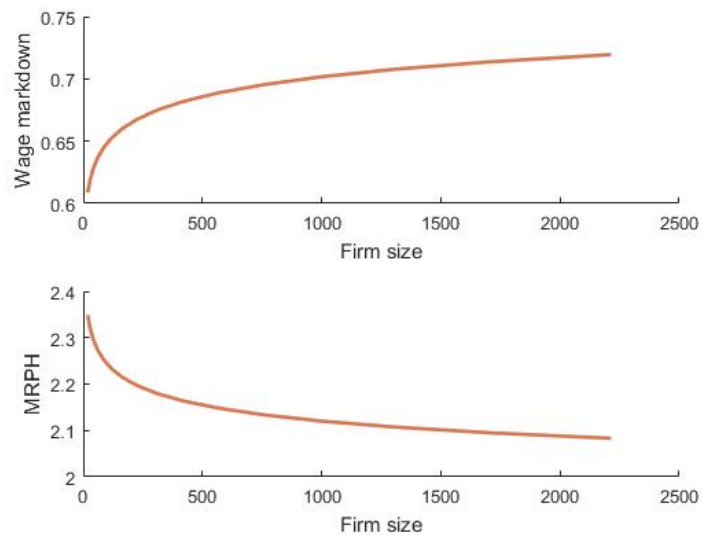
Figure 2.2: Firm size, wage markdowns, and MRPH under high monitoring costs ($\theta = 1$, $\eta = 1.1$)

are low ($\theta = 0.1$, $\eta = 0.1$), efficiency wage considerations are negligible. Therefore, as figure 2.1 shows, the model makes the standard prediction that larger firms have a higher MRPH and pay a lower fraction of MRPH as wages. This is because, while more productive firms wish to grow larger, labor market frictions partially hold back their growth, implying that they have a higher $MRPH$ due to diminishing marginal returns. At the same time, they face lower labor market competition locally since they are the higher-paying firms. When monitoring costs are high ($\theta = 1$, $\eta = 1.1$), efficiency wage considerations become important. Figure 2.2 shows that in this case larger firms pay a higher fraction of $MRPH$ as wages and they also have lower $MRPH$. This is because more productive firms are now not only willing to pay a higher wage to grow larger, they are also willing to pay a higher wage to reduce the costs of monitoring worker effort. This attracts even more jobseekers to more productive firms' vacancies. This effect further increases their size, which reduces their $MRPH$ through diminishing marginal returns.

Firm size and wage markdowns with effort compensation. An alternative, but similar, hypothesis for the positive wage-size relationship advanced in the literature is that larger firms pay higher wages, not to discourage workers from shirking, but to encourage workers to exert more effort (Brown and Medoff, 1989; Oi and Idson, 1999; Colonelli et al., 2018). Similarly, higher effort requirements may boost worker morale (Akerlof, 1982). In such a model, if compensation for exerting a higher level of effort at larger firms is the main driver of the wage-size relationship, then the fraction of $MRPH$ paid as wages is increasing over firm size. To see this, I start by assuming that effort can be monitored costlessly by firms, effort is a continuous variable (as opposed to binary), and it is a choice of the firm. The value function of employed workers become:

$$rV_{E,j} = \Phi_j - \epsilon e_j + \delta(V_{U,j} - V_{E,j}^N)$$

where e_j denotes the effort requirement of firm j. The value function of unemployed workers remain the same (see equation (3). In equilibrium, the relationship between wages (Φ_j), the job-filling rate (m_j), and effort requirement (e_j) is:

$$\Phi_j = b + \epsilon e_j + \rho + \left(\frac{(r+\delta)\rho}{f_j}\right)$$

This equation shows that wages are increasing in firms' effort requirements. The wage markdown becomes:

$$WM_j = \left(1 + \frac{1}{\epsilon_{\Phi,j}^H}\right)^{-1} = \left(1 + \frac{\left(\frac{1-\kappa}{\kappa}\right)(r+\delta)\rho m_j^{\frac{1-\kappa}{\kappa}}}{b + \epsilon e_j + \rho + (r+\delta)\rho m_j^{\frac{1-\kappa}{\kappa}}}\right)^{-1}$$

Since firm productivity X and worker effort e are complements in production, larger and more productive firms demand greater effort from workers. Therefore, just as before, the wage markdown equation shows that efficiency wage considerations tend to require larger firms to pay a higher fraction of the $MRPH$ as wages.

2.3   Data and Estimation

The empirical methodology of this paper is based on Wong (2020).[7] The approach consists of first estimating a twoway fixed effect regression to partition out the contributions of worker and firm heterogeneity to the wage-size relationship, and then estimating firms' wage markdowns and marginal revenue product of labor to learn about the channels that drive a positive relationship between firm size and the firm-specific component of wages. I briefly describe the approach below and refer interested readers to Wong (2020) for the full details of the estimation procedure.

2.3.1   Separating the contributions of worker and firm heterogeneity

I implement the statistical model of Abowd et al. (1999) following Bonhomme et al. (2019):

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

with i representing the individual worker, j the employer, $g(j)$ the classification of employer j, a_i the worker time-invariant characteristics, $\phi_{g(j(i,t))}$ are firm-class effects, and X_{it} the time-varying worker characteristics such as age and part-time status. Firms are classed based on the similarity of their internal wage distribution, a method aimed at mitigating the noisiness of firm effect estimates due to the lack of firm-to-firm mobility (Bonhomme et al., 2019).

The Abowd et al. (1999), or AKM, regression is a highly tractable statistical model. However, it entails a number of restrictions. In particular, it abstracts from interactions between worker and firm fixed characteristics, requires random mobility of workers conditional on worker (fixed and time-varying) characteristics and firm effects, and suffers from noisy firm effect estimates due to limited worker mobility.[8] Using the same sample, Wong (2020) assesses these concerns in the context of France.

2.3.2   Estimating components of the firm-specific wage premium

I now briefly describe how the wage markdown and the marginal revenue product of effective labor are estimated. As discussed in Wong (2020), when product markets are imperfectly competitive and worker productivity enter the production function in efficiency units, the marginal revenue product of effective labor is equal to a multiplicative function of the inverted price-cost markup (PM), the output elasticity with respect to labor inputs (LEO), and revenue per unit of effective labor ($\frac{P_{jt}Y_{jt}}{H_{jt}}$).

$$MRPH_{jt} = PM_{jt}^{-1} LEO_{jt} \frac{P_{jt}Y_{jt}}{H_{jt}}$$

The revenue per unit of effective labor is constructed using revenue per hour, adjusted

---

[7]This approach builds on, and combines, the important work on twoway fixed effect regressions by Abowd et al. (1999), Bonhomme et al. (2019), and Lentz et al. (2019), and important work on production function and markup estimation by Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), De Loecker and Warzynski (2012), and De Loecker et al. (2020).

[8]These issues are discussed in Card et al. (2018), Andrews et al. (2008), Eeckhout and Kircher (2011), and Bonhomme et al. (2019).

for differences in unobserved worker heterogeneity across firms. This adjustment uses firm-level average worker effects from the twoway fixed effect regression described in the previous subsection, $H = \bar{E}L$, where H denotes the unit of effective labor, L the labor hours, and $\bar{E}$ the average worker effects.

To obtain wage markdowns, price-cost markups, and the output elasticity with respect to labor, my approach follows closely the approach of Ackerberg et al. (2015) and De Loecker and Eeckhout (2018). First, to obtain the output elasticity with respect to each input, I estimate two-digit sector-specific translog production functions:

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

where y, h, m, k, x, and ϵ represent the log of output, effective units of labor, intermediate inputs, capital inputs, Hicks-neutral total factor productivity, and an error term.

Second, I use the estimated output elasticity with respect to intermediate inputs and the expenditure share of intermediate inputs (MEO) to back out firm-specific price-cost markups (PM):

$$
PM_{jt} = MEO_{jt}\frac{P_{jt}Y_{jt}}{P_{M,jt}M_{jt}}
$$

under the assumption that intermediate inputs are fully flexible and firms take their prices as given (De Loecker and Warzynski, 2012).

Finally, I use the estimated output elasticities, the wage bill, and the intermediate input expenditure to obtain the wage markdown (WM):

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

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

Production function estimation entails a number of challenges. Production function parameter estimates are potentially biased for three reasons, discussed in detail in De Loecker and Goldberg (2014): endogeneity bias, output price bias, and input price bias. At the same time, when estimating a gross output production function, the control function approach does not generally separately identify price-cost markups and returns to scale (Gandhi et al., 2019; Flynn et al., 2019). Wong (2020) addresses these challenges in the context of France.

2.3.3 Administrative microdata from France

The analyses in the paper is based on linked worker-firm and firm balance sheet panel data from France between 1995 and 2014. These datasets provide key information on the population of workers and firms in the French business sector. I use the *Fichier de comptabilité unifié dans SUSE* (FICUS) and *Fichier approché des résultats d'Esane* (FARE) firm balance sheet datasets provided by the fiscal authority of France, *Direction Générale des Finances Publiques* (DGFiP), to estimate the firm characteristics necessary for the decomposition of the wage-firm

size relationship. These datasets contain information on sales, employment, intermediate input and capital expenditure.

I use the *Déclarations Annuelles de Données Sociales* (DADS) matched worker-firm data provided by the French statistical authority, *Institut National de la Statistique et des Études Économiques* (INSEE), to estimate the worker and firm-specific component of wages in a twoway fixed effect regression Abowd et al. (1999). The key information I use from this dataset are age, gender, earnings, hours, and occupational category, all of which are at the job-level.

2.4 Decomposing the Wage-Size Relationship

This sections presents the main findings. To summarize, the results show that the wage-size relationship differs between manufacturing and non-manufacturing sectors, with worker composition differences and rent-sharing being the main drivers.

The wage-firm size profile is flatter in the non-manufacturing sectors than in the manufacturing sector. To see the relationship between firm size and wages, I compute the average (log) hourly wage within 10 firm size categories (5-19, 20-49, 100-199, 200-299, 300-499, 500-999, 1000-1999, 2000-4999, and 5000 or more employees), weighted by firms' employment size. Figure 2.3 displays the relationship between firm size and (log) hourly wages by sector in 2014, with wages for firms with less than 20 employees normalized to 0. Table 2.2 reports the estimates of the wage-size relationship. Figure 2.8 shows the relationship between firm size and wages for manufacturing and non-manufacturing sectors while controlling for 4-digit sector fixed effects within these broad sectors. The results are similar to those in figure 2.3.
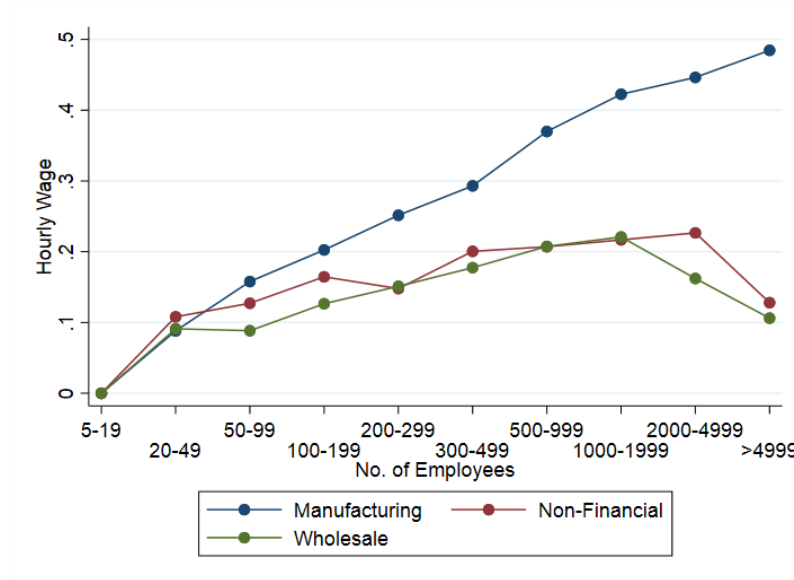
Figure 2.3: Firm size and log hourly wages by sector in 2014

In the manufacturing sector, the relationship between firm size and wages is monotonically increasing over firm size. Manufacturing firms with 100 to 200 employees pay wages that are 20% higher than firms with less than 20 employees on average, while manufacturing firms with

between 1000 and 2000 employees pay approximately 42% higher wages.

In the non-financial service and wholesale-retail sectors, the firm size profile of wages is hump-shaped, though increasing for the most part. In both sectors, table 2.2 shows that a firm with 100-200 employees paid 17% higher wages than firms with less than 20 employees. For firms with 1000 to 1999 employees, the corresponding number is 22%. However, large firms with 5000 employees or more pay a wage similar to firms with 50 to 99 employees. This difference between the manufacturing and non-financial service sectors in the profile of wages over firm size is consistent with evidence that the wage costs of job displacement is larger in the former than the latter (Helm et al., 2020).

Differences in worker composition account for most of the wage-firm size relationship. Figure 2.4 shows how the average worker productivity changes by firm size. Comparing this figure with figure 2.3, one pattern to notice is the similarity of the profile of wages and average worker ability – it is monotonically increasing in manufacturing, but significantly flatter and hump-shaped in non-manufacturing. Figure 2.4 also shows that a large share of the wage-firm size relationship can be accounted for by differences in average worker productivity. In each sector, about 75% to 80% of the relationship can be accounted for by average worker productivity differences. These results are robust to controlling for 4-digit sector fixed effects within the manufacturing and non-manufacturing sectors, as shown in figure 2.9.
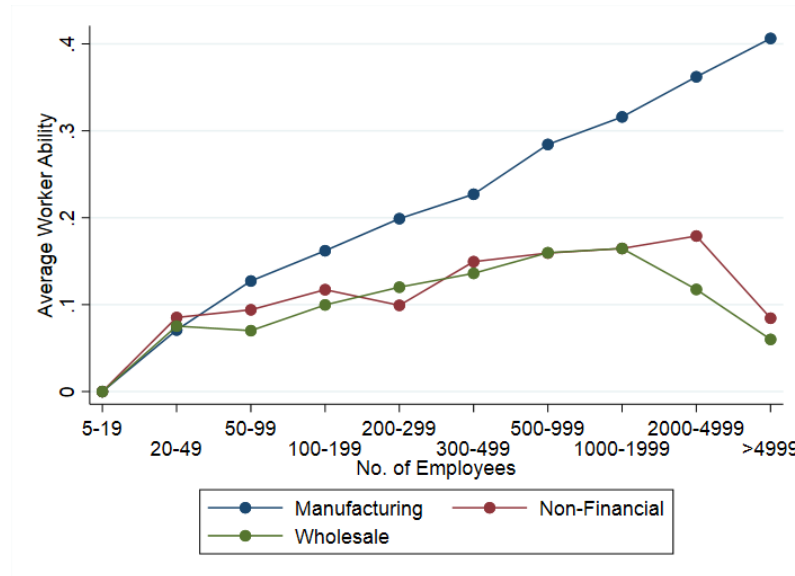
Figure 2.4: Firm size and worker productivity by sector in 2014

Nonetheless, the firm-specific component of wages accounts for a non-trivial share of the wage-size relationship. Figure 2.5 and table 2.3 summarize the size profile of the firm-specific component of wages, also known as the firm-specific wage premium (Abowd et al., 1999; Bonhomme et al., 2019, 2020). Figure 2.5 shows that the profile of the firm wage premium over firm size has a diverging pattern between the manufacturing and non-manufacturing sector, similar to the profile of hourly wages. However, the firm wage premium is not monotonically

increasing over firm size in the manufacturing sector. In the wholesale and retail sector, the firm wage premium increases slowly over firm size. In the non-financial service sector, the firm wage premium plateaus for firms with 200 or more employees. These findings are similar after controlling for 4-digit sector fixed effects within the manufacturing and non-manufacturing sectors, as shown in figure 2.10.
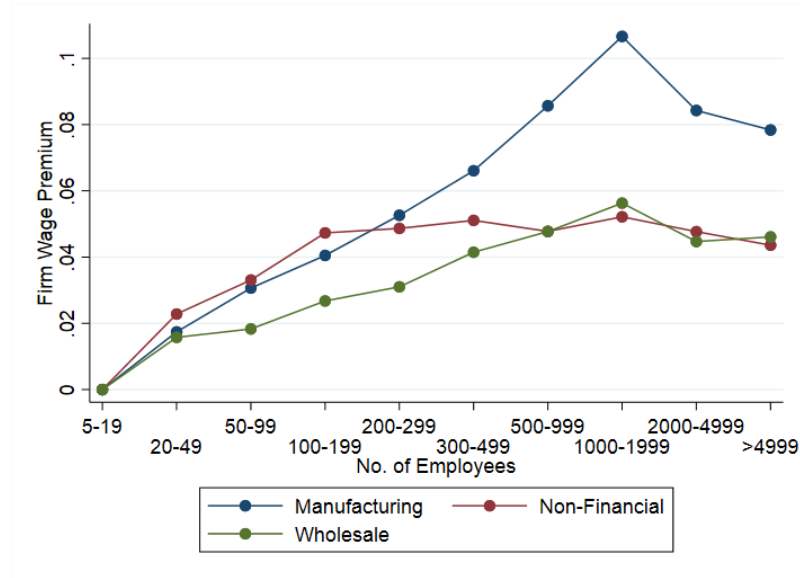
Figure 2.5: Firm size and the firm-specific wage premium by sector in 2014

Rent-sharing appears to be the main driver of the profile of the firm-specific component of wages over firm size. As discussed in the frictional labor market model in section 2, I interpret the wage-size relationship as driven by rent-sharing when (i) the marginal revenue product of labor (net of worker composition differences) is increasing with firm size and (ii) larger firms markdown wages more. Figure 2.6 displays how the marginal revenue product of effective labor evolves over firm size. The MRPH diverges over firm size between the manufacturing and the non-financial service sector, similar to the pattern of the firm-specific wage premium. At the same time, figure 2.7 shows that for these two sectors the wage markdown is larger for larger firms, which partially offsets the positive wage effects of a higher MRPH. This pattern is consistent with models of rent-sharing or wage-posting models (Burdett and Mortensen, 1998; Cahuc et al., 2006; Gouin-Bonenfant, 2018; Elsby et al., 2018).

However, the same is not true in the wholesale and retail sector. While the firm wage premium is monotonically increasing in firm size, the MRPH displays an inverted-U shaped profile over firm size while the wage markdown displays a U-shaped profile. Figures 2.11 and 2.12 show that the same patterns occur after controlling for 4-digit sector fixed effects within the manufacturing and non-manufacturing sectors. This pattern is inconsistent with rent-sharing as the main driver. While rent-sharing remains a potential explanation for the wage-size relationship among smaller firms (firms with less than 300 employees), this channel cannot explain the increasing profile of wage markdowns and decreasing profile of $MRPH$ among larger firms (firms with more than 300 employees). Other mechanisms, such as efficiency
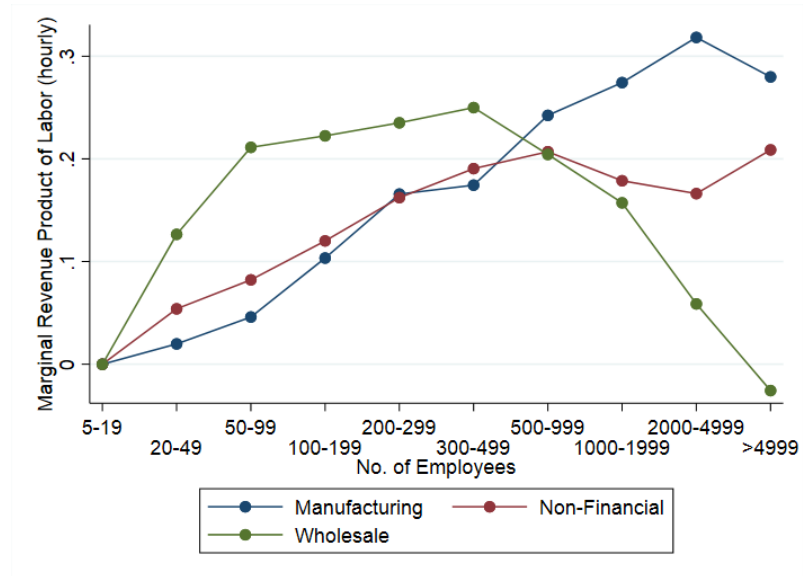
Figure 2.6: Firm size and the marginal revenue product of labor by sector in 2014
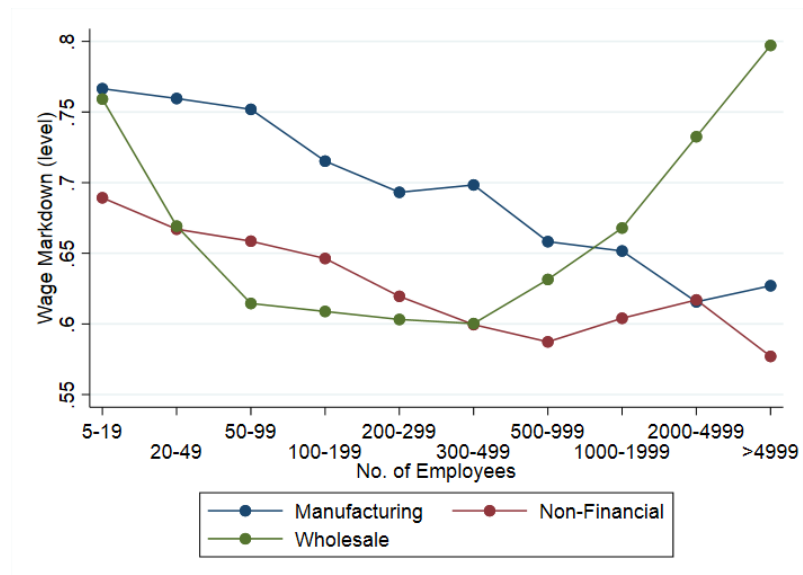
Figure 2.7: Firm size and wage markdowns by sector in 2014

95

wage considerations, are more promising potential explanations for the wage-size relationship among larger firms in the wholesale and retail sector. As shown in the model in section 2, high effort-monitoring costs can lead to an increasing profile for wage markdowns and declining profile for the $MRPH$ over firm size. Although the model shows how one could distinguish between rent-sharing and efficiency wage drivers of the wage-size relationship, non-wage amenities unrelated to efficiency wage considerations could also play a role in explaining the wage-size relationship. To empirically examine whether non-wage amenities play a role, I will estimate the firm-specific value of non-wage amenities inferred from systematic worker mobility to lower-paying firms (in a way that cannot be explained by mass layoffs) to assess the extent to which this can explain the wage-size relationship (Sorkin, 2018).[9]

2.5   Conclusion

This paper revisits the well-established positive relationship between firm size and wages. I use administrative microdata from France to study how this relationship differs across sectors and use a structural decomposition to learn about the potential underlying drivers. The administrative data allows one to study hourly wages as opposed to earnings, non-manufacturing sectors in addition to manufacturing, separately identify the contributions of worker characteristics (observed and unobserved), and investigate the mechanisms underlying the wage-size relationship. The structural decomposition makes it possible to assess the importance of worker productivity, rent-sharing, and efficiency wage considerations.

The main findings are: (i) the profile of wages over firm size is significantly flatter in the non-financial service and wholesale-retail sectors than in the manufacturing sector, (ii) the profile of average worker productivity over firm size accounts for the bulk of the wage-size relationship across sectors, (iii) firm-specific characteristics still play a non-trivial role in the wage-size relationship, with rent-sharing appearing to be the main driver in the manufacturing and non-financial service sector, but not in the wholesale-retail sector.

---

[9]Work on this front will resume as soon as circumstances allow, as the current travel restrictions and university closures due to the COVID-19 pandemic has hindered progress by disrupting data access.

2.6 Appendix
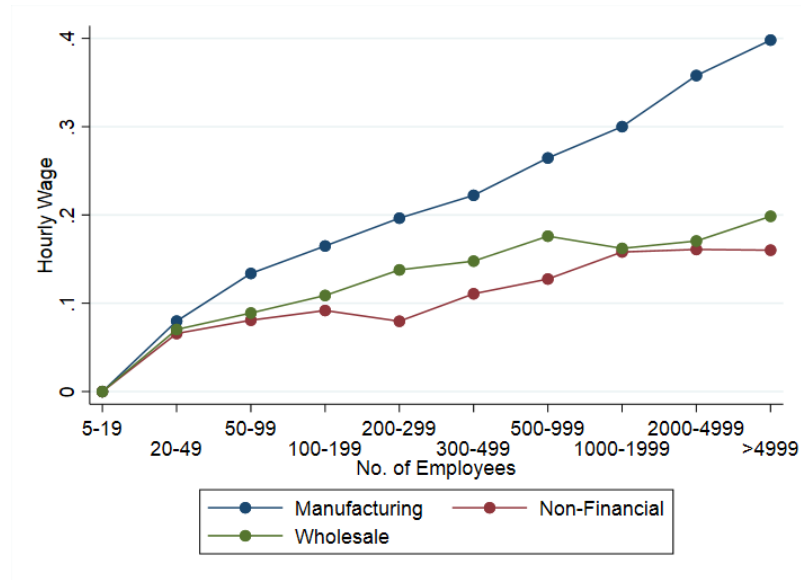
Figure 2.8: Firm size and log hourly wages by sector in 2014

The relationship between firm size and average log hourly wages displayed in this figure accounts for 4-digit sector fixed effects.
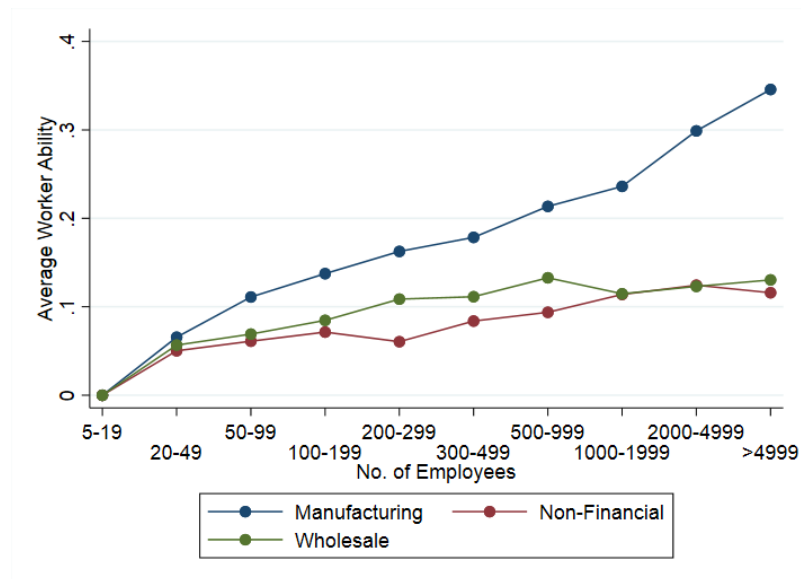
Figure 2.9: Firm size and worker productivity by sector in 2014

The relationship between firm size and average worker productivity displayed in this figure accounts for 4-digit sector fixed effects.
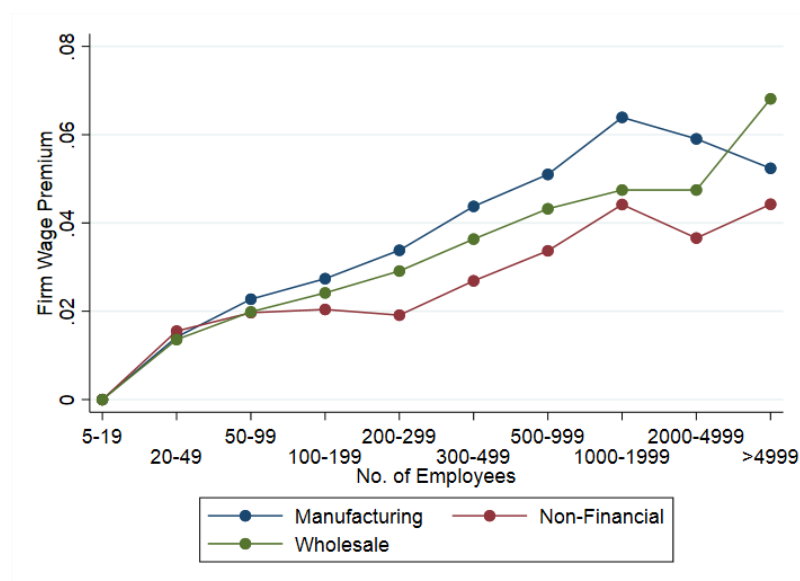
Figure 2.10: Firm size and the firm-specific wage premium by sector in 2014

The relationship between firm size and the firm-specific component of wages displayed in this figure accounts for 4-digit sector fixed effects.
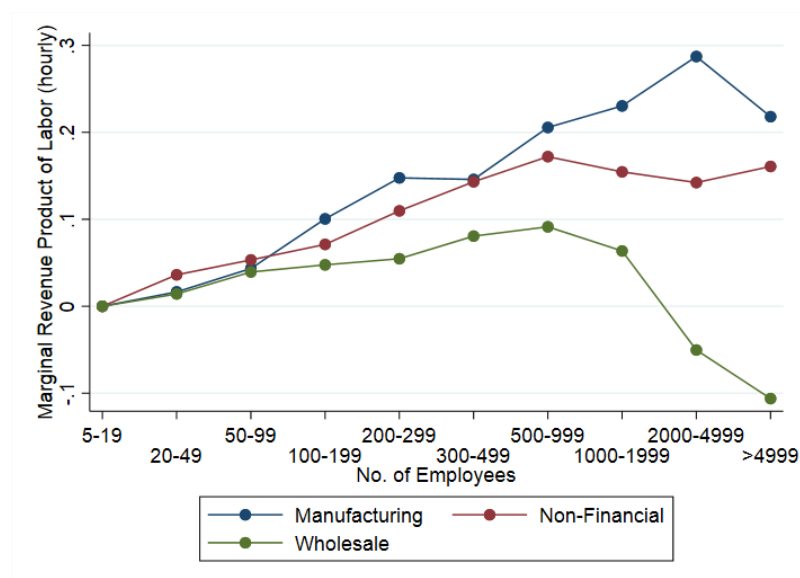
Figure 2.11: Firm size and the marginal revenue product of labor by sector in 2014

The relationship between firm size and the marginal revenue product of labor displayed in this figure accounts for 4-digit sector fixed effects.

98
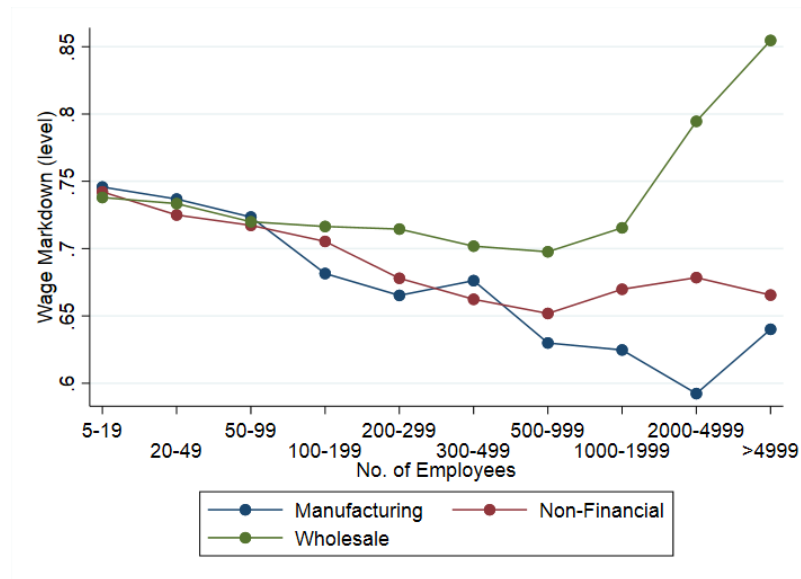
Figure 2.12: Firm size and wage markdowns by sector in 2014

The relationship between firm size and the wage markdown displayed in this figure accounts for 4-digit sector fixed effects.

Parameter	
Real interest rate (r)	0.05
Separation rate (δ)	0.10
unemployment benefits (b)	0.70
Returns to scale (α)	0.80
Vacancy posting cost - scale (γ)	0.01
Vacancy posting cost - convexity (μ)	3
Utility cost of effort (ϵ)	0.2
Elasticity of job-filling rate w.r.t. queue length (κ)	0.5
Fixed entry cost (f_e)	1
Functional form	
Vacancy posting cost	$c(V) = \gamma V^{\mu}$

Table 2.1: Calibrated parameters for wage markdown and MRPH illustration

	Manufacturing	Non-financial services	Wholesale-retail
Firm size category		Log hourly wage	
20-49 employees	0.089	0.108	0.091
	(0.003)	(0.003)	(0.003)
50-99 employees	0.157	0.127	0.088
	(0.004)	(0.005)	(0.005)
100-199 employees	0.203	0.165	0.126
	(0.005)	(0.007)	(0.007)
200-299 employees	0.252	0.148	0.151
	(0.008)	(0.011)	(0.012)
300-499 employees	0.293	0.201	0.178
	(0.011)	(0.014)	(0.015)
500-999 employees	0.369	0.207	0.207
	(0.014)	(0.018)	(0.019)
1000-1999 employees	0.423	0.217	0.221
	(0.019)	(0.029)	(0.027)
2000-4999 employees	0.446	0.227	0.162
	(0.033)	(0.043)	(0.037)
5000 or more employees	0.485	0.128	0.106
	(0.087)	(0.075)	(0.074)
Adjusted-R^2	0.16	0.26	0.12
No. of firms	34,422	96,726	80,812

Table 2.2: Relationship between firm size and wages by sector in 2014

This table presents estimates from a regression of log hourly wages on indicators of the firm size category. Heteroskedasticity robust standard errors are reported in parentheses.

	Manufacturing	Non-financial services	Wholesale-retail
Firm size category		Firm wage premium	
20-49 employees	0.017	0.023	0.015
	(0.001)	(0.001)	(0.001)
50-99 employees	0.031	0.033	0.018
	(0.002)	(0.001)	(0.002)
100-199 employees	0.040	0.047	0.027
	(0.002)	(0.002)	(0.003)
200-299 employees	0.053	0.049	0.031
	(0.003)	(0.003)	(0.004)
300-499 employees	0.066	0.051	0.041
	(0.004)	(0.004)	(0.006)
500-999 employees	0.086	0.048	0.048
	(0.005)	(0.005)	(0.007)
1000-1999 employees	0.107	0.052	0.056
	(0.007)	(0.008)	(0.010)
2000-4999 employees	0.084	0.048	0.045
	(0.012)	(0.011)	(0.014)
5000 or more employees	0.078	0.044	0.046
	(0.033)	(0.019)	(0.027)
Adjusted-R^2	0.14	0.18	0.10
No. of firms	34,422	96,726	80,812

Table 2.3: Relationship between firm size and the firm wage premium by sector in 2014

This table presents estimates from a regression of the firm-specific component of wages on indicators of the firm size category. Heteroskedasticity robust standard errors are reported in parentheses.

Chapter 3

Firms and International Trade: Productivity or Geographical Location?

This chapter contains a number of empirical results based on confidential administrative data from France, some of which are not and cannot be published at the point of thesis submission due to the COVID-19 pandemic. Our request for permission from the French statistical authorities to publish these results has been substantially delayed as ongoing university closures have blocked access to the datasets we use. These datasets can only be accessed at a specific location within the university. The relevant empirical results will be published as soon as the university reopens and clearance is granted.

3.1 Introduction

Firms differ enormously in their export participation and performance (Bernard et al., 2007). Some firms become exporters while others do not. Conditional on exporting, some exporters have larger export revenues and export to more foreign markets than others. On the one hand, a large literature explains these facts by showing that firm characteristics, such as size and productivity (Melitz, 2003b), are key determinants. On the other hand, a growing literature shows that there are large spatial disparities in intranational and international trade participation and performance (Atkin and Donaldson, 2015; Donaldson, 2018; Bellone et al., 2016).

How important are firms' geographical locations, compared to their productivity, in determining their export participation and performance? What determines the geographical distribution of exporters? The answers to these questions have important implications for quantifying the gains and losses from changing trade frictions and for place-based policies that aim to attract high productivity firms (Behrens et al., 2014; Gaubert, 2018; Fajgelbaum and Gaubert, 2018; Caliendo et al., 2019). Yet, much of the existing work on international trade focuses on either, but not both, internal spatial frictions and the role of firm heterogeneity (Overman et al., 2003; Melitz, 2003b). At the same time, work on the spatial distribution of firms tends to abstract from international trade (Puga, 2010; Combes et al., 2012).

In this paper, we document and study how the relationship between firm characteristics and export performance varies with firms' geographical locations within a country. We use large administrative datasets that cover the universe of firms and trade flows into and out of France and we collect data on the distance between locations in France and export markets. Our datasets cover over 2 million firms, over 180 export destinations, and over 18 million goods-level international transactions per year. We then build a heterogeneous-firm trade model in the tradition of Melitz (2003b) augmented with an intranational spatial dimension to interpret both the new empirical patterns that we document and other patterns documented in the literature. In our model, internal space is characterized by iceberg trade costs between locations.

We have two main findings. Empirically, we find that exporters in locations farther away from export markets tend to be larger and more productive, but the fraction of firms who are exporters declines with distance to export markets. Theoretically, we find that our model can rationalize these empirical patterns and also the well-known fact that more productive firms are located in larger cities (Puga, 2010). While there is ample evidence for two key ingredients needed for workhorse spatial models to generate the latter prediction – larger cities make firms more productive (agglomeration externalities) and that these externalities benefit more productive firms disproportionately (log-supermodularity) – we obtain this prediction in a model without these ingredients. Instead, in our model, the tension between access to domestic and international markets shapes the spatial distribution of firm types, export participation and performance, and city size. Intuitively, more productive firms are located in larger cities because these cities are closer to international markets and exporters, who are larger and more productive than non-exporters, benefit disproportionately from better access to international markets.

We start by documenting how export participation and firm characteristics differ by their distance from international markets. The data requirements are demanding, as we need information on firm characteristics, where they are located within the country, and which countries they trade with. We use administrative balance sheet data for the population of firms in France to obtain information on firms' export status, size, and productivity. We also use a dataset of all trade flows into and out of France, which provides information on what, where, and how much firms export. Next, we obtain data on the kilometer distance between a firm's location and the capital city of a Foreign country at the postcode level using Google maps. We focus on manufacturing firms and trade within the European Union to ensure that we measure distance-related trade costs as reliably as possible.

Our empirical findings show that firms' export participation and size differ systematically by the distance between their location within the country and international markets. We find that the fraction of firms who are exporters fall with their distance from foreign markets. At the same time, the difference between the average exporter and the average firm in terms of revenue and measured labor productivity is larger in locations farther away from export markets.[1] Models of international trade with firm heterogeneity, largely based on Melitz (2003b), abstract

---

[1]We also find that exporters in locations farther away from export markets are larger and have higher measured labor productivity. However, due to UK-wide university shutdowns driven by the COVID-19 pandemic, access to the administrative datasets used, which is only possible at a specified location within the university, have been temporarily blocked. As such, we have not been able to submit these additional empirical results to the French statistical agency to request for permission to publish.

from internal space. Models of intranational trade largely abstract from firm heterogeneity and international trade (Overman et al., 2003; Allen and Arkolakis, 2014; Atkin and Donaldson, 2015). Models in urban economics that focus on the relationship between firm heterogeneity and city size largely abstract from internal space and international trade (Combes et al., 2012; Gaubert, 2018). Hence, existing models do not speak directly to the spatial distribution of exporters within a country and their characteristics.

To make sense of our empirical findings, we then build a heterogeneous-firm international trade model with intranational space. Our modelling approach combines the classic Melitz (2003b) *international* trade model with ingredients of workhorse *intranational* trade models (Allen and Arkolakis, 2014; Gaubert, 2018). Relative to a standard Melitz (2003b) model, trade between cities within a country entails an iceberg trade cost, which defines the meaning of internal space in our model. Therefore, in contrast with Melitz (2003b), the iceberg trade costs between the home country and a foreign market depends on where exporters are located within the home country. As a result, the distribution of firm types by productivity and export status across locations depends on which location provides greater access to domestic and foreign markets. We begin by considering an internal spatial structure in which three cities (locations) are arranged symmetrically on a line. There is also a foreign export market whose distance to locations within the home country is governed by a relative distance parameter. In the model, workers are freely mobile, equalizing utility across cities. Firms choose a city to locate in and face mobility costs – they draw a city-specific productivity. Cities with a higher population density creates higher congestion costs, for example, due to a fixed land supply. Workers in denser cities are compensated in the form of higher wages. Therefore, firms' location decisions balance local labor costs and access to domestic and foreign goods markets.

Our main theoretical result is that cities farther from export markets are harsher in selecting which firms are sufficiently productive to become exporters, while these cities are also more lenient in terms of selection at entry. This implies that exporters in cities that are farther from export markets are larger and more productive, and that the productivity and size differential between the average exporter and the average firm is also larger at these locations. These predictions are in line with the patterns we observe in the data.[2] Our model also predicts the well-known empirical pattern that denser cities are more productive and have higher wages while additionally suggesting that denser cities are those with greater access to foreign markets.

The intuition behind our model's predictions is that, if export markets are sufficiently large, cities closer to export markets are more central because they face lower iceberg trade costs, implying a more lenient selection criterion for exporting. In equilibrium, more exporters set up in these cities, attracting workers to move there, and driving up local wages due to congestion. The higher wages imply higher labor costs, which leads to a harsher productivity selection criterion for entrant firms. At the same time, if locations closer to export markets are farther from internal markets, the lower market access to internal markets for non-exporters contributes to the harsher selection criterion for non-exporters. Therefore, the distribution of firm types differs across cities. This result is driven by a selection-into-exporting mechanism similar to

---

[2]These empirical results have yet to be published, since they need to be submitted to French statistical agencies for permission to publish. This process has been severely hindered due to the COVID-19 pandemic.

Melitz (2003b): proximity to export markets only benefits more productive firms, since these are the firms who find it profitable to export. In an autarky equilibrium, we show that the selection criterion for entrant firms is the same across cities, despite more central locations providing better market access to all other locations within the country. Therefore, the distribution of firm types is the same across cities. This is because better market access is offset by higher wages due to more congestion, as more central locations attract more workers.

This paper contributes to two related bodies of work. First, a large literature in international trade links firm characteristics such as size and productivity to their export participation and performance. Bernard et al. (2007) systematically document this relationship, while Melitz (2003b) provides a highly tractable structural framework that links firm heterogeneity to international trade. Eaton et al. (2011) use French firm-level data to show that firm productivity accounts for more half of the differences in export participation across firms, but little of the differences in export revenues for a given market. Our contribution to this literature is to bring in the role of the internal geographical location of firms. In particular, we document new empirical patterns that show that firms located farther from export markets are less likely to export, but those who do are more productive than exporters located closer to export markets. Our empirical results suggest that part of the differences in export participation and export revenue across firms come from differences in the location choice of exporters. The empirical patterns we find are related to the work of Bellone et al. (2016) and Crozet and Koenig (2010), who show that there are large differences in the share of exporters across French regions. We also present a spatial trade model that shows how the tension between export opportunities and domestic sales opportunities can explain the spatial distribution of exporters and their characteristics. We are not the first to link internal geography to international trade. Cosar and Fajgelbaum (2016) incorporate internal geography into a trade model based on comparative advantages to study the spatial impact of trade integration. Pflueger and Tabuchi (2019) show that comparative advantages lead to spatial dispersion of economic activity while increasing returns lead to agglomeration, with trade costs governing their relative importance. In contrast, our focus is on how internal geography affects the spatial distribution of heterogeneous exporters and their characteristics, and in turn, how internal geography affects regional export participation, city size, productivity, and wages.

A vast and growing body of research in spatial economics seeks to understand why productivity and wages are higher in denser cities, why the composition of workers and firms differ across cities, and what determines the spatial distribution of economic activity. In general, these are jointly determined by local fundamentals, agglomeration externalities, congestion externalities, selection, and the sorting of workers and firms to cities of different sizes. Donaldson (2018), Donaldson and Hornbeck (2016), and Atkin and Donaldson (2015) show that internal trade costs are substantial. Allen and Arkolakis (2014) show that when there are internal trade costs, better connected cities grow larger, leading to higher wages through congestion costs. Rosenthal and Strange (2004) and Puga (2010) provide a survey of the literature on agglomeration externalities, for example, due to knowledge spillovers and improved labor market matching (Davis and Dingel, 2019; Martellini, 2020). A number of studies provide empirical evidence for agglomeration and congestion externalities (Ahlfeldt et al., 2015; Dauth et al., 2019; Helm, 2020; Heblich et al., 2020). Eeckhout et al. (2014), Martellini (2020),

Combes et al. (2012), Behrens et al. (2014), Gaubert (2018) study the role of the sorting of workers and firms to cities of different sizes in explaining why cities differ from each other. We add to this body of work by showing that the internal geography of cities, characterized by trade costs, determines the relative profitability of domestic sales and exporting, which provides a further explanation for why larger cities have more productive firms. Our model makes these predictions without agglomeration externalities or differences in local fundamentals besides geographical location.

3.2   Data Description

To make our analysis possible, we require information on firm characteristics, their export conduct, firms' locations, and the distance between each firm location and export destination pair. We obtain this information from firm balance sheet data, customs trade flow data, and matched employer-employee panel data from France. We also construct distance data between firm location and export destination.

We have access to the *Fichier approché des résultats d'Esane* (FARE), which are firm balance sheet datasets. These data are collected by the fiscal authority of France, *Direction Générale des Finances Publiques* (DGFiP), for tax purposes. They cover the population of firms in France between 2008 and 2014 without firm size restrictions, consisting of over 2 million firms per year. From these datasets, we obtain information about firms' revenue, employment, and commune codes.

We obtain information about firms' exports and imports at the product-level from trade flow data provided by the customs authority of France. This dataset covers the years 2010 to 2014 and contains detailed information about product prices, units, firm identifiers, six-digit product codes, and over 180 import origin or export destination countries for over 18 million transactions a year.

Next, we combine French commune identifiers from the FARE dataset with export destinations from the customs dataset to calculate the distance between each commune-export-market pair. There are over 36000 communes We first obtain the names of the communes associated with a commune identifier and then obtain the corresponding postcode of a given commune. We next obtain the postcode of the capital city of an export destination country. Finally, we calculate the geodesic distance, in kilometers, of each French commune-export-destination pair using Google maps.

We restrict firm level observations from the FARE balance sheet data to the manufacturing sector. We do not include French overseas territories. We keep only export destinations in the European Union (EU) and exclude multi-plant and multi-product firms. This is to ensure that our measure of the distance between firm location and export destination meaningfully captures the associated trade costs (for example, transportation costs). Our findings are robust to the inclusion of multi-plant and multi-product firms. We drop observations with invalid firm and export market identifiers. We refer the interested reader to Bergounhon et al. (2018) for a set of important guidelines to cleaning the French customs trade data. The observations of our cleaned sample are at the level of firm and export destination pairs.

	Exporter	Non-Exporter	All Firms
Number of firms	7,583	81,552	89,135
Number of communes	3,775	15,549	15,874
Number of EU export destinations	27	-	-
Number of firm-market observations	-	-	2,208,191
Average employment size	79	10	16
Average revenue size (in 000's of Euros)	15,996	1,379	2,623
Average fraction of revenue from exports	0.21	-	-

Table 3.1: Summary statistics for manufacturing firms in 2014.

3.3 Motivating Evidence: Firms and Proximity to Export Markets

Existing research relating firms' export performance with their characteristics largely abstract from exporters' geographical locations. In this section, we document several novel empirical patterns relating firm characteristics and export participation to the distance of the firm's location from export markets. We then seek to explain these empirical patterns using an international trade model with intranational trade frictions in the next section.

To document these patterns, we compute the average of firm characteristics and export participation by distance to export markets while controlling for sectoral and regional differences. We first regress the outcome variables of interest on sector-region fixed effects to obtain the predicted residuals ($\hat{\epsilon}_{i,j,t}$):

$$Y_{i,j,t} = D_{s,r,t} + \epsilon_{i,j,t}$$

Denote i as the firm, j as the export market, t as the year, s as the four-digit sector, and r as the region. The variable $D_{s,r,t}$ is a sector-region-year indicator. The variable $Y_{i,j,t}$ represents the natural logarithms of labor productivity (value-added per worker), revenue, employment, or wages. We categorize distances into 10 discrete bins. The distance bins are as follows: 0-300km, 301-600km, 601-900km, 901-1200km, 1201-1500km, 1501-1800km, 1801-2100km, 2101-2400km, 2401-2700km, and 2701km and above. We then compute the within-distance-bin average of the outcome variable of interest using the predicted residuals ($\hat{\epsilon}_{i,j,t}$), and compare exporters with non-exporters within each distance bin.

First, we find that the proportion of firms who are exporters declines with distance to export markets. Figure 3.1 shows the relationship between the proportion of exporting firms and proximity to export markets. This is consistent with predictions of our model, presented in the next section.

Second, we find that the difference in productivity and size between the average exporter and the average non-exporter increases with distance to export markets. Figures 3.2 and 3.3 display the difference in (log) labor productivity between exporters and non-exporters, and the difference in (log) sales revenue between exporters and non-exporters by distance bins.
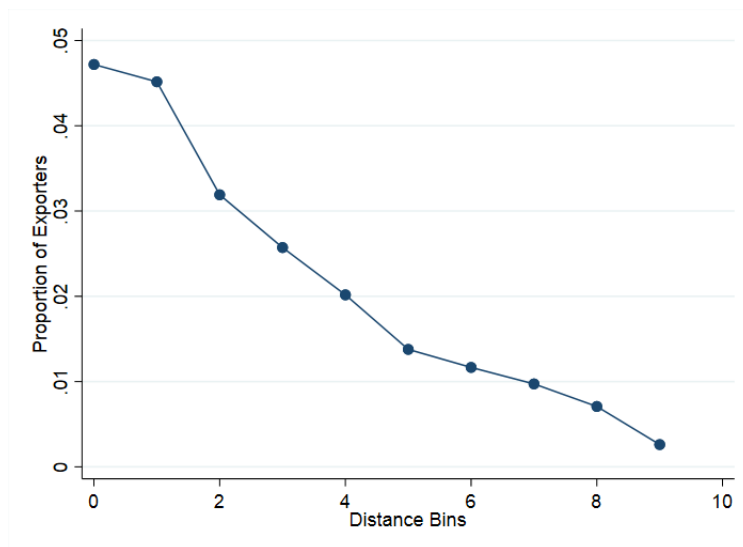
Figure 3.1: Data: Fraction of firms that are exporters by distance.
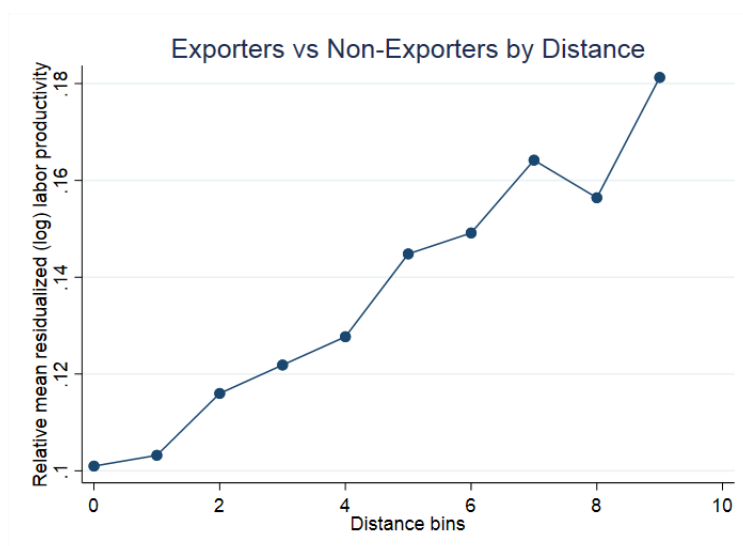
Figure 3.2: Data: Relative productivity of exporters compared to non-exporters by distance.
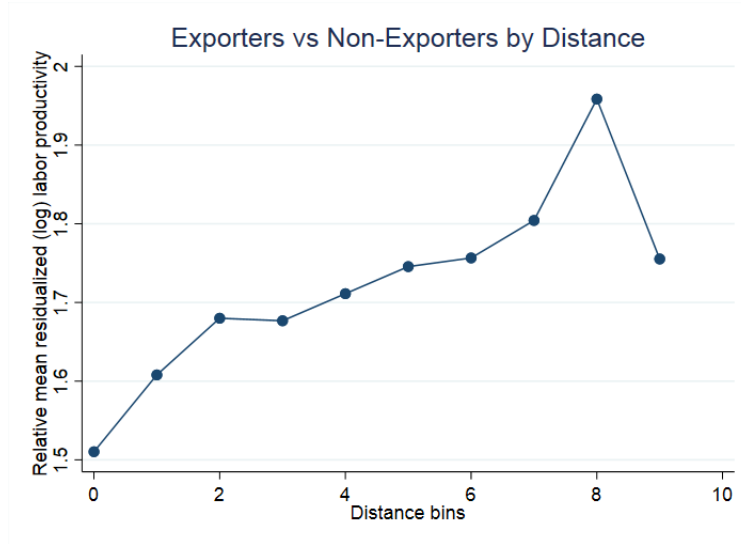
Figure 3.3: Data: Relative revenue size of exporters compared to non-exporters by distance.

We also document that both exporters and non-exporters in locations farther from export markets tend to be more productive and larger in terms of employment and revenue. [3]

3.4 An International Trade Model with Internal Space

We now present a model of international trade with heterogeneous firms in the Melitz (2003b) tradition with internal space, drawing on Allen and Arkolakis (2014) and Gaubert (2018). We use the model to interpret the empirical patterns presented in the previous section. In our model, cities differ ex-ante in terms of their location within a country, which determines their internal distance from other cities and also their distance from export markets. Our main result is that the relative attractiveness of firms' market access to domestic and foreign markets determines the distribution of population size, wages, and firm types by productivity and export participation. Our results do not depend on standard assumptions on firms' production functions: firm productivity is increasing in city size (agglomeration externalities) and more productive firms benefit more from larger city sizes (log-supermodularity). Instead, our results come from the presence of export opportunities, which only benefits firms who are sufficiently productive to participate in export markets. Once we shut down the possibility of exporting, our model predicts that the distribution of firm types are identical across cities.

Below, we describe a model in which there are three locations within the home country, arranged on a real line. The two locations at the periphery are symmetric. There is one foreign country with which the home country trades. Within the home country, locations at the center and the periphery differ in their distance from each other and their distance to the foreign export market.

---

[3]These additional empirical results and the relevant regression tables require permission from French statistical authorities before publication. The unfortunate onset of the COVID-19 pandemic and associated university shutdowns have prevented this process from being completed, as access to French administrative data is not permitted outside of the university. These results will be published as soon as permission is requested and granted.

3.4.1 Foreign country

Consumers. There is a representative Foreign consumer who supplies $N(F)$ units of labor inelastically, capturing the size of the export market. This consumer maximizes a constant-elasticity-of-substitution utility function:

$$U(F) = \left[(1-\beta)^{\frac{1}{\sigma}} \int_{\omega \in \Omega(F)} q(\omega, F, F)^{\frac{\sigma-1}{\sigma}} \, d\omega + \beta^{\frac{1}{\sigma}} \sum_{j=1}^{3} \int_{\omega \in \Omega(j)} q(\omega, j, F)^{\frac{\sigma-1}{\sigma}} \, d\omega \right]^{\frac{\sigma}{\sigma-1}\eta} N(F)^{-b(1-\eta)}$$

subject to a budget constraint:

$$p(F)q(F) + \sum_{j=1}^{3} \int_{\omega \in \Omega(j)} p(\omega, j, F)q(\omega, j, F) \, d\omega = W(F)N(F)$$

where $\sigma > 1$ is the elasticity of substitution between varieties. Wages in the foreign country $W(F)$ is determined in a perfectly competitive labor market, $q(F)$ denotes the CES composite good produced by foreign firms, and $p(F)$ denotes the CES price index of this composite good. The price and quantity of a good ω imported from a location j in the Home country are represented by $p(\omega, j, F)$ and $q(\omega, j, F)$. The parameter β governs the relative utility weight on imported goods. There is a utility cost of congestion $N(F)^{-b}$ in which the parameter b controls the degree of congestion, following Allen and Arkolakis (2014) and Behrens et al. (2014). Foreign firms with homogenous productivity produce a differentiated variety in a monopolistically competitive goods market. The price index at Foreign is:

$$\bar{P}(F) = \left[(1-\beta)M(F)p(F)^{1-\sigma} + \beta \sum_{j=1}^{3} \frac{1-G(z_x(j))}{1-G(\tilde{z}(j))} M(j) \int_{\tilde{z}(j)} p(\omega, j, F)^{1-\sigma} \, f_j^x(z)dz \right]^{\frac{1}{1-\sigma}}$$

where $M(j)$ denotes the mass of active firms and $\frac{1-G(z_x(j))}{1-G(\tilde{z}(j))}$ represents the fraction of firms who are exporters in location j.

Firms. Foreign firms compete in a monopolistically competitive goods market. They operate a constant returns to scale production technology with only labor as inputs and homogenous productivity $\bar{\Psi}(F)$. Firms pay a fixed operating cost f in labor units. To maximize profits, they choose the amount of labor to employ:

$$\max_{l(\omega,F,F),l(\omega,F,j)} p(\omega, F, F)q(\omega, F, F) - W(F)l(\omega, F, F)$$

$$+ \sum_{j=1}^{3} (p(\omega, F, j)q(\omega, F, j) - \tau(F, j)W(F)l(\omega, F, j)) - fW(F)$$

subject to their production technology and product demands they face from consumers in the domestic and international goods markets:

$$q(\omega, F, j) = \bar{\Psi}(F)l(\omega, F, j)$$

$$q(\omega, F, F) = (1 - \beta)p(\omega, F, F)^{-\sigma}W(F)N(F)\bar{P}(F)^{\sigma-1}$$

$$q(\omega, F, j) = \beta p(\omega, F, j)^{-\sigma}W(j)L(j)\bar{P}(j)^{\sigma-1}$$

where $L(j)$ is the population size of workers in location j in the home country. I assume that the iceberg trade costs are symmetric, $\tau(i,j) = \tau(j,i)$ for any pair (i,j). The mass of firms in the foreign country $M(F)$ is determined by a free entry condition:

$$\bar{\Pi}(F) = f_e W(F)$$

where $\bar{\Pi}(F)$ denotes average profits in the foreign country and f_e denotes the sunk entry cost.

3.4.2 Home country

Consumers. There is a mass $N(H)$ of identical workers, each endowed with one unit of labor. Workers are freely mobile across locations. They choose a consumption bundle and a location to maximize utility subject to a budget constraint. The utility function of a consumer in a location i is identical to that of the foreign consumer:

$$U(i) = \left[\beta^{\frac{1}{\sigma}} q(F)^{\frac{\sigma-1}{\sigma}} + (1-\beta)^{\frac{1}{\sigma}} \sum_{j=1}^{3} \int_{\omega \in \Omega(j)} q(\omega, i, j)^{\frac{\sigma-1}{\sigma}} d\omega \right]^{\frac{\sigma}{\sigma-1}\eta} L(i)^{-b(1-\eta)}$$

Each worker's budget constraint is:

$$p(F)q(F) + \sum_{j=1}^{3} \int_{\omega \in \Omega(j)} p(\omega, j, i)q(\omega, j, i) \, d\omega = W(i)$$

Free worker mobility equalizes workers' utility across locations, $U(i) = \bar{U}$ for each i. Solving the worker's utility maximization problem yields the following location-specific equilibrium wage equation:

$$W(i) = \kappa_0 \bar{U}^{\frac{1}{\eta}} \bar{P}(i) L(i)^{b\frac{1-\eta}{\eta}}$$

where $\bar{U}^{\frac{1}{\eta}}$ is an economy-wide constant to be determined in general equilibrium and κ_0 is a constant defined as $\kappa_0 \equiv (1-\eta)^{b\frac{1-\eta}{\eta}}$. The equilibrium wage equation shows that workers' wages depend on two location-specific characteristics – population size $L(i)$ and price index $\bar{P}(i)$. Population size matters for wages as workers need to be compensated for congestion externalities. All else equal, larger cities pay higher wages due to congestion. Workers in more expensive cities (higher $\bar{P}(i)$) also need to be compensated for the higher cost of living. The price index $\bar{P}(i)$ takes the form:

$$\bar{P}(i) = \left(\frac{\sigma}{\sigma-1} \right) \left[\beta \sum_{j=1}^{3} M(j) \left(\frac{W(j)\tau(j,i)}{\bar{\Psi}(j)} \right)^{1-\sigma} + (1-\beta)M(F) \left(\frac{W(F)\tau(F,i)}{\bar{\Psi}(F)} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}$$

where $\bar{\Psi}(j)$ is the average productivity of firms in location j. It is location-specific due to internal trade frictions. All else equal, the price index is lower in locations with greater consumer

market access – more central locations – as these locations face lower internal trade frictions. Therefore, all else equal, wages are lower at more central locations.

Firms. Home firms are heterogeneous in terms of their productivity. Firms' productivity $\Psi(z, L)$ depends on their idiosyncratic efficiency z, drawn from a distribution $G(z)$, and the population size of the city L. Following the literature, I assume that firm productivity is increasing in population size ($\Psi_L > 0$) and that it is log-supermodular in city size ($\frac{\partial \log \Psi}{\partial \log z \log L} > 0$) (Puga, 2010; Combes et al., 2012; Gaubert, 2018). To maximize profits, firms choose the amount of labor inputs to employ and whether to participate in export markets subject to a fixed cost of exporting f_F, with $f_F > f$. The profit maximization problem for a firm with efficiency z in location i is:

$$\max_{l(z,i,j),l(z,i,F)} \sum_{j=1}^{3} \left(p(z,i,j)q(z,i,j) - \tau(i,j)W(i)l(z,i,j) \right)$$
$$+ p(z,i,F)q(z,i,F) - \tau(i,F)W(i)l(z,i,F) - W(i)(f + f_F)$$

subject to a constant returns-to-scale production technology and domestic and international demand for its goods:

$$q(z,i,j) = \Psi(z, L(i))l(z,i,j)$$

$$p(z,i,j) = \bar{P}(j) \left(\frac{q(z,i,j)}{(1-\beta)W(j)L(j)} \right)^{-\frac{1}{\sigma}}$$

$$p(z,i,F) = \bar{P}(F) \left(\frac{q(z,i,F)}{\beta W(F)N(F)} \right)^{-\frac{1}{\sigma}}$$

The equilibrium profits for a non-exporter is:

$$\pi_d(z,i) = \kappa_1 \left(\frac{\Psi(z,L(i))}{W(i)} \right)^{\sigma-1} \mu(i) - W(i)f$$

where $\kappa_1 \equiv \left(\frac{1}{\sigma} \right) \left(\frac{\sigma-1}{\sigma} \right)^{\sigma-1}$ and $\mu(i) = \sum_j \tau(i,j)^{1-\sigma}(1-\beta)W(j)L(j)\bar{P}(j)^{\sigma-1}$ represents domestic market access for firms' in location i. The equilibrium profit for an exporter is:

$$\pi(z,i) = \kappa_1 \left(\frac{\Psi(z,L(i))}{W(i)} \right)^{\sigma-1} \mu(i,x) - W(i)(f + f_F)$$

where $\mu(i,x) = (1-\beta)\sum_j \tau(i,j)^{1-\sigma}W(j)L(j)\bar{P}(j)^{\sigma-1} + \beta\tau(F)^{1-\sigma}W(F)N(F)\bar{P}(F)^{\sigma-1}$ represents market access to domestic and foreign markets from location i. This shows that producer market access now also depends on access to Foreign markets.

Selection at entry and selection into exporting. The presence of a fixed operating cost f and a fixed exporting cost f_F imply that there is an efficiency cutoff above which firms select into operating ($z_e(i)$) and an efficiency cutoff above which firms select into exporting ($z_x(i)$) for each location i. The selection criterion for entry is such that $\pi_d(z_e(i), i) \geq W(i)f$, which

translates into:

$$\Psi(z, L(i)) \geq \Psi(z_e(i), L(i)) \equiv \left(\frac{fW(i)^\sigma}{\kappa_1 \mu(i)} \right)^{\frac{1}{\sigma-1}}$$

The location-specific selection criterion for exporting is such that $\pi_x(z_x(i), i, F) \geq W(i)f_F$, which implies:

$$\Psi(z, L(i)) \geq \Psi(z_x(i), L(i)) \equiv \left(\frac{f_F W(i)^\sigma}{\kappa_1 (\mu(i, x) - \mu(i))} \right)^{\frac{1}{\sigma-1}}$$

The efficiency cutoffs depend on three factors: city size, labor costs, and market access. All else equal, a larger city implies lower efficiency cutoffs since agglomeration externalities imply that more firms become sufficiently productive to operate and to export. Similarly, higher labor costs imply higher efficiency cutoffs. Locations with greater market access face lower iceberg trade costs, which imply that goods produced in these locations are cheaper when shipped to other locations. All else equal, the central location has greater market access than the periphery.

The average productivity of active firms in location i is:

$$\bar{\Psi}(i) = \left[\int_{z_e(i)}^{z_{max}} \Psi(z, L(i))^{\sigma-1} f_i(z) \, dz \right]^{\frac{1}{\sigma-1}}$$

where $f_i(z)$ is the probability density function of firm efficiency in location i:

$$f_i(z) = \frac{g(z)}{1 - G(z_e(i))} \quad \text{if } z \geq z_e(i)$$

Similarly, the average productivity of exporters in location i is:

$$\bar{\Psi}(i, x) = \left[\int_{z_x(i)}^{z_{max}} \Psi(z, L(i))^{\sigma-1} f_i^x(z) \, dz \right]^{\frac{1}{\sigma-1}}$$

where the probability density function for exporter efficiency is:

$$f_i^x(z) = \frac{g(z)}{1 - G(z_x(i))} \quad \text{if } z \geq z_x(i)$$

Therefore, the average productivity of exporters and non-exporters will differ across locations if there is differential selection in equilibrium and if city size differs.

Firms' location choice. Our baseline assumption is that potential entrants to each location pay a sunk entry cost f_e to draw a location-specific efficiency z, similar to Melitz (2003b). This assumption captures the idea that firms face mobility costs: in order to move to a different location, firms must first shut down, and draw another location-specific efficiency z. A rationale for firm mobility costs is that firms' productivity in a different location depends on costs of understanding the tastes of that particular market or an element of luck such as meeting the right person at the right time (Behrens et al., 2014). For as long as firms expect to make a profit, they will choose to enter a location i. In equilibrium, the efficiency cutoffs determine the type

of active firms and a free entry condition pins down the mass of active firms in each location:

$$[1 - G(z_e(i))]\bar{\pi}(i) = W(i)f_e$$

where $\bar{\pi}(i)$ is the average profit in location i and f_e is the fixed entry cost in labor units.

3.4.3 Distribution of exporter and non-exporter types across cities

In this model, the relative importance of market access to domestic and foreign markets determines the distribution of firm types by efficiency and export participation. To see this, the equilibrium efficiency cutoff for entry can be written as:

$$\frac{\Psi(z_e(i), L(i))}{\bar{\Psi}(i)} = \left(\frac{1 + \frac{1 - G(z_x(i))}{1 - G(z_e(i))} \left(\frac{\bar{\Psi}(i, x)}{\bar{\Psi}(i)} \right)^{\sigma - 1} \left(\frac{\mu(i, x) - \mu(i)}{\mu(i)} \right)}{1 + \frac{1 - G(z_x(i))}{1 - G(z_e(i))} \left(\frac{f_F}{f} \right) + \frac{f_e/f}{1 - G(z_e(i))}} \right)^{\frac{1}{\sigma - 1}}$$

and the efficiency cutoff for exporting can be written as:

$$\frac{\Psi(z_x(i), L(i))}{\bar{\Psi}(i, x)} = \left(\frac{\frac{1 - G(z_x(i))}{1 - G(z_e(i))} + \left(\frac{\mu(i)}{\mu(i, x) - \mu(i)} \right) \left(\frac{\bar{\Psi}(i)}{\bar{\Psi}(i, x)} \right)^{\sigma - 1}}{\frac{f}{f_F} + \frac{1 - G(z_x(i))}{1 - G(z_e(i))} + \frac{f_e/f_F}{1 - G(z_e(i))}} \right)^{\frac{1}{\sigma - 1}}$$

where $\frac{\mu(i, x) - \mu(i)}{\mu(i)}$ captures the key tension between access to domestic and foreign markets. As discussed above, these cutoffs also depend on city size through agglomeration externalities and congestion externalities.

To see how relative market access determines the distribution of firm types, we make several functional form assumptions about the cumulative distribution function for firm efficiency ($G(z)$), the icerberg trade cost ($\tau(i, j)$), and the productivity function ($\Psi(z, L)$). The distribution of firm efficiency is Pareto ($G(z) = 1 - \left(\frac{z_{min}}{z} \right)^{\theta}$), with support $[z_{min}, +\infty)$ and θ is the shape parameter. We make a standard parameter restriction $\theta > \sigma - 1$ in order to guarantee that trade flows are finite. Let $i = \{c, p\}$ denote central and peripheral locations. The iceberg trade cost functions are $\tau(i, j) = (1 + d)^{\delta}$ and $\tau(i, j) = (1 + 2d)^{\delta}$ for adjacent and non-adjacent locations within the home country, where d is the distance between adjacent locations. Trade with the foreign country comes at a cost $\tau(i = c, F) = (1 + \gamma d_F)^{\delta}$ for the central location and $\tau(i = p, F) = (1 + \gamma d_F)^{\delta}$ for peripheral locations. The parameter γ then governs the proximity of the central location to the foreign market, relative to the distance between peripheral locations and the foreign market. Finally, following Gaubert (2018), the productivity function is translog $\Psi(z, L) = z^{(1 + s \log(\frac{L(i)}{L_{min}}))} L^{\alpha}$, where L_{min} is the smallest existing city size, $\alpha > 0$ governs the strength of agglomeration externalities, and $s \geq 0$ governs the strength of the log-supermodularity between firm efficiency and city size. When $s = 0$, there is no log-supermodularity and agglomeration externalities benefit firms of any efficiency level in the same way (Combes et al., 2012).

Proposition 1: Suppose there are no agglomeration externalities and the cumulative distribution of efficiency draws is Pareto. Then, the entry and exporting selection cutoffs can be written as:

$$\left(\frac{z_e(i)}{z_{min}}\right)^{\theta} = \frac{(\sigma-1)}{\theta+(\sigma-1)}\left(\frac{f}{f_e}\right)\left[1+\left(\frac{f}{f_F}\right)^{\frac{\theta-(\sigma-1)}{\sigma-1}}\left(\frac{\mu(i,x)-\mu(i)}{\mu(i)}\right)^{\frac{\theta}{\sigma-1}}\right] \qquad (3.1)$$

$$\left(\frac{z_x(i)}{z_{min}}\right)^{\theta} = \frac{(\sigma-1)}{\theta+(\sigma-1)}\left(\frac{f_F}{f_e}\right)\left[1+\left(\frac{f_F}{f}\right)^{\frac{\theta-(\sigma-1)}{\sigma-1}}\left(\frac{\mu(i)}{\mu(i,x)-\mu(i)}\right)^{\frac{\theta}{\sigma-1}}\right] \qquad (3.2)$$

Locations with greater relative market access $\frac{\mu(i,x)-\mu(i)}{\mu(i)}$ have a harsher selection criterion for exporting $z_x(i)$, but a more relaxed selection criterion for entry $z_e(i)$. The derivation is provided in the appendix.

Suppose $\gamma > 1$, so that the central location is farther away from the foreign market than peripheral locations are. Then, this result implies that the entry cutoff is more relaxed in the center than the periphery ($z_e(c) < z_e(p)$), but the exporting cutoff is harsher in the center than the periphery ($z_x(c) > z_x(p)$). The intuition is as follows. Since the central location will have better market access to the home markets, but worse market access to the foreign market ($\frac{\mu(c,x)-\mu(c)}{\mu(c)} < \frac{\mu(s,x)-\mu(s)}{\mu(s)}$), all else equal, firms in the center find it more costly to trade externally, but less costly to trade internally. This implies that exporters in the center are more productive firms compared to exporters at the periphery, while some less productive non-exporters are able to survive in the center, but not the periphery. With free worker mobility and a sufficiently large foreign market, the positive selection of exporters and negative selection of non-exporters in the center relative to the periphery will be compounded, as a larger population size in the periphery relative to the center will drive up periphery wages compared to the center due to congestion costs. We illustrate these effects in figure 3.4 in the next section.

The presence of exeral trade opportunities is key to our result on differential selection across cities. In autarky, our model shows that there can be no difference in firm types across cities, despite the ex-ante difference in the position of each location. To see this, suppose that we shut down external trade in our model. Then, there is no exporting cutoff, and the entry cutoff for any location i becomes:

$$\left(\frac{z_e(i)}{z_{min}}\right)^{\theta} = \frac{(\sigma-1)}{\theta+(\sigma-1)}\left(\frac{f}{f_e}\right)$$

The reason that the entry selection criterion is the same across locations in this case is that all firms benefit from better market access (being in the center compared to the periphery) in the same way. Therefore, in equilibrium, adjustments in wages perfectly offset the benefits of better market access. In order to generate different compositions of firm productivity across cities, two key ingredients are required: agglomeration externalities ($\Psi_L(z,L) > 0$) and log-supermodularity ($s > 0$). These ingredients are standard and well-evidenced (Rosenthal and Strange, 2004; Puga, 2010; Combes et al., 2012; Gaubert, 2018). In this case, the entry cutoff

becomes:

$$\left(\frac{z_e(i)}{z_{min}}\right)^\theta = \frac{(\sigma-1)(1 + s\log\frac{L(i)}{L_{min}})}{\theta + (\sigma-1)(1 + s\log\frac{L(i)}{L_{min}})}\left(\frac{f}{f_e}\right)$$

This shows that larger cities comprise of firms with higher efficiency. The reason is that a larger city size benefits more efficient firms disproportionately more.

In our model, the opportunity to participate in export markets leads to different composition of firm efficiency across locations, because export opportunities only benefit more efficient firms. Only firms that are sufficiently efficient will find it profitable to export since exporting entails a fixed cost on top of a fixed operating cost, $f_F > f$. This driver of different firm compositions across locations comes from (i) fundamental differences between cities in terms of their spatial location and (ii) selection into exporting, which is similar to the one driven by (i) agglomeration externalities and (ii) log-supermodularity (Gaubert, 2018).

Model simulation: how does the relative accessibility to domestic and foreign markets shape the spatial distribution of firm types, population size, productivity, and wages? We illustrate the impact of changes in the distance between the central location and foreign markets compared to the distance between the periphery and foreign markets (γ). The numeraire is wages in the peripheral locations. When $\gamma > 1$, the center is farther from the foreign market than the periphery, while the opposite is true if $\gamma < 1$. The calibrated values for this exercise is in table 3.2 and figure 3.4 displays the results from our simulation.

The first row of figure 3.4 shows that if the periphery is sufficiently close to export markets compared to the center, then the periphery has a larger population size, higher wages, and a lower price index. There are two reasons for this. First, for locations sufficiently close to export markets, the profitability of exporting opportunities start to outweigh the profitability of domestic sales and more firms will become exporters, leading them to demand more labor and grow larger in terms of revenue and employment.[4] At the same time, if the periphery is sufficiently close to export markets relative the the center, the price index in the periphery will be lower than in the center. This is because the periphery is now more "central" than the center – it incurs a lower iceberg trade cost. While the lower price index reduces wages, a larger population density increases wages in equilibrium due to congestion costs.

The second row of figure 3.4 shows that for γ sufficiently large – when access to foreign markets is sufficiently attractive compared to access to home markets – the periphery has a higher average productivity, harsher entry cutoff, and more relaxed exporting cutoff than the periphery. The exporting selection cutoff is lower in the periphery because it has better access to foreign export markets. However, the entry selection cutoff is higher in the periphery for two reasons. First, for a non-exporter, the periphery has lower market access to all domestic markets than the center. Second, since the population density is higher in the periphery in equilibrium, labor costs are higher in the periphery.

Since the exporting cutoff is lower in the periphery and firms in the periphery have better

---

[4]Equivalently, one can obtain the same results by varying the size of export markets, for a given level of $\gamma > 1$. That is, set a $\gamma > 1$, for a sufficiently large export market, the attractiveness of exporting opportunities in the periphery outweighs the attractiveness of domestic sales opportunities.
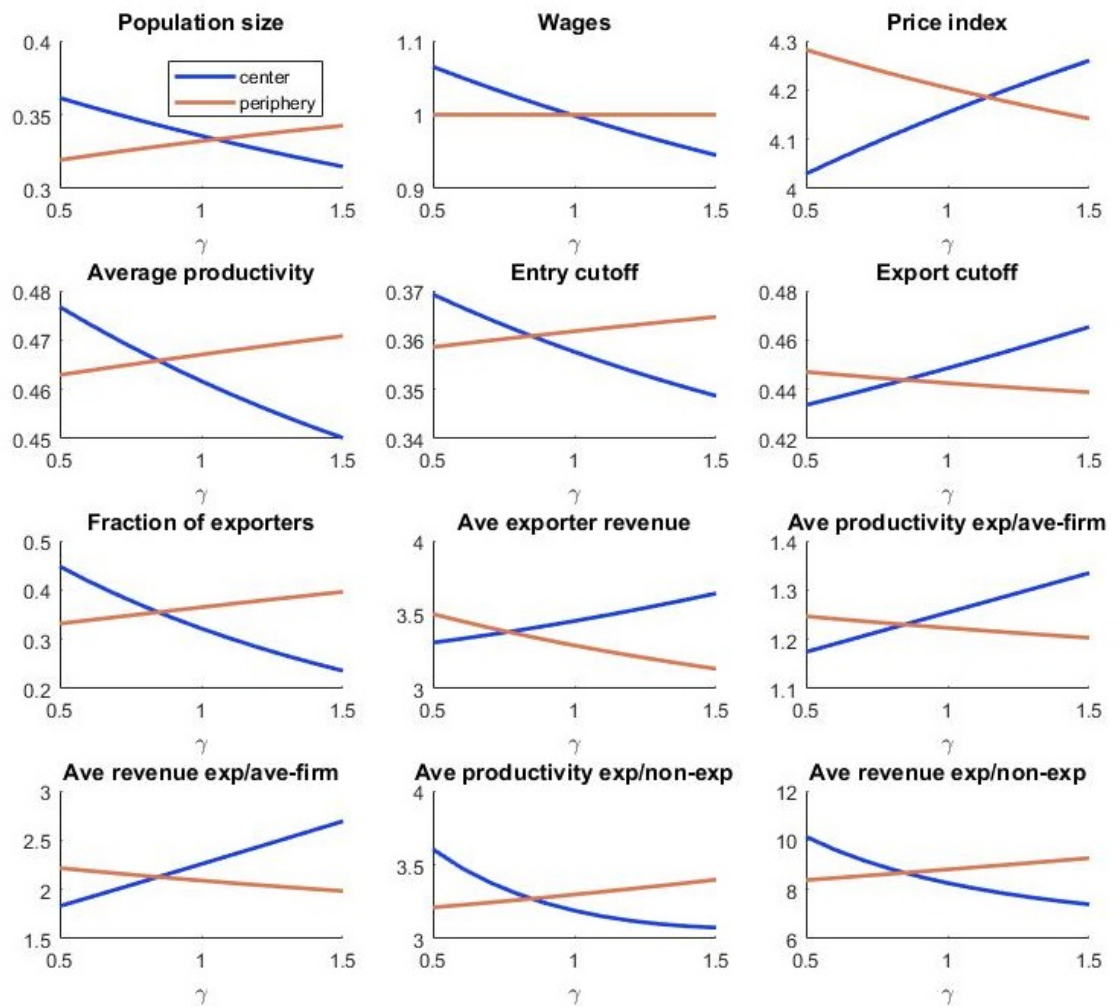
Figure 3.4: Relationship between location characteristics and distance to export markets

market access to foreign markets, the fraction of active firms who are exporters is higher and the average size of exporters is smaller there, as shown in the third row figure 3.4. This is consistent with our empirical finding in figure 3.1. Therefore, as shown in the fourth row, the difference in productivity and size (revenue) between the average exporters and the average firm is smaller in the periphery than in the center. These predictions of the model are consistent with what we observe in the data.[5] The final row shows that the difference in productivity between the average exporter and the average non-exporter is smaller for locations farther from export markets. This is inconsistent with the pattern in figure 3.2. The (in)ability of the model to make this prediction likely depends on the assumption on the distribution of firm efficiency draws $G(z)$. We currently assume that $G(z)$ is Pareto. The assumption on $G(z)$ matters for this prediction in the following sense. In the center, where the entry cutoff is lower and exporting cutoff is higher, one effect is that lower efficiency entrants can survive while another effect is that even highly efficient firms may not become exporters. Therefore, the model is ambiguous about how the difference in productivity between the average exporter and non-exporter differs by proximity to export markets.

3.5 Conclusion

What explains the large differences across firms in export participation and performance? A large literature shows that firm heterogeneity in productivity drives differences in export performance, but abstract from internal space. We study how firms' geographical location within the country simultaneously determine their export participation and performance, in addition to the spatial distribution of exporters.

We contribute by documenting new empirical patterns showing that exporters farther from export markets are systematically more productive, larger, and comprise of a smaller fraction of active firms. Existing trade models do not speak to these patterns, since these models feature at most two features among the following: heterogeneous firms, external trade, and internal space. We build a trade model with these features to interpret the empirical patterns we document as well as other well-established patterns in the literature. Here, we show that the tension between foreign and domestic trade opportunities across internal locations can help explain the data. Further, our model predicts the well-known positive relationship between city size, productivity, and that more efficient firms locate in larger cities. Existing models explain this pattern with agglomeration externalities that benefit more efficient firms disproportionately, mechanisms for which there is ample evidence. We obtain these predictions while abstracting from agglomeration externalities.

As we continue our work on this project, we seek to improve on multiple avenues. For one, we will further explore the robustness of our empirical findings by using alternative measures of distance while focusing on particular sectors. For example, we plan to use the amount of time needed to travel by road transport from an origin to a destination and focus on the furniture manufacturing sector, a sector that relies heavily on road transport to move goods. We also

---

[5]Our request for permission from the French statistical authorities to publish these results have been substantially delayed as university shutdowns have prevented access to the datasets we use. These empirical patterns will be published as soon as clearance is granted.

plan to look into the empirical relationship between location population density and exporter characteristics, a relationship for which our model provides predictions. We are currently also working on the structural estimation of a quantitative version of the model to study: (i) how important internal geography is compared productivity for firms' export performance, and (ii) the spatial implications of our model for trade liberalization policy, transport infrastructure policy, and spatial development policies aiming to attract high productivity exporters.

3.6 Appendix

3.6.1 Tables

Parameters	Values
Congestion externality (b)	1
Utility weight on tradables (η)	0.5
Utility weight on imported goods (β)	0.5
Trade cost elasticity (δ)	1
Distance between adjacent regions (d)	0.1
Distance between Home and Foreign (d_F)	0.2
Relative distance to Foreign (γ)	1.1
Agglomeration externality strength (α)	0
Log-supermodularity (s)	0
Own trade cost ($\tau(i,i)$)	1
Fixed entry cost in labor units (f_e)	0.10
Fixed operating cost in labor units (f)	0.30
Fixed exporting cost in labor units (f_F)	0.50
Firms' raw efficiency (z)	1
Total population at Home ($N(H)$)	1
Total population at Foreign ($N(F)$)	1
Periphery wage ($W(i=p)$)	1

Variable	Functional form
Adjacent trade cost	$\tau(i,j) = (1+d)^\delta$
Non-adjacent trade cost	$\tau(i,j) = (1+2d)^\delta$
Periphery-Foreign trade cost	$\tau(i,F) = (1+d_F)^\delta$
Center-Foreign trade cost	$\tau(i,F) = (1+\gamma d_F)^\delta$
Productivity ($\Psi(z,L)$)	$\Psi(z,L) = z^{(1+s\log(\frac{L(i)}{L_{min}}))}L^\alpha$

Table 3.2: Calibration and functional forms

3.6.2 Derivation: proposition 1

Start with the free entry condition is:

$$[1 - G(\tilde{z}(i))]\bar{\pi}(i) = W(i)f_e$$

where $\bar{\pi}(i)$ is the average profit and f_e is the fixed entry cost in labor units. The average profits can be expressed as:

$$\bar{\pi}(i) = \int_{\tilde{z}(i)} \pi(z, i)\, f_i(z)\, dz + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))} \int_{z_x(i)} \pi(z, i, F)\, f_i^x(z)\, dz$$

Using the above expressions for exporter and non-exporter profits, the free entry condition can be expressed as:

$$\kappa_1 W(i)^{-\sigma}\bar{\Psi}(i)^{\sigma-1}\mu(i) \; + \; \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}\kappa_1 W(i)^{-\sigma}(\mu(i, x) - \mu(i))\bar{\Psi}(i, x)^{\sigma-1}$$
$$= f + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}f_F + \frac{f_e}{1 - G(\tilde{z}(i))} \tag{3.3}$$

Next, the labor market clearing condition in a location i is:

$$\frac{L(i)}{M(i)} = \int_{\tilde{z}(i)}^{z_{max}} [\left(\sum_{j=1}^{3} l(z, i, j)\right) + f]\, f_i(z)\, dz$$
$$+ \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))} \int_{z_x(i)}^{z_{max}} [l(z, i, F) + f_F]\, f_i^x(z)\, dz$$

$$\Rightarrow \left(\frac{\sigma}{\sigma-1}\right)^\sigma \left(\frac{L(i)}{M(i)} - \left(f + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}f_F + \frac{f_e}{1 - G(\tilde{z}(i))}\right)\right) W(i)^\sigma$$
$$= \mu(i)\bar{\Psi}(i)^{\sigma-1} + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}(\mu(i, x) - \mu(i))\bar{\Psi}(i, x)^{\sigma-1} \tag{3.4}$$

Combining (3) and (4) gives the mass of firms in linear in population size:

$$M(i) = \frac{1}{\sigma\left(f + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}f_F\right) + \frac{f_e(\sigma-1)}{1 - G(\tilde{z}(i))}}L(i)$$

Substituting this into the entry cutoff gives:

$$\frac{\Psi(\tilde{z}(i), L(i))}{\bar{\Psi}(i)} = \left(\frac{1 + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}\left(\frac{\bar{\Psi}(i, x)}{\bar{\Psi}(i)}\right)^{\sigma-1}\left(\frac{\mu(i, x) - \mu(i)}{\mu(i)}\right)}{1 + \frac{1 - G(z_x(i))}{1 - G(\tilde{z}(i))}\left(\frac{f_F}{f}\right) + \frac{f_e/f}{1 - G(\tilde{z}(i))}}\right)^{\frac{1}{\sigma-1}} \tag{3.5}$$

122

Similarly, substituting this into the exporting cutoff gives:

$$\frac{\Psi(z_x(i), L(i))}{\bar{\Psi}(i,x)} = \left(\frac{\frac{1-G(z_x(i))}{1-G(\tilde{z}(i))} + \left(\frac{\mu(i)}{\mu(i,x)-\mu(i)}\right)\left(\frac{\bar{\Psi}(i)}{\bar{\Psi}(i,x)}\right)^{\sigma-1}}{\frac{f}{f_F} + \frac{1-G(z_x(i))}{1-G(\tilde{z}(i))} + \frac{f_e/f_F}{1-G(\tilde{z}(i))}} \right)^{\frac{1}{\sigma-1}} \tag{3.6}$$

When the productivity function is translog $\Psi(z, L) = z^{(1+s\log(\frac{L(i)}{L_{min}}))} L^\alpha$ and the cumulative distribution function for firms' efficiency draws is Pareto $G(z) = 1 - \left(\frac{z_{min}}{z}\right)^\theta$, the average firm productivity $\bar{\Psi}(i)$ and average exporter productivity $\bar{\Psi}(i,x)$ in location i is:

$$\bar{\Psi}(i) = \left(\frac{\theta}{\theta - (\sigma-1)\left(1 + s\log\frac{L(i)}{L_{min}}\right)} \right)^{\frac{1}{\sigma-1}} L(i)^\alpha z_e(i)^{(1+s\log\frac{L(i)}{L_{min}})}$$

$$\bar{\Psi}(i,x) = \left(\frac{\theta}{\theta - (\sigma-1)\left(1 + s\log\frac{L(i)}{L_{min}}\right)} \right)^{\frac{1}{\sigma-1}} L(i)^\alpha z_x(i)^{(1+s\log\frac{L(i)}{L_{min}})}$$

substituting these into equations (5) and (6) gives equations (1) and (2).

References

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.

Ahlfeldt, G., S. Redding, D. Sturm, and N. Wolf (2015). The economics of density: Evidence from the Berlin wall. *Econometrica 83*(6), 2127–2189.

Akerlof, G. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics 97*(4), 543–569.

Allen, T. and C. Arkolakis (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics 129*(3), 1085–1140.

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.

Arellano-Bover, J. (2020). Career consequences of firm heterogeneity for young workers: First job and firm size. *Working paper*.

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.

Atkin, D. and D. Donaldson (2015). Who's getting globalized? the size and implications of intra-national trade costs. *Working paper*.

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*.

Behrens, K., G. Duranton, and F. Robert-Nicoud (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy 122*(3), 507–553.

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

Bellone, F., P. Musso, L. Nesta, and F. Warzynski (2016). International trade and firm-level markups when location and quality matter. *Journal of Economic Geography 16*, 67–91.

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.

Bergounhon, F., C. Lenoir, and I. Mejean (2018). A guideline to French firm level trade data. *working paper*.

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

Berlingieri, G., S. Calligaris, and C. Criscuolo (2018). The productivity-wage premium: Does size still matter in a service economy? *OECD Science, Technology and Industry Working Papers*.

Bernard, A., B. Jensen, S. Redding, and P. Schott (2007). Firms in international trade. *Journal of Economic Perspectives 21*(3), 105–130.

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

Bloom, N., F. Guvenen, B. Smith, J. Song, and T. Von Wachter (2018). Inequality and the disappearing large firm wage premium. *working paper*.

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

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

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

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

Brown, C. and J. Medoff (1989). The employer size-wage effect. *Journal of Political Economy 97*(5), 1027–1059.

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*.

Caliendo, L., M. Dvorkin, and F. Parro (2019). Trade and labor market dynamics: General equilibrium analysis of the China trade shock. *Econometrica 87*(3), 741–835.

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

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

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

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

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

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

Colonelli, E., J. Tag, M. Webb, and S. Wolter (2018). A cross-country comparison of dynamics

in the large firm wage premium. *American Economic Review: Papers and Proceedings 108*, 323–327.

Combes, P., G. Duranton, L. Gobillon, D. Puga, and S. Roux (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica 80*(6), 2543–2594.

Cosar, A. and P. Fajgelbaum (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics 8*(1), 24–56.

Crozet, M. and P. Koenig (2010). Structural gravity equations with intensive and extensive margins. *Canadian Journal of Economics 43*(1), 41–62.

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum (2019). Matching in cities. *Working paper*.

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

Davis, D. and J. Dingel (2019). A spatial knowledge economy. *American Economic Review 109*(1), 153–170.

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

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

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

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

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

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

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

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review 108*(4), 899–934.

Donaldson, D. and R. Hornbeck (2016). Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics*, 799–858.

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.

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from French firms. *Econometrica 79*(5), 1453–1498.

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.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. *Journal of Political Economy 133*(3).

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*.

Fajgelbaum, P. and C. Gaubert (2018). Optimal spatial policies, geography and sorting. *NBER working paper No. 24632*.

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*.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review 108*(11), 3117–3153.

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.*

Haltiwanger, J., H. Hyatt, L. Kahn, and E. McEntarfer (2017). Cyclical job ladders by firm size and firm wage. *Working paper.*

Heblich, S., S. Redding, and D. Sturm (2020). Making of the modern metropolis: Evidence from London. *Quarterly Journal of Economics.*

Helm, I. (2020). National industry trade shocks, local labour markets, and agglomeration spillovers. *Review of Economic Studies 87*(3), 1399–1431.

Helm, I., A. Kuegler, and U. Schoenberg (2020). Displacement effects in manufacturing. *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.

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

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

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

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

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, K. and B. Kimasa (2019). Firm dynamics with frictional product and labor markets. *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.*

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

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.

Main, B. and B. Reilly (1993). The employer size-wage gap.

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.

Martellini, P. (2020). The city-size wage premium: Origins and aggregate implications. *Working paper*.

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

Melitz, M. J. (2003b, November). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 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.

Moore, H. (1911). Laws of wages: An essay in statistical economics. *Macmillan*.

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

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

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

Moscarini, G. and F. Postel-Vinay (2012). The contribution of large and small employers to

job creation in times of high and low unemployment. *American Economic Review 102*(6), 2509–2539.

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*.

Oi, W. and T. Idson (1999). Firm size and wages. *Handbook of Labor Economics 3*, 2165–2214.

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

Overman, H., S. Redding, and A. Venables (2003). The economic geography of trade, production, and income: A survey of empirics. *Handbook of International Trade (eds E.K. Choi and J. Harrigan).*

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

Pflueger, M. and T. Tabuchi (2019). Comparative advantage, agglomeration economies and trade costs. *Journal of Urban Economics 109*, 1–13.

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

Piyapromdee, S. (2018). Residual wage dispersion with efficiency wages. *International Economic Review 59*(3), 1315–1343.

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

Puga, D. (2010). The magnitude and causes of agglomeration economies. *Journal of Regional Science 50*(1), 203–219.

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*.

Rosenthal, S. and W. Strange (2004). Evidence on the nature and sources of agglomeration economies. *Chapter 49, Handbook of Regional and Urban Economics, edited by J.V. Henderson and J.F. Thisse 4.*

Salop, S. (1979). A model of the natural rate of unemployment. *American Economic Review 69*(1), 117–125.

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

Schmidt, C. and K. Zimmermann (1991). Work characteristics, firm size and wages. *Review of Economics and Statistics 73*(4), 705–710.

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

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

Shapiro, C. and J. Stiglitz (1984). Unemployment as a worker discipline device. *American Economic Review*, 433–444.

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*.

Troske, K. (1999). Evidence on the worker-firm size premium from worker-establishment matched data. *Review of Economics and Statistics 81*(1), 15–26.

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

Weiss, A. (1980). Job queues and layoffs in labor markets with flexible wages. *Journal of Political Economy 88*(3), 526–538.

Wong, H. (2020). Understanding high-wage and low-wage firms. *Working paper*.

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*.