THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par l'Université Toulouse Capitole

École doctorale : **Sciences Economiques-Toulouse School of Economics**

Présentée et soutenue par

AZKARATE-ASKASUA Miren

le 10 septembre 2020

Essays on Macroeconomics and Labor Markets

Discipline : **Sciences Economiques**

Unité de recherche : **TSE-R (UMR CNRS 5314 – INRA 1415)**

Directeur de thèse : Monsieur Christian HELLWIG, Professeur, Université Toulouse Capitole

JURY

Rapporteurs Monsieur Jean-Marc ROBIN, Professeur, Science-Po et UCL
Madame Uta SCHÖNBERG, Professeur, University College London (UCL)

Suffragants Monsieur Patrick FÈVE, Professeur, Université Toulouse Capitole
Monsieur Christian HELLWIG, Professeur, Université Toulouse Capitole

Essays on Macroeconomics and Labor Markets

Ph.D. Thesis

Miren Azkarate-Askasua[1]

Toulouse School of Economics

September 2020

[1]Address: 1, Esplanade de l'Université, 31080 Toulouse, France; email: miren.askasua@gmail.com

Acknowledgements

I would like to thank my advisor Christian Hellwig for his invaluable guidance and help during my PhD as well as for his support during the job market. I am also very grateful to Fabrice Collard and Patrick Fève for their time and perspective at times that I could not have it myself. I am also indebted to other current and past members of the macro group at TSE, in particular, Tiziana Assenza, Martial Dupaigne, Tim Lee, Robert Ulbricht and Nicolas Werquin for building an encouraging environment. I am grateful for their engagement at the workshops and seminars where most of my training happened.

I am also thankful to the rest of macro students and friends for their support that helped me go through my PhD. In particular, I am grateful to Miguel Zerecero with whom I built up most of my thesis and I spent so many fruitful discussions and fun moments. I also want to thank for making this journey more joyful and contributing to my work to Rodrigo Arnabal, Hussein Bidawi, Jacopo Bregolin, Paloma Carrillo, Filippo Maria D'Arcangelo, Joana Duran, Alberto Grillo, Elia Lapenta, Paula Martinatto, Celia Ruiz and Ling Zhou. I am also thankful to friends outside TSE for keeping me in contact with reality.

I am more than grateful to my family, and to my parents in particular, for their continuing support. Finally, special thanks to Nadia Meneses for being there in all the ups and downs of the journey. This thesis would not be here if it was not for her. Nire bizipoza elikatzen egunero laguntzeagatik, eskerrik asko.

Abstract - English

This thesis contains three essays on the macroeconomic effects of labor markets with a special emphasis on market power and the determination of wages.

In the first chapter, Miguel Zerecero and I study the efficiency and welfare effects of employer and union labor market power. We use data of French manufacturing firms to first document a negative relationship between employment concentration and wages and labor shares. At the micro-level, we identify the effects of employment concentration thanks to mass layoff shocks to competitors. Second, we develop a bargaining model in general equilibrium that incorporates employer and union labor market power. The model features structural labor wedges that are heterogeneous across firms and potentially generate misallocation of resources. We propose an estimation strategy that separately identifies the structural parameters determining both sources of labor market power. Furthermore, we allow different parameters across industries which contributes to the heterogeneity of the wedges. We show that observing wage and employment data is enough to compute counterfactuals relative to the baseline. Third, we evaluate the efficiency and welfare losses from labor market distortions. Eliminating employer and union labor market power increases output by 1.6% and the labor share by 21 percentage points translating into significant welfare gains for workers. Workers' geographic mobility is key to realize the output gains from competition.

In the second chapter, Miguel Zerecero and I propose a bias correction method for estimations of quadratic forms in the parameters of linear models. It is known that those quadratic forms exhibit small-sample bias that appears when one wants to perform a variance decomposition such as decomposing the sources of wage inequality. When the number of covariates is large, the direct computation for a bias correction is not feasible and we propose a bootstrap method to estimate the correction. Our method accommodates different assumptions on the structure of the error term including general heteroscedasticity and serial correlation. Our approach has the benefit of correcting the bias of multiple quadratic forms of the same linear model without increasing the computational cost and being very flexible. We show with Monte Carlo simulations that our bootstrap procedure is effective in correcting the bias and we compare it to other methods in the literature. Using administrative data for France, we apply our method by doing a variance decomposition of a linear model of log wages with person and firm fixed effects. We find that the person and firm effects are less important in explaining the variance of log wages after correcting for the bias.

In the third chapter, I study peer effects at the workplace. I focus on how potential peers determine a worker's location and her future wage profile. I empirically disentangle if workplace peers affect each other through learning or network effects. Similarly to the literature, I document the importance of learning which is more pronounced for the youngest cohorts arguably with no networks. I propose a structural model to understand the mechanism behind learning. The final goal of the model is to quantify the impact of peer learning the firm geographical allocation of workers, and on the rising between firm wage inequality.

Abstract - French

Ce travail de thèse est composé de trois chapitres traitant du marché du travail et de macroéconomie avec une emphase particulière sur le pouvoir du marché et la détermination des salaires.

Dans le premier chapitre, Miguel Zerecero et moi étudions les effets du pouvoir du marché des employeurs et les syndicats sur l'efficience et le bien-être. Nous utilisons des données du secteur de la production industrielle française pour documenter premièrement la relation négative entre concentration d'emploi avec les salaires et la partie de la valeur ajoutée qui va au paiement du travail. Au niveau micro, nous identifions les effets de la concentration d'emploi grâce à un choque de licenciement aux compétiteurs. À la suite nous construisons un modèle de négociations en équilibre général avec pouvoir de marché des employeurs et les syndicats. Ce modèle délivre des wedges structurelles hétérogènes à travers des entreprises que génère potentiellement une mis-allocation des ressources. Nous proposons une estimation qu'identifie séparément chaque source de pouvoir du marché au marché de travail. En outre nous permettons que les paramètres soient flexibles à travers des secteurs ce qui contribue à l'hétérogénéité des wedges. Nous montrons que l'observation des salaires et niveau d'emploi est suffisant pour calculer des contrefactuelles relatives à la base. Nous évaluons le coût des distorsions du marché du travail. Éliminer le pouvoir du marché des employeurs et les syndicats augmente la production en 1.6% et la partie qui va au paiement de la main d'oeuvre en 21 points pourcentuelles ce qui signifie une augmentation significative du bien-être des salariés. La mobilité géographique est la clé pour réaliser les gains de la compétition.

Dans le second chapitre, Miguel Zerecero et moi proposons une méthode de correction de biais qui apparait dans les estimations des formes quadratiques des paramètres de modèles linéaires. Ce biais de faible échantillonnage apparait quand nous voulons faire une décomposition de variance comme par exemple pour décomposer les sources des inégalités salariales. Quand le nombre de variables indépendantes est grand, le calcul directe du biais n'est pas faisable. Nous proposons une méthode de bootstrap pour corriger le biais. Notre méthode s'adapte à différentes hypothèses de la structure des erreurs comme heteroscdecasticité et autocorrélation. Nous pouvons corriger le biais de plusieurs formes quadratiques d'un modèle linéaire sans augmenter le coût des calculs. Nous montrons à travers de simulations de Monte Carlo que notre procédure de bootstrap effectivement corrige le biais et nous le comparons à d'autres méthodes de la littérature. Nous misons en application notre méthode avec des données administratives françaises pour faire une décomposition de la variance des salaires avec effets fixes de travailleur et entreprise. Nous trouvons que les effets de personne et entreprise sont moins importants une fois nous avons corrigé pour le biais.

Dans le dernier chapitre, j'étudie l'effet des collègues au lieu de travail. En particulier, comment collègues potentielles déterminent l'emplacement et les salaires futures des travailleurs. Je démêle empiriquement entre les effets d'apprentissage et réseau. De la même façon que la littérature je documente l'importance de l'apprentissage pour les plus jeunes qui n'ont pas eu le temps de former leur réseau. Je propose un modèle

structurale pour comprendre les mécanismes d'apprentissage. Le but est quantifier l'effet de l'apprentissage des collègues sur l'allocation entre firmes, l'allocation géographique et l'augmentation des inégalités salariales à travers des entreprises.

Table of Contents

List of Tables

List of Figures

Chapter 1

The Aggregate Effects of Labor Market Concentration

Miren Azkarate-Askasua and Miguel Zerecero[1]

Abstract

What are the efficiency and welfare effects of employer and union labor market power? We use data of French manufacturing firms to first document a negative relationship between employment concentration and wages and labor shares. At the micro-level, we identify the effects of employment concentration thanks to mass layoff shocks to competitors. Second, we develop a bargaining model in general equilibrium that incorporates employer and union labor market power. The model features structural labor wedges that are heterogeneous across firms and potentially generate misallocation of resources. We propose an estimation strategy that separately identifies the structural parameters determining both sources of labor market power. Furthermore, we allow different parameters across industries which contributes to the heterogeneity of the wedges. We show that observing wage and employment data is enough to compute counterfactuals relative to the baseline. Third, we evaluate the efficiency and welfare losses from labor market distortions. Eliminating employer and union labor market power increases output by 1.6% and the labor share by 21 percentage points translating into significant welfare gains for workers. Workers' geographic mobility is key to realize the output gains from competition.

JEL Codes: J2, J42, D24

Keywords: Labor markets, market power, misallocation

[1]We are indebted to our advisor Christian Hellwig for his invaluable guidance throughout this project. We are grateful to Fabrice Collard and Patrick Fève for their extensive suggestions. We also thank Tiziana Assenza, Sergio Feijoo, Simon Fuchs, Elia Lapenta, Nour Meddahi, Federico Mendelman, François Poinas, Sébastien Roux, Robert Ulbricht and Nicolas Werquin for their comments and suggestions. We finally thank the participants at the TSE Macro workshop and researchers at the Federal Reserve Bank of Atlanta, Universitat de Girona and University of Tilburg for useful comments. First version: October 1, 2018. All errors are our own.

1.1 Introduction

There is growing evidence, especially for the United States, linking lower wages to labor market concentration.[2] Indeed, if this concentration reflects monopsony power in the labor market, standard theory predicts that establishments *mark down* wages by paying workers less than their marginal revenue product of labor. On the other hand, if labor market institutions enable workers to organize and have a say over the wage setting process, bargaining can mitigate, or even reverse, the effect of establishments' market power on wages.

In this paper we quantify the efficiency and welfare losses from labor market power in the French manufacturing sector. The French case stands out over other developed countries, especially with respect to the U.S., for having regulations that significantly empower workers over employers.[3] We therefore provide a framework that incorporates both, employer and union labor market power. Our main result is that, holding the total labor supply constant, removing employer and workers' labor market power increases French manufacturing output by 1.6 percent. Even if productivity and output gains are relatively small, distributional effects are important as the labor share increases by 21 percentage points and the aggregate wage rises by 45 percent. This wage increase translates into median expected welfare gains of 42 percent for workers.

We proceed in three steps. First, we establish empirically that, within a same firm, establishments with *higher* local employment shares pay *lower* wages for same occupations. We identify this effect by using a competitors national mass layoff shock as an external source of variation to an establishment's local employment share. Second, in line with the previous empirical result and the French labor institutional setting, we build and estimate a model where labor market power arises from: (i) employers that face upward sloping labor supplies, and (ii) workers that bargain over the wages. Third, we use the model to quantify the efficiency and welfare consequences of employers and workers' labor market power.

We start by documenting the link between concentration and wages/labor shares. We use data on French manufacturing firms from 1994 to 2007. Employer labor market power is related to the notion of local labor markets. We define those as a combination of commuting zone, industry, and occupation, and measure concentration at the local labor market level using the Herfindahl-Hirschman Index.[4] We find that concentrated industries have on average lower labor shares. Passing from the first to the third quartile of local labor market concentration, the labor share is reduced by 1 percentage point.

At the establishment-occupation level, our proxy for the strength of labor market power is the employment share within the local labor market. To explore a link between concentration and labor payments, we need to overcome the potential endogeneity of the employment share and the wages. Therefore, we instrument employment shares with negative employment shocks or mass layoffs to competitors. Identification comes from residual within firm-occupation-year variation across local labor markets. Depending on the specification, the estimated elasticity ranges from -0.17 to -0.04. That is, a 1 percentage point increase of employment share lowers the establishment wage by up to 0.17 percent.[5]

After presenting the reduced form evidence, we build a general equilibrium model that incorporates two

[2] See for example Berger et al. (2019), Jarosch et al. (2019b), Benmelech et al. (2018) among others.

[3] French labor market is characterized by having low unionization rates but high coverage of collective agreements. This is due to the institutional setting of the labor market that empowers union representation depending on the firm size. Section 1.3.4 provides more detail on the French institutional setting.

[4] The Herfindahl-Hirschman Index is defined as the sum of the squares of *employment* shares.

[5] This corresponds to a reduction of roughly 1000 euros (at 2015 prices) per year if we pass from the first to the third quartile of the employment share distribution.

elements: employer and union labor market power. First, we borrow from the trade and urban economics literature (e.g. Eaton and Kortum, 2002; Ahlfeldt et al., 2015) and assume workers have stochastic preferences to work at different workplaces. Heterogeneity of workers' tastes implies individual establishment-occupations face an upward sloping labor supply curve which gives rise to employer labor market power. In the absence of bargaining, as there is a discrete set of establishment-occupations per local labor market, employers act strategically and compete for workers in an oligopsonistic fashion. Wages are therefore paid with a markdown which is a function of the *perceived* labor supply elasticity. Similarly to Atkeson and Burstein (2008), this elasticity in turn depends on the employment share within the local labor market. The framework without bargaining is similar to Berger et al. (2019) under Bertrand competition. The second element is collective wage bargaining. We assume wages are set at the establishment-occupation level and workers force a wage setting process where they bargain over the status-quo scenario, the oligopsonistic competition outcome. In doing so, they internalize that if bargaining were to fail establishments compete oligopsonistically on the local labor market. Workers' ability to extract rents over that outside option depends on their bargaining power in a reduced form Nash bargaining.

This wage-setting process leads to a distortion that is reflected in a wedge between the equilibrium negotiated wage and the marginal revenue product of labor. This wedge summarizes both sides of market power as it is a combination of both, a markdown due to the oligopsony power, and a markup due to wage bargaining. The smaller this wedge is, the larger the market power of employers relative to workers and vice-versa. Heterogeneity of the labor wedge across establishments distorts relative wages and potentially generate misallocation of resources that decrease aggregate output. Heterogeneity comes from two sources: (i) the dependence of the markdown on industry specific labor supply elasticities and employment shares; and (ii) the across industry differences in the markup due to diversity of bargaining powers. Our model nests as special cases both a full bargaining setting or a model with oligopsonistic competition only.

Our framework features a large number of different prices, the establishment-occupation wages plus the product prices. We show how to solve for the general equilibrium of the model in two steps. We solve first for wages in each local labor market normalizing aggregate prices. Second, we show how to build industry level fundamentals and solve for aggregate prices. This two-step procedure eases the solution because the model can be rewritten at the industry level.[6] We provide an analytical characterization of the equilibrium at the industry level and along the way prove the existence and uniqueness of the equilibrium. This allows us to use the model to back out fundamentals that rationalize the observed data and perform counterfactuals on actual data without worrying about multiple equilibria.

After the model set-up, we discuss how to identify and estimate the model parameters. We have two types of parameters: the ones related to the labor supply and bargaining, and the ones related to technology. Regarding the labor supply, we assume that workers face a sequential decision: in a first stage, they observe their preferences for different local labor markets and choose the one that maximizes their expected utility; in a second stage, they observe their preferences to work for different employers and choose the establishment. Therefore, these labor supplies depend on two key parameters that jointly determine the magnitude of employers' labor market power: a *within* local labor market elasticity and an *across* local labor market elasticity.

[6]The intuition behind this is that after solving for wages for given industry and economy-wide constants, we can fully characterize the allocation of labor and capital *within* each industry. This fact, combined with the information about the establishment-level fundamentals, allows us to aggregate the model at the industry level with corresponding industry-level fundamentals.

They govern, respectively, the intensity of how workers respond to changes in *establishment* wages *within* a local labor market, and how workers react to changes in *average* utilities (which are in turn a function of establishment wages) *across* local labor markets.

The main challenge is to separately identify the union bargaining powers from the within and across local market labor supply elasticities. We propose a strategy to estimate the labor supply elasticities that is independent from the underlying wage setting process. Therefore, our identification strategy is readily applicable to set-ups with or without bargaining. In the first step we estimate the *across* local labor market elasticity and the inverse labor demand elasticity adapting the identification through heteroskedasticity of Rigobon (2003). We use the insight that the across local labor market elasticity is the only relevant elasticity for the establishments that are alone in their local labor markets, the full monopsonists. Their local labor market equilibrium boils down to a standard system representing the labor supply and demand equations. Ordinary least squares estimates present the traditional problem of other price-quantity systems as the estimated elasticities are biased towards zero. Rather than instrumenting to get exogenous variation in labor supply and demand, we identify using a restriction on the variance-covariance of structural shocks across occupations and their heteroskedasticity.[7] The identifying assumption is that the covariances between the labor demand and supply shifters, productivities and amenities respectively, are the same across occupations but not the variances. To gain intuition, let's fix the labor demand constant and assume different variances of the labor supply shifter, the amenity, across occupations. Increasing the variance of the labor supply shifter helps to identify the other side of the market, the labor demand.

In a second step we estimate the *within* local labor market labor supply elasticities by directly estimating the labor supply equation. We instrument for the wages by using revenue productivities as labor demand shifters and estimate by conditioning on within local labor market variation. This requires the inverse labor demand elasticity estimated in the first step. Finally, we calibrate the industry specific technology parameters (capital and labor elasticities) and bargaining powers to match the capital and labor shares.

Once the parameters are identified, we back out model primitives to perform counterfactuals. Ideally, we would like to have the distribution of fundamentals, in particular of physical productivities, at the establishment-occupation level that rationalizes the observed data on wages and employment. We back out amenities to match employment shares. However, the model only allows us to identify *revenue* productivity, which is a function of two objects: the physical productivity and the price of the good. These unobserved prices are equilibrium objects and the inability to identify the non-parametric distribution of productivities has prevented most studies (e.g. Hsieh and Klenow, 2009) from conducting full blown general equilibrium counterfactuals.

We show that the general equilibrium counterfactual can be computed using only revenue productivities. We do that by writing the model in terms of relative changes with respect to the current equilibrium. This approach, borrowed from the trade literature, allows us to solve for changes of equilibrium variables relative

[7]To see the notion behind Identification through Heteroskedasticity, consider the following system: $y = \alpha x + u$ and $x = \beta y + v$, with $var(\epsilon) \equiv \sigma_\epsilon$ and $cov(u, v) = 0$. The system is under-identified as the variance-covariance matrix of (x, y) yields three moments (σ_x, σ_y and $cov(x, y)$) while we have to solve for four unknowns: $(\alpha, \beta, \sigma_u, \sigma_v)$. Now suppose we can split the data into two sub-samples with the same parameters (α, β) but different variances. Now the two sub-samples give us 3+3=6 data moments with only six unknowns: the two parameters (α, β) and the four variances of structural errors. This system is identified under the additional assumption that the variances σ_u^2, σ_v^2 are different across sub-samples.

to a baseline scenario.[8] We are able to do that because changes in revenue productivities are completely driven by changes in prices and not the physical productivity part which is fixed.

We quantify the efficiency losses of employers and workers' labor market power by removing those distortions in a counterfactual economy while keeping workers' preferences fixed. This is a counterfactual scenario where employers are price takers and workers have no bargaining power. We find that output increases by 1.6 percent while the labor share rises by 21 percentage points. This increased labor share goes together with wage gains that in turn translate into 42 percent median welfare gains for workers. Removing the heterogeneity of wedges improves the allocation of labor by increasing the employment of more productive establishments. The counterfactual gains in the labor share suggest that employer labor market power is stronger than the one of the unions. This is a consequence of the estimated low labor supply elasticities that are in the range lower than the estimates of Berger et al. (2019) for the U.S.

Additionally, we find that geographic mobility is the key margin of adjustment to achieve the baseline counterfactual productivity gains, rather than within local labor market or within industry mobility. The intuition behind this is that there are a handful of concentrated and productive firms in the rural areas and removing labor market power increases their wage and employment more relative to the urban areas. We find that labor market distortions account for 13 percentage points – about a third – of the urban/rural wage gap. Consequently, the total employment decreases in urban areas relative to the baseline, which changes the geographical composition of manufacturing employment in France.

Finally, we incorporate two extensions to the model. First, we introduce an endogenous labor force participation decision by assuming that workers may voluntarily stay out of the labor force. Output gains in this case are slightly higher than in the baseline because wage gains increase the labor force participation. Second, we allow for agglomeration forces within the local labor market that also improve the output gains from the baseline counterfactual.

Literature. This paper speaks to several strands of the literature. First, and most closely related, is the literature on employer labor market power. Several empirical papers have documented the importance of labor market concentration on wages, employment and vacancies (Benmelech et al., 2018; Azar et al., 2017, 2018). These focus on aggregate measures of concentration as the Herfindahl-Hirschman Index. Our contribution to this empirical literature is to focus on establishment level concentration and use exogenous variations to show the existence of employer labor market power in France. We argue that firms having mass layoffs constitute a quasi-natural variation on the employment shares of the non-shocked establishments. This allows us to causally identify the effect of the employment share at the local labor market, our proxy of the strength of employer labor market power, on wages.

This paper also contributes to structural work on employer labor market power. We depart from the traditional monopsony power framework (e.g. Burdett and Mortensen, 1998; Manning, 2011; Card et al., 2018; Lamadon et al., 2018) by having heterogeneous markdowns and by extending it to allow for wage bargaining. The paper is complementary to Jarosch et al. (2019b) in the sense that they consider employer labor market power in a search framework. We contribute to those papers by incorporating unions. In contemporaneous and independent work, Berger et al. (2019) build a structural model with oligopsonistic competition in local

[8] Costinot and Rodríguez-Clare (2014) refer to this method as "exact hat algebra". They use this approach to compute welfare effects of trade liberalizations using easily accessible macroeconomic data.

labor markets. We share the objective of measuring the efficiency effects of labor market distortions and reach similar quantitative conclusions, but our contribution differs from theirs in several dimensions: (i) our framework nests theirs as an special case without bargaining; (ii) we incorporate occupations and use them for the identification of the structural parameters; (iii) we allow for differences in structural parameters across industries. In particular, within local labor market elasticities and bargaining powers are diverse across industries. Importantly, this adds heterogeneity to the labor wedges and employment misallocation; (iv) on the empirical evidence, they instrument with tax changes across states in the U.S. whereas we use labor shocks to competitors; (v) we show that counterfactuals can be computed without the need to back out underlying productivities and we perform the counterfactuals using actual establishment data.

Second, the paper is related to the literature on Nash bargaining. We take the axiomatic approach (Osborne and Rubinstein, 1990) rather than the sequential or strategic approach (Binmore et al., 1986; Stole and Zwiebel, 1996; Brügemann et al., 2018) with offers and counter-offers. In our framework, collective bargaining happens at the establishment-occupation level and the employer cannot discriminate against different workers. Therefore collective bargaining applies universally even if only a subset of workers is unionized. Regarding the union bargaining power, our estimates relate to the estimates for manufacturing from Cahuc et al. (2006) in a framework with on the job search.

Third, the paper relates to the literature on imperfect competition in general. Our approach is similar to Edmond et al. (2018) and Morlacco (2018) in trying to quantify the effect of heterogeneous market power on aggregate output. They study, output and intermediate input market powers respectively while we focus on the effects of labor market power. Karabarbounis and Neiman (2013) documented the falling trend of the labor share and Barkai (2016) and Gutiérrez and Philippon (2016) the rising trend of the profit share for different countries. Output market power has been pointed out as an explanation for the decline of labor payments out of GDP (e.g. De Loecker and Eeckhout, 2017, 2018). Contrary to the evidence on output market power, other studies suggest that employer labor market power is not the driver behind the decreasing trends of the U.S. labor share (e.g. Lipsius, 2018; Berger et al., 2019). The focus of this paper is therefore not on labor share trends but on the effects employer and union labor market power in a given cross section of firms, markets and industries.

Our model builds on the trade (Eaton and Kortum, 2002) and urban economics (Redding, 2016; Ahlfeldt et al., 2015) literature. The establishment perceived elasticity has the same functional form as the perceived demand elasticities in Atkeson and Burstein (2008) under Bertrand competition. Diversity of perceived elasticities is the main source of heterogeneity of the labor wedge and is at the origin of resource misallocation as emphasized by Hsieh and Klenow (2009) and Restuccia and Rogerson (2008).

Finally, the paper contributes to micro-estimates of labor supply elasticities. Staiger et al. (2010), Falch (2010) and Berger et al. (2019) use quasi-experimental variation on the wages to estimate the firm labor supply elasticities that go from 0.1 (Staiger et al., 2010) to 5.4 (Berger et al., 2019). Both our within and across local labor market labor supply elasticities lie in that range.

The rest of the paper is organized as follows. Section 1.2 introduces the data. Section 1.3 shows the stylized facts and our empirical strategy. Section 1.4 introduces the model. Section 1.5 discusses details about parameter estimation. Section 1.6 discusses the results from counterfactual exercises. Section 1.7 presents extensions of the model and Section 1.8 concludes.

1.2 Data

We use two main data sources. Our first and primary source of data are firm-level fiscal records consisting of balance sheet information including wage bill, capital stock, number of employees and value added. This dataset is known as *FICUS* and it includes all French firms except for the smallest firms declaring at the micro-BIC regime and some agricultural firms. We also use *DADS Postes*, an employer-employee dataset with the universe of salaried employees. It provides firm and establishment identifiers (SIREN and SIRET respectively). We recover the location, occupation classification, wages and employment. This source is necessary to know how employment and wages are distributed across different establishment-occupations of a given firm. The sample covers private manufacturing firms in France from 1994 to 2007. A break in the industry classification series prevents us from extending the time span of the sample.[9] Additionally we use data relating the city codes to commuting zones and Consumer Price Index data to deflate nominal variables.[10] We define four broad categories of occupations: top management, supervisor, clerical and operational.[11] We define a local labor market as the intersection between commuting zone, 3-digit industry and occupation. On average throughout the sample there are 57.900 local labor markets per year.

Our sample consists of approximately 4 million establishment-occupation-year observations that belong to around 1.25 million firms. Details about sample selection are in Appendix 1.E.3.

1.2.1 Summary Statistics

Table 1 presents the final sample establishment-occupation level summary statistics. The median occupation at a given establishment has 2 employees and pays 27,439 euros per worker. Certain firms have occupations in different locations, which we denote as multilocation occupations. The micro evidence in the next Section focuses on multilocation firm-occupations.[12] Panels (a) and (b) of Table 1 have the summary statistics of occupations belonging to monolocation and multilocation firms. Occupations of firms with plants or establishments at multiple locations have larger average (median) size of 27 employees than the 7 employees (4 versus 2) of monolocation occupations. Firms with multilocation occupations pay wages per capita that are 15% higher than the monolocation ones.

Manufacturing firms belong to 97 3-digit industries or sub-industries that are present in 364 different commuting zones. We denote the 3-digit industries as h and the commuting zones as n. Summary statistics of sub-industries at 2007, the baseline year for the counterfactuals, are in Table 2. Average 3-digit industry labor share is 52% and the share of capital is 26%. Taking those averages, the profit share would be around 22%. We see that variation across sub-industries in size and labor productivity is important but more limited in average wage per establishment $\overline{w}_h$. Number of establishments N_h and total employment L_h are about 5 times higher passing from the first to the third quartile (from percentile 25 to 75), average wage increases by 27%.

[9] Before 1994 the wage data was imputed and after 2007 the industry classification (APE) is not consistent with previous versions. On the contrary, the classification change between the 1993 and 2003 codes are consistent at the 3-digit level.

[10] The sources are https://www.insee.fr/fr/information/2114596 and https://www.insee.fr/fr/statistiques/serie/001643154 respectively.

[11] The classification is very similar to the one in Caliendo et al. (2015). We group together their first two categories (firm owners receiving a wage and top management positions) into top management because the distinction between the two was not stable in 2002.

[12] The multilocation definition is occupation specific. A firm can have both monolocation and multilocation occupations.

Table 1 – Establishment-Occupation Summary Statistics

Statistic	Obs.	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
L_{iot}	4,151,892	11.077	1.058	2.261	6.216	59.456
$w_{iot}L_{iot}$	4,151,892	367.155	31.566	71.813	196.554	2,379.449
w_{iot}	4,151,892	34.029	20.857	27.439	39.517	117.055
$s_{io\|m}$	4,151,892	0.203	0.011	0.051	0.238	0.306
			(a) Monolocation			
L_{iot}	3,359,236	7.411	1.032	2.083	5.140	29.688
$w_{iot}L_{iot}$	3,359,236	216.710	29.636	64.480	159.624	925.159
w_{iot}	3,359,236	32.843	20.299	26.641	38.478	35.478
$s_{io\|m}$	3,359,236	0.182	0.009	0.042	0.193	0.292
			(b) Multilocation			
L_{iot}	792,656	26.612	1.294	4.101	15.061	120.345
$w_{iot}L_{iot}$	792,656	1,004.734	45.711	139.315	532.979	5,052.361
w_{iot}	792,656	39.052	23.601	30.692	43.750	257.690
$s_{io\|m}$	792,656	0.290	0.023	0.113	0.480	0.347

Notes: The top panel shows summary statistics for the whole sample. Panels (a) and (b) present respectively summary statistics of monolocation and multilocation firm-occupations. L_{iot} is full time equivalent employment at the establishment-occupation io, $w_{iot}L_{iot}$ is the wage bill, w_{iot} is establishment-occupation wage or wage per FTE, $s_{io|m}$ is the employment share out of the local labor market. All the nominal variables are in thousands of constant 2015 euros.

Table 2 – Sub-industry Summary Statistics.

Variable	Obs.	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
N_h	97	2,840.000	493	1,261	2,639	4,530.496
L_h	97	30,466.030	7,559	15,070	50,036	33,899.330
$\overline{w}_h$	97	34.607	29.562	32.990	37.531	6.902
LS_h	97	0.520	0.482	0.527	0.581	0.098
KS_h	97	0.261	0.165	0.233	0.316	0.133

Notes: N_h is the number of establishments per 3-digit industry h, L_h is total employment of h, $\overline{w}_h$ is the average establishment wage of h, LS_h is the labor share and KS_h is the capital share. We calibrate the interest rate following Barkai (2016). All the nominal variables are in thousands of constant 2015 euros.

We define a local labor market based on location, industry and occupation combinations. The choice is guided by the observed transition rates in the data where conditional on changing one of the dimensions, occupational transitions are the most common followed by changes in industry. Table 17 in Appendix 1.E.1 shows the transition rates along the location, industry and occupation dimensions. Following those transition rates, the local labor market, denoted by m, is a combination of commuting zone n, 3-digit industry h and occupations o. Table 3 presents summary statistics for local markets in 2007. The median local market is small and has only 2 establishments and 10 employees. This is a consequence of the handful of manufacturing firms that are present in the countryside demanding certain occupations. One example of a local labor market are the blue collar workers working in the food industry in Lourdes, close to the Pyrenees. The median local labor market is concentrated with a Herfindahl-Hirschman Index (HHI henceforth) of 0.68.[13] The HHI is very similar (0.69) if we consider wage bill shares $s^w_{io|m}$ instead of employment shares $s_{io|m}$. High median local labor market concentrations do not imply that most of the workers are in highly concentrated environments but rather that there are few local labor markets with low concentration levels and high employment. Further summary statistics on establishment and firm level are in Appendix 1.E.1.

Table 3 – Local Labor Market Summary Statistics. Baseline Year

Variable	Obs.	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
N_m	57,940	4.755	1	2	4	14.400
L_m	57,940	51.005	2.786	9.421	34.912	196.201
$\overline{w}_m$	57,940	36.619	24.264	30.224	42.492	36.078
$\widehat{w}_m$	57,940	36.189	24.081	30.028	42.179	25.556
HHI($s_{io\|m}$)	57,940	0.671	0.384	0.683	1.000	0.320
HHI($s^w_{io\|m}$)	57,940	0.676	0.392	0.698	1.000	0.318

Note: N_m is the number of competitors in the local labor market m, L_m is total employment in m, $\overline{w}_m$ is the mean w_{iot} of the establishment-occupations in m, $\widehat{w}_m$ is the weighted average wage at m with weights equal to employment shares, HHI($s_{io|m}$) and HHI($s^w_{io|m}$) are respectively the Herfindahls with employment and wage shares. All the nominal variables are in thousands of constant 2015 euros.

1.3 Empirical Evidence

This section provides suggestive evidence of employer labor market power in France and presents the French institutional setting. We start by documenting some stylized facts on labor market concentration and the labor share at the industry level. Those are complemented with establishment level estimates that explore a causal link between wages and concentration. Finally, we present evidence on the institutional framework of French labor market and the importance of wage bargaining.

[13] The Herfindahl of local labor market m ranges from the inverse of the number of competitors ($1/N_m$) if all the establishments have the same shares to 1. A local labor market can have a HHI of almost one if one establishment has virtually all the employment.

1.3.1 Concentration and the Labor Share

A standard measure of concentration is the Herfindahl Hirschman Index (HHI). From our definition of local labor market m, the HHI of market m, HHI_{mt}, is the sum of the squared employment shares of the plants present in m. Labor share at the 3-digit industry level, LS_h, is the ratio of the wage bill over value added. Due to data restrictions of observing value added only at the firm level, we cannot compute labor shares at the local labor market level. We build a sub-industry concentration index $\overline{HHI}_{ht}$ by taking the employment weighted mean of HHI_{mt} across different local labor markets.[14]

We use the following specification:

$$\log(LS_{h,t}) = \delta_{b,t} + \beta \log(\overline{HHI}_{h,t}) + \varepsilon_{h,t}. \tag{1.1}$$

Table 4 presents the results. Column (3) shows that the negative correlation between employment concentration and the labor share is robust to controlling for industry and industry-year fixed effects. Industry fixed effects capture differences across industries in the usage of capital. The focus of the paper being the cross sectional allocation of resources we also take industry-year fixed effects to only use cross sectional variation.

This regression gives a sense of the importance of the labor wedge heterogeneity to generate output and labor share losses. At face value, the estimate with industry fixed effects (Column (2)) imply a reduction of 1 percentage point of the labor share when passing from the first to the third quartile of concentration (quartiles of $HHI(s_{io|m})$ in Table 3). Estimates in Column (3) with industry-year fixed effects are very similar.

[14]The HHI index at market m and year t is: $\sum_{i\in\mathcal{I}_m} s_{io|m}^2$ where shares at the market are accounted as shares of full time equivalent employees and $\mathcal{I}_m$ is the set of all firms in the sub-market m. The sub-industry concentration index $\overline{HHI}_{ht}$ is:

$$\overline{HHI}_{ht} = \frac{1}{|\mathcal{M}_h|} \sum_{m\in\mathcal{M}_h} HHI_{mt} \frac{L_{mt}}{L_{ht}},$$

where $|\mathcal{M}_h|$ is the number of local labor markets that belong to h, L_m is the local labor market employment and L_h is the 3-digit industry employment.

Table 4 – Concentration and Labor Share

	$\log(LS_{h,t})$		
	(1)	(2)	(3)
$\log(\overline{HHI}_{h,t})$	-0.064^{***}	-0.054^{***}	-0.056^{***}
	(0.013)	(0.013)	(0.014)
Industry FE	N	Y	N
Industry-year FE	N	N	Y
Observations	1357	1357	1357
R^2	0.017	0.290	0.343
Adjusted R^2	0.017	0.280	0.170

Notes: This table presents estimates of equation (1.1). Column (1) presents the estimate without any fixed effect. Column (2) shows the exercise with industry fixed effects and Column (3) has industry-year fixed effects. The dependent variable is the logarithm of 3-digit industry h labor share $\log(LS_{h,t})$ at time t. $\log(\overline{HHI}_{h,t})$ is the logarithm of the employment weighted average of the local labor market Herfindahl Index. *p<0.1; **p<0.05; ***p<0.01

The small estimated coefficient is most likely a result of two effects: the averaging across different local labor markets and level effects. The regression does not take into account the effect of concentration on the average level of the labor share as this is absorbed by the fixed effects. Below we use this empirical exercise to validate our model.

1.3.2 Concentration and Wages

This section explores a causal relationship between employer labor market power and wages. The challenge is finding a source of exogenous variation in our proxy of local labor market power, the employment share $s_{io|m}$, that will allow to estimate the effect of employer market power on wages or labor shares. Given our restriction of not observing value added at the plant level, we focus on wages. We briefly discuss the type of shocks we account for in the main specification and later on present our instrumental variable (IV henceforth) estimates with two different instruments. We focus on multi-location occupation for both exercises and the effects are estimated using residual variation across local labor markets within a firm-occupation-year.

The baseline specification is:

$$\log(w_{io,t}) = \beta\, s_{io|m,t} + \psi_{\mathbf{J}(i),o,t} + \delta_{\mathbf{N}(i),t} + \varepsilon_{io,t}\,, \tag{1.2}$$

where $\log(w_{io,t})$ is the log average wage at plant i of firm j and occupation o at sub-market m in year t, $s_{io|m,t}$ is the employment share of the plant out of the market m, $\psi_{\mathbf{J}(i),o,t}$ is a firm-occupation-year fixed effect, $\delta_{\mathbf{N}(i),t}$ is a commuting zone-year fixed effect and $\varepsilon_{io,t}$ is an error term. Our parameter of interest is β.

The specification controls for industry labor demand shocks with firm-occupation-year fixed effects $\psi_{\mathbf{J}(i),o,t}$. These include for example trade shocks either to manufacturing as a whole or for a particular industry. Shocks to occupation labor demand at the aggregate or firm level are captured by the fixed effects $\psi_{\mathbf{J}(i),o,t}$. Lastly, the commuting zone times year fixed effects $\delta_{\mathbf{N}(i),t}$ control for permanent differences across locations and also for potential geographical spillovers of mass layoff shocks as stressed by Gathmann et al. (2017).

Establishment i and occupation o employment share, $s_{io|m,t}$, is very likely to be endogenous to the wages themselves. On the one hand, everything else equal, higher wages attract more workers and therefore increase the employment share. On the other hand, if there is labor market power on the employer side, we expect two establishments with the same fundamentals to pay differently depending on their local labor market power. That is, everything else equal, we expect the plant with higher employment share to pay relatively less than the one at a more competitive local labor market. Given these endogeneity issues, we propose two different instruments for the employment share. First, we instrument for the employment share by using lagged measures of concentration and second, we use a quasi experimental variation of the employment shares coming from mass layoff shocks to competitors.

Lagged Concentration Measures

We start by instrumenting the employment share by lagged concentration measures. More specifically, we instrument the employment share $s_{io|m,t}$ by the lagged inverse of the number of competitors at the local labor market $1/N_{m,t-1}$. Lagged concentration measures exclude potential endogeneity of the market structure to current period shocks. The correlation between employment shares and lagged concentration measures is 0.77.

Table 5 shows the results. The first two columns recover estimates of the specification (1.2) with commuting zone (CZ) fixed effects and the last two columns with commuting zone-year fixed effects. Columns (1) and (3) present the Ordinary Least Squares (OLS) estimates. This econometric model reflects both labor demand and supply therefore a direct OLS estimation of (1.2) is theoretically problematic and expected to be biased towards zero. We indeed find that both OLS estimates are very close to zero and positive. Columns (2) and (4) present the results once we instrument for the employment share. Both specifications (with CZ and CZ-year fixed effects) give the same point estimates. Those imply that an increase of one percentage point (p.p. henceforth) of the local labor market share is associated with a decrease of 0.03% of the plant wage. This implies that the same establishment passing from the first to the third quartile of the employment share distribution reduce 0.68% the wages. This elasticity translates into a reduction of roughly 190 euros of the median yearly establishment-occupation wage.

Labor Shock to Competitors

We propose a second reduced form estimation to provide further evidence on the causal link between labor market concentration and wages. We now instrument the endogenous employment shares by using quasi-experimental variation coming from mass layoffs to competitors. The instrument is built by the presence of a firm having a *national* mass layoff in the same local labor market as non affected establishments. We expect that a national level shock is exogenous to the residual within firm-occupation variation across local labor markets that identifies the effect. Here we provide some detail of the construction of the instruments that is

Table 5 – Wage Regression. Multi-location firms

	Dependent variable:			
	$log(w_{io,t})$			
	OLS	IV	OLS	IV
$s_{io\|m,t}$	0.010^{***}	-0.030^{***}	0.007^{***}	-0.030^{***}
	(0.001)	(0.002)	(0.001)	(0.002)
Firm-Occ-Year FE	Y	Y	Y	Y
CZ FE	Y	Y	N	N
CZ-Year FE	N	N	Y	Y
Observations	792,656	733,576	792,656	733,576
R^2	0.833	0.861	0.853	0.862
Adjusted R^2	0.763	0.802	0.790	0.802

Notes: Columns (1) and (2) present estimates with commuting zone (CZ) fixed effects for the ordinary least squares (OLS) and instrumental variable (IV) exercises. The instruments in this table are lagged concentration measures $\frac{1}{N_{m,t-1}}$. Columns (3) and (4) present the analogous with commuting zone-year fixed effects. The dependent variable $log(w_{io,t})$ is the logarithm of establishment-occupation wage at time t. $s_{io|m,t}$ is the establishment-occupation employment share at time t. *p<0.1; **p<0.05; ***p<0.01

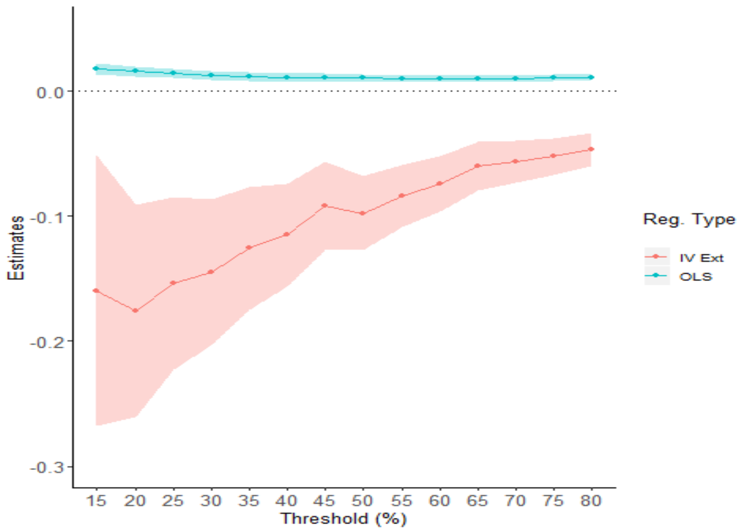
complemented in Appendix 1.F.

We first need to identify the firms suffering from a mass layoff. We classify a firm-occupation as having a mass layoff if the establishment-occupation employment at t is less than a threshold κ% of the employment last year for all the firm establishments. Ideally we would like to identify firms that went bankrupt ($\kappa = 0$). Unfortunately, we cannot externally identify if a firm disappears because it went bankrupt or changes identifiers keeping the number of competitors at the local market constant. Our instrument is a proxy to capture the impact of a firm's bankruptcy into the competitors.[15] We restrict the sample to non affected firm-occupations with establishments in local labor markets with and without a competitor suffering a mass layoff. In particular, we use the subsample of firms that have establishments at local labor markets hit by a mass layoff shock to a competitor and without mass layoff shocks.

There is a trade-off when choosing κ. A lower threshold leads to considering stronger negative shocks and the generated instrument will be cleaner, but it reduces the number of events considered. This creates a bias-variance trade-off in the selection of the threshold. Lacking a clear candidate for κ, we try different cut-off values.[16]

Results with commuting zone fixed effects are in Figure 1. OLS estimates of β from (1.2) are in blue slightly above zero and IV estimates are in red.[17] Both are plotted with 95% confidence intervals.[18]

Figure 1 – Impact of Employment Share on Wages



Notes: This figure presents the point estimates and 95% confidence bands of the OLS and IV exercises on the y-axis. The x-axis presents different thresholds κ that define a mass layoff shock. The instrument is the presence of a mass layoff shock firm in the local labor market. We focus on non-affected competitors (not suffering a mass layoff shock). The specification is as (1.2) with commuting zone fixed effects. Results with commuting zone-year fixed effects are in Section 1.3.3.

The employment share being endogenous, the estimated effect with OLS is biased up and closer to zero. OLS estimates are in line with the column (3) of Table 5. The Figure shows clearly the trade-off in the selection of the cutoff κ. The lower the threshold, the stronger the impact but higher the variance of the

[15] See Appendix 1.F for a graphical illustration of the identification.

[16] A standard value in the literature is κ =70%. That is a 30% loss of employment.

[17] We are restricting to firms classified as not having a mass layoff. The regression sample therefore changes depending on κ which is why the OLS estimates change slightly with κ.

[18] Details of the point estimates and confidence bands are in Appendix 1.F.

estimated effect. With $\kappa = 20\%$ we estimate an elasticity of 0.17. A one p.p. increase in the employment share causes a 0.17% decrease of the establishment wage. This translates into a wage loss of roughly 1000 euros when passing from the first to the third quartile of employment shares.[19] For the more standard threshold of $\kappa = 70\%$ (reduction of 30% employment) the elasticity is almost divided by 4 to 0.06 which implies a twice as big reduction as with the first instrument. This is twice the estimated loss with lagged concentration measures. As we increase the threshold the estimated coefficient converges to the OLS estimate.

1.3.3 Robustness Checks

We perform several robustness checks by changing the instrument, the fixed effects and the definition of local labor market. Results are qualitatively unchanged.

Instrument. Panel (a) of Figure 9 in Appendix 1.F.2 shows a robustness check where the new instrument is not binary any more and takes into account the original employment share of the mass layoff establishments. Panel (b) of the same Figure shows the results from the specification with commuting zone times year fixed effects. Results are qualitatively unchanged from the baseline in both cases.

Local Labor Market. Figure 10 in Appendix 1.F.2 does the same exercise as in the main empirical strategy but changing the definition of local labor market. Local labor markets are here defined with 2-digit industries.[20]

The empirical evidence up to now focused on establishing the presence of employer labor market power of French manufacturing firms. We found that more concentrated industries have lower labor shares and firms pay lower wages in local labor markets where they have relatively higher labor market power. The last part of the empirical evidence aims to motivate the importance of unions in France.

1.3.4 Unions

The institutional framework of the French labor market is characterized by legal requirements that give unions an important role even in medium sized firms.

French labor market is known to be one where unions are relevant players, despite the fact that trade union affiliation in France is among the lowest of all the OECD countries.[21] According to administrative data, the unionization rate in France was 9% in 2014.[22] This unionization rate is slightly below to the one in the U.S. (10.7%) and well below the ones in Germany (17.7%) or Norway (49.7%).

Low affiliation rates do not translate into low collective bargaining coverage for the French case. Collective bargaining agreements extend almost automatically to all the workers, unionized or not. That is, if an agreement is reached in a particular sector, all the workers within the sector are covered. Table 6 presents the unionization and collective bargaining coverage rates for several countries. This institutional framework implies that coverage of collective agreements was in 2014 as high as 98.5% in France despite the low union

[19] This computation is done taking the employment share differences between the percentile 75 and 25 from Table 1 for the median wage. The analogous computation with the average wage gives a wage reduction of roughly 1300 euros.

[20] That is, a local labor market is defined as a combination between commuting zone, 2-digit industry and occupation.

[21] Article in The Economist 'Why French unions are so strong' The Economist.

[22] Source OECD data https://stats.oecd.org/Index.aspx?DataSetCode=TUD. Unionization rate is also denoted as union density.

affiliation rates.[23] This is in stark contrast to the U.S. collective bargaining agreements that only apply to union members and therefore coverage is very similar to the unionization rate.

Table 6 – Union Density and Collective Bargaining Coverage

Country	Union Density	Coverage
Australia	15.10	59.91
Austria	27.70	98.00
Canada	29.30	30.40
Chile	15.30	19.33
Finland	67.60	89.30
France	9.00	98.46
Germany	17.70	57.80
Ireland	26.30	33.52
Italy	36.40	80.00
Japan	17.50	16.90
Korea	10.00	11.90
Netherlands	18.10	85.93
Norway	49.70	67.00
Spain	16.80	80.16
Switzerland	16.10	49.23
Turkey	6.90	6.63
United Kingdom	25.00	27.50
United States	10.70	12.30

Notes: Year 2014. All the variables are in percents. *Union Density* is the unionization rate which is unionized workers relative to total employment. *Coverage* is the collective agreement coverage; the ratio of employees covered by collective agreements divided by all wage earners with the right to bargain. The sources are administrative data except for Australia, Ireland and United States which are based on survey data.

Collective bargaining can happen at different levels. Firms and unions can negotiate at some aggregate level (e.g. industry, occupation, region) and also at economic units such as the group, firm or plant.[24] When wage bargaining happens at the firm level it affects all the workers. Most firms that explicitly bargaining over the wages do so at the firm level (rather than at the plant or occupation level). 92% of mono-establishment firms with a specific collective bargaining agreement in 2010, negotiated it at the firm level. Only 9% of the multi-establishment firms with specific agreements negotiated exclusively at the establishment level.[25]

Legal requirements regarding union representation depend on firm or plant size. First requirements start when the establishment reaches 10 employees and there is an important tightening of duties when reaching

[23] The source of collective bargaining agreements is the OECD as for unionization rates.
[24] Several collective agreements can coexist at a given establishment.
[25] Source DARES.

the threshold of 50 employees.[26] As a consequence, firm level wage bargaining is common even at relatively small establishments. 52% (51%) of establishments with at least 20 employees bargained over the wages in 2010 (in 2004).[27]

Theoretically, workers organize into unions to extract rents from the firm through bargaining. The French institutional setting does not clearly guide about bargaining power differences across the different layers. We build a proxy of rents at the firm level and then compare the correlation of wages with rents depending on idustries and occupations. In particular we compute rents at the firm level $y_{\mathbf{J}(i),t}$ by computing value added minus capital expenditures per worker. The reduced form model is the following:

$$\ln w_{io,t} = \gamma_k \ln y_{\mathbf{J}(i),t} + \varepsilon_{io,t},$$

where γ_k is the elasticity of wages and k denotes either 2-digit industry b or occupation o, $y_{\mathbf{J}(i),t}$ is the proxy of rents at the firm level and $\varepsilon_{io,t}$ is the error term.

Results in Appendix 1.G.1 find that the elasticities at the industry level range from 0.14 for *Metallurgy* to 0.4 for *Food*. On the contrary, when running the same regressions per occupation the elasticities range from 0.27 for *Supervisor* to 0.38 for *Top management*. Given the higher dispersion of the elasticities at the industry level, we will assume differentiated bargaining powers depending on the industry later on in the model.

The prevalence of wage bargaining in the French labor market suggests it is an important element to incorporate into the structural model. Having established the existence of employer labor market power and the importance of unions, next section lays out a model in line with the stylized facts and the French labor market institutions.

1.4 Model

The economy consists of discrete sets of establishments $\mathcal{I} = \{1, ..., I\}$, locations $\mathcal{N} = \{1, ..., N\}$ and industries $\mathcal{B} = \{1, ..., B\}$. Each establishment can have several occupations $o \in \mathcal{O} = \{1, ..., O\}$. Each establishment i is located in a specific location n and belongs to sub-industry h in a particular industry b. We define a local labor market m as the combination between location n, sub-industry h and occupation o, i.e. $m = n \times h \times o$.

We denote the set of establishments that are in local labor market as $\mathcal{I}_m$ with cardinality N_m. We define the set of all local labor markets m as $\mathcal{M}$ and the set of all sub-markets in industry b (in sub-industry h) as $\mathcal{M}_b$ ($\mathcal{M}_h$). The distribution of establishments across local labor markets is determined exogenously. Every establishment can only belong to one location and one sub-sector but can have several occupations and therefore belong to different local labor markets. We define the set of sub-markets that have at least one establishment of sector b as $\mathcal{N}_b$.

The economy is populated by an exogenous measure L of workers who are homogeneous in ability but heterogeneous in tastes for different workplaces. They decide their workplace (establishment-occupation) in two steps without any restriction on mobility. First, workers choose in which local labor market m they would like to be employed, and second, they choose in which establishment i of that sub-market they will work. Workers do not save so they do not own any capital.

[26] The Appendix of Caliendo et al. (2015) provide a comprehensive summary of size related legal requirements in France.

[27] The prevalence of wage bargaining was 44% for establishments with 11 employees or more.

Capital and output markets are competitive. Industry specific rental rates of capital R_{bt} are exogenous. Establishments are owned by entrepreneurs who rent the capital and collect the profits.[28] Those are not explicitly modeled and therefore are excluded from the welfare analysis.

We propose a 'right-to-manage' model where firms and workers bargain over the wages at the establishment-occupation io level. The equilibrium bargained wage is the solution to a reduced form Nash bargaining problem. Once they are hired, workers force a negotiation process over the wages. They internalize that if bargaining were to fail, employers compete in an oligopsponistic fashion. We therefore assume that workers' outside options are oligopsonistic competition outcome wages. This means that if bargaining were to fail, workers would earn wages with a markdown over their marginal revenue product. On the contrary, the threat point of employers when entering the negotiation is having zero profits. If they were not able to agree on the wage setting process and cannot hire anyone, their production and profits would be null.

If bargaining were to fail, establishments post wages per occupation in order to attract workers taking into account the labor supply they face. Having a discrete set of establishments per local labor market means they internalize the effect of their wages on the labor supply of their most immediate competitors. This reflects the idea that competition for labor is mostly local. Geography in our model is only important to define local labor markets.

Below we first set up the production side of the economy and workers' labor supply decisions. Second we present equilibrium wages in the absence of bargaining (wages in the oligopsonistic competition case) and finally we incorporate bargaining to the model.

Production

The final good c is produced by a representative firm with an aggregate Cobb-Douglas production function using as inputs a composite good Y_b for each industry b:

$$Y = \prod_{b \in \mathcal{B}} Y_b^{\theta_b}, \tag{1.3}$$

where θ_b is the elasticity of the intermediate good produced by firms in sector b and $\sum_b \theta_b = 1$. Profit maximization implies that the representative firm spends a fixed proportion θ_b on the industry composite Y_b:

$$P_b Y_b = \theta_b P Y. \tag{1.4}$$

The final good price, which we choose as the numeraire, is equal to:

$$P = 1 = \prod_{b \in \mathcal{B}} \left(\frac{P_b}{\theta_b} \right)^{\theta_b}.$$

Firms produce in a perfectly competitive goods market. P_b is the price of the homogeneous good produced by every firm in sector b, Y_b is their production and P is the price of the final good which we take as a numeraire. Y_b is the aggregate of output of all the firms in that sector:

$$Y_b = \sum_{i \in \mathcal{I}_b} y_i, \tag{1.5}$$

[28] It is not important whether the entrepreneurs own capital or not. As it is a small open economy, the rental rate of capital is fixed and entrepreneurs rent capital from abroad until the marginal product is equal to the cost.

where $\mathcal{I}_b$ is the set of establishments that belong to industry b. The establishment production function y_i is an aggregate of occupation productions. Establishment i produces using occupation o specific inputs, labor L_{io} and capital K_{io}, with a decreasing returns to scale technology. Output elasticity with respect to labor β_b and capital α_b are industry specific and establishment-occupations are heterogeneous in their total factor productivity. We assume that occupations are perfect substitutes and their output is aggregated linearly. That is, total establishment output y_i is the sum of occupation specific outputs y_{io}. Decreasing returns to scale in the occupation output y_{io} generate an incentive to produce using several occupations.

Establishment i's output, y_i, is defined as:

$$y_i = \sum_{o=1}^{O} y_{io} = \sum_{o=1}^{O} \widetilde{A}_{io} K_{io}^{\alpha_b} L_{io}^{\beta_b}. \tag{1.6}$$

The choice of this particular production function is motivated by theoretical and empirical reasons. The linearity of the aggregation within establishments allows for the separability of different local labor markets.[29] The second reason is data motivated. The absence of a particular occupation in an establishment can be rationalized by having null productivity in that particular occupation. An alternative specification where labor is a Cobb-Douglas composite of occupations is at odds with the pervasive prevalence of missing at least one occupation category. The median establishments lacks at least one occupation. Lacking a particular occupation, those establishments would not be able to produce if labor is a Cobb-Douglas composite of occupations. Appendix 1.H lays out the model and proofs with this alternative production function.

The separability of local labor markets comes from restricting the inverse elasticity of labor demand to be equal across different industries. We assume that output elasticities with respect to capital α_b and labor β_b are such that: $\frac{\beta_b}{1-\alpha_b} = 1-\delta$, where $\delta \in [0,1]$ is a constant across sectors. This specification nests constant returns to scale when $\delta = 0$. As long as $0 < \delta < 1$ the establishment faces decreasing returns to scale within occupations. This assumption together with the linearity of the production function give us separability of the local labor markets. This is further discussed in Section 1.4.4.

Substituting optimal demand for capital, the establishment-occupation production is:

$$y_{io} = F_b^{\alpha_b(1+\varepsilon_b\delta)} A_{io} L_{io}^{1-\delta}, \quad A_{io} \equiv \widetilde{A}_{io}^{\frac{1}{1-\alpha_b}} \left(\frac{\alpha_b}{R_b}\right)^{\frac{\alpha_b}{1-\alpha_b}}, \tag{1.7}$$

A_{io} is a transformed productivity of io that incorporates elements coming from the optimal demand of capital and F_b is a transformed industry b price.[30] Details of these derivations are in Appendix 1.A. From now on we work with the production function with optimal demand for capital.

Labor Supply

We now present worker preferences that give rise to upward sloping establishment-occupation specific labor supplies. A worker k receives utility by consuming a single final good c and by the product of two idiosyncratic utility shocks: one establishment-occupation specific preference shifter z_{kio} and another one common for all establishments in local labor market m, u_{km}. The utility of a worker k working for establishment i at occupation o in local labor market m is:

$$\mathcal{U}_{kio} = c_k z_{kio} u_{km}. \tag{1.8}$$

[29] The solution and characterization of the model are in Section 1.4.4.

[30] $F_b = P_b^{\frac{1}{\chi_b}}$, $\chi_b = (1-\alpha_b)(1+\varepsilon_b\delta)$ is the transformed industry price.

Following Eaton and Kortum (2002) in the trade literature and Redding (2016) and Ahlfeldt et al. (2015) in urban economics literature we assume that the idiosyncratic utility shocks are drawn from a Fréchet distribution:

$$P(z) = e^{-T_{io}z^{-\varepsilon_b}}, \quad T_{io} > 0, \varepsilon_b > 1 \tag{1.9}$$

$$P(u) = e^{-u^{-\eta}}, \quad \eta > 1, \tag{1.10}$$

where the parameter T_{io} determines the average utility derived from working in establishment i and occupation o. In contrast, we normalize these parameter to 1 for the sub-market specific shock u. The shape parameters ε_b and η control the dispersion of the idiosyncratic utility. They are inversely related to the variance of the preference shifters. We name the parameters ε_b and η as the within and across labor market elasticities. If both have high values workers have similar tastes for different local labor markets and establishment-occupations. This in turn implies that their labor supply is more elastic and will react more to changes in wages.

The labor supply elasticities in this framework are different from the ones studied by public economists. Our baseline model features a constant level of aggregate employment and workers do not decide the *amount* of hours to work but rather the *workplace* to which they want to supply their labor. The Frisch elasticity of labor supply is zero in our baseline environment but yet workers do not supply their labor inelastically to any establishment.

We assume that establishments cannot discriminate workers based on their taste shocks. This implies that establishment i for occupation o pays the same wage w_{io} to all its employees, leaving the marginal worker indifferent between working in io or moving. Small wage reductions induce the movement of the marginal worker but infra-marginal workers stay. One can view these taste shocks as mobility costs in a static model.

The only source of worker income are wages, therefore the indirect utility of worker k is:

$$\mathcal{U}_{kio} = w_{io} z_{kio} u_{km}, \tag{1.11}$$

where the last two elements are the taste shocks. A worker chooses where to work in two steps: first, they choose their local labor market after observing local labor market shocks u_{km}. After picking a local labor market, the worker then observes the establishment idiosyncratic shocks and chooses the establishment that maximizes expected utility. Following the usual derivations as in Eaton and Kortum (2002), the probability of a worker choosing establishment i and occupation o is a product of two terms: the employment share of the establishment-occupation within the local labor market $s_{io|m}$ and the employment share of the local labor market itself s_m. We develop the derivations in Appendix 1.A. The probability $\Pi_{io} = s_{io|m} \times s_m$ writes as:

$$\Pi_{io} = \frac{T_{io} w_{io}^{\varepsilon_b}}{\sum_{j \in I_m} T_{jo} w_{jo}^{\varepsilon_b}} \times \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\sum_{m' \in \mathcal{M}} \Phi_{m'}^{\eta/\varepsilon_b} \Gamma_b^{\eta}}, \tag{1.12}$$

where $\Phi_m = \sum_{j \in I_m} T_j w_{jo}^{\varepsilon_b}$ is a local labor market aggregate, the functional Γ_b is independent of the endogenous variables and the economy wide constant Φ is $\Phi = \sum_{m \in \mathcal{M}} \Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}$. In equilibrium, the first fraction is equal to $s_{io|m}$ and the second term in (1.12) is s_m.

Integrating over the continuous measure of workers L, the labor supply L_{io} for establishment and occupation o is:

$$L_{io}(w_{io}) = \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi} L = \Pi_{io} L. \tag{1.13}$$

The inverse of this labor supply is upward sloping as long as the within and across local labor market elasticities are bounded. In the limit where both tend to infinity, workers are indifferent across workplaces and the inverse labor supply becomes flat.

1.4.1 Absence of Bargaining

In this section we characterize equilibrium wages in the absence of bargaining. Given the labor supply curves with bounded elasticities, establishments post wages taking into account the labor supply curves (1.13) they face. This monopsony power translates into a markdown between the wages and the marginal revenue products of labor. When the establishments solve their wage posting problem, they look at probability Π_{io} and take into account the effect of wages on the establishment-occupation term $T_{io} w_{io}^{\varepsilon_b}$ and also on the local labor market aggregate Φ_m. However, they take as given economy wide aggregates (Φ and L).[31] The finite set of establishments per local labor market generates strategic interaction among the competitors. The strategic interaction within a local labor market induces oligopsonistic competition that features a heterogeneous markdown.

The first order condition for the establishment-occupation wage io under oligopsonistic competition is:

$$w_{io}^{MP} = \frac{e_{io}}{e_{io}+1} \beta_b A_{io} L_{io}^{-\delta} P_b^{\frac{1}{1-\alpha_b}}, \tag{1.14}$$

where $e_{io} = \varepsilon_b \left(1 - s_{io|m}\right) + \eta \, s_{io|m}$ is the perceived labor supply elasticity. This expression is similar to Card et al. (2018) with the difference that we have variable perceived elasticities that arise from the strategic interaction between establishments. We denote with a subscript MP the equilibrium wage when there is only employer labor market power. The fraction $\frac{e_{io}}{e_{io}+1}$ in equation (1.14) is the markdown and it is defined as:

$$\mu(s_{io|m}) = \frac{\varepsilon_b \left(1 - s_{io|m}\right) + \eta \, s_{io|m}}{\varepsilon_b \left(1 - s_{io|m}\right) + \eta \, s_{io|m} + 1}. \tag{1.15}$$

In the absence of bargaining, the wedge between the marginal revenue product of labor and the wages boils down to a markdown (1.15).[32] We denote this object in short notation as μ_{io}.

As long as workers are less elastic across local labor markets than across establishments within a given local labor market (i.e. as long as $\eta < \varepsilon_b$), the markdown (1.15) is a decreasing function of the share of employment $s_{io|m}$. Once an establishment is big with respect to the nearby competitors, it internalizes that it is facing a more inelastic labor supply and applies a more important markdown. In the limit where ε_b and η tend to infinity, establishments face an infinitely elastic labor supply and a perfectly competitive labor market rises with $\mu(s_{io|m}) = 1$.

Heterogeneous markdowns distort relative wages across establishment-occupations and therefore the labor supplies. This implies that the labor allocation to a particular establishment-occupations is different to the one if the markdowns were absent. Distorting the labor allocation across the production units, the heterogeneous markdown generates misallocation of resources and potentially reduces aggregate output even at the case where total employment is fixed. We formalize the source of misallocation in Section 1.4.4.[33]

[31] Similar to Atkeson and Burstein (2008), this type of behavior could be rationalized either by assuming a myopic behavior of the establishment or by having a continuous of local labor markets.

[32] Appendix 1.A derives this expression.

[33] Appendix 1.G provides an illustration of the distributional and efficiency consequences.

When the markdown is constant and total labor supply fixed, labor market power does not have efficiency consequences as it only affects the division of output into the labor share and the profit share. This is not any more true if we were to allow an endogenous leisure or labor force participation decisions. Counterfactually increasing wages would increase total labor supply L and therefore total output.[34]

1.4.2 Bargaining

We now introduce the bargaining between employers and unions. We assume that bargaining happens at the establishment-occupation level and involves only wages rather than indirect utilities because workers do not know each others' taste shocks. Given the perfect substitutability of occupations in the production function, bargaining at the occupation level is equivalent to a situation where bargaining happens at the establishment level but there are different wage agreements per occupation.

When they are hired, workers force the negotiation over the wages in order to earn above the status-quo. Workers understand the nature of employer labor market and take the wages under oligopsonistic competition as their threat points. Their reservation wage is therefore: $w_{io}^r = \mu_{io} \times MRPL$. We assume that firms on the contrary act naively and take as threat points a situation without production or profits.

The bargained equilibrium wage is the solution to a reduced form Nash bargaining where union's bargaining power is φ_b and the one of the establishment is $1 - \varphi_b$. Appendix 1.A.4 gives more detail on the bargaining set up and discusses other situations that lead to the same negotiated equilibrium wages.

The equilibrium bargained wage is:

$$w_{io} = \underbrace{\left[(1-\varphi_b)\,\mu_{io} + \varphi_b\,\frac{1}{1-\delta}\right]}_{\text{Wedge } \lambda(\mu_{io},\varphi_b)} \times \underbrace{\beta_b A_{io} L_{io}^{-\delta} P_b^{\frac{1}{1-\alpha_b}}}_{\text{MRPL}}. \tag{1.16}$$

The wedge between equilibrium wages and the marginal revenue product of labor, $\lambda(\mu_{io}, \varphi_b) \equiv (1-\varphi_b)\mu_{io} + \varphi_b \frac{1}{1-\delta}$, is a combination of two parts. First, the markdown μ_{io} coming from the oligopsonistic competition in the absence of bargaining, and second, the markup $\frac{1}{1-\delta}$ coming from the bargaining process. The markup is a consequence of the ability of the union to extract quasi-rents coming from the decreasing returns to scale $1 - \delta < 1$.[35] Bargained wages will be above or below the marginal revenue product depending on the union's bargaining power φ_b and the relative strength of markdowns and markups. This comes from the fact that the term inside brackets is a convex combination between $\mu_{io} < 1$ and $\frac{1}{1-\delta} > 1$.

In our calibrated model, labor supply elasticity e_{io} is decreasing in the local labor market employment share. Hence, even with bargaining ($0 < \varphi_b < 1$), one would observe a negative relationship between employment shares $s_{io|m}$ and wages w_{io}. A desirable feature of the model is that it nests the oligopsonistic competition only and bargaining only as special cases. The former is equivalent to a situation where union's bargaining power is zero $\varphi_b = 0$. Equilibrium wages would be equal to a markdown times the marginal revenue product of labor $w^{MP} = \mu_{io} \times MRPL$. A bargaining model without employer labor market power is encompassed when worker's outside option is the competitive wage. The wedge in that case is equal to:

[34] The industry constant $\mu_b = \frac{\varepsilon_b}{\varepsilon_b + 1}$ drives down the wages. If labor supply is endogenous, workers' decision between consumption c and leisure l would be distorted. Denote by w the wage under monopsonistic competition and by $\tilde{w}$ the wage under competitive labor market. Worker's maximization under endogenous labor supply leads the marginal rate of substitution to be equal to the wage rate. $w < \tilde{w}$ and therefore $MRS_{c,l} \equiv \frac{U_l}{U_c} = w < \tilde{w}$. Meaning that workers would supply less labor than in the perfectly competitive case.

[35] The last part $\frac{1}{1-\delta}$ is a markup only under the assumption of decreasing returns to scale. That is, when $\delta > 0$.

$1 - \varphi_b + \varphi_b \frac{1}{1-\delta} = 1 + \varphi_b \frac{\delta}{1-\delta}$. The bargained wages incorporate a markup over the marginal product and become $w^B = (1 + \varphi_b \frac{\delta}{1-\delta}) \times MRPL$. Workers are not only paid their marginal product but are also able to extract rents that come from the decreasing returns to scale. Rent extraction from the workers is governed by their bargaining power φ_b.

1.4.3 Equilibrium

For given industry rental rates of capital $\{R_b\}_{b=1}^{B}$, the general equilibrium of this economy is a set of wages $\{w_{io}\}_{io=1}^{IO}$, output prices $\{P_b\}_{b=1}^{B}$, a measure of labor supplies to every establishment and occupation $\{L_{io}\}_{io=1}^{IO}$, capital $\{K_{io}\}_{io=1}^{IO}$ and output $\{y_{io}\}_{io=1}^{IO}$, industry $\{Y_b\}_{b=1}^{B}$ and economy wide outputs Y, such that equations (1.3)-(1.13) and (1.16) are satisfied $\forall$ $io \in \mathcal{I}_m$, $m \in \mathcal{M}$ and $b \in \mathcal{B}$.

1.4.4 Characterization of the Equilibrium

Solving the model amounts to finding establishment wages, industry prices and allocations. In order to simplify the solution, we restrict the labor demand elasticity to be the same across industries. That is, we assume $\frac{\beta_b}{1-\alpha_b} = 1 - \delta$, where $\delta \in [0,1]$. This restriction implies the separability of the different local labor markets which allows us to split the solution in two. First, we take a partial equilibrium approach and solve for establishment-occupation components normalizing aggregates above the local labor market and show existence and uniqueness of the system of normalized wages. Second, we show that the model can be rewritten at the 2-digit industry level with the solution to these normalized wages and deep parameters. This last aggregate model is in turn enough to solve for industry prices.

Substituting the labor supply into (1.16) and simplifying we obtain:

$$w_{io} = \left(\beta_b \lambda(\mu_{io}, \varphi_b) \frac{A_{io}}{\left(T_{io}\Gamma_b^{\eta}\right)^{\delta}} \right)^{\frac{1}{1+\varepsilon_b \delta}} \Phi_m^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi}{L}\right)^{\nu_b} F_b, \quad \nu_b \equiv \frac{\delta}{1+\varepsilon_b \delta} \tag{1.17}$$

where $\nu_b = \frac{\delta}{1+\varepsilon_b \delta}$ is just an auxiliary parameter to ease notation.

To gain intuition on the allocation distortions from the heterogeneous wedges we focus on two establishments in the same local labor market. From (1.17), their relative wages are:

$$\frac{w_{io}}{w_{jo}} = \left(\frac{\lambda(\mu_{io}, \varphi_b)}{\lambda(\mu_{jo}, \varphi_b)}\right)^{\frac{1}{1+\varepsilon_b \delta}} \left(\frac{A_{io}}{A_{jo}} \frac{T_{jo}^{\delta}}{T_{io}^{\delta}}\right)^{\frac{1}{1+\varepsilon_b \delta}}. \tag{1.18}$$

The ratio of heterogeneous labor wedges $\frac{\lambda(\mu_{io},\varphi_b)}{\lambda(\mu_{jo},\varphi_b)}$ distorts the relative wages of the establishments at the same local labor market and consequently the labor supply (1.13). It is important to note that even in the absence of the labor wedge, in equilibrium, establishments pay different wages. This is a consequence of the workers' idiosyncratic taste shocks. In the limit where workers are infinitely elastic across establishments within the local labor market $\varepsilon_b \to \infty$, wages would be equalized. The same logic applies for differences across local labor markets and the respective elasticity η.

The first order condition (1.17) separates the establishment wage into terms constant for every establishment in sub-market m ($\Phi_m^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi}{L}\right)^{\nu_b} F_b$) and establishment-occupation specific components of wages.

We denote the latter as $\widetilde{w}_{io}$ and are defined as:

$$\widetilde{w}_{io} = \left(\beta_b \lambda(\mu_{io}, \varphi_b) \frac{A_{io}}{\left(T_{io}\Gamma_b^{\eta}\right)^{\delta}} \right)^{\frac{1}{1+\varepsilon_b \delta}}, \tag{1.19}$$

The real wage w_{io} is therefore $w_{io} = \widetilde{w}_{io} \Phi_m^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi}{L}\right)^{\nu_b} F_b$.

We can now establish existence and uniqueness of the system of equations (1.17) in partial equilibrium:

Proposition 1. *For given parameters $\{\alpha_b, \beta_b, \varphi_b \quad s.t. \quad 0 \leq \alpha_b, \beta_b, \varphi_b < 1, \forall b \in \mathcal{B}\}$ and $1 < \eta < \varepsilon_b \forall b \in \mathcal{B}$, $0 \leq \delta \leq 1$, transformed price F_b, constants $\{\Phi_m\}$, Φ, total labor supply L and non-negative vectors of productivities $\{A_{io}\}_{io \in m}$ and amenities $\{T_{io}\}_{io \in m}$, there exists a unique vector of wages $\{w_{io}\}_{io \in I_m}$ for every local labor market m that solves the system formed by* (1.17).

Proof. See Appendix. □

Proposition 1 tells us that if we take these aggregate terms as constants, then the solution for this system exists and is unique. Employment shares $s_{io|m}$ are not affected if all local labor market wages are scaled up or down. This is a result of the wedges $\lambda(\mu_{io}, \varphi_b)$ being homogeneous of degree zero with respect to local labor market constants. System (1.17) has a unique solution as we can use Proposition 1 with $\Phi_m = \Phi = L = F_b = 1$.

We now turn to the second step of the model solution. Given the solutions to the establishment-occupation components we build industry level productivity measures and write the model at the industry b level.

Starting from the lowest production unit (1.7) we aggregate up to industry output:

$$Y_b = F_b^{\alpha_b(1+\varepsilon_b\delta)} A_b L_b^{1-\delta}, \quad A_b = \sum_{io \in \mathcal{I}_b} A_{io} s_{io|m}^{1-\delta} s_{m|b}^{1-\delta}, \tag{1.20}$$

where A_b is an employment weighted productivity and F_b is the transformed industry price. Solving the model now amounts to solving the system of intermediate good demand (1.4) to find industry prices. Using the final good production function (1.3) and the intermediate input demand (1.4),

$$F_b^{1+\varepsilon_b\delta} A_b L_b(\mathbf{F})^{1-\delta} = \theta_b \prod_{b' \in \mathcal{B}} \left(F_{b'}^{\alpha_{b'}(1+\varepsilon_{b'}\delta)} A_{b'} L_{b'}(\mathbf{F})^{1-\delta} \right)^{\theta_{b'}}. \tag{1.21}$$

Steps to get to this expression are in Appendix 1.A.5. Having the solution for normalized wages we can leave the industry labor supply L_b and total output Y as a function of the transformed prices $\mathbf{F} = \{F_b\}_{b \in \mathcal{B}}$.

Collecting all these expressions for the different industries forms a system of B equations with B unknowns.[36] Solving for the vector of transformed prices $\mathbf{F}$ we can back out the rest of the variables in the model. Note that the system of equations is unchanged irrespective of the aggregate level of employment L because the final good production function being constant returns to scale and industry employment L_b is linear on aggregate labor supply.

Given the solution for normalized wages, we can think of industry productivity A_b and industry level normalized wages $\widetilde{\Phi}_b$ as additional parameters at the industry level. The following proposition characterizes the solution for this system as a function of these parameters.

Proposition 2. *For any set of parameters $\{\beta_b, \theta_b \quad s.t. \quad 0 \leq \beta_b, \theta_b < 1, \forall b \in \mathcal{B}\}$, $0 \leq \delta \leq 1$, $\{\psi_b \equiv \frac{1+\varepsilon_b\delta}{1+\eta\delta}\}_{b \in \mathcal{B}}$, non-negative vectors $\{A_b\}_{b \in \mathcal{B}}$ and $\{\widetilde{\Phi}_b\}_{b \in \mathcal{B}}$, there exists a unique vector of transformed prices* $\mathbf{F}$ *such that solves the*

[36] B is the number of different 2-digit industries.

system formed by (1.21) *and it's characterized by:*

$$F_b = X_b C^{\frac{1}{\psi_b(1+\eta)}}, \tag{1.22}$$

$$X_b = \left(\frac{\theta_b}{A_b(\widetilde{\Phi}_b \Gamma_b^{\eta})^{(1-\delta)}} \right)^{\frac{1}{\psi_b(1+\eta)}}, \quad C = \left(\prod_{b' \in \mathcal{B}} \left(\theta_{b'} X_{b'}^{-\chi_{b'}} \right)^{\theta_{b'}} \right)^{\frac{1+\eta}{\sum_{b' \in \mathcal{B}} \theta_{b'} (1-\alpha_{b'})(1+\eta\delta)}}$$

for all $b \in \mathcal{B}$.

Proof. See Appendix. □

Proposition (2) provides an analytical solution for the (transformed) industry prices. Given the aggregations of the establishment-occupation components up to the industry level, the solution of the prices is unique and is characterized in closed form.

Proposition 1 showed the existence and uniqueness of the establishment-occupation components. A useful characteristic of those components is that they are homogeneous of degree zero with respect to local labor market aggregates. We therefore have that the normalized wages (or establishment-occupation components) are independent of industry prices. By taking together Propositions 1 and 2 we therefore can then conclude that there exists a unique solution to the model for any set of valid parameters and vectors of productivities and amenities.

1.5 Estimation

In this section, we describe the estimation procedure and present the results. The parameters to estimate are the within and across local labor market elasticities ($\{\varepsilon_b\}_{b=1}^B$ and η respectively), the inverse elasticity of the labor demand (δ), the industry output elasticities ($\{\alpha_b\}_{b=1}^B, \{\beta_b\}_{b=1}^B$) and the workers' bargaining powers ($\{\varphi_b\}_{b=1}^B$). Given our restriction δ, we only need to calibrate either the capital elasticities $\{\alpha_b\}_{b=1}^B$ or the labor ones $\{\beta_b\}_{b=1}^B$.

We estimate the model in three steps. First, by exploiting differences in the variance-covariance matrix of structural shocks across occupations we identify the across local labor market labor supply elasticity η and the inverse elasticity of labor demand δ. Then, we calibrate the output elasticities of capital to match industry capital shares. Second, we estimate the within local labor market labor supply elasticities $\{\varepsilon_b\}_{b=1}^B$ by estimating the labor supply equation while instrumenting for the wages. Finally, we calibrate the union's bargaining powers $\{\varphi_b\}_{b=1}^B$ to match the industry labor shares.

We take advantage of the presence of establishment-occupations with $s_{io|m} = 1$ in the data. We name those establishment-occupations that are alone in a particular local labor market as full monopsonists. We restrict the sample to full monopsonists for the first estimation step. Being alone in their local labor markets, the only firm specific labor supply elasticity in play is the across local labor market one η. Identification of the within local labor market elasticities ε_b requires to focus on the establishment-occupations competing with others in their local labor markets.

We start the estimation by restricting to full monopsonists to perform the first step of the estimation procedure. Being the only players in the local labor market, the labor wedge they apply is constant and equal to $\mu(s=1) = \frac{\eta}{\eta+1}$. Their labor demand is:

$$w_{io} = \left[(1-\varphi_b)\frac{\eta}{\eta+1} + \varphi_b \frac{1}{1-\delta}\right] \beta_b P_b^{\frac{1}{1-\alpha_b}} A_{io} L_{io}^{-\delta}, \tag{1.23}$$

and the labor supply they face is:

$$L_{io} = \frac{T_{io}^{\eta/\varepsilon_b} w_{io}^{\eta} \Gamma_b^{\eta}}{\Phi} L. \tag{1.24}$$

Similar labor supply and demand systems can be formed for each occupation. This system suffers from standard identification issues when we have simultaneous equations. Independent identification of each of the equations requires different instruments shifting only one of them.

Lacking such instruments, we follow the *identification through heteroskedasticity* approach of Rigobon (2003) to identify the across local labor market labor supply elasticity η and the inverse elasticity of labor demand δ. Our identification strategy is based on restrictions on the variance-covariance matrix of structural shocks. In our preferred specification, we group the occupations into two categories and assume that the covariance between the demand and supply shifters (productivity and amenity respectively) are constant within the occupation category. This assumption is in line with the idea that amenities such as working hours, repetitiveness of the tasks or more general working environments are similarly related to productivity. In our main specification we group occupations into white collar workers (top management and clerical) and blue collar (supervisor and operational). The assumption states that occupations within those two categories share the same relationship between productivity and amenities.

Taking logarithms and demeaning by substracting the industry b average per year, the system for occupation o is:

$$\begin{pmatrix} \ln(L_{io}) \\ \\ \ln(w_{io}) \end{pmatrix} = \frac{1}{1+\eta\delta} \begin{pmatrix} 1 & -\eta \\ \\ \delta & 1 \end{pmatrix} \begin{pmatrix} \frac{\eta}{\varepsilon_b} \ln(T_{io}) \\ \\ \ln(A_{io}) \end{pmatrix}$$

We estimate the variance covariance matrix of employment and wages per occupation from the data. The restriction we impose on the variance-covariance matrix of the structural shocks is that the covariance between the labor demand shifter (the productivity) and the labor supply shifter (the amenity) is constant across occupations within the same category. Equalizing the covariances we obtain a system of equations that do not depend on the within local labor market labor supply elasticity ε_b anymore. More details about the estimation are in Appendix 1.D.

The second step is devoted to the calibration of the output and the within local labor market labor supply elasticities. We start by calibrating the capital elasticities. We follow Barkai (2016) to construct the industry interest rates or required rates $\{R_{bt}\}_{b=1}^{B}$ per year and target the average industry capital shares.[37] From the first order condition for capital, the industry b capital share of output is:[38]

$$\frac{R_{bt}K_{bt}}{P_{bt}Y_{bt}} = \alpha_b.$$

We calibrate α_b such that $\mathbb{E}_t\left[\frac{R_{bt}K_{bt}}{P_{bt}Y_{bt}}|b\right] = \alpha_b$. Given our restriction of constant inverse labor demand elasticity δ, we back out the output elasticities with respect to labor by using $\frac{\beta_b}{1-\alpha_b} = 1-\delta$.

[37] Details are in Appendix 1.E.4.

[38] This is derived in Appendix 1.A.

The within market labor supply elasticities ε_b are estimated exploiting the labor supply equation of non full monopsonists. The labor supply they face (1.13) in logs is:

$$\ln(L_{io}) = \varepsilon_b \ln(w_{io}) + f_m + \ln(T_{io}),$$

where f_m is a local labor market constant. At this point of the estimation the amenities T_{io} are unobserved. The usual exclusion restrictions when running this regression requires that the conditional expectation of the error term (here, the amenity) is equal to zero. Everything else equal, higher amenity establishments pay lower wages. We instrument for the wages using a proxy $\widehat{A}$ of firm productivity.

$$\widehat{A} = \frac{P_b Y_J}{\sum L_{io}^{1-\delta}}.$$

The first estimation step did not require independence of the structural shocks. In order to minimize the potential of endogeneity bias of our instrument, we use the lag instrument instead of the contemporaneous one.

Finally, the union bargaining powers are pinned down by industry labor shares. In the model, labor share of any establishment i and occupation o at period t is:

$$LS_{io} = \frac{w_{io} L_{io}}{P_b y_{io}} = \beta_b \lambda(\mu_{io}, \varphi_b), \tag{1.25}$$

where the only parameter left is φ_b in the wedge function $\lambda(\mu_{io}, \varphi_b) = (1-\varphi_b)\frac{\varepsilon_b(1-s_{io|m})+\eta s_{io|m}}{\varepsilon_b(1-s_{io|m})+\eta s_{io|m}+1} + \varphi_b \dfrac{1}{1-\delta}$. Writing the analogous at the industry level, the union bargaining power φ_b is pinned down by the average industry labor share. When constructing the theoretical labor share, we assume that given the estimated parameters, we later perfectly match the observed wages of establishments and labor allocations. We do not target the unobserved establishment-occupation value added and therefore neither the industry value added measures.[39] For now, we assume that we match the wages and labor allocations in equilibrium. Details of how we back out amenities T_{io} to ensure that are in Appendix 1.E.5.

We additionally need to calibrate the elasticities of the final good production function in order to be able to compute counterfactuals. Table 15 in Appendix 1.D.3 has the calibrated elasticities and interest rates for 2007, our baseline year for the counterfactuals. The next Section presents the estimation results and the goodness of the fit.

1.5.1 Estimation Results

Table 7 recovers the estimation results of the main parameters. The most important parameters of the estimation are arguably the firm specific labor supply elasticities and the union bargaining powers.

The estimated across local labor market elasticity is $\widehat{\eta} = 0.42$ and the industry specific local labor market labor supply elasticities $\widehat{\varepsilon}_b$ range from 1.22 to 4.05.[40] η and ε_b are inversely related to the variances of the taste shocks. The across local labor market elasticity being lower than the within ones ($\widehat{\varepsilon}_b > \widehat{\eta} \quad \forall b$), workers are more likely to change workplaces within than across local labor markets. This implies that the markdown μ_{io} is more relevant (further away from 1) for establishments having higher employment shares out of the local labor market. Consequently, the structural labor wedge $\lambda(\mu_{io}, \varphi_b)$ of our calibrated model is decreasing in employment shares $s_{io|m}$. This feature is in line with the empirical evidence from Section 1.3.

[39] We could in principle also do the reverse if the occupation specific value added were observed in the data.

[40] Table 14 in Appendix 1.D.3 provides details of industry estimates.

Table 7 – Main Estimates

Param.	Name	Estimate	Identification
η	Across labor market elast.	0.42	Heteroskedasticity
δ	1 - Returns to scale	0.04	Heteroskedasticity
$\{\varepsilon_b\}$	Within labor market elast.	1.2 - 4	Labor supply
$\{\beta_b\}$	Output elast. labor	0.57 - 0.85	Capital share and δ
$\{\varphi_b\}$	Union bargaining	0.06 - 0.7	Industry LS

Comparing our labor supply elasticities to the recent estimates for the U.S. from Berger et al. (2019) for the US, they are qualitatively similar. Their analogous estimate of the across local labor market elasticity η is 0.66 (compared to our estimate of 0.42) and their estimated within local labor market elasticity is 5.38. The across local labor market estimates are very similar. On the contrary, all of our industry specific within local labor market elasticities lie below their estimate. This might be a consequence of the low mobility that characterizes the French labor market.[41]

The estimates of union bargaining power range from 0.06 for Chemical to 0.73 for Telecommunications. According to our estimates, there is an important heterogeneity of bargaining power across industries. Lacking direct estimates of bargaining power within manufacturing we validate our estimates by two comparisons. First, French labor law imposes more restrictive legal duties regarding union representation for larger establishments. We compute the correlation between the bargaining power estimates $\widehat{\varphi}_b$ and average plant or firm size (in terms of employment) per industry. We find a positive correlation of 0.33 between average establishment employment per industry and union's bargaining power φ_b.[42] Second, Cahuc et al. (2006) provide manufacturing bargaining power estimates for France in a framework of search and matching with on the job search. Our estimated bargaining power for manufacturing as a whole is 0.37.[43] This is close to the estimate of Cahuc et al. (2006) for top management workers of 0.35.

The estimate of the inverse labor demand elasticity, δ, is $\widehat{\delta} = 0.04$. This parameter is also related to the average returns to scale of the production function which are about 0.97. The combination of δ and the estimated capital elasticities per industry $\{\alpha_b\}$ allow us to recover the values for the output elasticity with respect to labor. We have that $\{\beta_b\}$ is equal to $\beta_b = (1-\alpha_b)(1-\delta)$. Labor elasticities go from 0.56 for *Transport* to the 0.85 for *Shoe and leather production*.

1.5.2 Estimation Fit

Using the point estimates we check the fit of the model for non-targeted moments. Figure 2 depicts the fit of the model and the non-targeted data. In panel (a) we have industry labor shares per year. On the horizontal axis we have the model generated moments while on the vertical axis we observed the corresponding moment in the data. If the fit was perfect, each dot would be on the 45 degree line. Each color represents an industry.

[41] See Jolivet et al. (2006) for a comparison of French mobility against the U.S.

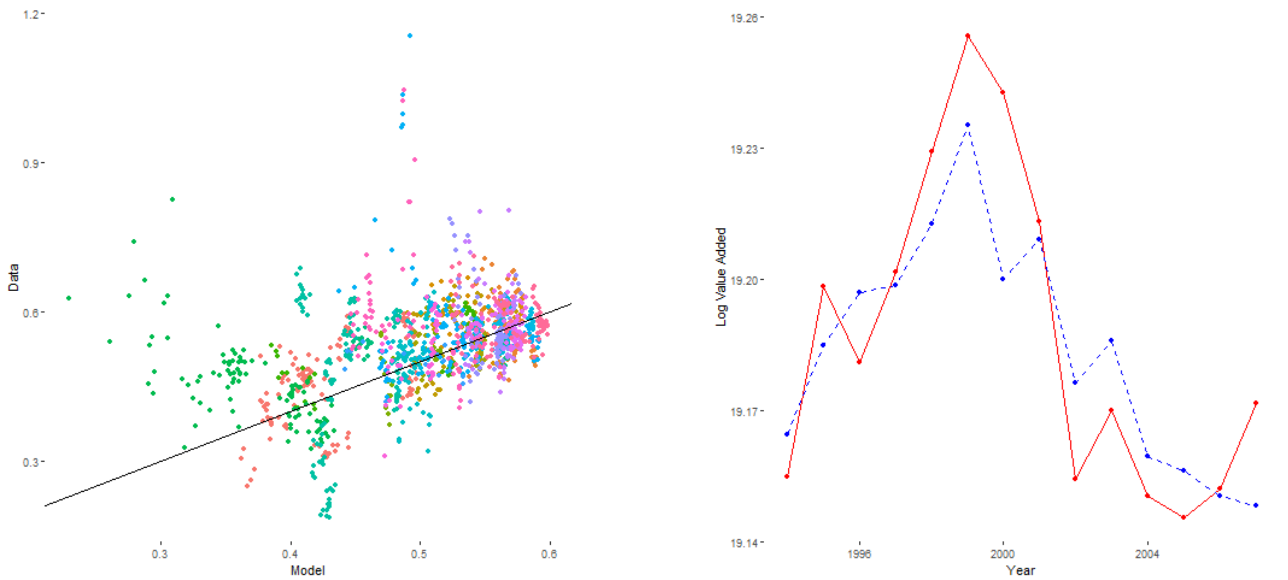
[42] The correlation between average per industry firm size and our estimated bargaining power is 0.31.

[43] This is an employment weighted average of the industry estimates. The direct average of industry bargaining powers is 0.41.

We see that most of the dots are aligned around the 45 degree line. Next to it, in panel (b), we show the fit to aggregate value added.

We can check the model does against other non-targeted moments. Panel (a) of Figure 2 shows the model matches value added per industry. This in fact might not be surprising as there is a very strong relationship between establishment's production and wage bill in the model and in the data. Since the model exactly matches the establishment's wages and labor allocations, it also has a good fit of the value added. The second non-targeted moment is the evolution of the aggregate value added, shown in panel (b) of the same figure. The model also does a very good job following the actual data.

Figure 2 – Model Fit Non Targeted Moments



(a) Sub-industry Labor Share

(b) Aggregate Value Added. Model in dashed blue, data in red.

Table 8 – Concentration and Labor Share: Data vs. Model

	Data: $\log(LS^{D}_{h,t})$		Oligopsony: $\log(LS^{M,MP}_{h,t})$		Model: $\log(LS^{M}_{h,t})$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\overline{HHI}_{h,t})$	-0.054^{***}	-0.056^{***}	-0.388^{***}	-0.416^{***}	-0.175^{***}	-0.161^{***}
	(0.013)	(0.013)	(0.009)	(0.003)	(0.007)	(0.005)
Ind FE	Y	N	Y	N	Y	N
Ind-Year FE	N	Y	N	Y	N	Y
Obs.	1357	1357	1357	1357	1357	1357
R^2	0.29	0.343	0.901	0.903	0.946	0.909
Adj. R^2	0.280	0.172	0.899	0.878	0.945	0.936

Notes: The dependent variable of the first two Columns are the logarithm of 3-digit industry labor share at year t, $\log(LS^{D}_{h,t})$. These present the results from Table 4 with fixed effects. Next two Columns present the model generated log labor shares $\log(LS^{M,MP}_{h,t})$ when the model does not incorporate wage bargaining. This is a framework where the labor wedge λ boils down to $\lambda(\mu_{io}, 0) = \mu_{io}$. Last two Columns present the analogous regressions with our framework where bargaining is incorporated $\log(LS^{M}_{h,t})$. Throughout the different frameworks Column (1) presents estimates with industry fixed effects and Column (2) results with industry-year fixed effects. *p<0.1; **p<0.05; ***p<0.01

To further investigate the model fit to non-targeted moments we repeat the aggregate empirical evidence of Section 1.3. Table 8 presents the empirical evidence of Table 4 with fixed effects (Columns (2) and (3)) in the first 2 Columns and the rest of the rest are devoted to compare two alternative models. Model results present the same regressions as the ones for the data for the model with oligopsonistic competition only $LS^{M,MP}_{h,t}$ (Columns (3) and (4)) and for our model with collective wage bargaining $LS^{M}_{h,t}$ (Columns (5) and (6)). The negative relationship between labor share and concentration in the model with oligopsonistic competition is about 8 times higher than in the data. Comparing now the last two Columns that correspond to our model, the negative relationship is still too strong but it is half of the model without bargaining. Models with bargaining only and with employer labor market power without strategic interactions would not match the data as the effect of concentration on the labor shares would be null.

1.6 Counterfactuals

In this section we evaluate efficiency and welfare effects of the labor wedges. We compute the main counterfactuals for the last year of our sample, 2007. We start by showing that counterfactuals can be computed observing establishment Revenue Total Factor Productivities (TFPRs) instead of the underlying productivities. Second, we perform our main counterfactual where we completely eliminate the structural labor wedges and compute output and welfare gains under free mobility of workers. We also consider other counterfactual situations where labor wedges remain and are equal to the bargaining only or oligopsonistic competition only cases.

Our baseline counterfactuals assume free mobility of labor. We perform three additional counterfactuals relaxing the free mobility assumption to evaluate if output gains can be attained when mobility is restricted. First, in the most restrictive case, we allow movements only within local labor markets. This is equivalent to assuming infinite mobility costs across locations, industry and occupations. Second, we fix employment at the 2-digit and occupation level and let labor move across locations and 3-digit industries. Third, we fix employment at the 2-digit level. Compared to the previous case, labor is mobile across occupations.

We finally use the model to study the incidence of labor market power on the pass-through of productivity to wages, the urban-rural wage gap and de-industrialization process over time.

1.6.1 Fundamentals

This section shows that is possible to compute the counterfactuals in general equilibrium by just backing out the Revenue Total Factor Productivities (TFPRs), which are a function of prices determined in general equilibrium, rather than the underlying physical productivities. A priori, the issue is that counterfactually changing the labor wedge changes equilibrium prices and therefore the 'fundamental' TFPRs.

The literature has used the TFPRs, together with a modeling assumption on the industry price, to compute the normalized within industry productivity distribution. This has prevented to compute full blown general equilibrium counterfactuals that also take into account productivity differences across industries.[44] We show that we can perform counterfactuals in general equilibrium by writing the model in relative differences from a baseline scenario and also compute the movement of production factors across industries.

We observe employment and wages at the establishment-occupation level from the data. The method is based on recovering establishment-occupation TFPRs using the wages' first order conditions. Equation (1.16) in nominal terms is:

$$Pw_{io} = \beta_b \lambda(\mu_{io}, \varphi_b)\ PF_b^{1+\varepsilon_b \delta} A_{io} L_{io}^{-\delta}, \tag{1.26}$$

where Pw_{io} and L_{io} are observed and $\beta_b \lambda(\mu_{io}, \varphi_b)$ depends on the estimated parameters and observed employment shares. Equation (1.26) makes clear that given the observed nominal wages and employment, one can only back out the transformed TFPRs $Z_{io} = PF_b^{1+\varepsilon_b \delta} A_{io}$ that are a function of the establishment-occupation physical productivity A_{io} and prices $PF_b^{1+\varepsilon_b \delta}$.[45]

Our approach is to write counterfactual industry prices relative to the baseline and fix the transformed revenue productivities.[46] Using the definition of the transformed revenue productivities, the above equation

[44]For example, Hsieh and Klenow (2009) conduct a counterfactual where they remove distortions at the firm level and compute the productivity gains at the *industry* level. The productivity gains are a result of factors of production reallocating to more productive firms *within* each industry. This allows them to compute a *partial* equilibrium effect on total factor productivity, i.e. keeping the production factors constant *across* industries. A general equilibrium effect on total factor productivity takes into account, not only the reallocation of inputs within, but also across industries. They can't do this as they can only identify relative productivity differences within each industry while normalizing average differences across industries. For more details, see equation (19) and the discussion below in their paper.

[45]Revenue Total Factor Productivities are defined as $PP_b A_{io}$. With some abuse of notation, we name the transformed revenue total factor productivities $PF_b^{1+\varepsilon_b \delta} A_{io}$ as TFPRs. Given that one cannot observe industry prices P_b, backing out productivities A_{io} from the data requires performing some normalizations to get rid of industry prices.

[46]Solving the counterfactuals in level as stated in Section 1.4 would require to back out the productivities. It would be possible to do so by making some additional normalizations per industry. For example, one could assume that the minimum physical productivity (or

(1.26) is:

$$Pw_{io} = \beta_b \lambda(\mu_{io}, \varphi_b) Z_{io} L_{io}^{-\delta}.$$

We denote with a prime the variables in the counterfactual (e.g. F_b') and with hat the relative variables (e.g. $\widehat{F}_b = \frac{F_b'}{F_b}$). We have that $Z_{io}' = P'(F_b')^{1+\varepsilon_b\delta} A_{io} = \widehat{P}\widehat{F}_b^{1+\varepsilon_b\delta} Z_{io}$. Fixing the transformed TFPR's observed in the data, we can compute Z_{io}. Denoting by λ_{io}' the counterfactual wedge, the counterfactual real wages are:

$$\begin{aligned} w_{io}' &= \beta_b \lambda_{io}' Z_{io}' L_{io}'^{-\delta} \frac{1}{P'} \\ &= \beta_b \lambda_{io}' Z_{io} \frac{\widehat{F}_b^{1+\varepsilon_b\delta}}{P} L_{io}'^{-\delta}, \end{aligned} \tag{1.27}$$

where in the last step we used the definition of the transformed TFPRs. In the counterfactuals Z_{io} is fixed and we have to solve for industry prices relative to the baseline $\widehat{F}_b$.

Substituting the labor supply and solving for the wages the system becomes:

$$w_{io}' = \left(\beta_b \lambda_{io}' \frac{Z_{io}}{(T_{io}\Gamma_b^{\eta})^{\delta}} \right)^{\frac{1}{1+\varepsilon_b\delta}} \frac{\widehat{F}_b}{P^{\frac{1}{1+\varepsilon_b\delta}}} \Phi_m'^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi'}{L'} \right)^{\nu_b}, \tag{1.28}$$

The establishment-occupation component in the counterfactual ω_{io} is: $\omega_{io} = \left(\beta_b \lambda_{io}' \frac{Z_{io}}{(T_{io}\Gamma_b^{\eta})^{\delta}} \right)^{\frac{1}{1+\varepsilon_b\delta}}$.

Finally, the counterfactual conditional employment shares up to the industry level, $s_{io|mo}' s_{m|b}'$ and industry employment L_b' can be computed. Following the same steps as in the baseline, the industry level system of equations is analogous to (1.21) but with relative variables.[47] Propositions 1 and 2 apply and therefore the solution for the relative counterfactuals exists and is unique.

1.6.2 Main Counterfactuals

We consider four different situations. First, the main counterfactual presents a situation where labor wedges disappear and establishments and workers acts as price takers. Second, the limit case of our framework where there is only bargaining. Third, the limit case where employer labor market power is the only one present, and finally, a situation where unions collect all the profits.

Table 9 shows results of different counterfactuals under the free mobility assumption. The first Column present labor shares in the baseline and the counterfactuals and the rest of the Columns recover the percentage gains of the counterfactuals with respect to the baseline. Output gains are in Column 2 of Table 9. Eliminating labor wedges coming from employer and union labor market power increases aggregate output by 1.6%.

The counterfactual without employer labor market power but keeping the one of unions almost attains the output gains from eliminating both distortions. This counterfactual is a situation where establishments would not internalize movements along the labor supply and the labor wedges become $\lambda(1, \varphi_b) = 1 + \varphi_b \frac{\delta}{1-\delta}$. It is important to note that this is due to the assumed institutional framework for the unions. The bargaining only case features a reduced heterogeneity of labor wedges (only different across industries) that is behind the result of almost attaining output gains of the main counterfactual.

Comparing now to the counterfactual with employer labor market power, we see that output is reduced by 0.21% with respect to the baseline. This result is despite the fact that total employment is fixed. The

Total Factor Productivity, TFP) is constant across industries and get rid of industry relative prices by normalizing the minimum TFP per industry.

[47] Appendix 1.A provides the steps for the computation of the relative counterfactuals.

mechanism behind this result is that labor wedges would be slightly more heterogeneous than in the baseline. Finally, output gains when there is full bargaining and workers extract all the profit rents are the same as in the main counterfactual as wedges would be constant.

Table 9 – Counterfactuals: Efficiency and Distribution

		Gains (%)		
	LS (%)	ΔY	Δ Wage	Δ Welfare (L)
Baseline	50.62	-	-	-
Counterfactuals				
No wedges $\lambda(\mu, \varphi_b) = 1$	72.26	1.62	45.06	42.07
Not internalize $\lambda(1, \varphi_b) = 1 + \varphi_b \frac{\delta}{1-\delta}$	73.38	1.60	47.27	44.34
Oliposonistic $\lambda(\mu, 0) = \mu_{io}$	40.94	-0.21	-19.29	-20.53
Full bargain $\lambda(\mu, 1) = 1 + \frac{\delta}{1-\delta}$	75.47	1.62	51.51	48.38

Notes: First Column presents the aggregate labor share (in percent) for the baseline and the different counterfactuals. The last three Columns changes with respect to the baseline in percentages. ΔY is the change of aggregate output, Δ *Wage* is the change in aggregate wage. Aggregate wage is an employment weighted average of establishment-occupation wages. Δ *Welfare (L)* is the change of the median expected welfare of the workers. The main counterfactual is the one without wedges $\lambda = 1$. The second counterfactual *Not internalize* is the counterfactual where the workers' outside options are the competitive wages. *Oligopsonistic* is the counterfactual where the wedge is equal to the equilibrium markdown under oligopsonistic competition and *Full bargain* is the counterfactual where $\varphi_b = 1$ workers earn all the profits. Counterfactuals are performed in 2007.

Getting now to the split of output into the labor and profit shares, the aggregate labor share in the model can be constructed from industry level labor wedges Λ_b. Those are sufficient statistics to compute the aggregate labor share which is simply a value added weighted sum of industry labor shares. Aggregate labor share is:[48]

$$LS = \sum_{b \in \mathcal{B}} \beta_b \Lambda_b \theta_b.$$

Aggregate labor share is equal in all the variations of the main counterfactual without labor wedges as $\Lambda_b = 1$, for all industries b.

Column (1) of Table 9 presents the aggregate labor shares of the different counterfactuals. We find that completely removing structural labor wedges increases the labor share by 21 percentage points, passing from 50.62% in the baseline to 72.26% in the counterfactual. Aggregate labor share increases slightly more in the counterfactuals where employer labor market power disappears (up to 75% where there is full bargain) and is reduced by 9 p.p. in the counterfactual with oligopsonistic competition.

Labor share changes imply changes in aggregate wages and worker welfare. Column 3 presents the relative change of wages with respect to the baseline. Wages go up by 45% in the price taking case and are

[48]The derivation of the theoretical labor share is in Appendix 1.A.5.

reduced by 19% when the wedge becomes $\lambda(\mu, 0) = \mu_{io}$. Increases in the aggregate wage do not imply that wage inequality is reduced. Figure 13 in Appendix 1.G shows that the demeaned wage distributions on the baseline and the price taking counterfactuals (in Panel (a) and (b) respectively) are very similar. This Figure highlights that even in the absence of labor wedges, wages across establishments are not equalized. This result is due to differences in productivities and amenities across establishments.

We can also analyze how the median expected welfare changes for workers. This median expected utility is:[49]

$$Median(\mathcal{U}_{iok}) \propto \Phi^{\frac{1}{\eta}}.$$

Column (4) of Table 9 present counterfactual gains of the median worker utility. The median expected worker utility is 42% greater in the scenario without labor wedges compared to the baseline. Unsurprisingly, welfare gains are greater than output gains as the workers not only benefit from the productivity boost but also from the redistribution of pure rents that the owners were taking. Unsurprisingly, gains in wages are higher than gains in median welfare. Given the taste shocks, welfare gains go hand in hand with wage gains. Nevertheless, wages need to increase slightly more than welfare to induce labor reallocation.

We perform three additional counterfactuals to locate the output gains in a more realistic environment with mobility costs. They differ in restrictions imposed on mobility. First, we limit mobility to be only within industry, industry-occupation and local labor market. Table 10 compares the free mobility case with restricted mobility cases. Comparing Column (1) across the different scenarios, we find that the key margin of adjustment is geographical mobility. Fixing employment at the industry-occupation level accounts for 82% of the gains of the free mobility case. Restricting workers to stay in their particular local labor market output gains are 0.49% which constitute only 30% of the gains under free mobility.

These results underscore the importance of free mobility of labor across locations as the main driver for output gains. Figure 3 shows the percentage change of manufacturing employment in the free mobility case. Each block is a commuting zone and we aggregate all local labor markets. The main conclusion from the counterfactual analysis is that, in the absence of labor wedges, manufacturing employment in big cities as Paris, Lyon, Marseille or Toulouse would be reduced. The counterfactual reveals that there are a handful rural productive establishments in concentrated local labor markets. In the baseline these have lower wage markdowns and lower employment. Moving to the counterfactual, those are the ones with the biggest wage and employment gains.[50]

Turning now to the source of the output gains, we can use the aggregate production function and the relative industry output from Appendix 1.A (equation (41)), and decompose the logarithm of the relative final output into three terms:

$$\ln \widehat{Y} = \underbrace{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{F}_b^{\alpha_b(1+\varepsilon_b \delta)}}_{\Delta \text{ GE}} + \underbrace{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{Z}_b}_{\Delta \text{ Productivity}} + \underbrace{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{L}_b^{1-\delta}}_{\Delta \text{ Labor}}. \tag{1.29}$$

The first term on the right hand side corresponds to the capital effects or general equilibrium effects of capital flowing to different sectors as a response to changes in relative prices. The second term, arguably the most

[49] As the across local labor market elasticity η being smaller than 1, the expected value of the Fréchet distribution is not defined. We therefore can only compute the median and the mode of the worker welfare.

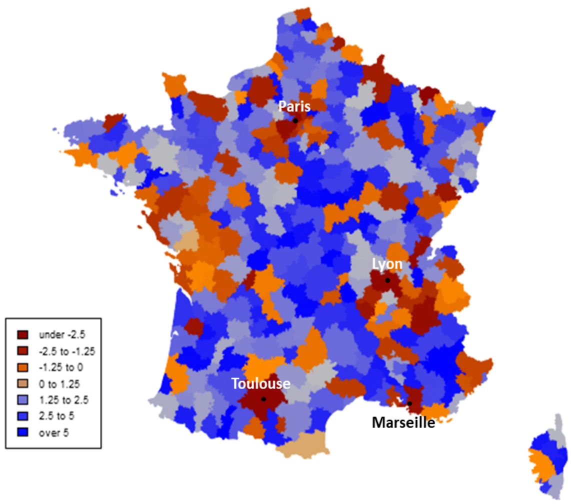
[50] Another potential reason is the differential in the amenities. The reduction of manufacturing labor in the big cities could be magnified if they have in general worse amenities. We leave this analysis to future work.

Table 10 – Counterfactuals: Limited Mobility

	ΔY (%)	Δ Prod (%)	Contribution (%)		
			Sh. GE	Sh. Prod	Sh. Labor
Free mobility	1.62	1.33	9	83	8
Mobility within					
Industry	1.32	1.33	-1	101	0
Industry-occ	1.33	1.35	-2	102	0
Local market	0.49	0.49	-2	102	0

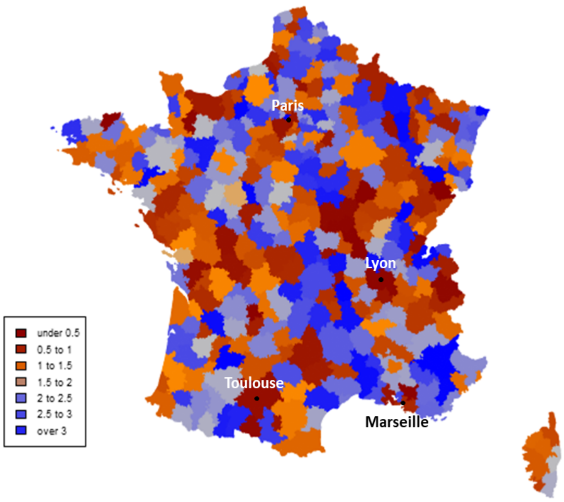
Notes: All the table presents results in percentages. First Column presents the ΔY is the change of aggregate output with respect to the baseline, Δ *Prod* is the change in aggregate productivity from decomposition (1.29). Last three Columns present the contribution of each of the elements of the decomposition (1.29) to output gains. *Free Mobility* presents the main counterfactual without wedges and under free mobility of labor. *Industry* is the counterfactual where mobility is restricted to be only within industry, *Industry-occ* fixes employment at the industry-occupation and allows for mobility across locations, and *Local market* allows for mobility only across establishments within local labor markets. Counterfactuals are performed in 2007.

Figure 3 – Employment Change (%) with Counterfactual



Notes: The map presents employment changes with respect to the baseline economy in percentages. Each block constitutes a commuting zone. Local labor markets are aggregated up to the commuting zone. Counterfactuals are performed in 2007.

Figure 4 – Productivity Change (%) with Counterfactual



Notes: The map presents productivity changes with respect to the baseline economy in percentages. Each block constitutes a commuting zone. Local labor markets are aggregated up to the commuting zone. Commuting zone productivity is an employment weighted average of individual productivities. Following the discussion in Section 1.6.1, keeping fixed the baseline revenue productivities, any change in the counterfactual comes from changes in productivities. Counterfactuals are performed in 2007.

important, represents total productivity gains. This term suffers the most from labor market concentration as big productive firms are shrinking their relative participation, therefore reducing overall productivity. The third term corresponds to how labor is allocated across sectors.

Columns (3) to (5) of Table 10 show the relative changes of output with respect to these three terms.[51] The main source of output gains come from productivity. Industry productivity is an employment weighted sum of establishment-occupation productivities (that are unchanged). The original source of productivity and output gains is the reallocation of workers towards productive firms.

Column (2) shows the productivity gains in the different mobility cases. Those are similar as long as labor is mobile at the industry level. General equilibrium effects determine the reallocation of employment across industries and total output gains. Mobility restrictions below the industry level prevent the reallocation towards productive establishments and reduce the productivity gains.

Figure 4 shows geographical differences of productivity gains in the free mobility case. The Figure is similar to Figure 3 in the sense that most significant gains of the counterfactual productivity happen outside urban areas. The largest gains in terms of productivity, wages and employment are in commuting zones without big cities.

1.6.3 Pass Through

The structural wage equation (1.28) relates our recovered measure of productivity Z_{io} to equilibrium wages. Taking logs, equilibrium wage in the baseline economy is:

$$\log w_{io} = \frac{1}{1+\varepsilon_b \delta} \left(\log Z_{io} - \delta \log T_{io} + \log \lambda(\mu_{io}, \varphi_b)\right) + f_m, \tag{1.30}$$

[51] Note that $\Delta Y = \widehat{Y} - 1 \approx \ln \widehat{Y}$. The decomposition is with respect to $\ln \widehat{Y}$. The share of the gains that come from productivity (Sh. Prod) is simply $\frac{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{Z}_b}{\ln Y}$. Each row from Columns 3 to 5 sums up to 1.

where f_m is a fixed effect at the local labor market level. We use this equation to study the incidence of labor market power on the pass through of the transformed revenue productivity Z. The elasticity of wages with respect to Z is:

$$\epsilon_Z^W = \frac{\partial \log w_{io}}{\partial \log Z_{io}} = \underbrace{\frac{1}{1+\varepsilon_b \delta}}_{\text{Pass Through No Wedge}} + \frac{1}{1+\varepsilon_b \delta} \underbrace{\epsilon_s^\lambda}_{<0} \underbrace{\epsilon_Z^s}_{>0},$$

where ϵ_s^λ and ϵ_Z^s denote respectively the elasticity of the wedge λ_{io} with respect to the employment share s and the elasticity of the employment share s with respect to our measure of productivity Z. The equation above emphasizes the origin of potential distortions coming from labor market power. When the wedge is constant, the last term becomes zero because $\epsilon_s^\lambda = 0$. In that case, the pass through of productivity to wages is the same as in the price taking case and the labor allocations are not distorted.

We estimate the following:

$$\log w_{iot} = f_{mot} + \beta_b^Z \log Z_{iot} + \beta_b^T \log T_{iot} + u_{iot}$$

Table 21 in Appendix 1.I presents the estimates of the productivity pass through in the baseline β_b^Z and the one in the absence of labor wedges. The average dampening due to labor market power is equal to 0.05. This means that when Z increases by 1%, 0.05% of that increase is not translated to wages due to labor market frictions.

1.6.4 Mobility and Wage Gap

Figure 3 suggests an important movement from cities to rural areas in the counterfactual. This section explores the impact of employer and union labor market power on the de-industrialization process and the urban-rural wage gap.

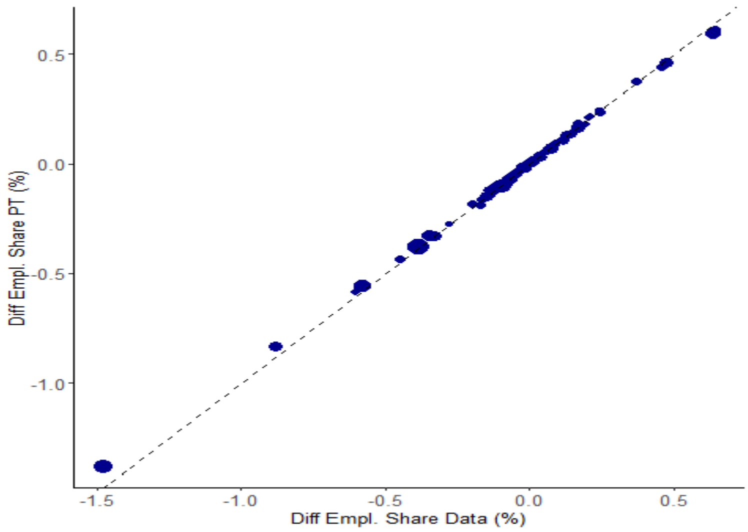
Mobility over time

Movements in Figure 3 suggest employment reallocation from cities to rural areas in the manufacturing sector. Here we compare the de-industrialization process observed in the data and the one from the counterfactuals.

In the data, de-industrialization or the reduction of manufacturing employment occurred primarily in cities. Figure 5 compares the de-industrialization process observed in the data to the one we would have in the counterfactuals.

First, we compute the commuting zone employment share out of total manufacturing for the initial and final years (1994 and 2007 respectively) and for the different scenarios. Then, we compute the differences in the data ($\Delta^D = S_{07}^D - S_{94}^D$) and in the counterfactual ($\Delta^M = S_{07}^{PT} - S_{94}^{PT}$). Figure 5 presents this comparison. The x axis shows Δ^D and the y axis shows Δ^M. The size of the dots are the initial population. The counterfactual de-industrialization process is very similar to the process observed in the data. Over time, de-industrialization is mostly guided by exogenous productivity and firm location decisions and not by labor market distortions.

Figure 5 – De-industrialization differences



Notes: The x-axis shows the percentage differences of employment shares over time in the data ($\Delta^D = S^D_{07} - S^D_{94}$). The y-axis presents the analogous for the counterfactual without wedges ($\Delta^M = S^{PT}_{07} - S^{PT}_{94}$). The initial period is 1994 and the final year is 2007.

The line generated by the largest population commuting zones is slightly flatter than the 45 degree line. The de-industrialization would have been a bit slower in the counterfactual. This is mostly explained by the closure of manufacturing firms in the largest cities that became more concentrated over time.

Wage Gap

Table 11 presents urban and rural wages besides the urban/rural wage gap.[52] Both experience important wage gains in the counterfactual. Gains are bigger outside cities, which reduces the wage gap from 36% to 23%. This reveals that labor market distortions account for more than a third of the urban/rural wage gap.

Table 11 – Wage Gap

	Rural Wage	Urban Wage	Gap (%)
Baseline	33.321	45.210	36
Counterfactual	49.486	60.675	23

Note: Wages in thousands of constant 2015 euros. We classify as *Urban* the 10 biggest commuting zones: Paris, Marseille, Lyon, Toulouse, Nantes, and the Paris surrounding, Boulogne-Billancourt, Creteil, Montreuil, Saint-Denis and Argenteuil. The rest are considered as *Rural*. Wages are employment weighted averages per category for the baseline and counterfactual for the year 2007.

[52] We consider urban the 10 biggest commuting zones: Paris, Marseille, Lyon, Toulouse, Nantes, and the Paris surrounding, Boulogne-Billancourt, Creteil, Montreuil, Saint-Denis and Argenteuil. Rural are the rest of the commuting zones.

1.7 Extensions

Total labor supply was fixed in the baseline counterfactual, workers were perfectly mobile and there were no agglomeration externalities. In this section, we propose two extensions. First, we allow for an endogenous labor participation. Second, we introduce agglomeration forces in the local labor markets.

1.7.1 Endogenous Participation

We briefly present an extension where we allow for endogenous labor force participation decisions. We assume workers can decide between working and staying at home. In the latter case, they earn some wages related to home production. In the model, staying at home is an endogenous choice that happens when the indirect utility of being out of the labor force is higher than the one being employed.

We lack detailed data on the geographical distribution of out of the labor force status. Labor force surveys provide only information at the region level. Basing our counterfactuals in those surveys would require the extreme assumption of constant rates of labor participation for entire regions. Instead, while acknowledging is not a perfect assumption, we use commuting zone level unemployment rates as out-of-the labor-force rates.

Defining out-of-the-labor-force, from now on OTLF, as a new sub-industry at every location, 2-digit industry and occupation combination, we have that the probability of being OTLF in a particular commuting zone n and 2-digit industry b is:

$$L_{uo} = \frac{(T_{uo} w_{uo}^{\varepsilon_b})^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi} L, \quad \Phi = \Phi_e + \Phi_u,$$

where $\Phi_e = \sum_{m \in \mathcal{I}_m} \Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}$ is the part of Φ that comes from the employed and $\Phi_u = \sum_{uo \in \mathcal{U}_m} (T_{uo} w_{uo}^{\varepsilon_b})^{\eta/\varepsilon_b} \Gamma_b^{\eta}$ is the part from the unemployed ($\mathcal{U}_m$ is the set of all OTLF local labor markets). L is the total labor supply of the both employed and OTLF workers. The proportion of OTLF workers in each local market identifies the home production wage $T_{uo} w_{uo}^{\varepsilon_b}$.[53] This wage is fixed in the counterfactuals while the real wages of firms change depending on the counterfactual wedges.

Table 12 shows the results of the counterfactuals with endogenous labor force participation. The counterfactual output gain is 1.98%. Introducing the endogenous labor participation margin induces higher output gains than in the baseline (Fixed L). In contrast with the results shown in Table 10, around 30% of the gains come from the increased total employment. Labor force increases 1% in the main counterfactual without wedges. This extensive margin adjustment in the total labor supply amplifies original differences in output gains across counterfactuals. In particular, output losses from oligopsonistic competition are as high as 1.29% because total labor force participation is reduced (-0.75%).

1.7.2 Agglomeration

In this section we present an extension of the model that includes agglomeration forces at the local labor market level. To keep the model tractable, we assume that the productivity is: $\widehat{A}_{io} = \widetilde{A}_{io} L_m^{\gamma(1-\alpha_b)}$. The agglomeration effect is a local labor market externality with elasticity $\gamma(1-\alpha_b)$. The wage first order condition is:

$$P w_{io} = \beta_b \lambda(\mu_{io}, \varphi_b) Z_{io} L_{io}^{-\delta} L_m^{\gamma}. \tag{1.31}$$

[53] Details on the theoretical model with endogenous participation are in Appendix 1.B.

Table 12 – Counterfactual: Endogenous Participation

				Contribution (%)		
	ΔY (%)	Δ Prod (%)	Δ L (%)	Sh. GE	Sh. Prod	Sh. Labor
Fixed L	1.62	1.33	0.00	9	83	8
Endogenous Part.						
No wedges $\lambda(\mu, \varphi_b) = 1$	1.98	1.18	1.00	11	60	29
Not internalize $\lambda(1, \varphi_b) = 1 + \varphi_b \frac{\delta}{1-\delta}$	2.04	1.18	1.04	10	58	32
Oligopsonistic $\lambda(\mu, 0) = \mu(s)$	-1.29	-0.59	-0.75	2	46	53
Full bargain $\lambda(\mu, 1) = 1 + \frac{\delta}{1-\delta}$	2.09	1.18	1.12	10	57	33

Notes: All the table presents results in percentages. First Column ΔY is the change of aggregate output with respect to the baseline, Δ *Prod* is the change in aggregate productivity from decomposition (1.29) and ΔL is the counterfactual change in total employment. Last three Columns present the contribution of each of the elements of the decomposition (1.29) to output gains. *Fixed L* is the main counterfactual without wedges, under free mobility of labor and fixed total labor supply. The main counterfactual is the one without wedges $\lambda = 1$. All the other counterfactuals in this table allow for endogenous labor force participation. *No wedges* is the analogous to the main counterfactual without wedges. *Not internalize* is the counterfactual where the workers' outside options are the competitive wages. *Oligopsonistic* is the counterfactual where the wedge is equal to the equilibrium markdown under oligopsonistic competition and *Full bargain* is the counterfactual where $\varphi_b = 1$ workers earn all the profits.

Similarly to the baseline counterfactual, we back out the fundamental Z_{io} to perfectly match observed establishment-occupation wages w_{io}. In the case where employment for a given local labor market is high, the productivity of the establishments in that market m is lower than for the main counterfactual.[54]

[54]Following the steps described in Appendix 1.B.2, we can solve for the counterfactuals solving first the normalized wages and then for industry prices.

Table 13 – Counterfactuals: Agglomeration

			Contribution (%)		
	ΔY (%)	Δ Prod (%)	Sh. GE	Sh. Prod	Sh. Labor
No Agglomeration	1.62	1.33	9	83	8
Agglomeration					
$\gamma = 0.05$	1.73	1.40	8	82	10
$\gamma = 0.1$	1.84	1.48	7	81	12
$\gamma = 0.15$	1.96	1.57	6	81	13
$\gamma = 0.2$	2.08	1.66	5	80	15
$\gamma = 0.25$	2.22	1.75	3	80	17

Notes: All the table presents results in percentages. First Column ΔY is the change of aggregate output with respect to the baseline, Δ *Prod* is the change in aggregate productivity from decomposition (1.29). Last three Columns present the contribution of each of the elements of the decomposition (1.29) to output gains. *No Agglomeration* is the main counterfactual without wedges, under free mobility of labor, fixed total labor supply and no agglomeration forces. All the other counterfactuals in this table allow for agglomeration within the local labor market. Similarly to the main counterfactual, workers are freely mobile and total employment is fixed. We present different counterfactuals depending on the agglomeration elasticity γ.

Table 13 summarizes the counterfactual results for different values of γ. All the counterfactuals in Table 13 also assume price taking and free mobility but introduce agglomeration forces in local labor markets. As γ becomes higher, the more important are the agglomeration forces and the higher are the efficiency gains. The reason behind this result is that increasing γ the local labor market employment L_m becomes more important in (1.31)- Consequently, productivity differences across local labor markets with different employment are amplified. The movements towards small local labor markets are therefore bigger than in the main counterfactual. Output gains are monotonic in the importance of agglomeration externalities.

1.8 Conclusion

This paper measures efficiency and welfare losses generated by employer and union labor market power for French manufacturing establishments. We present stylized facts at the aggregate level that show higher employment concentration relates to lower labor shares for French manufacturing firms. We further document the relevance of heterogeneous labor market power at the establishment level. Our empirical strategy identifies a negative relationship between local labor market employment share and wages. This reduced form evidence suggests employer labor market power is relevant and heterogeneous across markets and firms.

We lay out a quantitative general equilibrium model that links structural labor wedges to employment shares and union's bargaining power. Our framework nests the cases with bargaining only and oligopsonistic competition only as special cases. We show existence and uniqueness of the equilibrium and provide its analytical characterization. We estimate parameters by exploiting the structural equations and the micro-data.

Finally, we evaluate the efficiency and welfare costs of employer and union labor market power. We find that removing structural labor wedges would increase output by 1.6%. Gains are slightly bigger, up to 1.98% when we allow for an endogenous labor force participation margin. The main mechanism behind the output gains is the reallocation of resources towards more productive firms.

Removing labor market distortions lead to significant labor share and wage gains. Those results imply that the markdown is more important on the labor wedge and in turn highlight the importance of employer labor market power in France. The potential insights for policy are clear. The framework suggests that the allocation without labor market distortions can be implemented by hiring subsidies that would eliminate the effect of the labor wedge. Those subsidies could be financed either by taxes on profits or on wage earnings.

1.A Derivations

In this section we provide the derivations of the model that are not presented in the main text. First, we show how to obtain the establishments labor supplies by solving the workers establishment choice problem. Later, we show how we obtain the markdown function from the establishments optimality conditions. We then show how to get a close form solution for the prices given the solution for the normalized wages.

1.A.1 Establishment-Occupation Labor Supply

To simplify the notation, we get rid of the occupation subscript o in this subsection. The indirect utility of a worker k that is employed in establishment i in sub-market m is:

$$u_{kim} = w_i z^1_{i|m} z^2_m,$$

where $z^1_{i|m}$ and z^2_m are independent utility shocks. They are both distributed Frèchet with shape and scale parameters ε_b and T_i for $z^1_{i|m}$, and η and 1 for z^2_m.

Workers first see the realizations of the shocks z^2_m for all local labor markets. After choosing to which labor market to go, the workers then observe the establishment specific shocks. Therefore, there is a two stage decision: first, the worker choose the local labor market that maximizes her expected utility, and later will choose the establishment that maximizes her utility conditional on the chosen sub-market.

The goal is to compute the unconditional probability of a worker going to establishment i in sub-market m. This probability is equal to:

$$\Pi_i = P\left(w_i z^1_{i|m} \geq \max_{i' \neq i} w_{i'} z^1_{i'|m}\right) P\left(\mathbb{E}_m(\max_i w_i z^1_{i|m}) z^2_m \geq \max_{m' \neq m} \mathbb{E}_{m'}(\max_i w_i z^1_{i|m'}) z^2_{m'}\right)$$

We first solve for the left term. Let's define the following distribution function:

$$G_i(v) = P\left(w_i z^1_{i|m} < v\right) = P\left(z^1_{i|m} < v/w_i\right) = e^{-T_i w_i^{\varepsilon_b} v^{-\varepsilon_b}}.$$

To ease notation, define *conditional* utility $v_i = w_i z^1_{i|m}$ for all i, i'. We need to solve for $P\left(v_i \geq \max_{i' \neq i} v_{i'}\right)$. Fix $v_i = v$. Then we have:

$$P\left(v \geq \max_{i' \neq i} v_{i'}\right) = \bigcap_{i' \neq i} P\left(v_j < v\right) = \prod_{i' \neq i} G_{i'}(v) = e^{-\Phi_m^{-i} v^{-\varepsilon_b}} = G_m^{-i}(v),$$

where $\Phi_m^{-i} = \sum_{i' \neq i} T_{i'} w_{i'}^{\varepsilon_b}$. Similarly, the probability of having at most conditional utility v is equal to

$$G_m(v) = P\left(v \geq \max_{i'} v_{i'}\right) = e^{-\Phi_m v^{-\varepsilon_b}},$$

where $\Phi_m = \sum_{i'} T_{i'} w_{i'}^{\varepsilon_b}$. Integrating $G_m^{-i}(v)$ over all possible values of v we then get:

$$\begin{aligned}
P\left(v_i \geq \max_{i' \neq i} v_{i'}\right) &= \int_0^\infty e^{-\Phi_m^{-i} v^{-\varepsilon_b}} dG_i(v) \\
&= \int_0^\infty \varepsilon_b T_i w_i^{\varepsilon_b} v^{\varepsilon_b - 1} e^{-\Phi_m v^{-\varepsilon_b}} dv \\
&= \frac{T_i w_i^{\varepsilon_b}}{\Phi_m} \int_0^\infty \varepsilon_b \Phi_m v^{\varepsilon_b - 1} e^{-\Phi_m v^{-\varepsilon_b}} dv \\
&= \frac{T_i w_i^{\varepsilon_b}}{\Phi_m} \int_0^\infty dG_m(v) = \frac{T_i w_i^{\varepsilon_b}}{\Phi_m}.
\end{aligned}$$

Now we need to find $P\left(\mathbb{E}_m(\max_i w_i z^1_{i|m})z^2_m \geq \max_{m'\neq m}\mathbb{E}_{m'}(\max_i w_i z^1_{i|m'})z^2_{m'}\right)$. So first, the expected utility of working in sub-market m is:

$$\mathbb{E}_m(\max_i w_i z^1_{i|m}) = \int_0^\infty v_i dG_m(v) = \int_0^\infty \varepsilon_b \Phi_m v^{-\varepsilon_b} e^{-\Phi_m v^{-\varepsilon_b}} dv.$$

We define this new variable:

$$x = \Phi_m v^{-\varepsilon_b} \qquad dx = -\varepsilon_b \Phi_m v^{-(\varepsilon_b+1)} dv.$$

Now we can change variable in the previous integral and obtain:

$$\int_0^\infty x^{-1/\varepsilon_b}\Phi_m^{1/\varepsilon_b} e^{-x} dx = \Gamma\left(\frac{\varepsilon_b - 1}{\varepsilon_b}\right)\Phi_m^{1/\varepsilon_b},$$

where $\Gamma(\dot{)}$ is just the Gamma function. Defining $\Gamma_b \equiv \Gamma\left(\frac{\varepsilon_b-1}{\varepsilon_b}\right)$, we can then rewrite:

$$P\left(\mathbb{E}_m(\max_i w_i z^1_{i|m})z^2_m \geq \max_{m'\neq m}\mathbb{E}_{m'}(\max_i w_i z^1_{i|m'})z^2_{m'}\right) = P\left(\Phi_m^{1/\varepsilon_b}\Gamma_b z^2_m \geq \max_{m'\neq m}\Phi_{m'}^{1/\varepsilon_{b'}}\Gamma_{b'} z^2_{m'}\right).$$

Following the similar arguments as above, this probability is equal to:

$$P\left(\Phi_m^{1/\varepsilon_b}\Gamma_b z^2_m \geq \max_{m'\neq m}\Phi_{m'}^{1/\varepsilon_b}\Gamma_{b'} z^2_{m'}\right) = \frac{\Phi_m^{\eta/\varepsilon_b}\Gamma_b^\eta}{\Phi},$$

where $\Phi = \sum_{b'\in\mathcal{B}}\sum_{m'\in\mathcal{M}_{b'}}\Phi_{m'}^{\eta/\varepsilon_{b'}}\Gamma_{b'}^\eta$.

Finally, combining the two probabilities we obtain the same expression in the main text:

$$\Pi_i = \frac{T_i w_i^{\varepsilon_b}}{\Phi_m}\frac{\Phi_m^{\eta/\varepsilon_b}\Gamma_b^\eta}{\Phi}.$$

By integrating Π_i over the whole measure of workers L, we can obtain the labor supply for each establishment:

$$L_i = \frac{T_i w_i^{\varepsilon_b}}{\Phi_m}\frac{\Phi_m^{\eta/\varepsilon_b}\Gamma_b^\eta}{\Phi}L.$$

Workers' welfare. An obvious way to measure workers welfare would be to compute the average utility for workers. However this is not possible as the shape parameter η is smaller than 1. This implies that the mean for the Frechét distributed utilities is not defined. Instead, we compute the median utility agents expect to receive in each local labor market. This is equal to:

$$\text{Median}\left[\max_m \mathbb{E}_m(\max_i w_i z^1_{i|m})z^2_m\right] \propto \Phi^{\frac{1}{\eta}}.$$

1.A.2 Establishment Decision

In the absence of bargaining, the profit maximization problem of establishment i is:

$$\max_{w_{iot},K_{iot}} P_{bt}\sum_{o=1}^{O}\widetilde{A}_{iot}K_{iot}^{\alpha_b}L_{iot}^{\beta_b} - \sum_{o=1}^{O} w_{iot}L_{iot}(w_{iot}) - R_{bt}\sum_{o=1}^{O}K_{iot},$$

where $L_{iot}(w_{iot})$ is the labor supply (1.13) where they take Φ and L as given but internalize their effect on Φ_{io} and Φ_m. P_{bt} and R_{bt} are respectively the industry price and required rate.[55] Getting rid of the time index t, the first order conditions of this problem are:

$$\begin{aligned} w_{io} &= \beta_b \frac{e_{io}}{e_{io}+1}P_b\widetilde{A}_{io}K_{io}^{\alpha_b}L_{io}^{\beta_b-1}, \\ R_b &= \alpha_b P_b \widetilde{A}_{io}K_{io}^{\alpha_b-1}L_{io}^{\beta_b}. \end{aligned} \tag{32}$$

[55]The construction details of the rental rate of capital or the required rate are in Appendix 1.E.4.

$e_{io} = \varepsilon_b \left(1 - s_{io|m}\right) + \eta\, s_{io|m}$ is the perceived elasticity of supply for establishment i in occupation o.

We can use the first order conditions of capital to substitute it into the establishment's production function and obtain an expression that depends only in labor:

$$y_{io} = \left(\frac{\alpha_b}{R_b}\right)^{\frac{\alpha_b}{1-\alpha_b}} \widetilde{A}_{io}^{\frac{1}{1-\alpha_b}} L_{io}^{\frac{\beta_b}{1-\alpha_b}} P_b^{\frac{\alpha_b}{1-\alpha_b}}. \tag{33}$$

In order to gain tractability in the solution of the model we restrict the output elasticity with respect to capital such that $1 - \frac{\beta_b}{1-\alpha_b} = \delta$, where $\delta \in [0,1]$ is a constant across sectors. This specification would nest a constant returns to scale technology when $\delta = 0$. As long as $0 < \delta < 1$ the establishment faces decreasing returns to scale within occupations. Define a transformed productivity $A_{io} \equiv \widetilde{A}_{io}^{\frac{1}{1-\alpha_b}} \left(\frac{\alpha_b}{R_b}\right)^{\frac{\alpha_b}{1-\alpha_b}}$. The establishment-occupation production is:

$$y_{io} = P_b^{\frac{\alpha_b}{1-\alpha_b}} A_{io} L_{io}^{1-\delta}. \tag{34}$$

1.A.3 Markdown function

We derive the markdown function from the establishments optimality condition with respect to wages. The establishment post a wage and choose capital quantity in order to maximize profits subject to their individual labor supply. Establishments only take into account the effect on their local labor market. As explained in the main text, this can happen because of a myopic behavior from the establishments or if there is a continuum of local labor markets. The establishment problem is:

$$\max_{w_{io}, K_{io}} P_b \sum_{o=1}^{O} \widetilde{A}_{io} K_{io}^{\alpha_b} L_{io}^{\beta_b} - \sum_{o=1}^{O} w_{io} L_{io}(w_{io}) - R_b \sum_{o=1}^{O} K_{io},$$

The first order condition with respect to labor is:

$$P_b \frac{\partial F}{\partial L_{io}} \frac{\partial L_{io}}{\partial w_{io}} = L_{io}(w_{io}) + w_{io} \frac{\partial L_{io}}{\partial w_{io}},$$

where the derivative of the labor supply L_{io} with respect to the establishment-occupation wage w_{io} is:

$$\begin{aligned}
\frac{\partial L_{io}}{\partial w_{io}} &= \frac{L\Gamma_b^{\eta}}{\Phi} \left(\left[\frac{\varepsilon_b T_{io} w_{io}^{\varepsilon_b - 1} \Phi_m - T_{io} w_{io}^{\varepsilon_b} \varepsilon_b T_{io} w_{io}^{\varepsilon_b - 1}}{\Phi_m^2} \right] \Phi_m^{\eta/\varepsilon_b} + \eta \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} \Phi_m^{\eta/\varepsilon_b - 1} T_{io} w_{io}^{\varepsilon_b - 1} \right) \\
&= \frac{\varepsilon_b T_{io} w_{io}^{\varepsilon_b - 1}}{\Phi_m} \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi} L - \frac{\varepsilon_b T_{io} w_{io}^{\varepsilon_b - 1} \Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi_m \Phi} L \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} + \eta \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} \frac{T_{io} w_{io}^{\varepsilon_b - 1}}{\Phi_m} \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi} L \\
&= \varepsilon_b \frac{L_{io}}{w_{io}} - \varepsilon_b \frac{L_{io}}{w_{io}} \frac{L_{io}}{L_m} + \eta \frac{L_{io}}{w_{io}} \frac{L_{io}}{L_m} \\
&= \frac{L_{io}}{w_{io}} \left(\varepsilon_b (1 - s_{io|m}) + \eta s_{io|m} \right).
\end{aligned}$$

Substituting this last derivative into the first order condition we get:

$$\begin{aligned}
& L_{io} + L_{io} \left(\varepsilon_b (1 - s_{io|m}) + \eta s_{io|m} \right) = P_b \frac{\partial F}{\partial L_{io}} \frac{L_{io}}{w_{io}} \left(\varepsilon_b (1 - s_{io|m}) + \eta s_{io|m} \right) \\
\Rightarrow \quad & w_{io} = \frac{\varepsilon_b (1 - s_{io|m}) + \eta s_{io|m}}{\varepsilon_b (1 - s_{io|m}) + \eta s_{io|m} + 1} P_b \frac{\partial F}{\partial L_{io}} \\
& w_{io} = \mu(s_{io|m}) P_b \frac{\partial F}{\partial L_{io}}.
\end{aligned}$$

1.A.4 Bargaining Details

Workers of each occupation bargain with the establishment that retains the right-to-manage. Each establishment has different occupation profit functions $(1-\alpha_b)pF(L_{io}) - w^u_{io}L_{io}$, where the optimal capital decision has been taken. In what follows we abstract from the occupation index o for clarity. The threat point for workers is the wage they would obtain under oligopsonistic competition, w^{MP}_{io}, that we take as an status-quo scenario, where the threat point of firms is zero profits. The bargaining solution chooses wages to maximize:

$$\max_{w^u_{io}} (w^u_{io}L_{io} - w^{MP}_{io}L_{io})^{\varphi_b}((1-\alpha_b)pF(L_{io}) - w^u_{io}L_{io})^{1-\varphi_b},$$

with φ_b being the union's bargaining power, w^u_{io} the wage bargained with the unions at establishment-occupation io, L_{io} the number of workers employed at establishment-occupation io in equilibrium, w^{MP}_{io} is the threat wage of workers, $(1-\alpha_b)F(L_{io})$ is the output of the establishment-occupation after substituting for the optimal decision of capital. To ease exposition, in what follows we get rid of the establishment-occupation subscript io. The first order conditions of the above maximization problem are:

$$\varphi_b((1-\alpha_b)pF(L) - w^uL) = (1-\varphi_b)(w^{MP}L - w^uL).$$

Rearranging the first order condition:

$$w^u = w^{MP} + \frac{\varphi_b}{1-\varphi_b}\frac{((1-\alpha_b)pF(L) - w^uL)}{L}.$$

This is the standard expression on bargaining models, where workers earn a fraction $\frac{\varphi_b}{1-\varphi_b}$ of the quasi-rents of the establishment on excess of their reservation wage. Solving for w^u and substituting $w^{MP} = \mu(s) \times MRPL = \frac{e}{e+1} \times MRPL$ yields:

$$w^u = (1-\varphi_b)\frac{e}{e+1}(1-\alpha_b)p\frac{\partial F(L)}{\partial L} + \varphi_b\frac{(1-\alpha_b)pF(L)}{L}.$$

In the case of a Cobb-Douglas production function, the marginal revenue product of labor is proportional to the labor productivity, i.e. $(1-\alpha_b)p\frac{\partial F(L)}{\partial L} = \beta_b\frac{pF(L)}{L}$, where β_b is the elasticity of output with respect to labor. By the definition of δ, $\beta_b/(1-\alpha_b) = (1-\delta)$, the bargained wage becomes:

$$w^u = \underbrace{(1-\alpha_b)p\frac{\partial F(L)}{\partial L}}_{MRPL}\left[(1-\varphi_b)\frac{e}{e+1} + \varphi_b\frac{1}{1-\delta}\right],$$

where we recovered the expression from the main text.

1.A.5 Aggregate Model

Given the equilibrium definition, the model contains a very large number of variables that could make it unfeasible to be solved numerically. This is because each firm in every location and industry sets its own wage. So if in every sector location pair there would be H sub-industries, and each sub-industry would have I firms, there would be $N \times B \times H \times I$ wages to be solved in the model plus $B+1$ equations for the prices and final output. In comparison, quantitative spatial economic models that assume implicitly that all firms in the same location have the same amenity would only need to solve for N different wages. In this section we show how the fact that firms only take into account the effect of their wage decision on the local labor market helps to tackle this problem by separating it in two main parts. First, we show that we can solve for each sub-market wages by normalizing the sectoral prices and an economy wide constant. Later, we use this normalized wages to construct aggregate expressions that are just functions of sectoral prices and some economy wide constants. Finally, we provide a closed form solution of these prices and the final output conditional on having the solution for the normalized wages.

Following this path allows us to solve the model in a feasible way. Instead of solving a system of $(N \times B \times H \times I) + (B+1)$ equations, we can solve $N \times B \times H$ smaller and simpler systems of I equations each and later a system of $B+1$ equations.

Starting from the expression of wages (1.17),

$$w_{io} = \widetilde{w}_{io} \Phi_m^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi}{L}\right)^{\nu_b} F_b,$$

we can use the definition of $\Phi_m = \sum_{io \in I_m} T_{io} w_{io}^{\varepsilon_b}$ to find,

$$\Phi_m = \widetilde{\Phi}_m^{\psi_b} F_b^{\varepsilon_b \psi_b} \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \varepsilon_b}, \quad \widetilde{\Phi}_m = \sum_{i \in I_m o} T_{io} \widetilde{w}_{io}^{\varepsilon_b}, \quad \psi_b \equiv \frac{1+\varepsilon_b \delta}{1+\eta\delta} \geq 1, \tag{35}$$

Plugging the expression of Φ_m into the one above, and noticing that $\psi_b \nu_b = \delta/(1+\eta\delta)$ we can rewrite the equilibrium wage as,

$$w_{io} = \widetilde{w}_{io} \widetilde{\Phi}_m^{\frac{\psi_b - 1}{\varepsilon_b}} F_b^{\psi_b} \left(\frac{\Phi}{L}\right)^{\frac{\delta}{1+\eta\delta}}. \tag{36}$$

The establishment-occupation labor supply L_{io} can be written as $L_{io} = s_{io|m} s_{m|b} L_b$. Given the solution of normalized wages per sub-market $\widetilde{w}_{io}$, we can actually compute the employment share out of the local labor market $s_{io|m}$:

$$s_{io|m} = \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} = \frac{T_{io} \widetilde{w}_{io}^{\varepsilon_b}}{\widetilde{\Phi}_m}, \quad \widetilde{\Phi}_m = \sum_{i \in \mathcal{I}_m} T_{io} \widetilde{w}_{io}^{\varepsilon_b}.$$

We can also compute the employment share of the local labor market out of the industry $s_{m|b}$. Using the definition of $\Phi_b = \sum_{m \in \mathcal{M}_b} \Phi_m^{\eta/\varepsilon_b}$ and (35),

$$s_{m|b} = \frac{\Phi_m^{\eta/\varepsilon_b}}{\Phi_b} = \frac{\widetilde{\Phi}_m^{\psi_b \eta/\varepsilon_b}}{\widetilde{\Phi}_b}, \quad \widetilde{\Phi}_b = \sum_{m \in \mathcal{M}_b} \widetilde{\Phi}_m^{\psi_b \eta/\varepsilon_b}.$$

where $\mathcal{M}_b$ is the set of all local labor markets that belong to industry b. This just formalizes the notion that, as long as we know the relative wages within an industry, we can compute the measure of workers that go to each establishment conditioning on industry employment.

Turning now to output, we can compute output at the industry level by aggregating establishment-occupation ones according to (1.5):

$$Y_b = F_b^{\alpha_b(1+\varepsilon_b\delta)} A_b L_b^{1-\delta}, \quad A_b = \sum_{m \in \mathcal{M}_b} \sum_{io \in \mathcal{I}} A_{io} s_{io|m}^{1-\delta} s_{m|b}^{1-\delta}, \tag{37}$$

where we obtained an expression that represents the productivity at the industry level A_b. As it is evident from the definition, A_b is an employment weighted industry productivity. The covariance between those two is key in order to determine industry productivity. As long as market power distorts the employment distribution making more productive firms to constraint their size, the covariance between productivity and employment is lower than in the case with competitive labor markets. This decreases total industry productivity A_b.

Using (35), industry labor supply can be written as function of normalized (tilde) variables and transformed prices:

$$L_b = \frac{\Phi_b \Gamma_b^{\eta}}{\sum_{b' \in \mathcal{B}} \Phi_{b'} \Gamma_{b'}^{\eta}} L = \frac{F_b^{\psi_b \eta} \widetilde{\Phi}_b \Gamma_b^{\eta}}{\widetilde{\Phi}} L, \quad \widetilde{\Phi} = \sum_{b' \in \mathcal{B}} F_{b'}^{\psi_b \eta} \widetilde{\Phi}_{b'} \Gamma_{b'}^{\eta}. \tag{38}$$

This is where the simplifying assumption on the labor demand elasticity $\delta \equiv 1 - \frac{\beta_b}{1-\alpha_b}$ being constant across industries buys us tractability. We can factor out the economy wide constant from (35) and leave everything on terms of normalized wages and transformed prices.

In order to find equilibrium allocations, we need to solve for the transformed prices $\mathbf{F} = \{F_b\}_{b=1}^{\mathcal{B}}$. Using the intermediate input demand from the final good producer (1.4) and the above expression for industry labor supply L_b we get:

$$F_b^{\psi_b(1+\eta)} A_b \left(\widetilde{\Phi}_b \Gamma_b^{\eta}\right)^{1-\delta} = \theta_b \prod_{b' \in \mathcal{B}} \left(A_{b'} \left(\widetilde{\Phi}_{b'} \Gamma_{b'}^{\eta}\right)^{1-\delta}\right)^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \left(F_{b'}^{\alpha_{b'}(1+\varepsilon_b \delta)+\psi_b \eta(1-\delta)}\right)^{\theta_{b'}},$$

where we used $1 + \varepsilon_b \delta + \psi_b \eta (1-\delta) = \psi_b (1+\eta)$. Solving for F_b we get (1.22) from the main text.

Aggregate Labor Share

Here we present the steps to compute aggregate labor share, capital to labor expenditures and profit to labor expenditure shares.

Aggregating (1.16) to the industry level,

$$w_b L_b = \beta_b \Lambda_b P_b Y_b, \tag{39}$$

where, $w_b = \sum_{io \in \mathcal{I}_b} w_{io} s_{io|m} s_{m|b}$ is the labor weighted average of individual and $\mathcal{I}_b$ is the set of establishment-occupations that belong to industry b. The industry wedge $\Lambda_b = \sum_{io \in \mathcal{I}_b} \lambda_{io} \frac{P_b Y_{io}}{P_b Y_b}$ is just the value added weighted average of individual wedges. Using (33) and (1.20), the industry markdown Λ_b yields the following expression:

$$\Lambda_b = \frac{\sum_{io \in \mathcal{I}_b} \lambda_{io} A_{io} s_{io|m}^{1-\delta} s_{m|b}^{1-\delta}}{A_b}.$$

Industry and aggregate labor shares are:

$$LS_b = \beta_b \Lambda_b, \quad LS = \frac{\sum_{b \in \mathcal{B}} w_b L_b}{\sum_{b \in \mathcal{B}} P_b Y_b}. \tag{40}$$

Substituting (39) and realizing that industry b expenditure share is equal to θ_b,

$$LS = \sum_{b \in \mathcal{B}} \beta_b \Lambda_b \theta_b.$$

For given parameters, knowing the industry wedge Λ_b is enough to compute the aggregate labor share.

1.A.6 Hat Algebra

From the main text, we get that the counterfactual wage w'_{io} from (1.28) can be written as: $w'_{io} = \omega_{io} \frac{\widehat{F}_b}{P^{\frac{1}{1+\varepsilon_b \delta}}} {\Phi'_m}^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi'}{L'}\right)$ where we denote by ω_{io} the establishment-occupation component of the counterfactual wage. This variable ω_{io} contains the counterfactual equilibrium wedge λ'_{io}.

Summing $T_{io}(w'_{io})^{\varepsilon_b}$ and factoring out the industry or economy wide constants we find the following relation,

$$\Phi'_m = \widetilde{\Phi'}^{\psi_b}_m \frac{\widehat{F}_b^{\psi_b \varepsilon_b}}{P^{\frac{\psi_b \varepsilon_b}{1+\varepsilon_b \delta}}} \left(\frac{\Phi'}{L'}\right)^{\psi_b \nu_b \varepsilon_b} \quad , \quad \widetilde{\Phi}'_m = \sum_{io \in I_m} T_{io} \omega_{io}^{\varepsilon_b}.$$

Using the definition of $\Phi'_b = \sum_{m \in \mathcal{M}_b} \Phi'_m{}^{\eta/\varepsilon_b} \Gamma_b^{\eta}$, we have that Φ'_b and Φ' are:

$$\Phi'_b = \widetilde{\Phi}'_b \frac{\widehat{F}_b^{\psi_b \eta}}{P^{\frac{\psi_b \eta}{1+\varepsilon_b \delta}}} \left(\frac{\Phi'}{L'}\right)^{\psi_b \nu_b \eta} \quad , \quad \widetilde{\Phi}'_b = \sum_{m \in \mathcal{M}_b} (\widetilde{\Phi}'_m)^{\psi_b \eta / \varepsilon_b}$$

$$\Phi' = (\widetilde{\Phi}')^{1+\eta\delta} P^{-\eta} L'^{-\eta\delta}, \quad \widetilde{\Phi}' = \sum_{b' \in \mathcal{B}} \widetilde{\Phi}'_b \widehat{F}_{b'}^{\psi_{b'} \eta} \Gamma_{b'}^{\eta}.$$

Industry employment in the counterfactual is equal to:

$$L'_b = \frac{\widehat{F}_b^{\psi_b \eta} \widetilde{\Phi}'_b \Gamma_b^{\eta}}{\sum_{b' \in \mathcal{B}} \widehat{F}_{b'}^{\psi_b \eta} \widetilde{\Phi}'_{b'} \Gamma_{b'}^{\eta}} L'.$$

Establishment-occupation output in the counterfactual is:

$$\begin{aligned} y'_{io} &= (F'_b)^{\alpha_b(1+\varepsilon_b \delta)} A_{io} (L'_{io})^{1-\delta} \\ &= PP_b^{\frac{1}{1-\alpha_b}} A_{io} \frac{(F'_b)^{\alpha_b(1+\varepsilon_b \delta)}}{PP_b^{\frac{1}{1-\alpha_b}}} (L'_{io})^{1-\delta} \\ &= \frac{\widehat{F}_b^{\alpha_b(1+\varepsilon_b \delta)}}{PP_b} Z_{io} (L'_{io})^{1-\delta}. \end{aligned}$$

The analogue expression for the baseline is: $y_{io} = \frac{1}{PP_b} Z_{io} L_{io}^{1-\delta}$. Aggregating up to industry b level, the counterfactual industry output Y'_b is ,

$$Y'_b = \frac{\widehat{F}_b^{\alpha_b(1+\varepsilon_b \delta)}}{PP_b} Z_b(s') (L'_b)^{1-\delta}, \quad Z_b(s') \equiv \sum_{io \in \mathcal{I}_b} Z_{io} (s'_{io|m})^{1-\delta} (s'_{mo|b})^{1-\delta}.$$

The analogue expression for the baseline is: $Y_b = \frac{1}{PP_b} Z_b(s) L_b^{1-\delta}$ with $Z_b(s)$ analogue to the one defined for the counterfactual but with baseline employment shares, $Z_b(s) \equiv \sum_{io \in \mathcal{I}_b} Z_{io} s_{io|m}^{1-\delta} s_{m|b}^{1-\delta}$. Taking the ratio, counterfactual industry output relative to the baseline, $\widehat{Y}_b$ is:

$$\widehat{Y}_b = \widehat{F}_b^{\alpha_b(1+\varepsilon_b \delta)} \widehat{Z}_b \widehat{L}_b^{1-\delta}, \tag{41}$$

where $\widehat{Z}_b = \frac{Z_b(s')}{Z_b(s)}$. Using L'_b and equation (1.4) we get,

$$\widehat{F}_b^{\psi_b(1+\eta)} \widehat{Z}_b \left(\frac{\widetilde{\Phi}'_b \Gamma_b^{\eta}}{L_b}\right)^{1-\delta} = \prod_{b' \in \mathcal{B}} \left(\widehat{F}_{b'}^{\alpha_b(1+\varepsilon_b \delta)+(1-\delta)\psi_b \eta}\right)^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \widehat{Z}_{b'}^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \left(\frac{\widetilde{\Phi}'_{b'} \Gamma_{b'}^{\eta}}{L_{b'}}\right)^{(1-\delta)\theta_{b'}}. \tag{42}$$

By taking the ratio, the elasticities θ_b and the economy wide constants cancel out on both side. Rewriting, we get an expression very similar to (1.22) in Proposition 2 with hat variables:

$$\widehat{F}_b = \widehat{X}_b \widehat{C}^{\frac{1}{\psi_b(1+\eta)}}, \tag{43}$$

$$\widehat{X}_b = \left(\frac{L_b^{1-\delta}}{\widehat{Z}_b \left(\widetilde{\Phi'}_b \Gamma_b^{\eta}\right)^{1-\delta}}\right)^{\frac{1}{\psi_b(1+\eta)}} \quad , \quad \widehat{C} = \left(\prod_{b' \in \mathcal{B}} \left(\widehat{X}_{b'}^{-\chi_{b'}}\right)^{\theta_{b'}}\right)^{\frac{1+\eta}{\sum_{b' \in \mathcal{B}} \theta_{b'}(1-\alpha_{b'})(1+\eta\delta)}}.$$

Fixed Labor

In the case where employment is fixed at the industry level b, the counterfactual wage (1.28) becomes:

$$w'_{io} = \left(\beta_b \lambda_{io} \frac{Z_{io}}{T_{io}^{\delta}}\right)^{\frac{1}{1+\varepsilon_b \delta}} \frac{\widehat{F}_b}{P^{\frac{1}{1+\varepsilon_b \delta}}} (\Phi'_m)^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi'_b}{L'_b}\right)^{\nu_b}.$$

Fixing lower levels than b would only change the last element. Keeping total employment at the local labor market fixed, the last term would become: $\left(\frac{\Phi'_m}{L'_m}\right)^{\nu_b}$. The constant Γ_b does not appear in this case as workers can't move across industries and the functional Γ_b is the same for all the local labor markets within an industry. Also, fixing lower levels than b clearly implies that L'_b is known and equal to the baseline labor in the industry L_b.

The counterfactuals where employment at b or lower level employment is fixed will give rise to a condition similar to (42). Given that L'_b is known, we have that:

$$\widehat{F}_b^{1+\varepsilon_b \delta} \widehat{Z}_b = \prod_{b' \in \mathcal{B}} \left(\widehat{F}_{b'}^{\alpha_b (1+\varepsilon_b \delta)} \widehat{Z}_{b'}\right)^{\theta_{b'}}.$$

Propositions 1 and 2 therefore also apply in the relative counterfactuals with fixed labor at the industry level b (or at a lower level).

1.B Extensions

1.B.1 Endogenous Participation

We showed in the proof of Proposition 2 that the solution of transformed prices $\mathbf{F}$ is homogeneous of degree zero with respect to total employment level which we denote here as L_e. We have that,

$$L_{io}(w_{io}) = \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi} L = \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi_e} L_e.$$

We have that $L_e = \frac{\Phi_e}{\Phi} L$ with $\Phi_e = \sum_{m \in \mathcal{I}_m} \Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}$ is the part of Φ that comes from the employed and $\Phi_u = \sum_{uo \in \mathcal{U}_m} (T_{uo} w_{Ro}^{\varepsilon_b})^{\eta/\varepsilon_b} \Gamma_b^{\eta}$ is the part from the out of the labor force as in the main text.

The model aggregation steps are the same as in 1.A with the exception that L_b now is $L_{b,e}$.

Note that the markdown is the same as the TFP of the out-of-the-labor-force workers and is set to 0. From (35),

$$\Phi_{b,e} = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \sum_{m \in \mathcal{M}_b} \widetilde{\Phi}_m^{\psi_b \eta/\varepsilon_b} F_b^{\psi_b \eta} \Gamma_b^{\eta} = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \widetilde{\Phi}_{b,e} F_b^{\psi_b \eta} \tag{44}$$

$$\widetilde{\Phi}_{b,e} = \sum_{m \in \mathcal{M}_b} \widetilde{\Phi}_m^{\psi_b \eta/\varepsilon_b},$$

and,

$$\Phi_e = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \sum_{b \in \mathcal{B}} \widetilde{\Phi}_{b,e} F_b^{\psi_b \eta} \Gamma_b^{\eta} = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \widetilde{\Phi}_e \tag{45}$$

$$\widetilde{\Phi}_e = \sum_{b \in \mathcal{B}} \widetilde{\Phi}_{b,e} F_b^{\psi_b \eta} \Gamma_b^{\eta}.$$

Therefore,

$$L_{b,e} = \frac{\Phi_{b,e}}{\Phi_e} L = \frac{\widetilde{\Phi}_{b,e}}{\widetilde{\Phi}_e} L,$$

where L is total labor supply (employed and out-of-the-labor-force) and we can solve for the prices without knowing total employment level L_e. In order to get that, we need to solve for Φ_e in equation (45),

$$\Phi_e^{\frac{1+\eta\delta}{\eta\delta}} L = (\Phi_e + \Phi_u)\widetilde{\Phi}_e^{\frac{1+\eta\delta}{\eta\delta}}.$$

The solution is obviously unique as the left hand side is convex and the right hand side linear. With the solution for Φ_e one can construct all the aggregates back.

1.B.2 Agglomeration

Plugging the labor supply into (1.31), the wage in the baseline economy is,

$$w_{io} = \left(\beta_b \lambda(\mu_{io}, \varphi_b) \frac{Z_{io}}{(T_{io}\Gamma_b^{\eta})^{\delta}}\right)^{\frac{1}{1+\varepsilon_b\delta}} \Phi_m^{\nu_b - \frac{\eta}{\varepsilon_b}\widetilde{\nu_b}} P^{-\frac{1}{1+\varepsilon_b\delta}} \left(\frac{\Phi}{L}\right)^{\widetilde{\nu_b}}, \quad \nu_b = \frac{\delta}{1+\varepsilon_b\delta}, \quad \widetilde{\nu_b} = \frac{\delta - \gamma}{1+\varepsilon_b\delta}.$$

The baseline wage can be written as: $w_{io} = \widetilde{w}_{io}\Phi_m^{\nu_b - \frac{\eta}{\varepsilon_b}\widetilde{\nu_b}} P^{-\frac{1}{1+\varepsilon_b\delta}} \left(\frac{\Phi}{L}\right)^{\widetilde{\nu_b}}$. Analogously, the counterfactual wage is: $w_{io} = \omega_{io}\widehat{F}_b\Phi_m^{\nu_b - \frac{\eta}{\varepsilon_b}\widetilde{\nu_b}} P^{-\frac{1}{1+\varepsilon_b\delta}} \left(\frac{\Phi}{L}\right)^{\widetilde{\nu_b}}$. Aggregating to generate Φ_m,

$$\Phi_m = \widetilde{\Phi}_m^{\widetilde{\psi_b}} P^{-\frac{\widetilde{\psi_b}\varepsilon_b}{1+\varepsilon_b\delta}} \left(\frac{\Phi}{L}\right)^{\widetilde{\psi_b}\widetilde{\nu_b}\varepsilon_b}, \quad \widetilde{\psi_b} \equiv \frac{1+\varepsilon_b\delta}{1+\eta(\delta-\gamma)}. \tag{46}$$

The counterfactual Φ'_m is analogously $\Phi'_m = (\widetilde{\Phi'}_m)^{\widetilde{\psi_b}} P^{-\frac{\widetilde{\psi_b}\varepsilon_b}{1+\varepsilon_b\delta}} \widehat{F}_b^{\widetilde{\psi_b}\varepsilon_b} \left(\frac{\Phi'}{L}\right)^{\widetilde{\psi_b}\widetilde{\nu_b}\varepsilon_b}$.

In order to be able to find a solution to the model, we need that $\widetilde{\psi_b} < \infty$. This is equivalent to requiring $\gamma \neq \frac{1}{\eta} + \delta$. The parameter γ governs the strength of agglomeration forces within a local labor market, and δ and $\frac{1}{\eta}$ are related with dispersion forces. Those come from the decreasing returns to scale (δ) and from the variance of taste shocks ($\frac{1}{\eta}$). When the latter is high, the mass of workers having extreme taste shocks is higher. This implies that agglomeration forces will impact less as workers would be more inelastic to changes in wages. The standard condition for uniqueness of the equilibrium with agglomeration would be that is sufficiently weak ($\gamma \leq \frac{1}{\eta} + \delta$). In our context we do not find such inequality condition.

The counterfactual industry labor supply is:

$$L'_b = \frac{\widehat{F}_b^{\widetilde{\psi_b}\eta}\widetilde{\Phi}'_b\Gamma_b^{\eta}}{\sum_{b\in\mathcal{B}}\widehat{F}_{b'}^{\widetilde{\psi_b}\eta}\widetilde{\Phi}'_{b'}\Gamma_{b'}^{\eta}}.$$

Turning to production, the establishment-occupation output y'_{io} and local labor market output Y_m in the counterfactual and the baseline are respectively:

$$y'_{io} = \frac{Z_{io}\widehat{F}_b^{\alpha_b(1+\varepsilon_b\delta)}}{P_b P} {L'_{io}}^{1-\delta} {L'_m}^{\gamma}$$

$$Y'_m = \frac{Z_m(s')\widehat{F}_b^{\alpha_b(1+\varepsilon_b\delta)}}{P_b P} {L'_m}^{1-\delta+\gamma}, \quad Z_m(s') = \sum_{i\in\mathcal{I}_m} Z_{io}{s'_{io|m}}^{1-\delta}.$$

The expressions for the baseline are analogous but setting $\widehat{F}_b = 1$. The counterfactual output of industry b, Y'_b, when there are agglomeration forces is:

$$Y'_b = \frac{Z_b(s')\widehat{F}_b^{\alpha_b(1+\varepsilon_b\delta)}}{P_b P} {L'_b}^{1-\delta+\gamma}, \quad Z_b(s') = \sum_{m\in\mathcal{M}_b} Z_m {s'_{mo|b}}^{1-\delta+\gamma},$$

where γ changed the returns to scale of the industry production function and the aggregation of productivities $Z_b(s')$. The intermediate good demand in the counterfactual relative to the baseline is:

$$\begin{aligned}
&\widehat{F}_b^{1+\varepsilon_b\delta}\widehat{Z}_b\left(\frac{L'_b(\widehat{\mathbf{F}})}{L_b}\right)^{1-\delta+\gamma} = \prod_{b'\in\mathcal{B}}\widehat{F}_{b'}^{\alpha_{b'}(1+\varepsilon_b\delta)}\widehat{Z}_{b'}\left(\frac{L'_{b'}(\widehat{\mathbf{F}})}{L_{b'}}\right)^{1-\delta+\gamma}\\
\Leftrightarrow\ &\widehat{F}_b^{\widehat{\varphi_b}(1+\eta)}\widehat{Z}_b\left(\frac{\widetilde{\Phi}'_b\Gamma_b^\eta}{L_b}\right)^{1-\delta+\gamma} = \prod_{b'\in\mathcal{B}}\widehat{F}_{b'}^{\alpha_{b'}(1+\varepsilon_b\delta)+\widetilde{\varphi_b}\eta(1-\delta+\gamma)}\widehat{Z}_{b'}\left(\frac{\widetilde{\Phi}'_{b'}\Gamma_{b'}^\eta}{L_{b'}}\right)^{1-\delta+\gamma}.
\end{aligned}$$

Uniqueness of the solution to this system of equations is guaranteed by $\sum_{b\in\mathcal{B}}\alpha_b\theta_b < 1$. This condition being the same as for Proposition 2, uniqueness of the equilibrium with agglomeration forces only needs the additional requirement of $\gamma \neq \frac{1}{\eta}+\delta$.

1.C Proofs

Proof of Proposition 1.

Existence. We follow closely the proof by Kucheryavyy (2012). Define the right hand side of (1.17) as:

$$f_{io}(\mathbf{w}) = [\lambda(\mu_{io}(\mathbf{w}),\varphi_b)]^{\frac{1}{1+\varepsilon_b\delta}}c_{io}, f_{io}(\mathbf{w}) = [\lambda(\mu(s(\mathbf{w})))]^{\frac{1}{1+\varepsilon_b\delta}}c_{io},$$

where $\mathbf{w}$ denotes the vector formed by $\{w_{io}\}$, we simplified the notation of the wedge $\lambda(\mu_{io},\varphi_b)$ from the main text getting rid of the second argument and $c_{io} = \left(\beta_b\frac{A_{io}}{(T_{io}\Gamma_b^\eta)^\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}\Phi_m^{(1-\eta/\varepsilon_b)\nu_b}\left(\frac{\Phi}{L}\right)^{\nu_b}F_b$ is an establishment-occupation specific parameter. This means we take Φ_m and Φ as constants and not as functions of w_{io}.

Under the assumption $0<\eta<\varepsilon_b$, the function $\mu(s) = \frac{\varepsilon_b(1-s)+\eta s}{\varepsilon_b(1-s)+\eta s+1}$ is decreasing in s, the employment share out of the local labor market. We therefore also have that the wedge $\lambda(\mu(s)) = (1-\varphi_b)\mu(s)+\varphi_b\frac{1}{1-\delta}$ is also decreasing in s. The employment share has bounds $0\le s\le 1$, which implies $(1-\varphi_b)\frac{\eta}{\eta+1}+\varphi_b\frac{1}{1-\delta}\le\lambda(\mu(s))\le(1-\varphi_b)\frac{\varepsilon_b}{\varepsilon_b+1}+\varphi_b\frac{1}{1-\delta}$. Also, $1+\varepsilon_b\delta>0$. Therefore we have that $f_{io}(\mathbf{w})$ is bounded:

$$\left((1-\varphi_b)\frac{\eta}{\eta+1}+\varphi_b\frac{1}{1-\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}c_{io}\le f_i(\mathbf{w})\le\left((1-\varphi_b)\frac{\varepsilon_b}{\varepsilon_b+1}+\varphi_b\frac{1}{1-\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}c_{io}.$$

If the number of participants in sub-market m is N_m, we can define the compact set S where $f_{io}(\mathbf{w})$ maps into itself as:

$$\begin{aligned}
S = &\left[\left((1-\varphi_b)\frac{\eta}{\eta+1}+\varphi_b\frac{1}{1-\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}c_1,\left((1-\varphi_b)\frac{\varepsilon_b}{\varepsilon_b+1}+\varphi_b\frac{1}{1-\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}c_1\right]\times\ldots\\
&\times\left[\left((1-\varphi_b)\frac{\eta}{\eta+1}+\varphi_b\frac{1}{1-\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}c_{N_m},\left((1-\varphi_b)\frac{\varepsilon_b}{\varepsilon_b+1}+\varphi_b\frac{1}{1-\delta}\right)^{\frac{1}{1+\varepsilon_b\delta}}c_{N_m}\right].
\end{aligned}$$

The function $f_{io}(\mathbf{w})$ is continuous in wages on S. We can therefore apply Brouwer's fixed point theorem and claim that at least one solution exists for the system of equations formed by (1.19). □

Uniqueness. First we introduce the following Theorem and Corollary that we will use later to establish uniqueness in our proofs. These are just transcribed from Allen et al. (2016):

Theorem 1. *Consider $g:\mathbb{R}_{++}^n\times\mathbb{R}_{++}^m$ for some $n\in\{1,...,N\}$ and $m\in\{1,...,M\}$ such that:*

1. *homogeneity of any degree: $g(tx,ty) = t^k g(x,y)$, $t\in\mathbb{R}_{++}$ and $k\in\mathbb{R}$,*

2. *gross-substitution property:* $\frac{\partial g_i}{\partial x_j} > 0$ *for all* $i \neq j$,

3. *monotonicity with respect to the joint variable:* $\frac{\partial g_i}{\partial y_k} \geq 0$, *for all* i, k.

Then, for any given $y^0 \in \mathbb{R}^M_{++}$ *there exists at most one solution satisfying* $g(x, y^0) = 0$.

Proof. See the proof for Theorem 5 in Allen et al. (2016). □

Corollary 1. *Assume (i)* $f(x)$ *satisfies gross-substitution and (ii)* $f(x)$ *can be decomposed as* $f(x) = \sum_{j=1}^{\nu_f} g^j(x) - \sum_{k=1}^{\nu_g} h^k(x)$ *where* $g^j(x), h^k(x)$ *are non-negative vector functions and, respectively, homogeneous of degree* α_j *and* β_k, $\bar{\alpha} = \max \alpha_j \leq \min \beta_k$.

1. *Then there is at most one up-to-scale solution of* $f(x) = 0$.

2. *In particular, if for some* j, k $\alpha_j \neq \beta_k$, *then there is at most one solution.*

Proof. See the proof for Corollary 1 in Allen et al. (2016). □

In order to prove uniqueness we use Theorem 1 and Corollary 1 stated above.

Define the function $g : \mathbb{R}^n_{++} \to \mathbb{R}^n$ for some $n \in \{1, ..., N\}$ as:

$$g_{io}(\mathbf{w}) = f_{io}(\mathbf{w}) - w_{io}, \quad \forall i \in \{1, .., N\}.$$

We want to prove that the solution satisfying $g(\mathbf{w}) = 0$ is unique. In order to do so, we first need to show that $g(\mathbf{w})$ satisfies the gross substitution property ($\frac{\partial g_{io}}{\partial w_{jo}} > 0$ for any $j \neq i$).

Taking the partial derivative of g_{io} with respect to w_{jo} for any $j \neq i$:

$$\frac{\partial g_{io}}{\partial w_{jo}} = \frac{\partial f_{io}(\mathbf{w})}{\partial \lambda(\mu(s(\mathbf{w})))} \times \frac{\partial \lambda(\mu(s_{io|m}))}{\partial \mu(s_{io|m})} \times \frac{\partial \mu(s_{io|m})}{\partial s_{io|m}} \times \frac{\partial s_{io|m}}{\partial w_{jo}},$$

where $\frac{\partial f_{io}(\mathbf{w})}{\partial \lambda(\mu(s(\mathbf{w})))} = \frac{1}{1+\varepsilon_b \delta} \frac{f_{io}(\mathbf{w})}{\lambda(\mu(s(\mathbf{w})))} > 0$. We have that $\frac{\partial \lambda(\mu(s_{io|m}))}{\partial \mu(s_{io|m})} > 0$ and we previously established that, under the assumption that $0 < \eta < \varepsilon_b$, $\frac{\partial \mu(s_{io|m})}{\partial s_{io|m}} < 0$. The share of an establishment i with occupation o in sub-market mo is defined as:

$$s_{io|m} = \frac{T_{io} w_{io}^{\varepsilon_b}}{\sum_{j \in \mathcal{I}_m} T_{jo} w_{jo}^{\varepsilon_b}}.$$

Clearly, $\frac{\partial s_{io|m}}{\partial w_{jo}} < 0$ for any $i \neq j$. Therefore $\frac{\partial g_{io}}{\partial w_{jo}} > 0$ for any $i \neq j$ and g satisfies the gross-substitution property.

The remaining condition to use Corollary 1 is simply that $f_{io}(\mathbf{w})$ is homogeneous of a degree smaller than 1.[56] Clearly, $f_{io}(\mathbf{w})$ is homogeneous of degree 0 as a consequence that the markdown function itself $\mu(s_{io|m})$ is homogeneous of degree 0. Therefore, the function g satisfies the conditions of Corollary 1 and we can conclude that there exists at most one solution satisfying $g(\mathbf{w}) = 0$. □

Proof of Proposition 2.

Developing equation (1.21) we get

$$F_b^{1+\varepsilon_b \delta} A_b \left(\frac{F_b^{\psi_b \eta} \widetilde{\Phi}_b \Gamma_b^{\eta}}{\widetilde{\Phi}} L \right)^{1-\delta} = \theta_b \prod_{b' \in \mathcal{B}} \left(F_{b'}^{\alpha_{b'}(1+\varepsilon_b \delta)} \right)^{\theta_{b'}} \prod_{b' \in \mathcal{B}} A_{b'}^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \left(\frac{F_{b'}^{\psi_b \eta} \widetilde{\Phi}_{b'} \Gamma_{b'}^{\eta}}{\widetilde{\Phi}} L \right)^{(1-\delta)\theta_{b'}}$$

$$\Leftrightarrow F_b^{\psi_b(1+\eta)} A_b \left(\widetilde{\Phi}_b \Gamma_b^{\eta} \right)^{1-\delta} = \theta_b \prod_{b' \in \mathcal{B}} \left(A_{b'} \left(\widetilde{\Phi}_{b'} \Gamma_{b'}^{\eta} \right)^{1-\delta} \right)^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \left(F_{b'}^{\alpha_{b'}(1+\varepsilon_{b'}\delta)+\psi_{b'}\eta(1-\delta)} \right)^{\theta_{b'}}.$$

[56] The degree of homogeneity of $h_{io}(\mathbf{w}) = w_{io}$ is 1.

Define $f_b = \log(F_b)$ and $\mathbf{f}$ as a $B \times 1$ vector whose element b' is $f_{b'}$. Then, taking logs and rearranging the previous expression we obtain:

$$f_b = C_b + \mathbf{d}'\mathbf{f},$$

where

$$C_b = \frac{1}{\psi_b(1+\eta)}\left[\log(\theta_b) - \log(A_b) - (1-\delta)\log(\widetilde{\Phi}_b\Gamma_b^{\eta}) + \sum_{b'\in\mathcal{B}}\theta_{b'}\left(\log(A_{b'}) + (1-\delta)\log(\widetilde{\Phi}_{b'}\Gamma_{b'}^{\eta})\right)\right]$$

and $\mathbf{d}$ is a $B \times 1$ vector whose b' element $\mathbf{d}_{b'}$ is:

$$\begin{aligned}\mathbf{d}_{b'} &= \frac{1}{\psi_{b'}(1+\eta)}\left(\alpha_{b'}(1+\varepsilon_{b'}\delta) + \psi_{b'}\eta(1-\delta)\right)\theta_{b'} \\ &= \frac{\theta_{b'}}{1+\eta}\left(\alpha_{b'}(1+\eta\delta) + \eta(1-\delta)\right).\end{aligned}$$

Define the vector $\mathbf{C} = [C_1, ..., C_b, ..., C_B]$ that contains the constant terms and the matrix $\mathbf{D} = [d, ..., d]$ which repeats the $\mathbf{d}$ vector B times. We can stack all the terms for all $b \in \mathcal{B}$ from the previous expression and obtain the following system of equations:

$$\mathbf{f} = \mathbf{C} + \mathbf{D}'\mathbf{f}. \tag{47}$$

A solution to the system (47) exists if the matrix $\mathbf{I} - \mathbf{D}'$ is invertible. This matrix has an eigenvalue of zero if and only if the sum of the elements of the vector $\mathbf{d}$ is equal to 1. Additionally, this sum is equal to 1 if and only if $\sum_b \alpha_b\theta_b = 1$ as:

$$\begin{aligned}\sum_b \mathbf{d}_b = 1 &\Leftrightarrow \sum_b \left(\alpha_b(1+\eta\delta) + \eta(1-\delta)\right)\theta_b = 1+\eta \\ \Leftrightarrow \sum_b \alpha_b\theta_b(1+\eta\delta) = 1+\eta-\eta(1-\delta) &\Leftrightarrow \sum_b \alpha_b\theta_b = \frac{1+\eta-\eta(1-\delta)}{1+\eta\delta} \Leftrightarrow \sum_b \alpha_b\theta_b = 1.\end{aligned}$$

Therefore we can conclude that whenever $\sum_b \alpha_b\theta_b \neq 1$ the transformed prices $\mathbf{F}$ have a unique solution. This is always the case as long as $0 \le \beta_b, \theta_b < 1\ \forall b \in \mathcal{B}$ and $0 \le \delta \le 1$.

In order to obtain the closed form solution, rewrite (1.21) as:

$$F_b = \left(\frac{\theta_b}{A_b\left(\widetilde{\Phi}_b\Gamma_b^{\eta}\right)^{(1-\delta)}}\right)^{\frac{1}{\psi_b(1+\eta)}} C^{\frac{1}{\psi_b(1+\eta)}} = X_b C^{\frac{1}{\psi_b(1+\eta)}},$$

where C is a constant that is equal to:

$$C = \prod_{b'\in\mathcal{B}}\left(A_{b'}\left(\widetilde{\Phi}_{b'}\Gamma_{b'}^{\eta}\right)^{1-\delta}\right)^{\theta_{b'}} \prod_{b'\in\mathcal{B}}\left(F_{b'}^{\alpha_{b'}(1+\varepsilon_{b'}\delta)+\psi_{b'}\eta(1-\delta)}\right)^{\theta_{b'}}.$$

To solve for the constant, we use the ideal price index equation substituting the relative prices P_b for the transformed prices F_b:

$$1 = \prod_{b\in\mathcal{B}}\left(\frac{F_b^{\chi_b}}{\theta_b}\right)^{\theta_b}.$$

Substituting F_b into the price index and solving for C we recover the expression showed in Proposition 2. □

1.D Identification Details

1.D.1 Identification of η and δ

In order to identify the across markets labor supply elasticity η and the labor demand elasticity δ we exploit the fact that in local labor markets where there is only one establishment, the wedge $\lambda(\mu, \phi_b)$ is constant within industries b. We denominate this type of establishments as *full monopsonists*. Additionally, the effect of wages on the labor supply of full monopsonists is only affected by the parameter η as the within market labor supply elasticity ε_b is irrelevant in local labor markets with only one establishment. Taking the logarithm for the labor supply full monopsonists face (1.13) we get:

$$\ln(L_{io,s=1}) = \eta \ln(w_{io}) + \frac{\eta}{\varepsilon_b} \ln(T_{io}) + \ln(\Gamma_b^{\eta} L/\Phi).$$

As mentioned before, full monopsonists apply a constant markdown equal to $\mu(s=1) = \frac{\eta}{\eta+1}$ that in turn will imply a constant wedge $\lambda(\mu, \phi_b)$ within industry b. Their (inverse) labor demand (1.16) in logs is:

$$\ln(w_{io,s=1}) = \ln(\beta_b) + \ln(\frac{\eta}{\eta+1}) + \ln(A_{io}) - \delta \ln(L_{io}) + \frac{1}{1-\alpha_b} \ln(P_b).$$

In order to get rid of industry and economy wide constants, we demean $\ln(L_{io,s=1})$ and $\ln(w_{io,s=1})$ by removing the industry b averages per year. Denoting with $\overline{\ln(X)}$ the demeaned variables, we rewrite the labor supply and (inverse) demand equations as:

$$\begin{aligned} \overline{\ln(L_{io})} &= \eta \overline{\ln(w_{io})} + \frac{\eta}{\varepsilon_b} \overline{\ln(T_{io})}, \\ \overline{\ln(w_{io})} &= -\delta \overline{\ln(L_{io})} + \overline{\ln(A_{io})}. \end{aligned} \tag{48}$$

The above system is just a traditional demand and supply setting. As it is well known, the above system is under-identified. It is the classic textbook example of when a regression model suffers from simultaneity bias. The reason for this under-identification is the following: while the variance-covariance matrix of $\left(\overline{\ln(L_{io})}, \overline{\ln(w_{io})}\right)$ gives us three objects from the data, the system above has five unknowns, which are the elasticities, η and δ, plus the three components of the variance-covariance matrix of the structural errors $\frac{\eta}{\varepsilon_b} \overline{\ln(T_{io})}$ and $\overline{\ln(A_{io})}$. Therefore, in absence of valid instruments that would exogenously vary either the supply or demand equations in (48) we can not identify the elasticities.[57]

In order to identify the elasticities using the labor supply and demand equations in (48), we impose restrictions on the variance-covariance matrix of the structural errors while exploiting the differences in the variance-covariance matrix of the employment and wages across occupations. This way of achieving identification is known in the literature as *identification through heteroskedasticity* (see Rigobon (2003)). We classify our four occupations into two broader categories $S \in \{1, 2\}$. Our identification assumption is that the covariance between the transformed productivity $\overline{\ln(A_{io})}$ and amenities $\frac{\eta}{\varepsilon_b} \overline{\ln(T_{io})}$, that we denote σ_{TA} is constant within each category S. The fact that the elasticities are the same across occupational groups, in addition to the assumption of common covariance of the structural errors within broad categories, are the reason we can achieve identification. The reason is simple: while the four occupational categories give us $3 \times 4 = 12$ bits of information, the unknowns to be identified are 2, δ and η, plus 2, the broad category covariances, plus 8, the

[57] Also note that even if we would have available some valid instruments, we would only be able to identify δ and η but not ε_b.

variances of the transformed productivities and amenities for each of the four occupational categories.[58]

We can rewrite the system (48) in the following way:

$$\frac{\eta}{\varepsilon_b}\,\overline{\ln(T_{io})} = \overline{\ln(L_{io})} - \eta\,\overline{\ln(w_{io})},$$
$$\overline{\ln(A_{io})} = \delta\,\overline{\ln(L_{io})} + \overline{\ln(w_{io})}. \tag{49}$$

Denote the covariance matrix of the structural errors for occupation o in category S (meaning the left hand side of system (49)) by Ψ_{oS}. Denote the covariance matrix between employment and wages of the full monopsonists by $\widehat{V}_{oS}$. The covariance of system (49) writes as:

$$\Psi_{oS} = D\widehat{V}_{oS}D^T, \quad D = \begin{pmatrix} 1 & -\eta \\ \delta & 1 \end{pmatrix},$$

where T denotes the transpose. Formally, our identifying assumption is that $\sigma_{AT,oS} = \sigma_{AT,o'S}$ for occupations that belong to the same category S. Taking differences within category,

$$\Delta_S \equiv \Psi_{oS} - \Psi_{o'S} = D[\widehat{V}_{oS} - \widehat{V}_{o'S}]D^T, \quad \forall S \in \{1,2\}$$

where the differences of covariances in the left hand (element $\Delta_{S,[1,2]}$) is equal to zero. This gives us a just identified system (two equations with two unknowns) to find the parameters η and δ. More details are provided in Appendix 1.D.

The system (49) in matrix form is $\Omega_{oS} = D\widehat{V}_{oS}D^T$.. Defining an auxiliary parameter $\widetilde{\delta} = -\delta$, the system writes as:

$$\begin{pmatrix} (\frac{\eta}{\varepsilon_b})^2\sigma^2_{T,oS} & \frac{\eta}{\varepsilon_b}\sigma_{TA,S} \\ \frac{\eta}{\varepsilon_b}\sigma_{TA,S} & \sigma^2_{A,oS} \end{pmatrix} = \begin{pmatrix} 1 & -\eta \\ -\widetilde{\delta} & 1 \end{pmatrix} \begin{pmatrix} \sigma^2_{L,oS} & \sigma_{LW,oS} \\ \sigma_{LW,oS} & \sigma^2_{W,oS} \end{pmatrix} \begin{pmatrix} 1 & -\widetilde{\delta} \\ -\eta & 1 \end{pmatrix}$$

This system only allows us to identify η and δ. Denote by $\Omega_S \equiv \widehat{V}_{oS} - \widehat{V}_{o'S}$ the difference between the variance covariance matrix within category S and by $\Omega_{S,[1,2]} = \omega_{12,S}$ the element on first row and second column. The system of differences is:

$$\Delta_S = D\Omega_S D^T, \quad \forall S \in \{1,2\}$$

With the identification assumption of equal covariance within category, we have that:

$$\Delta_{S,[1,2]} = 0 = -\eta\omega_{22,S} + (1+\eta\widetilde{\delta})\omega_{12,S} - \widehat{\delta}\omega_{11,S}.$$

Solving for η,

$$\eta = \frac{\omega_{12,S} - \widetilde{\delta}\omega_{11,S}}{\omega_{22,S} - \widetilde{\delta}\omega_{12,S}}, \quad \forall S \in \{1,2\}$$

Equalizing the above across both occupation categories we get a quadratic equation in $\widehat{\delta}$ that solves:

$$\widetilde{\delta}^2[\omega_{11,1}\omega_{12,2} - \omega_{11,2}\omega_{12,1}] - \widetilde{\delta}[\omega_{11,1}\omega_{22,2} - \omega_{11,2}\omega_{22,1}] + \omega_{12,1}\omega_{22,2} - \omega_{12,2}\omega_{22,1} = 0. \tag{50}$$

[58] Of course we could have a more stringent identification assumption that would leave us with an overidentified system, for example, that all covariances are equal to zero. As an additional exercise we also estimated the parameters following a different identification strategy: we assume that the covariances of the structural errors were the same among all the occupational groups. This gives us a system with one overidentification restriction. The point estimates using this assumption and the one we mentioned above are pretty similar.

This is the same system as the simple case without covariance between the fundamental shocks in Rigobon (2003). Different to him, Ω_S is not directly the estimated variance-covariance matrix of each of the 4 occupations but rather the matrix of differences within category or state S. As mentioned by Rigobon (2003) there are two solutions to the previous equation. One can show that if $\widetilde{\delta}^*$ and η^* are a solution then the other solution is equal to $\widetilde{\delta} = 1/\eta^*$ and $\eta = 1/\widetilde{\delta}^*$. This means that the solutions are actually the two possible ways the original structural system (48) can be written. In order to identify which of the two possible solutions we are identifying, we have that by assumption η is positive while $\widetilde{\delta}$ is negative. Therefore as long as the two possible solutions for $\widetilde{\delta}$ have different signs, we just need to pick the negative one.

Given the identification strategy, in order to estimate the elasticities δ and η we just need to obtain the employment and wages covariance matrices directly from the data on establishments that are full monopsonists and solve for (50).

1.D.2 Identification of φ_b

In order to identify the industry workers bargaining power, we need to construct the model counterparts of the industry labor share at every period t:

$$LS^M_{bt}(\varphi_b) = \frac{\beta_b \sum_{io \in \mathcal{I}_b} w_{iot} L_{iot}}{\sum_{io \in \mathcal{I}_b} w_{iot} L_{iot} / \lambda(\mu_{io}, \varphi_b)},$$

$\mathcal{I}_b$ being the set of all establishment-occupations that belong to 2-digit industry b. We target the average across time industry labor share. That is, we pick ϕ_b such that:

$$\mathbb{E}_t \left[LS^M_{bt}(\varphi_b) - LS^D_{bt} \right] = 0. \tag{51}$$

Given that the wedge $\lambda(\mu_{io}, \varphi_b)$ is increasing in φ_b, then $LS^M_{bt}(\varphi_b)$ is increasing in φ_b as well. Therefore, if a solution exists for (51) with $\varphi_b \in [0, 1]$ this has to be unique.[59]

1.D.3 Additional Results

Table 15 has the calibrated final good production function elasticities of the intermediate the $\{\theta_b\}_{b=1}^{\mathcal{B}}$ and the required rate $\{R_b\}_{b=1}^{\mathcal{B}}$ for the year 2007.

1.E Data Details

We provide additional summary statistics and details about sample selection and variable construction.

Table 14 – Industry Estimates

Industry Code	Industry Name	$\widehat{\beta}_b$	$\widehat{\varepsilon}_b$	$\widehat{\varphi}_b$
15	Food	0.74	1.69	0.22
17	Textile	0.74	1.49	0.51
18	Clothing	0.84	1.41	0.31
19	Leather	0.85	2.09	0.26
20	Wood	0.77	1.51	0.42
21	Paper	0.61	3.06	0.55
22	Printing	0.84	1.52	0.18
24	Chemical	0.67	3.25	0.06
25	Plastic	0.73	2.51	0.35
26	Other Minerals	0.65	1.62	0.43
27	Metallurgy	0.61	3.77	0.59
28	Metals	0.81	1.22	0.38
29	Machines and Equipments	0.79	2.18	0.32
30	Office Machinery	0.81	3.33	0.20
31	Electrical Equipment	0.65	3.02	0.67
32	Telecommunications	0.62	3.54	0.73
33	Optical Equipment	0.75	1.91	0.45
34	Transport	0.57	4.05	0.69
35	Other Transport	0.72	3.49	0.44
36	Furniture	0.81	1.57	0.43

Notes: All the estimated parameters are 2-digit industry specific. $\widehat{\beta}_b$ are the estimated output elasticities with respect of labor, $\widehat{\varepsilon}_b$ are the within local labor market elasticities and $\widehat{\varphi}_b$ are union bargaining powers.

Table 15 – Calibrated $\{\theta_b\}$ and $\{R_b\}$

Industry Code	Industry Name	θ_b	R_b
15	Food	0.13	0.11
17	Textile	0.02	0.14
18	Clothing	0.01	0.14
19	Leather	0.01	0.14
20	Wood	0.02	0.13
21	Paper	0.02	0.13
22	Printing	0.06	0.13
24	Chemical	0.14	0.16
25	Plastic	0.06	0.15
26	Other Minerals	0.05	0.15
27	Metallurgy	0.03	0.14
28	Metals	0.10	0.14
29	Machines and Equipments	0.09	0.17
30	Office Machinery	0.00	0.17
31	Electrical Equipment	0.04	0.23
32	Telecommunications	0.04	0.23
33	Optical Equipment	0.04	0.23
34	Transport	0.04	0.19
35	Other Transport	0.06	0.19
36	Furniture	0.03	0.14

Notes: All the calibrated parameters are 2-digit industry specific for the year 2007. θ_b are the intermediate good elasticities in the final good production function and R_b are the capital rental rates for 2007. We construct the rental rates following Barkai (2016).

1.E.1 Additional Summary Statistics

Table 18 – CZ Summary Statistics. Baseline Year

Variable	Obs.	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
N_n	356	773.798	266.8	461	861.2	1,168.407
L_n	356	8,300.567	2,567.403	5,244.300	10,086.210	11,322.000
$\overline{L}_n$	356	11.389	8.148	10.878	13.547	6.043
$\overline{w}_n$	356	34.399	32.707	34.161	35.593	3.242

Note: N_n is the number of establishments at the CZ, L_n is full time equivalent employment at CZ, $\overline{L}_n$ is the average L_{iot} of establishment-occupations at n, $\overline{w}_n$ is the mean w_{iot} of the establishment-occupations at n in thousands of constant 2015 euros.

[59] It can be the case that the solution does not exist. For example, given values of β_b, ε_b and η, even with $\varphi_b = 1$ the labor share generated by the model is too small to the one in the data. This does not happen with our data.

Table 16 – Estimated Within Elasticities for Different Lags

Industry Code	Industry Name	1 Lag $\widehat{\varepsilon}_b$	2 Lags $\widehat{\varepsilon}_b$
15	Food	1.69	1.99
17	Textile	1.49	1.83
18	Clothing	1.41	1.69
19	Leather	2.09	2.50
20	Wood	1.51	1.77
21	Paper	3.06	3.39
22	Printing	1.52	1.79
24	Chemical	3.25	3.56
25	Plastic	2.51	3.04
26	Other Minerals	1.62	1.77
27	Metallurgy	3.77	4.35
28	Metals	1.22	1.48
29	Machines and Equipments	2.18	2.63
30	Office Machinery	3.33	3.72
31	Electrical Equipment	3.02	3.61
32	Telecommunications	3.54	4.08
33	Optical Equipment	1.91	2.36
34	Transport	4.05	4.56
35	Other Transport	3.49	4.05
36	Furniture	1.57	1.90

Notes: All the estimated parameters are 2-digit industry specific. *1 Lag* $\widehat{\varepsilon}_b$ are the estimated within local labor market elasticities when we instrument for the wages with one lag and *1 Lag* $\widehat{\varepsilon}_b$ present the analogous when we instrument with two lags.

1.E.2 Sample Selection

Ficus. This data source comes from tax records therefore we observe yearly firm information. We exclude the source tables belonging to public firms.[60] Before 2000 we take table sources in euros and from 2001 onward we use data from consolidated economic units.[61] After excluding firms without firm identifier the raw data sample contains about 29 million firms from which about 2.8 million are manufacturing firms.[62] Manufacturing sector (sector code equal to *D*) constitutes on average 10% of the observations, 19.2% of value added and 27.2% of employment.

[60]We only use the Financial units (*FIN*) and Other units (*TAB*) tables and exclude Public administration (*APU*).

[61]The profiling of big groups consolidates legal units into economic units. In 2001 the Peugeot-Citroën PSA was treated, Renault in 2003 and the group Accor in 2005. This implies the definition of new economic entities and would therefore lead to the creation of new firm identifiers. Given the potential impact of big establishments in local labor markets we opted to maintain them.

[62]We consider a missing firm identifier (SIREN) also if the identifier equals to zero for all the 9 digits.

Table 17 – Transition Probabilities

Occup. Ch.	CZ Ch.	Ind. Ch.	Trans. Prob. FTE	Trans. Prob.
0	0	0	91.39	91.01
0	0	1	2.37	2.36
0	1	0	0.02	0.02
1	0	0	6.03	6.40
1	0	1	0.20	0.21
1	1	0	0.00	0.00
1	1	1	0.00	0.00

Note: The transition rates are computed over the whole sample period 1994-2007. *Occup. Ch.* is an indicator function of occupational change, *CZ. Ch.* is an indicator function of commuting zone change, *Ind. Ch.* is an indicator function of 3-digit industry change, *Trans. Prob. FTE* are the unconditional transition probabilities based on full time equivalent units and *Trans. Prob.* are the unconditional transition probabilities based on counts of working spells independently of duration and part-time status.

Postes. *DADS Postes* covers all the employment spells of a salaried employee per year. If a worker has several spells in a year we would have multiple observations. The main benefit of this employer-employee data source is that we can know the establishment and employment location of the workers. We exclude workers in establishments with fictitious identifiers (SIREN starting by F) and in public firms. For every establishment identifier (SIRET) we sum the wage bill and the full time equivalent number of employees.

Merged Data. After merging both data sources we finish with data with yearly establishment observations. After the filters and merging the sample consists of 1.3 million firms and 1.6 million establishment observations. In the process of filtering and merging about half of the original firms are lost. Wages and value added are deflated using the Consumer Price Index.[63]

Labor and wage data coming from the balance sheets (at the firm level) and the one from employee records needs to be consolidated. In order to be consistent with balance sheet information we assign labor and employment coming from *FICUS* to the establishments according to their respective shares. We proceed in several steps. First, we filter out observations with no wage or employment information from *Postes* from firms present at different commuting zones. Second, we do some additional cleaning by getting rid of observations with no labor, capital and wage bill information coming from *FICUS* and also observations with non existing or missing commuting zone. Third, we aggregate employee data to the firm times commuting zone level.[64] Then we compute the labor and wage shares of these entities out of the firm's aggregates. What we call establishment through out the text is the entity aggregated at the commuting zone level. Finally, we

[63]Nominal variables are expressed in constant 2015 euros.

[64]Data from 1994 and 1995 do not have commuting zone information. We therefore impute it using correspondence tables between city code and commuting zone. A city code has 5 digits coming from the department and city. Some commuting zone codes beyond the 2 missing years were modified or cleaned. City codes (*commune* codes) of Paris, Marseille and Lyon were divided into different *arrondissements*. We assign them codes 75056, 13055 and 69123 respectively. Then we proceed to the cleaning of commuting zones by assigning to the non existing codes the one corresponding to the city where the establishment is located. We get rid of non matched or missing commuting zone codes. We aggregate the data coming from *Postes* at the commuting zone level after this cleaning.

split firm data from the balance sheet according to those shares. This procedure leaves the firms in a unique commuting zone with their balance sheet data but allows to split wage bill and employment data coming from the balance sheet for multi-location firms. Establishment wage is simply the average wage. That is, wage bill over total full time equivalent employees.

We further exclude Tobacco (2-digits industry code 16), Refineries & Nuclear industry (code 23) and Recycling (code 37). We finally get rid of the outliers reducing the sample 1.5% and finish with 4,156,754 establishment-occupation-year observations that belong to 1.25 million firms.[65]

1.E.3 Variable Construction

Ficus:

- Value added: value added net of taxes (*VACBF*). We restrict to firms with strictly positive value added.[66]
- Capital: tangible and intangible capital without counting depreciation. It is the sum of the variables *IMMOCOR* and *IMMOINC*.
- Employment: full time equivalent employment at the firm (*EFFSALM*).
- Wage bill: gross total wage bills. Is the sum of wages (*SALTRAI*) and firm taxed (*CHARSOC*).[67]
- Industry: industry classification comes from *APE*. The sub-industries h are 3 digit industries and industries b are at two digits.

Postes:

- Occupation: original occupation categories come from the two digit occupations (*CS2*). We group occupations with first digits 2 and 3 into a unique occupation group.[68] This regrouping is done to avoid substantial changes in occupation groups between 1994 and 2007. Observations with missing occupation information are excluded.
- Employment: full time equivalent at the establishment-occupation level (*etp*).
- Wage: is the gross wage (per year) of individual worker (*sbrut*). If the spell is less than a year is the gross wage corresponding to the spell.
- Commuting zone: depending on the year, the commuting zone classification is taken from the variable *zemp* or *zempt*. Commuting zone information is missing for the years 1994 and 1995 and is imputed using the city codes.[69]

1.E.4 Construction of Required Rates

In order to construct the required rates for the different sectors we follow the methodology proposed by Barkai (2016) using the Capital Input Data from the EU KLEMS database, December 2016 revision. In this

[65]We get rid of wage per capita outliers by truncating the sample at the 0.5% below and 99.5%.
[66]We follow the advise of the French statistical instiute (INSEEE) in using net value added to perform comparisons across industries.
[67]For firms declaring at the BIC-BRN regime (*TYPIMPO*= 1) we only take *SALTRAI* .
[68]Occupations with first digit 1 and 7 are excluded. They constituted less than 0.05% of the matched sample.
[69]City codes are the concatenation of department (*DEP*) and city (*COM*).

dataset one can find, for a given industry, different depreciation rates and price indices for different types of capital. The types of capital that are present in the manufacturing sector are: Computing Equipment, Communications Equipment, Computer Software and Databases, Transport Equipment, Buildings and structures (non-residential), and Research and Development. We construct a required rate for each of the industries at the 2 digit level defined in the NAF classification. However, unlike the NAF classification, that we have up to 20 different industries, there are only 11 industries classified within manufacturing within the EU KLEMS database. Any industry classification in EU KLEMS is just an aggregation of the 2 digit industry classification in NAF. Therefore there are industries within the NAF classification that will have the same required rate of return on capital.

For a type of capital s and sector b, we define the the required rate of return R_{sb} as:

$$R_{sb} = \left(i^D - \mathbb{E}\left[\pi_{sb}\right] + \delta_{sb} \right),$$

where i^D is a the cost fo debt borrowing in financial markets, and π_{sb} and δ_{sb} are, respectively, the inflation and depreciation rates of capital type s in sector b.

Then we define the total expenditures on capital type s in sector b as:

$$E_{sb} = R_{sb} P^K_{sb} K_{sb},$$

where $P^K_{sb} K_{sb}$ is the nominal value of capital stock of type s. Summing over all types of capital within a sector we can obtain the total expenditures of capital of such sector:

$$E_b = \sum_{sb} R_{sb} P^K_{sb} K_{sb}.$$

Multiplying and dividing by the total nominal value of capital stock we obtain the following decomposition:

$$\sum_s R_{sb} P^K_{sb} K_{sb} = \underbrace{\sum_s \frac{P^K_{sb} K_{sb}}{\sum_{s'} P^K_{s'b} K_{s'b}} R_{sb}}_{R_b} \underbrace{\sum_s P^K_{sb} K_{sb}}_{P^{Kb} K_b},$$

where the first term R_b is the interest rate that we use in the model.

1.E.5 Amenities

In order to preform some counterfactuals we still need to compute other policy invariant parameters, or fundamentals, from the data. In particular we need to recover establishment-occupation amenities and TFPRs, while ensuring that in equilibrium the wages and labor allocations are the same as in the data.

Using the establishments labor supply (1.13), we can back out amenities, up to a constant:

$$T_{io} = \frac{s_{io|m}}{w_{io}^{\varepsilon_b}} \Phi_m.$$

The sub-market level Φ_m is a function of the amenities of all plants in m. We proceed by normalizing one particular local labor market. Note that the allocation of resources is independent from this normalization. We denote the local labor market that we normalize as 1. The relative employment share of market m with respect to the normalized one is: $\frac{L_m}{L_1} = \frac{\Phi_m^{\eta/\varepsilon_b}}{\Phi_m^{\eta/\varepsilon_{b'}}} \frac{\Gamma_b}{\Gamma_1}$. The local labor market aggregate is then:

$$\Phi_m = \left(\frac{L_m}{L_1} \frac{\Gamma_1}{\Gamma_b} \Phi_1^{\frac{\eta}{\varepsilon_{b'}}} \right)^{\frac{\varepsilon_b}{\eta}}$$

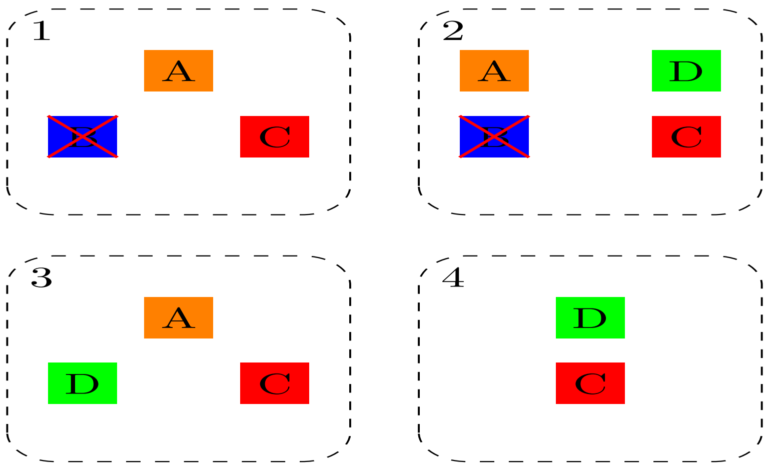
Substituting into the above we have that:

$$T_{io} \propto \frac{s_{io|m}}{w_{io}^{\varepsilon_b}} \left(\frac{L_m}{\Gamma_b}\right)^{\varepsilon_b/\eta}.$$

1.F Empirical Evidence

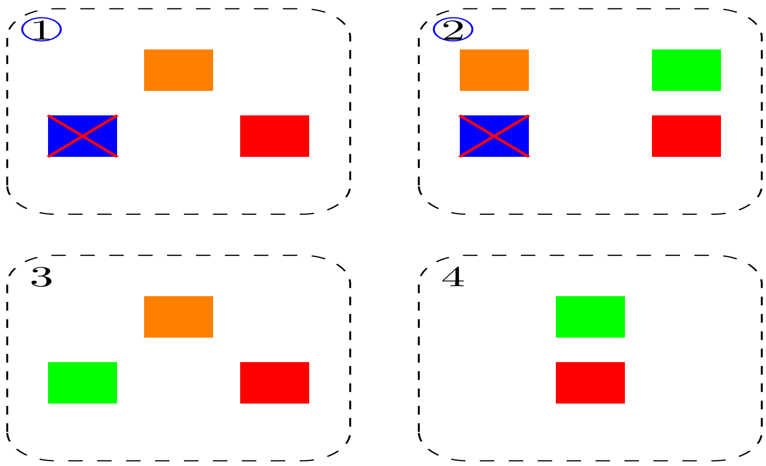
Example of an economy with four local labor markets and four firms identified by color. Each firm is multi-location with plants at different local labor markets. The blue firm is affected by a mass layoff at the national level (in all the local markets where it is present). Natural experiment on $s_{io|m}$ for non-blue establishments.

Figure 6 – Local Labor Markets with and without shock



The treated establishments are the ones in local markets 1 and 2.

Figure 7 – Treated Establishments



The first order condition for wages is:

$$P_b \frac{\partial F}{\partial L_{io}} = L_{io}(w_{io}) \frac{\partial w_{io}}{\partial L_{io}} + w_{io},$$

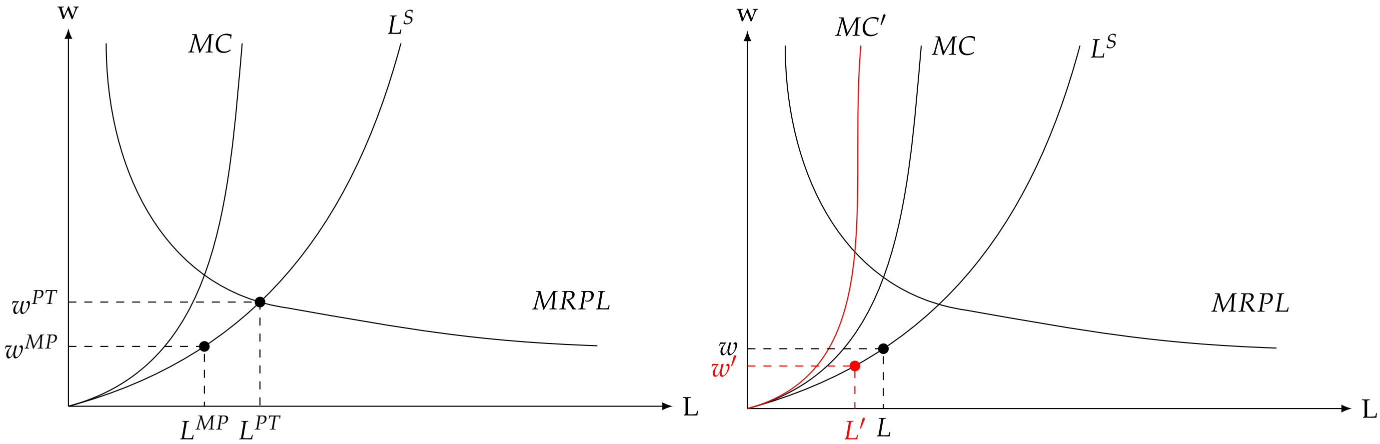
where the right hand side is the marginal cost ($\frac{\partial w_{io} L_{io}}{\partial L_{io}}$) when internalizing movements along the labor supply curve. Noting that $\frac{\partial w_{io}}{\partial L_{io}} = \frac{w_{io}}{L_{io}} \frac{1}{e_{io}}$ is the inverse of the labor supply elasticity e_{io}, the first order condition can be written as:

$$P_b \frac{\partial F}{\partial L_{io}} = w_{io}\left(1 + \frac{1}{e_{io}}\right).$$

When labor supply is infinitely elastic, the MRPL is equal to the wage. When $e_{io} < \infty$ the wage will be below the MRPL. Panel (a) of Figure 8 shows equilibrium wages and employment when the firm acts as a price taker (PT) and when it exerts labor market power (MP).

When firms have labor market power and do not act strategically, their perceived elasticity is constant, $e_{io} = e$. The last term above is therefore constant implying that conditional on a labor supply level, wages are independent to employment shares. When the perceived elasticity is a decreasing function of the employment share, shocks that increase employment share will move the marginal cost (MC) curve to the left. Panel (b) of Figure 8 gives an intuition of our instrument.

Figure 8 – Instrument



(a) Equilibrium wage. Price taking (PT) and oligopsonistic competition (MP)

(b) Instrument

1.F.1 Definition of Mass Layoff

Denote by ML the set of firms with a *national* mass layoff. That is, firms with all the establishments suffering a mass layoff. We instrument the employment share of the establishments of firms not suffering the national mass layoff $j \notin ML$ by the exogenous event of a firm present at the local labor market having a negative shock. We restrict the analysis to non-shocked firms present in different commuting zones with at least one establishment in a sub-market where a competitor has suffered a mass layoff and another plant whose competitors do not belong to firms in ML.

Local labor markets where a mass-layoff has occurred will take a value of $D_{m,t}$ equal to 1.[70] The first stage is:

$$s_{io|m,t} = \psi_{\mathbf{J}(i),o,t} + \delta_{\mathbf{N}(i)} + \gamma D_{m,t} + \epsilon_{io,t}$$

[70] A firm j at occupation o is hit by a negative shock if $\mathbb{1}\{L_{io,t}/L_{io,t-1} < \kappa \,\forall i$ where $\mathbf{J}(i), t = j\} = 1$. A local labor market is identified as shocked $D_{m,t} = 1$ if at least one establishment at the local market belongs to a firm in ML.

where as before, $\psi_{\mathbf{J}(i),o,t}$ is a firm-occupation-year fixed effect and $\delta_{\mathbf{N}(i)}$ is a commuting zone fixed effect. Using the fitted values we consider the following model for the second stage:

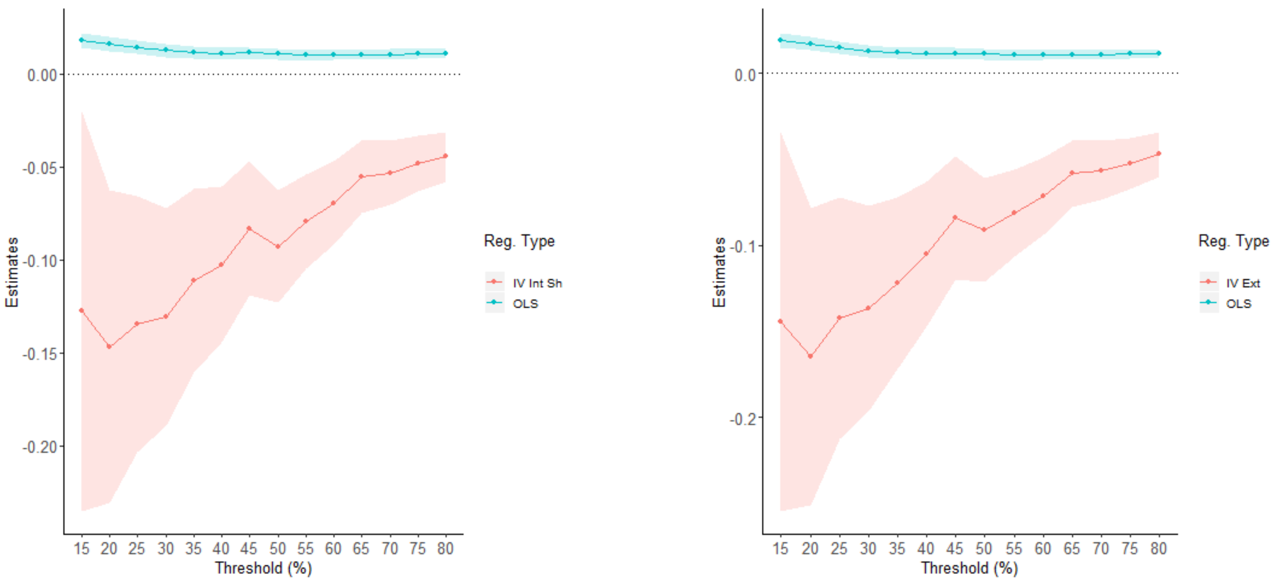
$$\log(w_{io,t}) = \psi_{\mathbf{J}(i),o,t} + \delta_{\mathbf{N}(i)} + \alpha\, \widehat{s_{io|m,t}} + u_{io,t} \tag{52}$$

Before generating the instrument, we need to identify the firms suffering from a mass layoff. Defining a cut-off value κ, we identify a firm-occupation $j \in LO$ if establishment-occupation employment at t is less than κ% employment last year. The best instrument would be identifying firms that went bankrupt ($\kappa = 0$). Given that we cannot externally identify if a firm disappears because it went bankrupt or change identifiers keeping the number of competitors at the local market constant. There is a trade-off when choosing κ. On the one hand, a lower threshold leads to considering stronger negative shocks and the generated instrument would be cleaner. On the other hand, we would classify less firms as having a negative shock reducing the number of events considered. This creates a bias-variance trade-off on the election of the threshold. Lacking a clear candidate for κ, we try with different cut-off values.[71]

1.F.2 Robustness Checks

Figure 9 shows robustness checks of the reduced form exercise. The former considers a different instrument for the employment shares and the latter is taking commuting zone-year fixed effects. The results in the main text are with commuting zone fixed effects.

Figure 9 – Robustness



(a) Instrument: Intensive Share

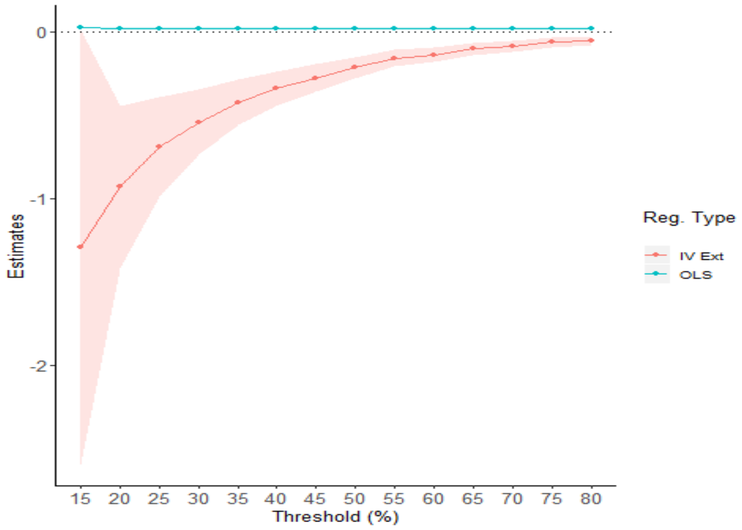
(b) CZ-year fixed effects

Notes: This figures present the point estimates and 95% confidence bands of the OLS and IV exercises on the y-axis. The x-axis presents different thresholds κ that define a mass layoff shock. In both cases we focus on non-affected competitors (not suffering a mass layoff shock). The instrument in Panel (a) is the presence of a mass layoff shock firm in the local labor market interacted with the employment share of the affected firm. Panel (b) presents the results with commuting zone-year fixed effects.

Instead of considering local labor markets with industries at the 3-digit level h as in the baseline, they are defined at the 2-digit level b.

[71] A standard value in the literature is κ =70%. That is a 30% lost of employment.

Figure 10 – Robustness. Local Labor Market at 2-digit Industry



Notes: This figure presents the point estimates and 95% confidence bands of the OLS and IV exercises on the y-axis. The x-axis presents different thresholds κ that define a mass layoff shock. We focus on non-affected competitors (not suffering a mass layoff shock). The instrument is the presence of a mass layoff shock firm in the local labor market. The definition of local labor market is a combination of commuting zone, 2-digit industry and occupation. The difference with respect to Figure 8 is that the local labor market is at 2-digit rather than 3-digit industry.

1.G Distributional and Efficiency Consequences

Here we illustrate the distributional and efficiency effects when the labor wedge λ is simply a markdown μ. Figure 11 illustrates the effect of labor market power on the distribution of value added into profits and wage payments. For simplicity, we illustrate with the case of a production function using only labor with a decreasing returns to scale technology. On the left panel, we have the case of perfect competition in the labor market where wages are equal to the marginal revenue product of labor and the firm earns quasi-rents generated from having decreasing returns. On the right panel, we illustrate the case with labor market power. Wages are below the marginal revenue product because the markdown μ. This generates additional profits for the firm, reducing wage bill payments and therefore the labor share.

Figure 11 – Distributional Consequences

(a) Perfect Competition (b) Labor Market Power

Figure 12 shows the efficiency consequences due to the misallocation of resources. The left panel shows two firms with the same markdown. For simplicity we assume that all firms and local labor markets have the same amenities so workers being indifferent, all establishments will have the same wage in equilibrium.

With homogeneous markdowns, the marginal revenue products are equalized across establishments. In particular, firm B is more productive and in equilibrium $L_B > L_A$. On the right panel we show an example with heterogeneous markdowns. Firm B being more productive is more likely to have a higher employment share at the local labor market and therefore a more important markdown. That is, $\mu_B < \mu_A$. Wages being equalized for all the establishments $MRPL_B < MRPL_A$. We illustrate the extreme case where the distortion generated by labor market power flips the employment size of both firms and we have $L_A > L_B$.

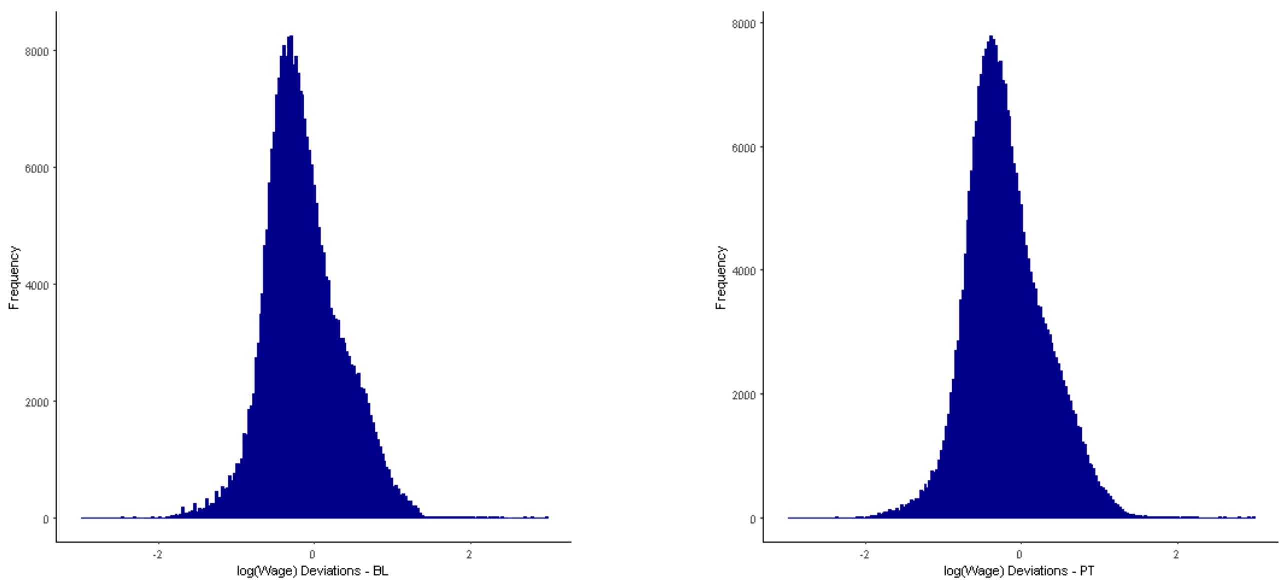
Figure 12 – Efficiency Consequences

(a) Perfect Competition

(b) Labor Market Power

The next Figure shows the baseline and counterfactual distribution of demeaned wages.

Figure 13 – Wage Distribution



(a) Baseline Demeaned Wage Distribution

(b) Counterfactual (PT) Demeaned Wage Distribution

1.G.1 Union

Tables 19 and 20 present respectively the rent sharing elasticities for industries and occupations.

Table 19 – Rent Sharing: Industry

Industry Code	Industry Name	Rent Sharing	SE Rent Sharing
15	Food	0.40	0.00
17	Textile	0.22	0.00
18	Clothing	0.31	0.00
19	Leather	0.31	0.00
20	Wood	0.32	0.00
21	Paper	0.22	0.00
22	Printing	0.34	0.00
24	Chemical	0.17	0.00
25	Plastic	0.23	0.00
26	Other Minerals	0.25	0.00
27	Metallurgy	0.14	0.00
28	Metals	0.37	0.00
29	Machines and Equipments	0.30	0.00
30	Office Machinery	0.33	0.01
31	Electrical Equipment	0.25	0.00
32	Telecommunications	0.23	0.00
33	Optical Equipment	0.32	0.00
34	Transport	0.22	0.00
35	Other Transport	0.31	0.00
36	Furniture	0.37	0.00

Table 20 – Rent Sharing: Occupation

Occupation	Rent Sharing	SE Rent Sharing
Top management	0.38	0.00
Supervisor	0.27	0.00
Clerical	0.29	0.00
Blue collar	0.30	0.00

1.H Alternative Production Function

In this section we denote the local labor market as in the main text. m denotes the combinations between commuting zone, 3-digit industry and occupations. That is: $m = n \times h \times o$. We denote as a location l the combinations of commuting zones and 3-digit industries $l = n \times h$.

Suppose that establishment i produces using some generic capital K_i and a labor composite H_i of different occupations:

$$y_i = \widetilde{A}_i K_i^{\alpha_b} H_i^{\beta_b} = \widetilde{A}_i K_i^{\alpha_b} \left(\prod_{o \in \mathcal{O}} L_{io}^{\gamma_o} \right)^{\beta_b}, \quad \sum_o \gamma_o = 1, \quad \alpha_b + \beta_b \leq 1. \tag{53}$$

The first order conditions are:

$$\begin{aligned} w_{io} &= \beta_b \gamma_o \lambda(\mu_{io} \varphi_b) P_b \frac{y_i}{L_{io}} \\ R_b &= \alpha_b \widetilde{A}_i K_i^{\alpha_b - 1} H_i^{\beta_b} \end{aligned}$$

Substituting the first order condition for capital into the production function, the wage first order condition becomes,

$$w_{io} = \beta_b \gamma_o \lambda(\mu_{io} \varphi_b) A_i H_i^{1-\delta} L_{io}^{-1} P_b^{\frac{1}{1-\alpha_b}}$$

where we plugged the labor supply and used the definition of $\delta = 1 - \frac{\beta_b}{1-\alpha_b}$ from the main text and $A_i = \widetilde{A}_i^{\frac{1}{1-\alpha_b}} \left(\frac{\alpha_b}{R_b} \right)^{\frac{\alpha_b}{1-\alpha_b}}$ as in the main text. Using those and solving for L_{io} we can write the labor composite H_i as function of wages:

$$H_i^{\delta} = P_b^{\frac{1}{1-\alpha_b}} \prod_{o \in \mathcal{O}} \beta_b \gamma_o \lambda(\mu_{io}, \varphi_b) w_{io}^{-1}$$

Substituting the wage first order condition with the labor supply (1.13) into this,

$$\begin{aligned} H_i^{1+\varepsilon_b \delta} &= P_b^{\frac{\varepsilon_b}{1-\alpha_b}} \prod_{o \in \mathcal{O}} \left(\beta_b \gamma_o \lambda(\mu_{io}, \varphi_b) A_i (T_{io} \Gamma_b^{\eta})^{1/\varepsilon_b} \right)^{\varepsilon_b \gamma_o} \prod_{o \in \mathcal{O}} \left(\Phi_m^{1-\eta/\varepsilon_b} \frac{\Phi}{L} \right)^{-\gamma_o} \\ &= P_b^{\frac{\varepsilon_b}{1-\alpha_b}} (\beta_b \Upsilon A_i)^{\varepsilon_b} T_i \Gamma \prod_{o \in \mathcal{O}} \lambda(\mu_{io}, \varphi_b)^{\varepsilon_b \gamma_o} \prod_{o \in \mathcal{O}} \left(\Phi_m^{1-\eta/\varepsilon_b} \frac{\Phi}{L} \right)^{-\gamma_o}, \end{aligned}$$

where $\Upsilon \equiv \prod_{o \in \mathcal{O}} \gamma_o$, $\Gamma \equiv \prod_{o \in \mathcal{O}} \Gamma_b^{\eta}$ and $T_i \equiv \prod_{o \in \mathcal{O}} T_{io}$. Plugging back into the wage equation and rearranging,

$$w_{io} = \left[\lambda(\mu_{io}, \varphi_b) \frac{\gamma_o}{T_{io} \Gamma_b^{\eta}} (\beta_b A_i)^{\frac{1+\varepsilon_b}{1+\varepsilon_b \delta}} (\Upsilon (T_i \Gamma)^{1/\varepsilon_b})^{\frac{\varepsilon_b(1-\delta)}{1+\varepsilon_b \delta}} \left(\prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \left(\prod_{o' \in \mathcal{O}} \Phi_{m'}^{(\eta/\varepsilon_b - 1)\gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \Phi_m^{1-\eta/\varepsilon_b} \right]^{\frac{1}{1+\varepsilon_b}} \tag{54}$$

$$\left(\frac{\Phi}{L} \right)^{\frac{1}{1+\varepsilon_b}} P_b^{1/\chi},$$

with $\chi_b = (1 - \alpha_b)(1 + \varepsilon_b \delta)$. Define the following:

$$\begin{aligned} c_{io} &\equiv \frac{\gamma_o}{T_{io} \Gamma_b^{\eta}} (\beta_b A_i)^{\frac{1+\varepsilon_b}{1+\varepsilon_b \delta}} (\Upsilon (T_i \Gamma)^{1/\varepsilon_b})^{\frac{\varepsilon_b(1-\delta)}{1+\varepsilon_b \delta}}, \\ C_l &\equiv \prod_{o' \in \mathcal{O}} \left(\Phi_{m'}^{(\eta/\varepsilon_b - 1)\gamma_o} \right)^{\frac{\delta}{1+\varepsilon_b \delta}} \left(\frac{\Phi}{L} \right)^{\frac{1}{1+\varepsilon_b}}, \\ F_b &\equiv P_b^{1/\chi}, \end{aligned}$$

where C_l is a location constant. Rearranging we have that:

$$w_{io} = \left[\lambda(\mu_{io}, \varphi_b) c_{io} \left(\prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \frac{\Phi_m^{1-\eta/\varepsilon_b}}{\prod_{o' \in \mathcal{O}} \Phi_{m'}^{(1-\eta/\varepsilon_b)\gamma'_o}} \right]^{\frac{1}{1+\varepsilon_b}} C_l F_b. \tag{55}$$

The last system is equivalent to the one in (54) and has the benefit to being able to write the wages: $w_{io} = \widetilde{w}_{io} C_m F_b$, where we want $\widetilde{w}_{io}$ to be homogeneous of degree zero with respect constants to m level. Note that the last term inside the brackets is homogeneous of degree zero with respect to location l constants shared by all the occupations of a establishments. Then, defining $\widetilde{\Phi}_m = \sum_{i \in \mathcal{I}_m} T_{io} w_{io}^{\varepsilon_b}$, the establishment-occupation or normalized wage is:

$$\widetilde{w}_{io} \equiv \left[\lambda(\mu_{io}, \varphi_b) c_{io} \left(\prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \frac{\widetilde{\Phi}_m^{1-\eta/\varepsilon_b}}{\prod_{o' \in \mathcal{O}} \widetilde{\Phi}_{m'}^{(1-\eta/\varepsilon_b)\gamma'_o}} \right]^{\frac{1}{1+\varepsilon_b}}. \tag{56}$$

$\widetilde{w}_{io}$ is homogeneous of degree zero with respect to location l constants shared by all occupations. This property, allows to solve for the normalized wages of every location l (combinations of commuting zone n and sub-industry h combinations) independently and then recover the aggregate constants. Aggregating (56) and solving for $\widetilde{\Phi}_m$,

$$\widetilde{\Phi}_m = \left[\frac{\sum_{i \in I_m} \left(\lambda(\mu_{io}, \varphi_b) c_{io} T_{io}^{\frac{1+\varepsilon_b}{\varepsilon_b}} \prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}}}{\prod_{o' \in \mathcal{O}} \widetilde{\Phi}_{m'}^{(1-\eta/\varepsilon_b)\gamma'_o}} \right]^{\frac{\varepsilon_b}{1+\eta}}.$$

Taking first all to the power $(1 - \eta/\varepsilon_b)\gamma_o$ and taking the product,

$$\mathcal{L}_l \equiv \prod_{o' \in \mathcal{O}} \widetilde{\Phi}_{m'}^{(1-\eta/\varepsilon_b)\gamma'_o} = \prod_{o' \in \mathcal{O}} \left[\sum_{i \in I_m} \left(\lambda(\mu_{io}, \varphi_b) c_{io} T_{io}^{\frac{1+\varepsilon_b}{\varepsilon_b}} \prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \right]^{\gamma_{o'} \frac{\varepsilon_b - \eta}{1+\varepsilon_b - \eta}},$$

which recovers all the constants inside $\widetilde{w}_m$.

In order to prove the existence and uniqueness of the solution of the system (56), define $\widehat{w}_{io}$ as:

$$\widehat{w}_{io} = \left[\lambda(\mu_{io}, \varphi_b) \left(\prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \right]^{\frac{1}{1+\varepsilon_b}} c_{io}^{\frac{1}{1+\varepsilon_b}}$$

$$w_{io} = \widehat{w}_{io} \left[\frac{\widetilde{\Phi}_m^{1-\eta/\varepsilon_b}}{\mathcal{L}_l} \right]^{\frac{1}{1+\varepsilon_b}} C_l F_b = \widehat{w}_{io} z_l = \widetilde{w}_{io} C_l F_b. \tag{57}$$

We can show that the system formed by (57) has a solution and is unique.

Proposition 3. *For given parameters $0 \leq \alpha_b, \beta_b < 1$, $1 < \eta < \varepsilon_b$, $0 \leq \delta \leq 1$, transformed price F_b, constants C_l, $\widetilde{\Phi}_m$, $\mathcal{L}_l$ and non-negative vectors of productivities $\{A_i\}_{i \in m}$ and amenities $\{T_{io}\}_{io \in m}$, there exists a unique vector of wages $\{w_{io}\}_{io \in I_m}$ for every location l (combination of commuting zone n and sub-industry h) that solves the system formed by* (57).

Sketch of the proof. For existence, first note that $\lambda(\mu_{io}, \varphi_b) \in \left[(1 - \varphi_b) \frac{\eta}{1+\eta} + \varphi_b \frac{1}{1-\delta}, (1 - \varphi_b) \frac{\varepsilon_b}{1+\varepsilon_b} + \varphi_b \frac{1}{1-\delta} \right], \forall i, o$. Define a vector $\mathbf{w}$ with wage of all the occupation-establishments at location l, $\mathbf{w} \equiv \{w_{11}, w_{12}, ..., w_{1O}, ..., w_{I1}, w_{I2}, ..., w_{IO}\}$. Taking for now the elements of z_l as constants. The system to solve is: $f_{io}(\mathbf{w}) = \widehat{w}_{io} z_l$. We have that

$$\mathbf{w} \in \mathcal{C} \equiv \left[\left((1 - \varphi_b) \frac{\eta}{1+\eta} + \varphi_b \frac{1}{1-\delta} \right)^{\frac{1}{1+\eta\delta}} c_{11}^{\frac{1}{1+\varepsilon_b}} z_{l1}, \left((1 - \varphi_b) \frac{\varepsilon_b}{1+\varepsilon_b} + \varphi_b \frac{1}{1-\delta} \right)^{\frac{1}{1+\eta\delta}} c_{11}^{\frac{1}{1+\varepsilon_b}} z_{l1} \right]$$
$$\times ... \times \left[\left((1 - \varphi_b) \frac{\eta}{1+\eta} + \varphi_b \frac{1}{1-\delta} \right)^{\frac{1}{1+\eta\delta}} c_{IO}^{\frac{1}{1+\varepsilon_b}} z_{IO}, \left((1 - \varphi_b) \frac{\varepsilon_b}{1+\varepsilon_b} + \varphi_b \frac{1}{1-\delta} \right)^{\frac{1}{1+\eta\delta}} c_{IO}^{\frac{1}{1+\varepsilon_b}} z_{IO} \right].$$

The system f_{io} is continuous on wages and maps into itself on $\mathcal{C}$. The last set being a compact set we can apply Brower's fixed point theorem.

For uniqueness, once the product of the wedges is substituted, $\widehat{w}_{io}$ is:

$$\widehat{w}_{io} = \left[\lambda(\mu_{io}, \varphi_b) c_{io} \prod_{o' \in \mathcal{O}} (w_{io'} c_{io}^{-\frac{1}{1+\varepsilon_b}})^{\gamma'_o \varepsilon_b (1-\delta)} \right]^{\frac{1}{1+\varepsilon_b}}$$

Define the function $g_{io}(\mathbf{w}) = f_{io}(\mathbf{w}) - w_{io}$. Gross substitution is fulfilled if $\frac{\partial g_{io}(\mathbf{w})}{\partial w_{jo}} > 0, \forall j \neq i$ with $j \in \mathcal{I}_l$ and $\frac{\partial g_{io}(\mathbf{w})}{\partial w_{io'}}, \forall o'$. Gross substitution resumes to taking the partial derivatives of $\widehat{w}_{io}$ which are positive by similar reasoning as in the main proof. Finally, $\widehat{w}_{io}$ is homogeneous of degree $\frac{\varepsilon_b}{1+\varepsilon_b}(1-\delta) < 1$. Therefore the solution to the system (57) exists and is unique. □

Finally, the model can be aggregated up to the industry level following similar steps as in the baseline. Steps to write the industry model are in Appendix 1.A.5.

1.I Pass Through

Table 21 – Pass Through of Z

Industry Code	Industry Name	ϵ_Z^W PT	$\widehat{\beta}_b^Z$	Diff	SE $\widehat{\beta}_b^Z$
15	Food	0.933	0.890	0.043	0.000
17	Textile	0.940	0.916	0.024	0.000
18	Clothing	0.943	0.925	0.018	0.000
19	Leather	0.918	0.842	0.076	0.000
20	Wood	0.939	0.888	0.052	0.000
21	Paper	0.885	0.835	0.050	0.000
22	Printing	0.939	0.914	0.025	0.000
24	Chemical	0.879	0.720	0.159	0.000
25	Plastic	0.904	0.856	0.048	0.000
26	Other Minerals	0.935	0.887	0.048	0.000
27	Metallurgy	0.862	0.777	0.085	0.001
28	Metals	0.951	0.932	0.019	0.000
29	Machines and Equipments	0.915	0.861	0.054	0.000
30	Office Machinery	0.876	0.760	0.116	0.001
31	Electrical Equipment	0.886	0.848	0.039	0.000
32	Telecommunications	0.869	0.840	0.029	0.000
33	Optical Equipment	0.925	0.894	0.031	0.000
34	Transport	0.853	0.802	0.051	0.000
35	Other Transport	0.871	0.788	0.083	0.000
36	Furniture	0.938	0.909	0.029	0.000

Notes: This table presents the estimation results of equation (1.30) in Column (4) $\widehat{\beta}_b^Z$ and its comparison to the pass through without the labor wedges in Column (3) ϵ_Z^W *PT*. *Diff* in Column (5) shows the difference between the pass thorough without the wedges and the estimated one and *SE* $\widehat{\beta}_b^Z$ in Column (6) presents the standard error of the estimated parameters $\widehat{\beta}_b^Z$.

Chapter 2

Correcting Small Sample Bias in Linear Models with Many Covariates

Miren Azkarate-Askasua and Miguel Zerecero[1]

Abstract

Estimations of quadratic forms in the parameters of linear models exhibit small-sample bias. When the number of covariates is large, the direct computation for a bias correction is not feasible. We propose a bootstrap method for correcting this bias. Our method accommodates different assumptions on the structure of the error term including general heteroscedasticity and serial correlation. Our approach can correct for the bias of multiple quadratic forms of the same linear model without increasing the computational cost. We show with Monte Carlo simulations that our bootstrap procedure is effective in correcting the bias and we compare to other methods in the literature. Using administrative data for France, we apply our method by doing a variance decomposition of a linear model of log wages with person and firm fixed effects. We find that the person and firm effects are less important in explaining the variance of log wages after correcting for the bias.

JEL Codes: C13, C23, C55, J30, J31

Keywords: Variance components, many regressors, fixed effects, bias correction.

[1]Azkarate-Askasua: miren.askasua@gmail.com; Zerecero: miguel.zerecero@tse-fr.eu. First version: May 2018. We thank our advisor Christian Hellwig for his guidance throughout this project. We thank Patrick Fève, Elia Lapenta, Tim Lee and Thierry Magnac for helpful comments. All errors are our own.

2.1 Introduction

With the increased availability of large panel data sets, researchers have been interested in understanding to what extent unobserved heterogeneity can explain the variation of an outcome of interest. Usually, econometricians include fixed effects in a standard linear model to control for this unobserved heterogeneity and then perform a variance decomposition. These methods have been used in the context of education to study the importance of classroom effects (e.g. Chetty et al. (2011)) and extensively in the labor market context where log-additive models of wages are used to study the determinants of labor income (e.g. Abowd et al. (1999); Card et al. (2013); Iranzo et al. (2008); Lopes de Melo (2018)).

The elements of a variance decomposition of a linear model are quadratic objects in the parameters. As long as the parameters are estimated with noise, these quadratic objects are subject to small-sample bias. This bias can be substantial and can even change the sign of estimated covariances and correlations. Moreover, this bias does not fade away by increasing the sample size when using panel data, as the number of parameters to estimate, i.e. the number of fixed effects, grows with the sample size.

Focusing on the context of labor economics, researchers have used employer-employee matched datasets to study sorting patterns of workers into firms. Various papers have estimated a linear model of log wages with person and firm fixed effects, following the seminal work of Abowd, Kramarz, and Margolis (1999) (AKM henceforth). These studies compute the correlation between the person and firm fixed effects to determine the degree of sorting in the labor market. Most studies have found zero or negative correlations, casting doubt on whether there is sorting in the labor market. However, as first noted by Abowd et al. (2004) this correlation is likely to suffer from small-sample bias, dubbed *limited mobility* bias in their paper. Andrews et al. (2008) derive formulae for correcting the bias when the errors are homoscedastic. Gaure (2014) provides formulae for more general variance structures. Unfortunately, the direct implementation of these corrections in high dimensional models is infeasible. The reason is that the corrections entail computing the inverse of an impractically large matrix.[2] This has prevented the direct application of the correction formulae.

In this paper we propose a bootstrap method to correct for small-sample bias in quadratic forms in the estimated parameters of linear models with a large number of covariates. The method is very similar to MacKinnon and Smith Jr (1998) when the bias is flat or independent of the initial estimates. Our method is more efficient than the one they propose and easily implementable while allowing for a flexible error structure. Using Monte Carlo simulations we show that our method successfully corrects the bias of quadratic forms in the parameters in cases where the error term is homoscedastic or heteroscedastic. We extend the approach to other covariance structures of the error term such as clustering and serial correlation. For the moment our procedure using block bootstrap is not as successful as for the cases with diagonal covariance matrix and we are exploring the use of a Sieve bootstrap as proposed by Davidson and MacKinnon (2006) that consists on estimating the error dependence in a intermediate step. As an alternative, the reader dealing with serial correlation within group can reconvert the data to the group level and then use our standard method accommodating heteroscedastic error terms.

We apply our method to French administrative data and perform a variance decomposition of an estimated AKM type model. Consistent with Andrews et al. (2008) formulation, we find that sample variances of person and firm effects are reduced and their covariance increased after the correction. The estimated cor-

[2] By large matrix we mean a matrix with dimension in the order of hundreds of thousands or millions.

relation between person and firm fixed effects increases from -0.19 to -0.11 after the correction. Abowd et al. (2004) using also French data but a different sample, found a correlation of -0.28. Compared to estimates from other countries, the correlation obtained with French data has been more negative and farther from zero than the ones found in these other studies. We believe the reason behind this is that the French data is a representative sub-sample of around 4% of the whole universe of workers. As identification of the fixed effects comes from workers moving across firms, the particular sampling procedure used to generate the French panel tends to deliver a sample with few workers moving across jobs, resulting in noisier estimates of the fixed effects. Indeed, in Table 1 of Lopes de Melo (2018) the correlation in the French data is the most negative and the ratio of workers to firms the smallest of all studies, suggesting that this dataset can exhibit substantial noise in the estimates making it harder to correct for the bias.

Our approach is similar to the ones proposed by Gaure (2014) and by Kline et al. (2019). All methods rely on iterative procedures to compute an estimate of the bias correction term. In general, the bias appears as the trace of a matrix, but when the number of covariates of the linear model is large, the explicit computation of this trace is not practical. Gaure exploits the fact that the trace can be represented as the expectation of a more manageable quadratic form in a random vector, which is estimated as a sample mean.[3] Gaure sketches the procedure to correct for the bias when the error terms are heteroscedastic but to the best of our knowledge does not implement it in his R package *lfe*.[4]

Kline et al. (2019) (KSS henceforth) follow a similar approach in estimating the small sample bias of second order moments. They compute the trace term leading to the bias based on leave-one-out estimates. For the applications with many covariates where the direct computation is unfeasible, they propose an approximation algorithm to estimate the bias. Similarly to us, they show their estimator is unbiased and consistent. Our approach is different in the way we estimate the trace term and also on the estimate of the covariance matrix we use. The main benefit of our method is its flexibility but KSS go one step ahead and propose how to perform inference in situations when the rank of the quadratic form depends on the sample size (e.g. when we have two-way fixed effects).

The computational cost in Gaure and KSS comes from building the quadratic forms, as it requires solving a large system of linear equations in each iteration. In contrast, we re-estimate the model with bootstrapped data and show that a sample mean of the *bootstrapped* moment estimates is an unbiased and consistent estimator of the ideal bias correction term. In our method, the computational cost comes from estimating the linear model in each bootstrap but does not increase depending on the number of moments to correct.

We believe our method to be faster in cases where one is looking to correct for several quadratic forms in the parameters, for example, when performing a variance decomposition of the model. Regardless of the number of moments to correct, we need to solve one system of linear equations per bootstrap, while with Gaure's and KSS methods one needs to solve as many systems of equations per iteration as needed corrections.[5] Following the work of MacKinnon and Smith Jr (1998) bootstrap methods have been used to

[3]In particular, the way Gaure estimates the trace is known as the Hutchinson method: denote a random vector $x \in \mathbb{R}^n$, where each individual entry is independently distributed Rademacher (entries can take values of 1 or -1 with probability 1/2). Then, for a square matrix $A \in \mathbb{R}^{n\times n}$ we have that $tr(A) = \mathbb{E}(x'Ax)$. The Hutchinson estimator of the trace of matrix A is $\frac{1}{M}\sum_{i=1}^{M} x_i' A x_i$, where x_i is the i-th draw of the random vector x. See Hutchinson (1989) and Avron and Toledo (2011).

[4]One can download the *lfe* package at: `https://cran.r-project.org/web/packages/lfe/index.html`. The function applying the correction is *bccorr*.

[5]For example, consider the linear model $y_t = X_{1t}\beta_1 + X_{2t}\beta_2 + \varepsilon_t$ where one is interested in doing a variance decomposition for each

correct for variance estimates in linear models with fixed effects (e.g. Kane and Staiger, 2008; Best et al., 2017). Our contribution with respect to those is to propose a more efficient bootstrap method and to compare it to others proposed in the literature.

Borovičková and Shimer (2017) (henceforth BS) provide an alternative method to compute the correlation of firm types and workers, which has the advantage of not requiring estimates of all the worker and firm fixed effects and directly computing the correlation. We perform two exercises to compare our method with theirs. First, we simulate labor market data that fulfills the key identifying assumptions of the AKM linear model and of BS. We find that both methods correct the bias but ours outperforms theirs in terms of accuracy of the estimation of each of the elements of the correlation, but is naturally more time consuming. Second, we apply their method to the French data. In order to do so, we need to deviate in two aspects from the original dataset used in our main application: first, we need to restrict the sample to workers that have at least two jobs and firms that have at least two workers; second, we need to take averages of every match between firm and workers.[6] Naturally, both approaches now yield different correlation estimates, ours being 0.16 and theirs 0.55. This difference clearly points out that the data generating process of French labor market data appears to violate the assumptions of either our bootstrap method, BS method or both. We discuss this in more detail in the next sections.

Labor economists have been aware of the small-sample bias problem with quadratic forms in the parameters and the difficulty in estimating a correction at least since Andrews et al. (2008). There have been several attempts to correct this bias when performing variance decompositions of estimated linear models. Some methods are based on leaving out part of the data, such as the panel jacknife estimator by Dhaene and Jochmans (2015) or the leave one out estimator by KSS already mentioned. Another method relies on reducing the dimensionality of the parameters to be estimated, thereby reducing the noise in the estimates and the small-sample bias in any quadratic form, like in Bonhomme et al. (2015). We leave the comparison to these methods for future research.

2.2 The Bias

Consider the following linear model:[7]

$$Y = X\beta + u, \tag{2.1}$$

where Y is a $n \times 1$ vector representing the endogenous variable, X is a matrix of covariates of size $n \times k$, and β is a vector of parameters. The error term u satisfies mean independence $\mathbb{E}(u|X) = 0$.

The OLS estimate of β is

$$\widehat{\beta} = \beta + Qu,$$

where $Q = (X'X)^{-1} X'$. We are interested in estimating the following quadratic form $\varphi = \beta' A \beta$ for some non-random matrix A of dimensions $k \times k$. From the expression for $\widehat{\beta}$ we can decompose the the plug-in

period t. This would yield three quadratic objects to correct ($Var(X_1\widehat{\beta}_1)$, $Var(X_2\widehat{\beta}_2)$, $Cov(X_1\widehat{\beta}_1, X_2\widehat{\beta}_2)$) *per period*.

[6] More precisely this would mean that if we observe one worker employed for a certain firms for some years, we would take the average wage of that worker in that firm as one observation

[7] We somewhat follow the notation in Kline et al. (2019) for the interested reader to compare the papers.

estimator $\widehat{\varphi}_{PI} = \widehat{\beta}' A \widehat{\beta}$ as

$$\widehat{\varphi}_{PI} = \beta' A \beta + u' Q' A Q u + 2u' Q' A \beta. \tag{2.2}$$

Using the general formula for the expectation of quadratic forms and the exclusion restriction $\mathbb{E}(u|X) = 0$ we obtain,[8]

$$\mathbb{E}\left(\widehat{\varphi}_{PI}|X\right) = \beta' A \beta + \text{trace}\left(Q' A Q \mathbb{V}(u|X)\right) = \varphi + \delta, \tag{2.3}$$

where the bias $\delta \equiv \text{trace}\left(Q' A Q \mathbb{V}(u|X)\right)$ comes from the fact that $\widehat{\beta}$ is estimated with noise. This bias is larger in cases where the sample size is small relative to the number of parameters to estimate. In the two-way fixed effects AKM model, the number of observations per worker/firm are usually small relative to the amount of fixed effects. Moreover, the observations identifying the firm fixed effects are the ones of firm movers which tend to be small in samples with low mobility.

The ideal bias correction term $\widehat{\delta}$ is defined as

$$\widehat{\delta} \equiv \text{trace}\left(Q' A Q \widehat{\mathbb{V}}(u|X)\right) \tag{2.4}$$

where $\widehat{\mathbb{V}}(u|X)$ is an estimator of the conditional variance of the error term $\mathbb{V}(u|X)$. The ideal bias correction $\widehat{\delta}$ is an unbiased estimate of the bias term δ if and only if $\widehat{\mathbb{V}}(u|X)$ is an unbiased estimator of $\mathbb{V}(u|X)$.[9] Therefore we need an unbiased estimate of the conditional variance to be able to compute the ideal bias correction $\widehat{\delta}$.[10] We can define then the following unbiased estimate of φ as[11]

$$\widehat{\varphi} = \widehat{\beta}' A \widehat{\beta} - \widehat{\delta}.$$

Unfortunately, when the number of covariates is very large, computing the ideal bias correction $\widehat{\delta}$ directly is computationally infeasible. This is because, in order to compute the trace, we need to calculate first the matrix Q, which is itself a function of the inverse of a very large matrix.[12] In the next section we propose a methodology to apply a computationally feasible correction. But first, we describe how the components of a variance decomposition of a linear model are indeed quadratic forms in the parameters.

2.2.1 Components of a variance decomposition as quadratic objects

When performing a variance decomposition of a linear model, one can think of each element as a particular form of $\widehat{\beta}' A \widehat{\beta}$ with the appropriate choice of A. To see this, we can rewrite (2.1) as

$$Y = X_1 \beta_1 + X_2 \beta_2 + u,$$

where X_1 and X_2 are matrices of covariates of size $n \times k_1$ and $n \times k_2$, $k = k_1 + k_2$ with $X = [X_1\, X_2]$ and $\beta' = [\beta_1'\, \beta_2']$.

[8]Given a random vector x and a symmetric matrix B we have that $\mathbb{E}(x'Bx) = \mathbb{E}(x')B\mathbb{E}(x) + \text{trace}(B\mathbb{V}(x))$.

[9]Proof: by the linearity of the trace and expectation operators we have that: $\mathbb{E}(\widehat{\delta}|X) = \mathbb{E}\left(\text{trace}\left(Q'AQ\widehat{\mathbb{V}}(u|X)\right)|X\right) = \text{trace}\left(Q'AQ\mathbb{E}\left(\widehat{\mathbb{V}}(u|X)|X\right)\right) = \text{trace}\left(Q'AQ\mathbb{V}(u|X)\right) = \delta$.

[10]For example, if we assume that the error term u is homocedastic, i.e. $\mathbb{E}(u^2|X) = \sigma_u^2 \mathbf{I}$, then we can use the variance estimator $\widehat{\sigma}_u^2 = \frac{n}{n-k}\sum \widehat{u}_i^2$ and construct the ideal bias correction as $\widehat{\delta} = \widehat{\sigma}_u^2 \times \text{trace}\left(A(X'X)^{-1}\right)$.

[11]Notice that we say "unbiased" and not "bias-corrected". The reason is that as long as $\mathbb{E}(\widehat{\delta}|X) = \delta$, then it follows that $\mathbb{E}(\widehat{\varphi}|X) = \varphi$.

[12]The dimension of this matrix is related with the number of covariates that are estimated in the linear model. In a typical AKM type model the data will typically comprise of hundreds of thousands of workers and tens of thousands of firms, each representing a covariate in the model.

We are interested in the sample variances ($\widehat{\text{var}}(X_1\beta_1)$, $\widehat{\text{var}}(X_2\beta_2)$) and covariance ($\widehat{\text{cov}}(X_1\beta_1, X_2\beta_2)$), denoted, respectively, as σ_1^2, σ_2^2 and σ_{12}.[13] Define $\mathbf{1}$ as a vector of ones with appropriate length. Then, denote the demeaning operator as $M_1 = \mathbf{I} - P_{\mathbf{1}} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$. We can then write the sample variances and covariances in matrix notation as

$$\sigma_j^2 = \beta' A_j \beta, \quad \text{for } j = \{1,2\} \text{ and}$$

$$\sigma_{12} = \beta' A_{12} \beta,$$

where the symmetric matrices A_1, A_2 and A_{12} are equal to

$$A_1 = \frac{1}{n-1}\begin{pmatrix} X_1'M_1X_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \qquad A_2 = \frac{1}{n-1}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & X_2'M_1X_2 \end{pmatrix}, \qquad A_{12} = \frac{1}{2(n-1)}\begin{pmatrix} \mathbf{0} & X_1'M_1X_2 \\ X_2'M_1X_1 & \mathbf{0} \end{pmatrix}.$$

The plug-in estimators of σ_1^2, σ_2^2 and σ_{12}, obtained by substituting β with the OLS estimate $\widehat{\beta}$, are just particular examples of $\widehat{\varphi}_{PI}$. Therefore, these estimates will also be biased.

2.3 Bootstrap Correction

Suppose that we have the residuals of our original regression $Y - X\widehat{\beta}$. Using these residuals we can construct an estimate of the covariance matrix, $\widehat{\mathbb{V}}(u|X)$. We can generate a new dependent variable Y^* as:

$$Y^* = v^*,$$

where v^* is a vector containing the bootstrapped residuals. How we construct v^* will depend ultimately in the assumption that we are making about the error term. In particular, we need that the variance of the bootstrapped errors $\mathbb{V}(v^*|X)$ to be equal to $\widehat{\mathbb{V}}(u|X)$. This is equivalent to performing a traditional bootstrap, while setting $\widehat{\beta} = \mathbf{0}$. The following proposition states the main result of the paper and all the proofs are left to Appendix 2.A:

Proposition 4. *Suppose the regression model (2.1) is correctly specified. Let n^* denote the number of bootstraps. Define β_j^* as the OLS estimate of regressing v_j^* over X for the j-th bootstrap iteration. If the conditional variance-covariance matrix of the bootstrapped residuals $\mathbb{V}(v_j^*|X)$ is equal to $\widehat{\mathbb{V}}(u|X)$, then*

$$\widehat{\delta}_b = \frac{1}{n^*}\sum_{j=1}^{n^*}\left(\beta_j^{*\prime} A \beta_j^*\right)$$

is an unbiased and consistent estimator of the ideal bias correction $\widehat{\delta}$.

The proposition tells us that instead of computing directly the ideal bias correction term $\widehat{\delta}$, which can be infeasible, we can estimate it using a sample average of estimated quadratic forms.

The intuition of why this procedure works, is that in every bootstrap iteration we are replicating the source of the bias, which is the noise embedded in the estimated parameters. The key for the bootstrap correction to work is that $\mathbb{V}(v^*|X)$ is equal to the sample variance-covariance matrix $\widehat{\mathbb{V}}(u|X)$, so the bootstrap correction $\widehat{\delta}_b$ is an unbiased and consistent estimator of the ideal bias correction term $\widehat{\delta}$. Therefore, the bootstrap procedure has to be consistent with the underlying assumption on the structure of the error term.

[13] The sample variance for a vector $\mathbf{x} = \{x_1, x_2, ..., x_n\}$ is $\widehat{\text{var}}(\mathbf{x}) = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})^2$, where $\bar{\mathbf{x}}$ is the sample mean. Similarly, the sample covariance for vectors $\mathbf{x}$ and $\mathbf{y}$ is $\widehat{\text{cov}}(\mathbf{x}, \mathbf{y}) = \frac{1}{n-1}\sum_{i=1}^{N}(x_i - \bar{\mathbf{x}})(y_i - \bar{\mathbf{y}})$.

MacKinnon and Smith Jr (1998) propose a similar bootstrap to correct for flat biases like the one we want to eliminate.[14] MacKinnon and Smith Jr (1998) propose to build the bootstrapped dependent variable by using the original estimate of β, $Y^* = X\widehat{\beta} + v^*$. In our application we would then compute the quadratic objects $\beta^{*\prime}_{j,MS} A \beta^*_{j,MS}$ and their correction would be: $\widehat{\delta}_{b,MS} = \frac{1}{n^*}\sum_{j=1}^{n^*}\left(\beta^{*\prime}_{j,MS} A \beta^*_{j,MS}\right) - \widehat{\beta}' A \widehat{\beta}$. They already note that one can estimate a flat bias correction by using any β to generate Y^*. In particular, the one we use $\widehat{\beta} = \mathbf{0}$. Nevertheless, analogously to equation (2.2) we have that in bootstrap j: $\beta^{*\prime}_{j,MS} A \beta^*_{j,MS} = \widehat{\beta}' A \widehat{\beta} + (v^*_j)' Q' A Q v^*_j + 2 v^{*\prime}_j Q' A \widehat{\beta}$. We have that the conditional variance of their estimator is: $\mathbb{V}(\widehat{\delta}_{b,MS}|X,u) = \dfrac{1}{n^{*2}}\sum_{j=1}^{n^*}\mathbb{V}((v^*_j)'Q'AQv^*_j|X,u) + \dfrac{4}{n^{*2}}\sum_{j=1}^{n^*}\mathbb{V}(v^{*\prime}_j Q' A\widehat{\beta}|X,u)$ where we used the fact that $\widehat{\beta}$ is not a random variable once we condition on X and u, and the independence of the bootstrap errors v^*. We therefore have that the conditional variance of their estimator:

$$\mathbb{V}(\widehat{\delta}_{b,MS}|X,u) = \mathbb{V}(\widehat{\delta}_b|X,u) + \frac{4}{n^{*2}}\sum_{j=1}^{n^*}\mathbb{V}(v^{*\prime}_j Q' A\widehat{\beta}|X,u),$$

is higher than ours due to the presence of the last term similarly to equation (2.2). Both methods are therefore unbiased and consistent but ours is more efficient.

The computational burden of our method comes from estimating β^*_j for each bootstrap. The advantage of our method is twofold. First, that within a single bootstrap loop, we can construct simultaneously several moments to correct. If we are interested in doing a variance decomposition exercise for each year using a linear model, we need a correction for the variances of each group of covariates and the covariance term for *every* year but estimate the effects only once. Second, to estimate β^*_j in every iteration one just needs to solve for a least squares regression. There are extremely efficient procedures to compute these regressions, especially in cases where the high dimensionality of the covariates is a result of a large number of fixed effects. This is the case in most applications.

The small sample properties of the bootstrap estimate $\widehat{\delta}_b$ would depend ultimately on the choice of estimate for the covariance matrix $\mathbb{V}(u|X)$. In particular, for the bias we have the following corollary of Proposition 4:

Corollary 2. *Conditioning on X, if $\widehat{\mathbb{V}}(u|X)$ is an unbiased estimator of $\mathbb{V}(u|X)$, then the bootstrap correction $\widehat{\delta}_b$ is an unbiased estimator of the bias δ.*

Given that the estimate of the covariance matrix is non-linear, in general, we would have a bias. In the next section we discuss the properties for some particular cases of popular choices for estimators of the covariance matrix and how to implement the correction.

2.3.1 Choice of covariance matrix estimate

We divide the discussion in this section in two parts. First, one when the researcher assumes that the covariance matrix is diagonal. This includes the cases where the error is homoscedastic or iid and cases with general heteroscedasticity. Then, we discuss some deviations from this assumption, in particular, when we have clustering or serial correlation.

[14] A flat bias is one that does not depend on the levels of the original estimates. In our notation, the bias is flat because the trace term in (2.3) is independent of $\widehat{\beta}$.

Diagonal covariance matrix

If a researcher assumes that the underlying covariance matrix $\mathbb{V}(u|X)$ is diagonal, with non-zero ith diagonal element equal to ψ_i, Proposition 4 suggests an algorithm to make simultaneous M corrections.[15] Let $\widehat{\psi}_i$ be the estimate of the variance for the ith observation error term. Algorithm 1 in Appendix 2.C takes as inputs X, $\{\widehat{\psi}_i\}_{i=1}^{N}$ and the different matrices $\{A_m\}_{m=1}^{M}$ associated with the different M quadratic forms that want to be computed. The output is a vector of bias corrections $\{\widehat{\delta}_{b,m}\}_{m=1}^{M}$ whose elements correspond to each quadratic form m.[16]

In the homoscedastic case we can use the well known unbiased estimate $\widehat{\psi}_i = n/(n-k)\widehat{\psi}$ with $\widehat{\psi} = \sum_{i=1}^{n} \widehat{u}_i^2$. Alternativelly, under the homoscedastic case we could replace steps (3) and (4) of Algorithm 1 by a residual bootstrap. That is, we can obtain the vector v^* by resampling with replacement from the estimated residuals and adjusting by the corresponding degrees of freedom.[17]

In the general hereoscedasticity case, several estimators have been proposed by the literature. The White (1980) estimator, while consistent, is biased:[18]

$$\mathbb{E}(\widehat{u}_i^2|X) = \psi_i - 2\psi_i h_{ii} + h_i'\mathbb{V}(u|X)h_i,$$

where h_i and h_{ii} are, respectively, the ith column and ith diagonal element of the projection matrix $H = X\left(X'X\right)^{-1}X'$. The latter term, h_{ii} is sometimes known as the *leverage* of observation i, because, as explained by Angrist and Pischke (2008), it tell us how much *pull* a particular observation exerts over the regression line. MacKinnon and White (1985) explore different variance estimates, including the original proposed by White HC_0, and compare their performance using simulations. The different estimators considered were:

$$HC_0 = \widehat{u}_i^2, \quad HC_1 = \frac{n}{n-k}\widehat{u}_i^2 \quad \text{and} \quad HC_2 = \frac{\widehat{u}_i^2}{1-h_{ii}}.$$

They acknowledge the existence of a bias in all three but denote HC_2 as an almost unbiased estimate of the variance. Recently, Kline et al. (2019) and Jochmans (2018) have proposed the following unbiased estimator of the ith conditional variance:

$$HC_U = \frac{Y_i\widehat{u}_i}{1-h_{ii}}.$$

Unfortunately, we can't use Algorithm 1 with this estimator as in practice, some observations have a negative estimated variance and that prevents us from taking the square root in step (4) of the Algorithm 1.[19] However, even though HC_U is unbiased, it might not minimize the mean squared error compared to other variance estimates. For example HC_U has a larger variance than the related estimator HC_2. Let $\widehat{Y}_i = h_i'Y$ be the fitted value for observation i. Then,

$$HC_U = \frac{Y_i\widehat{u}_i}{1-h_{ii}} = \frac{\left(\widehat{Y}_i + \widehat{u}_i\right)\widehat{u}_i}{1-h_{ii}} = \frac{\widehat{Y}_i\widehat{u}_i}{1-h_{ii}} + HC_2.$$

[15] M can be equal to 3 if we are interested only in the correlation between two variables but can be higher if the model has other covariates and we want to do a variance decomposition.

[16] One does not necessarily need to compute A_m and feed to the algorithm. If the matrix A is, for example, an operator to obtain a sample variance or covariance, one could just compute such sample variance or covariance within the algorithm.

[17] The bootstrap errors will be equal to $v^* = \sqrt{n/(n-k)}\,\widehat{u}^*$ where $\widehat{u}^*$ is the vector of resampled residuals.

[18] A textbook exposition of these issues can be found in Chapter 8 of Angrist and Pischke (2008).

[19] Negative estimates of individual variances are also prevalent in KSS.

The expectation of the first term is zero but HC_U has a higher variance than HC_2. $\widehat{Y}_i \widehat{u}_i$ is a random variable with positive covariance with $\widehat{u}_i^2$ and increases the variance of HS_2. We could define a mixed estimator, HC_M, that takes values of HC_U whenever they are positive and use HC_2 when the estimator HC_U is negative. In other terms,

$$HC_M = \begin{cases} HC_U & \text{if } HC_U \geq 0 \\ HC_2 & \text{otherwise.} \end{cases}$$

We need to compute the leverage h_{ii} for each observation for either the HC_2 or HC_M estimation and that moreover we need that they are smaller than 1 for every observation. In the following we describe how we ensure that the leverages are below the unity, how we estimate them and finally propose a diagnostic and an adjustment for our estimates.

Leave-one-out connected set. In two-way fixed effect models those are only estimated at the connected set. In the application to the labor market, firm fixed effects are only identified by firm movers. Those therefore determine the connected set of firms whose fixed effect can be identified. The need of having $h_{ii} < 1$ for all i requires that no single observation is necessary to estimate a particular fixed effect. This implies that eliminating any observation, the set of fixed effects in the connected set remain the same. We achieve this by pruning the data to get the leave-one-out connected set. It is analogous to Kline et al. (2019) and we leave the details for the Appendix.

Estimation of leverage. When the number of covariates is large, the direct computation of the leverage, by using the diagonal of the projection matrix H is computationally infeasible, as it is a function of the inverse of a very large matrix $(X'X)^{-1}$.

Given the definition of fitted values $\widehat{Y} = HY$, we have that the leverage of observation i is equal to

$$h_{ii} = \frac{\partial \widehat{y}_i}{\partial y_i}.$$

The following remark shows how to compute these leverages.

Remark 1. *Let $\tilde{Y}(i)$ be a vector of length n where every entry is equal to zero, except the ith entry that is equal to one. The leverage of observation i is equal to the fitted value $\widehat{y}_i$ of a linear regression of $\tilde{Y}(i)$ on X.*

The argument is as follows. h_{ii} being a linear function of y_i, the partial derivative $\frac{\partial \widehat{y}_i}{\partial y_i}$ is just a slope. Consider an initial scenario where all entries of the dependent variable Y are equal to zero. In that case all of the fitted values are equal to zero. Then, change the ith entry of Y to 1 and the rest are zero. We can compute a new vector of fitted values $\widehat{Y}'$. Then the leverage, that is partial derivative of a linear function is equal to $\frac{\partial \widehat{y}_i}{\partial y_i} = \frac{\widehat{y}'_i - 0}{y'_i - 0} = \frac{\widehat{y}'_i}{1} = \widehat{y}'_i$.

Recovering the estimates of a linear regression is very efficient nowadays and in principle we could compute the leverages one by one what would involve n regressions. When the data set is large, this is clearly not plausible. Instead, we propose a way to estimate the leverage of each observation that is similar to our bias estimator. Generate the endogenous variable ω where each entry is i.i.d. with (conditional) mean equal to zero and (conditional) variance equal to 1. Projecting it into X, we have

$$\mathbb{E}\left(\widehat{y}_i^2 | X\right) = x_i \left(X'X\right)^{-1} X' \mathbb{E}\left(\omega\omega' | X\right) X \left(X'X\right)^{-1} x'_i = x_i \left(X'X\right)^{-1} x'_i = h_{ii},$$

where x'_i is the ith row of matrix of covariates X. Let n_h the number of simulations for the vector ω used to estimate the leverages $\widehat{h}_{ii}$. Similarly to Proposition 4, we simulate different vectors of the dependent variable

ω, compute the fitted values for each simulation j and then take a sample mean across all the simulations $j = \{1, ..., n_h\}$ of ω.[20]

Diagnostic and adjustment. We use a diagnostic on the estimation by computing a lower bound for each leverage. Let $\tilde{X} = X\mathbf{1}$, meaning X_S is a vector of length n where each entry is the sum of the row elements of matrix X. The diagonal entries of $\tilde{H} = \tilde{X}\left(\tilde{X}'\tilde{X}\right)^{-1}\tilde{X}'$, which are equal to $\tilde{h}_{ii} = \tilde{x}_i^2 / \sum_{i=1}^{n} x_{S,i}^2$. We can perform a diagnostic of our estimates for the leverage by comparing them with $\tilde{h}_{ii}$. Those that are underestimated can then be directly computed using the result of Remark 1.

We can then use these estimated leverages to construct variance estimates HC_2 and HC_M by substituting $\frac{1}{1-h_{ii}}$ with $\frac{1}{1-\hat{h}_{ii}}\left(1 - \frac{1}{(1-\hat{h}_{ii})^2}\frac{\widehat{\text{var}}(\hat{\mathbf{y}}_i^2)}{n_h}\right)$, where the last term corrects for a non-linear bias with $\widehat{\text{var}}(\hat{h}_{ii})$ being a sample variance of the different estimates of the squared fitted values.

Algorithm 4 in Appendix 2.C shows takes as inputs the covariates X and gives output a combination of actual and estimated leverages, as well as the variance $\widehat{\text{var}}(\hat{h}_{ii})$ for the non-linearity adjustment.

Clustered errors and serial correlation

When the error terms are clustered or present serial correlation within group, the covariance matrix is not diagonal any more. We restrict our attention to dependence of the errors only within the group. The variance covariance matrix is not any more diagonal as there are non zero elements around the diagonal corresponding to the dependence of the errors within the group g.[21] One particular example is when the group is a worker-firm match and errors are autocorrelated within match. When the errors present dependence within the group we adapt the bootstrap from Algorithm 1 to a block bootstrap. Algorithm 2 describes the procedure for our bias estimator that keeps the dependence structure through a block bootstrap. At the moment we are exploring the Sieve bootstrap from Davidson and MacKinnon (2006) that would improve the performance of our current procedure.

[20] This is exactly the way Kline et al. (2019) estimate the leverage in their paper. However, they directly solve for the normal equations of the regression using the preconditioned conjugate gradient method, which can be less efficient than just computing the fitted values as we do.

[21] Assume that the the errors have a first order autocorrelation within group g and the true innovations are i.i.d. and therefore homoscedastic. We consider that the error term u of worker i at group g at time t in (2.1) is:

$$u_{i,g,t} = \rho u_{i,g,t-1} + \varepsilon_{i,g,t}, \quad \varepsilon_{i,g,t} \ \text{i.i.d.}$$

We denote the variance of the innovation ε as σ_ε^2. Ordering the data by group, suppose the first group has 3 observations and the second one two, $\mathbb{V}(u|X)$ is:

$$\mathbb{V}(u|X) = \frac{\sigma_\varepsilon^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \rho^2 & 0 & & \cdots & & 0 \\ \rho & 1 & \rho & \vdots & & \ddots & & \vdots \\ \rho^2 & \rho & 1 & 0 & & \cdots & & 0 \\ 0 & \cdots & 0 & 1 & \rho & 0 & \cdots & 0 \\ & & & \rho & 1 & 0 & \ddots & \\ \vdots & \ddots & \vdots & & & \ddots & & 0 \\ 0 & & 0 & & & & & 1 \end{pmatrix}.$$

The covariance matrix under clustering of the errors is similar but with al non-zero elements out of the diagonal equal to ρ.

2.3.2 Simple example

We do some Monte Carlo simulations to illustrate the effectiveness of our bias correction method. The model design is the same as in (2.1) with homoscedastic errors and sample size $n = 500$. The number of covariates is $k_1 = k_2 = 200$. We keep this number relatively low so we can compute the direct correction. In total we do 10,000 simulations. In each simulation, conditioning on X, we draw new error terms to form the dependent variable. We estimate $\widehat{\beta}$ and compute the ideal bias correction terms. After the estimation, we perform $n^* = 100$ bootstraps and use them to compute the estimation of the bootstrap correction terms.[22]

Figures 14 and 15 compare the distributions of the bias of the variance and covariance of the naive plug-in (i.e. non-corrected) estimates ($\widehat{\sigma}^2_{1,PI}$ and $\widehat{\sigma}_{12,PI}$) and the bootstrap corrected estimates ($\widehat{\sigma}^2_{1,b}$ and $\widehat{\sigma}_{12,b}$). The Figure shows that the distribution of the bias (i.e. the difference between the bootstrap corrected and the true moment) of the bootstrap corrected moment is centered at zero. We achieve an unbiased estimator using our correction but the distribution of the naive plug-in estimate is shifted to the right (left) for the variance (covariance).

Table 22 presents the mean and variance of the difference between our bootstrap method and the direct correction. The mean differences are very small as well as the variances, meaning that the estimated bootstrap correction is performing well in comparison to the direct correction in almost all simulations.

Table 22 also shows the Mean Squared Error (MSE) between the different estimated moments and the true ones. The MSE of naive plug-in estimators is larger than the one obtained with the directly corrected and bootstrap corrected moments. As our estimator is a random variable, the MSE of the directly corrected moments are always smaller than the ones with the estimated bootstrap correction, although very close.

2.3.3 Choosing the number of bootstraps

In the previous simple example we arbitrarily chose the number of bootstraps. In practice, given the computational burden of the procedure, we might want to discipline a little more how to choose this number. Our estimator $\widehat{\delta}_b$ is a sample mean estimate of the ideal bias correction term $\widehat{\delta}$. We can then use standard results from probability theory to guide our choice on the number of bootstraps. In particular, we exploit the information given by Chebyshev's inequality.

In Proposition 4 we show that $\mathbb{E}_{v^*}\left(\widehat{\delta}_{b,j}|X,u\right) = \widehat{\delta}_b$. Now assume that $\mathbb{V}(\widehat{\delta}_{b,j}|X,u) = \eta^2 < \infty$. As $\widehat{\delta}_b$ is a sample mean over a sequence of $\{\widehat{\delta}_{b,j}\}_{j=1}^{n^*}$, we have that $\mathbb{E}_{v^*}(\widehat{\delta}_b|X,u) = \widehat{\delta}$ (as shown in Proposition 4) and $\mathbb{V}(\widehat{\delta}_b|X,u) = \frac{1}{n^*}\eta^2$.[23] Then, by Chebyshev's inequality we have

$$\mathbf{P}\left(\left|\widehat{\delta}_b - \widehat{\delta}\right| \geq k\frac{\eta}{\sqrt{n^*}} \;\middle|\; X,u\right) \leq \frac{1}{k^2}.$$

Then one can choose the number of bootstraps n^* such that the distance between the bootstrap estimate $\widehat{\delta}_b$ and the ideal bias correction term $\widehat{\delta}$ is greater or equal than λ standard deviations with probability smaller than α. So, for arbitrary $\alpha > 0$ and $\lambda > 0$ we have

$$\frac{1}{k^2} = \alpha, \quad \frac{k}{\sqrt{n^*}} = \lambda.$$

[22] We use the covariance estimator HC_1 and therefore skip the part of computing the leave-one-out connected set.

[23] We have that $\mathbb{V}(\widehat{\delta}_b|X,u) = \frac{1}{n^{*2}}\mathbb{V}(\sum_j^{n^*}\widehat{\delta}_{b,j}|X,u) = \frac{1}{n^{*2}}\sum_j^{n^*}\mathbb{V}(\widehat{\delta}_{b,j}|X,u) = \frac{1}{n^*}\eta^2$ where we used the independence of different $\widehat{\delta}_{b,j}$ conditional on X and u.

Solving for n^* we get $n^* = \frac{1}{\alpha\lambda^2}$. So if, for example, we set $\alpha = 0.05$ and $\lambda = 1/2$ we get that the number of bootstraps such that the distance between the bootstrap estimate and the ideal correction term is greater than half a standard deviation is an event with a probability smaller than 5 per cent is $n^* = \frac{1}{0.05\times(1/2)^2} = 20 \times 4 = 80$. One could be more conservative and set $\lambda = 0.1$. In that case, we would obtain $n^* = 20 \times 1000 = 2000$ bootstraps.

Admittedly, the number of bootstraps suggested by inequality for any α and λ can be quite conservative. But this just reflects the generality of the result. Indeed, this criteria would work regardless the distribution of v^*, therefore regardless the choice of bootstrap.

2.4 Comparison of Methods

In this section we first compare our method to Gaure (2014), Kline et al. (2019) and Borovičková and Shimer (2017). The closest methods to ours are the ones by Gaure (2014) and Kline et al. (2019). All three aim to compute the trace term in equation (2.3). On the contrary, Borovičková and Shimer (2017) propose a method to compute the correlation of theoretically different worker and firm types. Second, we present results of Monte Carlo simulations of labor markets to compare the methods under different assumptions on the error terms.

The differences between Gaure, KSS and our method are on the scope of distribution of errors allowed, the covariance matrix estimation and the flexibility of application. All three methods are in principle suited to perform corrections with homoscedastic and heteroscedastic errors. Nevertheless, Gaure only implemented his bias correction method on the R package *lfe* under the assumption of homoscedastic errors. Moreover, KSS and us provide corrections under serial correlation or clustering of the errors. Second, the methods differ on the covariance matrix estimator they use. Gaure uses HC_0 directly estimating the variance from the residuals. As explained in Section 2.3.1, KSS estimate the covariance matrix by HC_U and our baseline application is with HC_2 even if we explore other covariance matrix estimates. Finally, our method is more flexible than Gaure and KSS in the correction of several second order moments at a time. Adding additional moments to correct (e.g. the variance of occupation fixed effects and their covariance with firm and worker types) does not increase the computational burden of the correction. KSS and Gaure on the contrary need to compute new sets of normal equations per additional correction.

In the following we give some detail of an alternative method to compute the correlation between the types of matched workers and firms by Borovičková and Shimer (2017). Their method completely bypasses the need to estimate a linear model and therefore avoids using noisy estimates, which are the source of the bias, to compute the correlation.

As explained by BS, the worker and firm types that they define are different than the types defined in the AKM model. In BS, a worker's type, denoted λ_i, is defined to be the expected log wage of the worker, while a firm's type, denoted $\mu_{J(i,t)}$, is defined to be the expected log wage that it pays. In contrast, in the AKM model, a worker and firm types $(\theta_i, \psi_{J(i,t)})$ are defined as such that a change in type will change the expected log wage while holding fixed the partner's type.[24]

[24] We refer to an older version of the Borovičková and Shimer where they provide a way to translate the variances and covariances of their worker and firm types to the ones in AKM. In the latest version, they slightly changed their estimator and do not provide this link any more.

BS show that their correlation and the AKM correlation, ρ, will be the same if the joint distribution of θ and ψ is elliptical (e.g. a bivariate normal) and $(\sigma_\lambda - \rho\sigma_\mu)(\sigma_\mu - \rho\sigma_\lambda) > 0$, where σ_λ and σ_μ are, respectively, the standard deviations of worker and firm types. With these assumptions, there is also a direct correspondance between the standard deviation of AKM types and BS types:[25]

$$\sigma_\theta = \frac{\sigma_\lambda - \rho\sigma_\mu}{1-\rho^2}, \quad \sigma_\psi = \frac{\sigma_\mu - \rho\sigma_\lambda}{1-\rho^2}.$$

The key identifying assumption in the BS method is that for each worker and firm they have two or more observations of the actual wage (received or payed) which are independent and identically distributed conditional on the types. In AKM, the identifying assumption is a standard exclusion restriction, i.e. that the error term has mean zero conditional on the types (and other covariates).

2.4.1 Labor market simulations

We compare the correction methods by simulating many labor markets under different assumptions on the error terms. We evaluate the methods in terms of computation time and mean squared errors. We also explore differences between the covariance estimation methods described in Section 2.3.1.

We compare all the methods under conditional homoscedasticity of the errors. Results are in Table 23. All the methods improve the initial bias of the plug-in estimate. The least accurate method is BS reducing by 78% the MSE of the naive estimates whereas the other three methods reduce it by 97%.[26] The objective of BS is to provide an estimate of the correlation and they base their estimation in different worker and firm types (λ and μ respectively). Table 23 presents their estimates of the AKM types. Under the assumption of linearity of conditional expectations, the correlation of their types $\rho_{\lambda,\mu}$ is a good estimator of of the correlation $\rho_{\theta,\psi}$ but their types are not suited to perform a variance decomposition. We find that the MSE taking their types are orders of magnitude bigger.[27] Gaure, KSS and our method are very similar in terms of MSE, Gaure being slightly more accurate than the other two.[28] Figure 17 shows the distribution of the bias of the firm variances for the naive estimate ($\sigma^2_{\theta,PI} - \sigma^2_\theta$) and the different correction methods. We see that our method is very similar to KSS and both are the ones with lowest biases. Even if the bias of Gaure is higher, his method has lower variance and outperforms KSS and ours in terms of MSE. Regarding the computation time, BS is the fastest one with computation time of less than a second. Gaure is the one performing best among the three closest competitors (Gaure, KSS and our method) as it has the lowest computing time and MSE yet the latter is comparable.[29]

Table 24 presents the comparison of our method to KSS under conditional heteroscedasticity for different degrees of mobility. Both methods are similar in accuracy and reduce by roughly 84% the MSE of the plug-in estimate in the low mobility case.[30] Our method is slightly more accurate for both mobility cases but also

[25]See Proposition 1 in Borovičková and Shimer (2017).

[26]We wrote the code for BS following Borovičková and Shimer (2017) and converting the data to the match level.

[27]The scaled MSE (MSE $\times 10^2$) of $\sigma^2_\lambda, \sigma^2_\mu$ and $\sigma_{\lambda,\mu}$ are respectively 51.4, 76.5 and 2.40.

[28]Gaure is corrected using the *bccor* with 300 maximum samples and tolerance of 1e-2. We run Version 2.15 of the KSS code eliminating observations (instead of matches) for the leave-one-out estimation and with *epsilon* parameter of 0.05. This translates into number of simulations p equal to 289. This guided our choice of 300 simulations to estimate the leverages and the bias corrections. We use the R package *fixest* for our regressions with default tolerance of 1e-5.

[29]KSS and our method do not incorporate the simplifications that come from having homoscedastic errors.

[30]Table 1 in Kline et al. (2019) shows that their connected set is similar to our low mobility scenario with 2.7 movers per firm and average firm size of 12.

more time consuming. Figure 18 shows the distribution of the bias of the plug-in estimate, KSS and our method. Both corrections are indistinguishable but the bootstrap method has smaller variance as shown in Section 2.3.1. Table 25 compares the different covariance matrix estimators applicable with our method. All the estimators have similar MSE but HC_2 outperforms the rest.

Table 26 presents results from a simulation with a non diagonal covariance matrix. In particular we assume that there is serial correlation of the wages within a given match. We compare the plug-in estimate to our bootstrap method with the match as unit of observation Boot, the block bootstrap method from Algorithm 2 Boot Block, and the KSS correction methods. We present results once we have eliminated 1% of the simulation due outliers in the KSS estimation.[31] The best performing correction method is Boot both in terms of time and MSE.[32] After the elimination of outliers, KSS improve the MSE of the naive estimates. On the contrary, the block bootstrap from Algorithm 2 worsen the naive estimation due to biases on the corrected variance of firm fixed effects and the covariance. We are currently exploring the Sieve bootstrap (e,.g. Davidson and MacKinnon (2006)) to improve the estimation under serial correlation. Results from Table 26 suggest that the interested reader working with serially correlated observations within group should transform the data to the group level and use our bootstrap bias correction method.

2.5 Application

An important application of two-way fixed effect models are the AKM type log wage regressions with worker and firm fixed effects. We closely follow Card et al. (2013) to implement the estimation of the following regression model for the log of the wage of worker i at time t:

$$w_{it} = \theta_i + \psi_{J(i,t)} + q_{it}\gamma + \varepsilon_{it}, \tag{2.5}$$

where the function $J(i,t)$ gives the identity of the unique firm that employs worker i at time t, θ_i is a worker fixed effect, $\psi_{J(i,t)}$ is the premium for all employees at firm $J(i,t)$, q_{it} are time varying observables (age and education interacted with year effects), and ε_{it} is the error term.

Equation (2.5) can be estimated by OLS. The person/firm fixed effect estimators have the same structure as the one in Section 2.2. Thus the second order moments have the same structure and the implementation of the correction is analogous.

The data is a panel from INSEE, the French statistical agency, from 2002 to 2014.[33] Our dependent variable is (log) gross daily wage of full time employees with ages between 20 and 60. To maximize the number of firms at the largest connected set, we estimate the model using both private and public firms but we exclude public firms from the analysis and correction.[34]

The goal is to use our bootstrap method to do a bias corrected variance decomposition of log wages. In order to do so we have to pick the number of bootstraps to perform. To get an idea of how many bootstraps

[31] KSS estimated negative variance of worker or firm effects in 0.8% of the simulations. Moreover, there are outliers where the estimated variance of worker fixed effects explodes. We keep fixed X and simulate 1000 serially correlated vectors of errors. The reason behind this instability is most likely the shortness of our sample together with their covariance matrix estimator HC_U.

[32] For Boot we use an HC_1 estimate of the covariance matrix and avoid the pruning of the data.

[33] In particular we use *Panel tous salariés-EDP* that consists of a random subsample of workers with firm identifiers and socio-demographic variables. The sample consists of workers born in October in certain days. The sample size was multiplied by 8 in 2002 so we took this as the starting year.

[34] We bundle all workers working at the public sector to have a unique firm identifier.

we should do, we perform some simulations with a fixed set of covariates that will give us approximately three movers per firm (as in our dataset for the application) and simulate one hundred datasets with different wages by just simulating the error. Then, with each dataset we perform corrections from one to one hundred bootstraps. We then compare the increase in precision by increasing the number of bootstraps to obtain the correction. Figure 16 shows the mean squared error between the true moment and the corrected one for different number of bootstraps for the covariance of person and firm fixed effects. This means that for all the samples we take the corrections obtained with one, two, three, etc. bootstraps and take the mean squared error against the true moment. The figure shows that with few bootstraps the MSE reduces significantly and around 10 it flattens. Given our sample size of around 60,000 observations and our sample size in the application of around 7 million, this gives us a benchmark of how many bootstraps to do for the application. This would be (10/60,000)*7 million $\approx$ 1000.

Table 27 shows the variance decomposition of log wages as well as the correlation between firm and worker fixed effects using the naive moments and the corrected ones.[35] The variance of the person and firm effects are both reduced and they explain a lower share of the total variance after the correction. The correlation becomes closer to zero and it approaches values that have been found in other countries with larger number of movers per firm, which should attenuate the bias, as reported by Table 1 of Lopes de Melo (2018). Naturally, the variance and covariance of the person and firm effects are the moments that change the most after the correction. The reason is that the underlying estimates of the person and fixed effects are very noisy. In contrast, when the underlying estimates of a particular moment are estimated with precision, as it is the case for the parameters $\widehat{\gamma}$ associated with the common covariates $\mathbf{q}$, the change between the naive and corrected moments is negligible.

To fully exploit the benefit of our bootstrap correction method we also perform a yearly variance decomposition. In Figure 19 we compare the year-to-year evolution of the different explained shares using the naive estimated moments and the corrected ones. The main takeaway from this figure is that the correction changes the levels but not the slopes of explained shares. This leads to a change in the relative importance of each component. In particular, the corrected variance of the residuals is relatively more important than the corrected variance of the firm effect in almost every year, while the opposite happens when considering uncorrected variances. A very interesting trend is the decline in explanatory power of the individual fixed effects for recent years. It might be just a feature of the French data. Explanations for this phenomenon are outside the scope of this paper.

2.5.1 Comparison of Methods

We compare our method to BS using the French data. Adapting to their method, instead of using annual wage data, we first average all the wage data to the worker-firm match level. We do this because, as mentioned in their paper, annual wage observations might not be independent conditional on type, especially for workers who do not switch firms. In order to accommodate for the extra covariates in the BS method, we first run a

[35] Due to COVID-19 related access restrictions to the university and therefore to the data, in the application we implement a previous version of our bootstrap correction. We use the HC_1 covariance matrix estimator and avoid pruning of the data. We use the MacKinnon and Smith Jr (1998) with an algorithm similar to 1. We generate a new bootstrap dependent variable $Y^* = X\widehat{\beta} + v^*$ where $\widehat{\beta}$ is the original estimate of β and project it into X. Due to the fact that the bootstrap replicates the structure of the original estimation, the estimator of the bias of moment m in bootstrap j is: $\widehat{\delta'}^{(j)}_{aux,m} = (\beta^*)' A_m \beta^* - \widehat{\beta}' A_m \widehat{\beta}$ for all $m \in \{1, ..., M\}$.

linear regression of log wage versus q_{it} (age and education interacted by year effects) and take the residual. Only after we use this residual-wage to average at the worker-firm match and use this as the dependent variable to compute the moments, both for the BS and our bootstrap method. Using averages at the match level is interesting for our bootstrap method as well, as it will take care of any serial correlation issues at the match level.

Table 28 compares the estimated moments using the BS method and the bootstrap correction method on the French data. Both columns report the moments using the AKM defined worker and firm types. In contrast with the simulated data, where we were satisfying the assumptions for both estimated correlations to be the same, when using French labor market data, both estimates differ by a large amount. The bootstrap corrected estimated correlation is 0.16, well below the estimated one using BS method, 0.55.[36] Looking at each of the components of the correlation, both variances are larger and the covariance is smaller when using the bootstrap corrected method instead of BS method.[37]

There are different reasons why both estimates might differ. To begin with, the types defined by BS are fundamentally different from the ones defined in the AKM model. They are only related when the assumptions stated at the beginning of this section are satisfied. It might be that the two correlations are not comparable because, even if the log-linear AKM model is correctly specified, these assumptions are violated, in particular, if the joint distribution of AKM types is not elliptical. Second, it might be that the identification assumption of at least one of the methods fail. It's hard to think of examples where an identifying assumption for a particular method holds while failing for the other. It's easier to think of examples where *both* identification assumptions are violated. For example, whenever there is selection of workers via the error term, some matches will be formed whenever this idiosyncratic component is high. This endogenous mobility would then violate both the AKM and BS identification assumptions.

Finally, a potential reason why the bootstrap estimated moments reported in Table 28 differ from the estimated moments previously reported in Table 27, is that if the annual wage error term exhibits serial correlation, then averaging at the match level and performing our correction should give a bias corrected moment. However, in section 2.5 we don't average at the match level, and the bootstrap procedure used, while consistent with general heteroscedasticity, is inconsistent with serial correlation. Then the assumption that the variance of the bootstrapped residuals $\mathbb{V}(v^*|X)$ is equal to sample variance $\widehat{\mathbb{V}}(u|X)$ and the estimated correction would be incorrect. The problem with averaging at the match level is that we can not perform yearly decompositions. One could then adapt the bootstrap procedure, such that is consistent with serially correlated errors without the need of averaging at the match level.

2.6 Conclusion

In this paper, we propose a computationally feasible bootstrap method to correct for the small-sample bias found in all quadratic forms in the parameters of linear models with a very large number of covariates. We show using Monte Carlo simulations that the method is effective at reducing the bias. Its application to a real labor market dataset showed that it increases the correlation between firm and worker fixed effects and changes the relative importance of the different components that explain the variance of log wages.

[36] The BS estimates are obtained by using the formulas of Section 5.2. in Borovičková and Shimer (2017).

[37] The estimated variances for BS defined worker and firm types are: $\sigma^2_\lambda = 0.0842$ and $\sigma^2_\mu = 0.0411$.

The only requirements to implement our correction is to have a bootstrap procedure that is consistent with the assumption on the variance-covariance matrix of the error term and being able to estimate the model several times. The correction can thus easily be applied to any study running an AKM type regression.The main advantages of the method we propose here are mainly its flexibility to make yearly corrections or to increase the number of moments of interest to correct without increasing the computational costs.

Comparisons to other models through Monte Carlo simulations show that, in terms of accuracy, our method is comparable to Gaure (2014) and Kline et al. (2019) but offers the advantage of flexibility to incorporate the correction of additional moments at no cost. The comparison to Borovičková and Shimer (2017) using Monte Carlo experiments showed that both are similar but our method is more accurate than theirs in simulated data that fulfill the assumptions of both approaches. However, when applied to French administrative data, the methods yield different estimates for the correlation and all its components. This suggests that the assumptions of one or both methods do not hold in the French labor market data. Further exploration would be required to disentangle the origin of this discrepancy.

2.A Proofs

Proposition 4.

Proof. First, note that for any bootstrap estimate of the quadratic form $\beta_j^{*\prime} A \beta_j^*$ we have that

$$\beta_j^{*\prime} A \beta_j^* = (v_j^*)' Q' A Q v_j^*.$$

Under the bootstrap, the only source of randomness is v_j^*. Taking expectations under the bootstrap of $\beta_i^{*\prime} A \beta_j^*$, conditionally on X and u, we get

$$\mathbb{E}_{v^*}\left(\beta_j^{*\prime} A \beta_j^* \,\middle|\, X, u\right) = \text{trace}\left(Q' A Q \mathbb{V}(v_j^* | X)\right).$$

By assumption $\mathbb{V}(v_j^*|X) = \widehat{\mathbb{V}}(u|x)$, then $\mathbb{E}_{v^*}\left(\beta_j^{*\prime} A \beta_j^* \,\middle|\, X, u\right) = \widehat{\delta}$.

Unbiased. Taking expectations over $\widehat{\delta}_b$ conditionally on X and u we obtain

$$\mathbb{E}_{v^*}(\widehat{\delta}_b | X, u) = \frac{1}{n^*} \sum_{j=1}^{n^*} \mathbb{E}_{v^*}\left(\beta_j^{*\prime} A \beta_j^* \,\middle|\, X, u\right) = \frac{1}{n^*} \sum_{j=1}^{n^*} \widehat{\delta} = \widehat{\delta}.$$

Consistent. From the definition of $\widehat{\delta}_b$, we have that

$$\frac{1}{n^*} \sum_{j=1}^{n^*} \left(\beta_j^{*\prime} A \beta_j^*\right) \xrightarrow{p} \mathbb{E}_{v^*}\left(\beta_i^{*\prime} A \beta_i^* \mid X, u\right) = \widehat{\delta}.$$

□

Corollary 2

Proof. Using the Law of Iterated Expectations we get

$$\mathbb{E}(\widehat{\delta}_b | X) = \mathbb{E}_u\left(\mathbb{E}_{v^*}(\widehat{\delta}_b | X, u) \,\middle|\, X\right) = \mathbb{E}_u(\widehat{\delta} | X) = \delta.$$

□

2.B Construction of Simulated Labor Market

We construct several simulated labor markets depending on the number of movers per firm and, the correlation between the worker and firm fixed effects. Here, we briefly describe the construction of the simulated labor markets.[38]

We start by determining the size of the labor market. We have 5000 unique workers and 400 unique firms at the beginning of the sample. This gives an average firm size of 12 workers which is similar to the average firm size in the data of Kline et al. (2019).[39] Their connected set with 2.7 movers per firm is similar to our low mobility simulations with 3 movers per firm. The sample runs for 7 years but we allow that workers randomly drop from the sample with a minimum of 2 observations per worker. This leads to total sample size of roughly 22000 observations.

[38] We thank Simen Gaure for sharing with us a piece of code that we used as a base for the simulations.

[39] See Table 1 in Kline et al. (2019) where each worker is observed twice.

Worker and firm fixed effects are random draws from normal distributions. We assume that there is sorting depending on the permanent types what leads to non negative correlations between worker and firm fixed effects while fulfilling exogenous mobility. That is workers a low type worker is more likely to match with a low type firm if we assume positive sorting but sorting does not depend on match specific shocks. Matches are formed either at the beginning of the sample or afterwards for the movers. Errors are i.i.d. and normally distributed in the baseline simulation with homoscedastic errors. Heteroscedastic errors are also normally distributed with a observation (worker-year) specific variance that is randomly drawn from a uniform distribution. Finally, serially correlated errors are simulated from a first order autoregressive process with persistence of 0.7 and homoscedastic innovations. The simulated log wage is like equation (2.5) without other covariates.

2.C Algorithms

Here we detail the implementation algorithms of our method. Algorithm 1 and 2 describe respectively the estimation of the bias correction for diagonal and non diagonal covariance matrices. Algorithm 3 describes how to prune the data to ensure that the maximum leverage is below 1 and Algorithm 4 details how to estimate the leverage.

Algorithm 1 Estimate $\{\widehat{\delta}_{b,m}\}_{m=1}^{M}$ when the covariance matrix is diagonal

```
for j = 1, ..., n* do
    Simulate a vector r* of length n of mutually independent Rademacher entries.
    Generate a vector of residuals v* of length n whose ith entry is equal to sqrt(ψ̂_i) × r*_i.
    Compute β* as the estimate of a regression of v* on X.
    Compute δ̂^(j)_{aux,m} = (β*)′ A_m β* for all m ∈ {1, ..., M}.
end for
Compute δ̂_{b,m} = (Σ_{j=1}^{n*} δ̂^(j)_{aux,m}) / n* for all m ∈ {1, ..., M}.
```

Algorithm 2 Estimate $\{\widehat{\delta}_{b,m}\}_{m=1}^{M}$ when covariance matrix is non diagonal

```
Let s_i be the set of matches with length i and S = {s_1, ..., s_T̄} be the set of all matches grouped by their
    duration where the maximum duration is T̄.
for j = 1, ..., n* do
    for i = s_1, ..., s_T̄ do
        Sample with replacement the estimated residuals of the whole group g for g ∈ s_i.
    end for
    Stack the different samples to form v*.
    Compute β* as the estimate of a regression of v* on X.
    Compute δ̂^(j)_{aux,m} = (β*)′ A_m β* for all m ∈ {1, ..., M}.
end for
Compute δ̂_{b,m} = (Σ_{j=1}^{n*} δ̂^(j)_{aux,m}) / n* for all m ∈ {1, ..., M}.
```

Algorithm 3 Leave-one-out connected set

Let Λ be the connected set.
$a = 1$.
while $a > 0$ **do**
    Compute the articulation points a.
    Eliminate articulation points a.
    Compute the new connected set Λ_1.
end while

Algorithm 4 Estimate leverages, diagnose and compute underestimated ones

$z_1^{(0)} = \mathbf{0}$ and $z_2^{(0)} = \mathbf{0}$, where $z_1^{(0)}$ and $z_2^{(0)}$ are vectors of length n.
for $j = 1, ..., n^*$ **do**
    Simulate a vector ω^* of length n of mutually independent Rademacher entries.
    Compute fitted values $\widehat{\omega^*}$ from a regression of ω^* on X.
    Compute $z_1^{(j)} = z_1^{(j-1)} + (\widehat{\omega^*})^2$ and $z_2^{(j)} = z_2^{(j-1)} + (\widehat{\omega^*})^4$.
end for
Compute $\widehat{h}_{ii} = z_{1,i}^{(n^*)} / n^*$ for all $i \in \{1, ..., n\}$.
Compute $\widehat{\text{var}}(\widehat{h}_{ii}) = \frac{n^*}{n^*-1} \left(\frac{z_{2,i}^{(n^*)}}{n^*} - \widehat{h}_{ii}^2 \right)$.
Compute $\tilde{X} = X\mathbf{1}$ and then the lower bounds $\tilde{h}_{ii} = \tilde{x}_i^2 / \sum_{i=1}^n x_{S,i}^2$ for all $i \in \{1, ..., n\}$.
for $i = 1, ..., n$ **do**
    if $\widehat{h}_{ii} < \tilde{h}_{ii}$ **then**
        Generate $\tilde{Y}(i) \in \mathbb{R}^n$, where $\tilde{Y}(i)_{j \neq i} = 0$, $\tilde{Y}(i)_i = 1$.
        Compute the fitted values $\widehat{\tilde{Y}}(i)$ of a regression of $\tilde{Y}(i)$ on X.
        Compute actual leverage $h_{ii} = \widehat{\tilde{Y}}(i)_i$.
    end if
end for

2.D Tables and Figures

Table 22 – Results of simple Monte Carlo simulations

	$\widehat{\delta} - \widehat{\delta}_b$		Mean Squared Error			True moment
	Mean	Variance	Naive	Ideal	Bootstrap	
$\widehat{\text{var}}(X_1\beta_1)$	-0.75$\times 10^{-3}$	0.005	37.06	14.15	14.15	61.50
$\widehat{\text{var}}(X_2\beta_2)$	-0.4$\times 10^{-3}$	0.005	60.67	36.93	36.92	368.65
$\widehat{\text{cov}}\,(X_1\beta_1, X_2\beta_2)$	0.52$\times 10^{-3}$	0.004	19.55	8.88	8.89	-4.36

Notes: The first two columns represent, respectively, the mean and the variance of the difference between the ideal correction and the bootstrap correction. Columns 3 to 5 just compute the MSE between the estimated moments and the true ones.

Table 23 – Monte Carlo simulations. Homoscedastic errors

		Mean Squared Error (MSE$\times 10^2$)			
	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.54321	0.23929	0.15156	2.31135
BS	0.3	0.99121	0.55175	0.00952	0.51749
Gaure	1.6	0.05251	0.11862	0.01526	0.06213
Boot	5.7	0.05017	0.12894	0.01684	0.06532
KSS	5.1	0.05297	0.13157	0.01767	0.06741

Notes: *Plug-in* is the naive plug-in estimator, *BS* refers to Borovičková and Shimer (2017), *Gaure* refers to the method Gaure (2014) implemented through the R package *lfe*, *Boot* refers to our method, and *KSS* is the Kline et al. (2019) method. The results of Borovičková and Shimer correspond to the AKM worker and firm types present in the cited version of the paper. The average firm has 10 movers and 12 employees. *Time* is the computing time in seconds. True moments are computed at the largest connected set. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors (MSE) of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled).

Table 24 – Monte Carlo simulations. Heteroscedastic errors

			Mean Squared Error (MSE$\times 10^2$)			
Mov/firm	Model	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Low Mobility						
3	Plug-in		19.74972	1.30053	8.31373	9.78799
3	Boot	5.2	0.20311	3.81797	0.49855	1.50654
3	KSS	4.2	0.23837	3.82256	0.53183	1.53092
Mid Mobility						
5	Plug-in		10.88573	0.62859	2.00381	4.50604
5	Boot	5.0	0.09920	0.99263	0.13651	0.40945
5	KSS	4.5	0.10504	0.98649	0.14078	0.41077

Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method, and *KSS* is the Kline et al. (2019) method. True moments are computed at the largest connected set. *Mov/firm* is the number of movers per firm and the average firm has 12 employees. *Time* is the computing time in seconds. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled).

Table 25 – Comparison of variance estimations

		Mean Squared Error (MSE$\times 10^2$)			
Model	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.39553	0.28303	0.11423	2.26426
Boot HC_0	2.2	0.48494	0.11808	0.02002	0.20768
Boot HC_1	2.2	0.47823	0.12019	0.02116	0.20652
Boot HC_2	6.2	0.48155	0.11799	0.01852	0.20602
Boot HC_M	6.2	0.48625	0.13513	0.02165	0.21434

Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method. True moments are computed at the largest connected set. *Model* is the model and type of variance estimator. *Time* is the computing time in seconds. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. All the MSE are multiplied by 100. *Average* is the average MSE (also scaled).

Table 26 – Monte Carlo simulations. Serial correlation

		Mean Squared Error			
	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		76.98325	3.12421	0.12754	26.74500
Boot	1.1	2.75531	0.30329	0.30241	1.12034
Boot Block	3.7	0.32975	96.19891	7.56496	34.69787
KSS	5.8	17.48475	0.43791	0.07621	5.99962

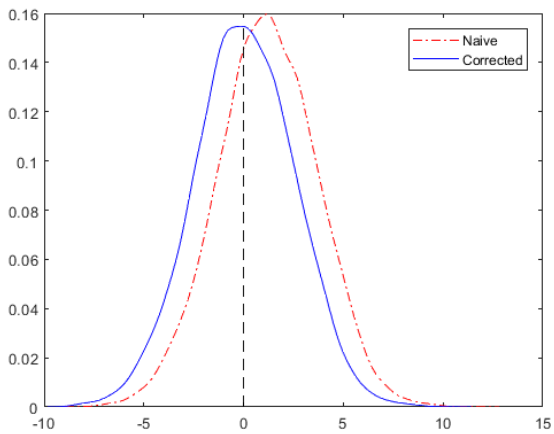
Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method with HC_1 covariance estimator where the unit of observation is the match and wages are transformed to average match wage. *Boot Block* refers to our method with a block bootstrap where each match defines a block. In both, *Boot* and *Boot Block* we skip the pruning of the data. *KSS* is the Kline et al. (2019) method. The average firm has 10 movers and 12 employees. *Time* is the computing time in seconds. True moments are computed at the largest connected set. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors (MSE) of the corrected estimates of the variance of the worker fixed effects, variance of the firm fixed effects and the covariance between worker and firm effects. *Average* is the average MSE.

Table 27 – Naive vs corrected decomposition of log wages

	Variance component		Explained shares	
	Naive	Corrected	Naive	Corrected
$Var(y)$	0.21	0.21	1	1
$Var(\widehat{\theta}_i)$	0.13	0.12	0.62	0.56
$Var(\widehat{\psi}_j)$	0.04	0.03	0.20	0.15
$Var(\mathbf{q}\widehat{\gamma})$	0.01	0.01	0.07	0.07
$Var(\widehat{\epsilon})$	0.027	0.033	0.13	0.16
$2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.03	-0.01	-0.12	-0.06
$2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$	0.02	0.02	0.10	0.10
$2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$	0.00	0.00	0.01	0.01
$Corr(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.19	-0.11		

Notes: *Naive* refers to the uncorrected estimates of each of the variance components and *Corrected* refers to the estimates after our bootstrapped correction. $Var(y)$ is the variance of log wages, $Var(\widehat{\theta}_i)$ the variance of worker fixed effects (naive $\hat{\sigma}^2_\theta$ or corrected $\tilde{\sigma}^2_\theta$), $Var(\widehat{\psi}_j)$ is the variance of firm fixed effects, $Var(\mathbf{q}\widehat{\gamma})$ is the variance of other covariates and $Var(\widehat{\epsilon})$ is the variance of the error term. The other terms of the decomposition are twice the covariances between the fixed effects and the covariates ($2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$, $2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$ and $2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$). Finally, $Corr(\widehat{\theta}_i, \widehat{\psi}_j)$ is the estimated correlation between worker and firm fixed effects.

Figure 14 – Density of $\widehat{\sigma}^2_{1,PI} - \sigma^2_1$ and $\widehat{\sigma}^2_{1,b} - \sigma^2_1$



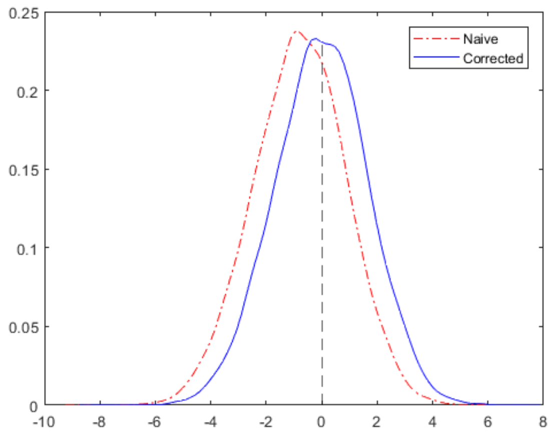
Notes: This figure presents the distributions of the differences between the true variance σ^2_1 and both, the naive plug-in estimated variance $\widehat{\sigma}^2_{1,PI}$ and the bias corrected estimated variance $\widehat{\sigma}^2_{1,b}$. The distribution of the difference between the true moment and the bias corrected estimated covariance is centered at zero.

Table 28 – Comparison of the Methods. French Data

	Borovičková & Shimer	Bootstrap HC_1
$\widehat{\sigma}^2_\theta$	0.066	0.1242
$\widehat{\sigma}^2_\psi$	0.0036	0.0189
$\widehat{\sigma}_{\psi,\theta}$	0.0085	0.0077
$\widehat{\rho}_{\theta,\psi}$	0.55	0.16

Notes: The results of Borovičková and Shimer correspond to the AKM worker and firm types. *Bootstrap* HC_1 are the results of our method under covariance matrix estimator HC_1 without leave-one-out. $\widehat{\sigma}^2_\theta \times 10^2$ and $\widehat{\sigma}^2_\psi \times 10^2$ are respectively the estimates of the variance of worker and firm fixed effects multiplied by 100. $\widehat{\sigma}_{\psi,\theta}$ is the covariance and $\widehat{\rho}_{\psi,\theta}$ the correlation between worker and firm fixed effects.

Figure 15 – Density of $\widehat{\sigma}_{12,PI} - \sigma_{12}$ and $\widehat{\sigma}_{12,b} - \sigma_{12}$



Notes: This figure presents the distributions of the differences between the true covariance σ_{12} and both, the naive plug-in estimated covariance $\widehat{\sigma}_{12,PI}$ and the bias corrected estimated covariance $\widehat{\sigma}_{12,b}$. The distribution of the difference between the true moment and the bias corrected estimated covariance is centered at zero.

Figure 16 – MSE of corrected $\widehat{cov}(\theta, \psi)$ by number of bootstraps

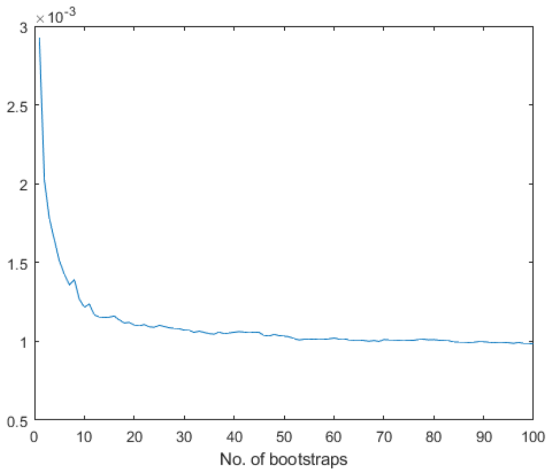
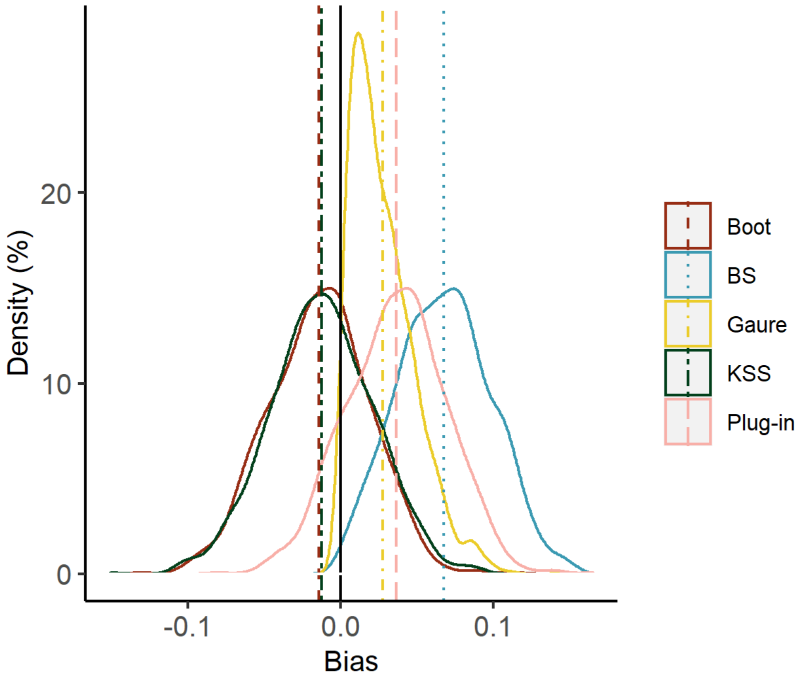
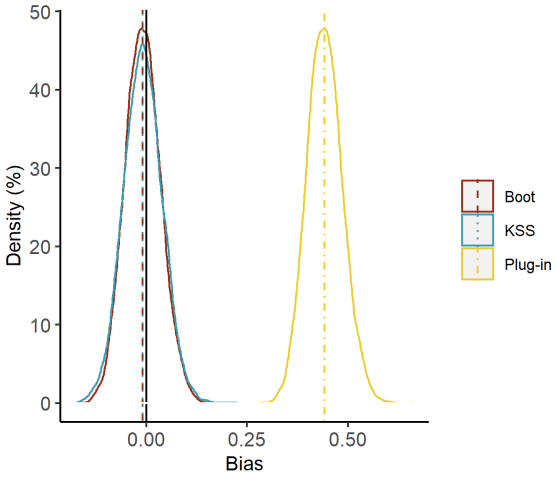


Figure 17 – Model Comparison: Homoscedastic Errors

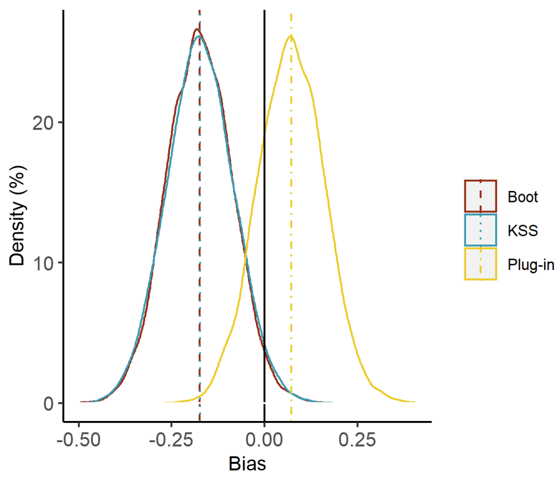


Notes: This figure presents the distributions of the bias of $\hat{\sigma}^2_\psi$ for the naive plug-in estimate and the corrected moments for the different methods. Simulated errors are homoscedastic and labor mobility is high.

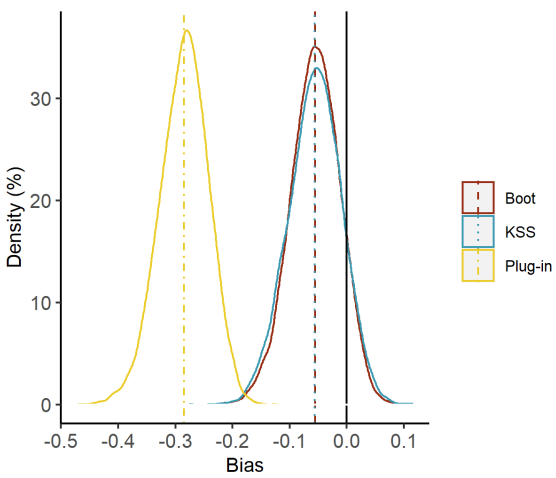
Figure 18 – Model Comparison: Heteroscedastic Errors



(a) Bias of $\widehat{\sigma}^2_{\theta}$



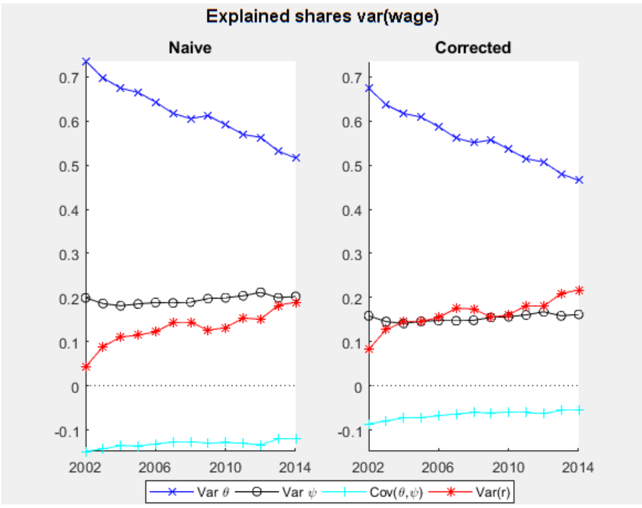
(b) Bias of $\widehat{\sigma}^2_{\psi}$



(c) Bias of $\widehat{\sigma}_{\theta,\psi}$

Notes: These figures present the distributions of the bias for the naive plug-in estimate and the bias of corrected moments for *KSS* and our method. Simulated errors are heteroscedastic and labor mobility is low.

Figure 19 – Evolution of the explained shares.



Notes: This figure presents the year-to-year evolution of the explained shares of the total log wage variance, with and without the correction, for the person and firm fixed effects, as well as the residual.

Chapter 3

Peer Effects: Network and Learning

Miren Azkarate-Askasua[1]

Abstract

How do workplace peers affect each other? How do peers determine a worker's location and future wage profile? This paper empirically disentangles if workplace peers affect each other through learning or network effects. Similarly to the literature, I document the importance of learning. In particular for the youngest cohorts arguably with no networks. I propose a structural model to understand the mechanism behind learning and its impact on the rising between firm wage inequality in the labor market and the geographical allocation of workers.

JEL Codes: J31, J24, E24

Keywords: Peer effects, wage setting, knowledge diffusion

[1] I am grateful to Christian Hellwig for his guidance through the project. I thank Isis Durrmeyer and participants at the TSE Macro workshop for useful comments. First version: March 2020. All errors are mine.

3.1 Introduction

How do workplace peers affect each other? How do peers determine a worker's location and future wage profile? This paper explores how interactions with coworkers affect the learning or human capital accumulation and how the network of coworkers improve the labor market status through an improvement in outside options. Understanding the nature of peer interaction is important as network search can lead to inefficient outcomes (see Chandrasekhar et al., 2020) and peer learning determines team formation, the location of those teams and finally wage inequality across firms. Both mechanisms can reinforce geographical differences of labor market conditions if they happen at the same time.

Learning leads to a gain in human capital that will be carried by the worker independently to future situations of the coworkers. A close example is that academics greatly improve their knowledge through interactions with their peers. On the contrary, the network effect can be in place only in situations where a worker hears from job opportunities thanks to the past coworkers being in a new firm. Going back to the example, given the difficulty to know the true type of the worker, it will be easier for a professor to move to a university where a past coworker is currently working.

First, I document the importance of coworker network and learning mechanisms in the French labor market. I build two measures of coworker network that proxy the labor demand that a worker has access to thanks to past peers. If networks are a mean to overcome information frictions over the quality of a worker we expect better connected workers to have improved mobility prospects. Also, a good coworker network should improve the outside options of a given worker which should lead to more mobility and a better bargaining position in front of the firm. I do not find consistent evidence supporting better mobility and wage prospects. Better connected workers have worse mobility prospects depending on the network measure and they seem to worsen wage growth prospects for both measures.

Second, I find evidence on the existence of peer learning. Following the literature, I proxy peer human capital by the average log hourly wage at the establishment. Similarly to previous literature I find that peer human capital has a positive impact on future wage growth and this is more start for the youngest cohorts. This is in line with learning happening mostly for workers with low human capital. The results are robust to considering other measures of peer human capital (like the 75th and 90th percentiles) motivated by the possibility of learning from the best peers. Finally, there are wage growth differences depending on age and location.

A natural challenge is to identify the contribution of each channel to the peer effects. As a first step to disentangle the importance of each of the channels I try to shut down one at a time and I find that the strength of each of the mechanisms is unchanged. Another challenge is to have exogenous network and human capital measures. As a work in progress, I am developing a structural model of peer learning that will first allow me to understand the human capital accumulation process through an structural estimation.[2] Second, I am interested in studying the importance of peer learning in determining wage inequality across firms and worker location.

Literature. This paper relates to two strands of the literature. First, it is related to empirical papers studying the effect of coworker networks (e.g. Cingano and Rosolia, 2012; Glitz, 2013; Saygin et al., 2014; Caldwell and Harmon, 2019) on worker mobility and wage profiles. Others have emphasized the role of

[2]An sketch of the model is in Appendix 3.C.

residence location in network formation. The closest paper in the empirical part of networks is Caldwell and Harmon (2019) from which I borrow one of the network measures. On the theoretical end, Matthew Jackson has extensively studied networks (e.g. Calvó-Armengol and Jackson, 2007; Bloch et al., 2017) and others have studied the effects of networks in structural models of search (e.g. Dustmann et al., 2016; Arbex et al., 2019).

Second, the paper relates to the literature studying coworker or peer effects. The difficulty on the empirical side is to overcome the selection and reflection effects of peers on one another. Mas and Moretti (2009) identify worker productivity and spillover effects using high frequency supermarket data. They focus on peer pressure effects whereas here the focus is on learning or human capital accumulation. Cornelissen et al. (2017) and Cardoso et al. (2018) use methods developed by Arcidiacono et al. (2012) to quantify peer effects or spillover effects among coworkers that extend the wage decomposition framework from Abowd et al. (1999). Applying these methods in my analysis is left for future research. The closest papers to the structural model I sketch in the Appendix are Herkenhoff et al. (2018) and Jarosch et al. (2019a). The former studies coworker learning in a search and matching framework with production function complementarities and proposes an optimal taxation. The latter show how to structurally estimate the learning parameters using continuation values that is similar in spirit to the method proposed by Scott et al. (2013) when the error terms are extreme value distributed.

To the best of my knowledge the literature has not tried to separate between coworker network and learning mechanisms. The contribution of this paper is to empirically disentangle both and to study location differences. In future work I plan to contribute to the mentioned structural papers by focusing on how coworker learning affects team generation and their geographical location.

3.2 Data

I mainly use three different data sources of French administrative data. First, the data source is *Panel DADS Tous salaries* where I can keep track of the worker movements over the years. The sample ranges from 2002 to 2015. This panel covers a representative subsample of the French labor force and includes a worker identifier, the establishment where she works, occupation, industry and some demographics like age. Second, aggregate firm data on employment comes from the universe of firms from the tax records *Ficus/Fare*. This data set is a panel of firm balance sheet information from which I construct the number of growing firms over time. Finally, I use the universe of workers from *DADS Postes* to compute distributional statistics at the establishment. *DADS Postes* is a repeated cross section and therefore is not suited to build network measure and wage paths.

I restrict the panel to workers in the private sector between the ages of 20 to 60 working full time in mainland France. Furthermore, I take different spells within a year only if they lasted at least 5 days and are consecutive (single job holders at all time). Details on data construction are in Appendix 3.A.

The main variables of study are worker mobility and wage growth or future wages. I define a job-to-job mobility if a worker changes establishments within 30 days. Wages are defined as log hourly wages in constant 1998 euros. We are interested in studying the effects of coworker networks and learning into mobility and wage growth. A network is a bipartite graph between the worker and the universe of workers in the panel. I define as peers the employees working at the same establishment in a given year. I build the coworker network using workers represented at the panel with less than 500 coworkers. Restricting the number of coworkers avoids the explosion of the measure and reduces the extent of considering in a network peers that are very unlikely to know each other. Every year I build the coworker network of each worker considering the previous 3 years. The network of worker i, Ω_{it}, will be the firms to which she has access through the past coworkers and where she did not work previously.[3]

Learning from coworkers requires measures of peer human capital. I use the universe of workers from *DADS Postes* and measure human capital of the coworkers based on log hourly wage of full time workers at the establishment. I assume that learning or human capital accumulation happens due to interactions with the peers. It is standard to focus on interactions with the average coworker as the most relevant (e.g. Jarosch et al., 2019a).[4] I also explore other measures describing the human capital distribution at the establishment such as the percentiles 75 and 90 of the wages. I only take into account establishments with at least 5 different spell observations over the year.

3.3 Empirical Evidence

This section presents the empirical evidence on the effects of peer networks and learning. First, I present results considering different coworker networks. Second, I show results on learning from coworkers. Finally I discuss ways of disentangling between both forces and potential explanations behind results.

3.3.1 Networks

I estimate the following econometric model to explore the network mechanism:

$$y_{it} = \gamma^k \Omega^k_{it} + \beta X_{it} + \theta_i + \alpha_{jt} + \delta_n + \varepsilon_{it}, \tag{3.1}$$

where y_{it} is an indicator variable of the outcome of interest (job-to-job mobility, connected job-to-job mobility, growth of net log hourly wage), Ω^k_{it} is a network measure, X_{it} are controls like age, squared age and occupation, θ_i is a person fixed effect, α_{jt} a industry-year δ_n a city fixed effect and ε_{it} is the error term.[5] The parameter of interest is γ^k. The superscript k refers to different measures of networks.

Identification comes from within individual and across time variation of network Ω_{it}. This variation comes either by changes on the labor demand of already connected firms or by changes on the set of connected firms over time. The main thread to the validity is the endogeneity of the network. This could happen due to unobserved changes in skill demand (leading to mobility or changes in wages) that are correlated with the measure of the network even after controlling for the several fixed effects. One example would be if there were geographically differentiated networks (e.g. cities vs country side) and there were temporary city/non-city shocks changing the shape of the coworker network and the dependent variable at the same time.

The objective of the network is to measure the potential access to job openings through a worker's past coworker network. I consider two measures of coworker network Ω_{it}. First, I follow Caldwell and Harmon (2019) computing the networks as a proxy of labor demand at the connected firms. Second, I measure the network by the number of connected firms.

[3]More details on variable construction are in Appendix 3.A.

[4]Differently to those authors I take the average log wage ($\mathbb{E}(\log(w_{lt})|\mathbf{J}(l,t) = \mathbf{J}(i,t))$) rather than the logarithm of the average wage ($\log(\mathbb{E}(w_{lt}|\mathbf{J}(l,t) = \mathbf{J}(i,t)))$). Similarly, I take the percentiles of establishment log hourly wages.

[5]When the dependent variable is wage growth I take out the industry-year fixed effects and put instead a city-year fixed effect.

Measurement

The first measure of worker i's network at time t, Ω_{it}^{LD}, captures the exposure to labor demand and is:

$$\Omega_{it}^{LD} = \sum_j \frac{I_{ijt}}{\sum_{j'} I_{ij't}} s_{jt},$$

where I_{ijt} is an indicator that takes value 1 if worker j worked with i at the same establishment-year at some point in the previous 3 years (up to $t-1$). s_{jt} is capturing labor demand at growing connected firms and is: $s_{jt} = (E_{jt} - E_{jt-1})^{+}$ where the position created for the connected mover is excluded from E_{jt}.[6]

The second measure of coworker network is the number of different connected establishments that a worker has access to through the past peers. It is defined as $\Omega_{it}^{NC} = \sum_j I_{ijt}$ where I_{ijt} is defined as above. Finally, as a robustness check I also present results of workers in firms facing a mass layoff.[7]

Results

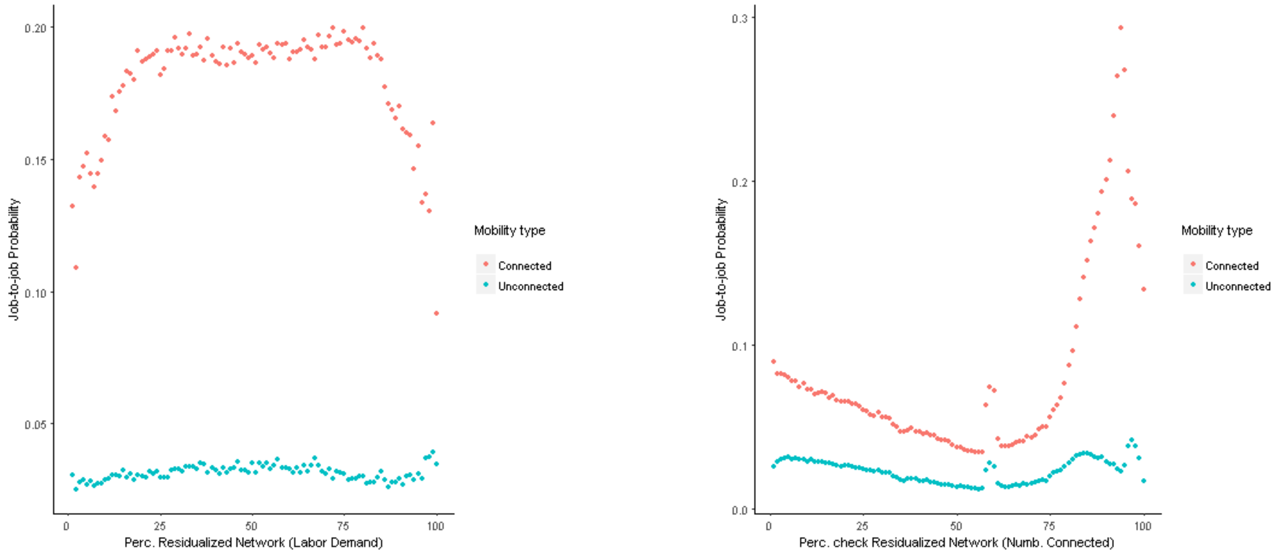
Figure 20 presents the results of the job-to-job mobility probabilities against residualized labor demand network in Panel (a) and against the network with the number of connected firms in Panel (b).[8] Both panels present the probability of a job-to-job mobility on the y-axis and the percentile of residualized network on the x-axis. Both show that workers move to a connected firm with higher probability. Panel (a) shows that this mobility is unrelated to the network measure capturing exposure to the labor demand Ω_{it}^{LD}. On the contrary, Panel (b) suggests that workers with the highest number of connected firms have a sizable increase on the job-to-job mobility to connected firms.

[6] I take yearly employment levels from *FICUS/FARE* using balance sheet data. Those are therefore defined at the firm level instead of the establishment.

[7] I define a mass layoff as a firm with at least 20 workers in the previous period reducing its workforce by at least 30%. Here I assume that all the establishments of the firm are affected.

[8] In both cases I residualize the network measure by taking the residuals from a regression into controls, person fixed effects, industry-year and location fixed effects. I use the residuals of the following model: $\Omega_{it}^{k} = \beta X_{it} + \theta_i + \alpha_{jt} + \delta_n + \epsilon_{it}$.

Figure 20 – Mobility and Network



(a) Labor Demand

(b) Number of Connected

Notes: Panels (a) and (b) present the results with Ω_{it}^{LD} and Ω_{it}^{NC} respectively. The y-axis is the probability of a job-to-job mobility and the the x-axis is the percentile of residualized network. I first take the residuals of $\Omega_{it}^{k} = \beta X_{it} + \theta_i + \alpha_{jt} + \delta_n + \epsilon_{it}$ with the variables defined as in the main text, and then generate the percentiles of these residuals.

The conclusions from Figure 20 are confirmed by results in Table 31 in the Appendix. The Table shows the estimates of γ^k of model (3.1). Contrary to what Caldwell and Harmon (2019) find, Column (3) shows that an increase of one unit of the network Ω_{it}^{LD} is related to a small but significant decrease in the probability of job-to-job mobility. It is also related to a decrease in mobility to a connected firm and the probability of having a wage growth. This would mean the coworker network effect is not relevant for mobility and furthermore, the workers having less human capital/productivity gains would be the ones having better networks. I discuss potential mechanisms behind the results in Section 3.3.3.

Results are different for the second network measure. Column (2) of Table 31 shows that having an additional connected establishment increases the probability of mobility by $\widehat{\gamma}^{NC} = 0.4\%$ and is also positively related to connected job-to-job mobility with the same estimated parameter.[9]

Table 32 in the Appendix presents the results when the dependent variable is the probability of having a gain of hourly wages. A better network is related to a decrease on the likelihood of having a wage increase.[10]

All in all, the empirical evidence of the importance of coworker network effects is mixed. The natural measure that proxies access to labor demand through the networks seems to be negatively related with mobility and wage growth measures. I briefly discuss a potential explanation of this finding in Section 3.3.3.

3.3.2 Learning

Restricting the attention to the learning mechanism the estimating equation is:

$$w_{i,t+h} = \phi_h \, \overline{w}_{\mathbf{J}(i,t)} + \rho \, w_{i,t} + \beta \, X_{it} + \mu_i + \delta_{nt} + \nu_{it}, \tag{3.2}$$

[9]This estimate is conditional on having at least one connected firm.

[10]This is robust when we restrict to the workers that have a job-to-job transition at some point in the sample.

where $w_{i,t+h}$ is the log hourly wage h years ahead, $\overline{w}_{\mathbf{J}(i,t)}$ is the average log hourly wage at the establishment in year t, $w_{i,t}$ is worker i's wage at t, X_{it} are the same controls as above, μ_i is a gender fixed effect, δ_{nt} is a city-year fixed effect and ν_{it} is the error term. The parameter of interest is ϕ_h where the subscript is the number of years forward.[11]

The main issue of the regression is the fact that due to data limitations I cannot identify worker i in the universe of workers at her establishment. This renders the average log hourly wages endogenous especially in small establishments. I only consider establishments with at least 5 different full time spells to mitigate this concern. Regarding the identification of the parameter of interest, the main challenge is the fact that future wage increases may reflect wage backloading or the presence of long lasting productivity shocks that affect both, the average log hourly wage of the peers and the future wage of worker i.

Results of the regression equation (3.2) with the average wage are in Table 29. Similarly to the previous literature, estimates of ϕ suggest positive learning effects from coworkers. Passing from the first to the third quartile of peer average wage increases the 1 period (10 periods) ahead wage by 0.05% (0.06%).

Table 29 – Learning: Average Wage

	Dependent variable:				
	w_{t+1}	w_{t+2}	w_{t+3}	w_{t+5}	w_{t+10}
	(1)	(2)	(3)	(4)	(5)
$\overline{w}_{\mathbf{J}(i,t)}$	0.120***	0.133***	0.137***	0.143***	0.152***
	(0.0004)	(0.0005)	(0.001)	(0.001)	(0.003)
Controls	Y	Y	Y	Y	Y
Person FE	N	N	N	N	N
Observations	6,056,021	4,975,770	3,996,992	2,426,980	265,425
R^2	0.825	0.804	0.790	0.771	0.741
Adjusted R^2	0.821	0.799	0.784	0.764	0.722

Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable w_{t+h} is the log hourly wage of worker i h spells ahead of time t, $\overline{w}_{\mathbf{J}(i,t)}$ is the yearly average log wage per hour of full time employees at the establishment where worker i worked in year t. Controls are age, squared age, and occupation, gender and city-year fixed effects.

The results in the short run are almost identical when we restrict to stayers.[12] In the longer run, stayers benefit less from having good coworkers as the estimates of ϕ_h 9% higher when we take the whole sample. Table 34 in the Appendix presents robustness checks. Repeating Table 29 with worker fixed effects the effect is 31% for one year ahead wages.[13] In the longer run peer human capital proxied by their average wage

[11] I do not restrict the attention to unique spells per year therefore future wages are defined as wages in the following spells

[12] I define stayers as workers not experiencing a job-to-job mobility over the sample.

[13] I replace the gender fixed effects by the person fixed effects as the former would not be identified.

does not have a positive effect on future wages.[14] The results are qualitatively similar when we consider peer human capital proxied by the 75th and 90th wage percentiles at the establishment. Table 34 shows that the results are weaker than for the average wage and the effects are insignificant in the long run.

Also in line with previous findings, there is a stark difference if we consider workers below and above that average wage.

Age Differences

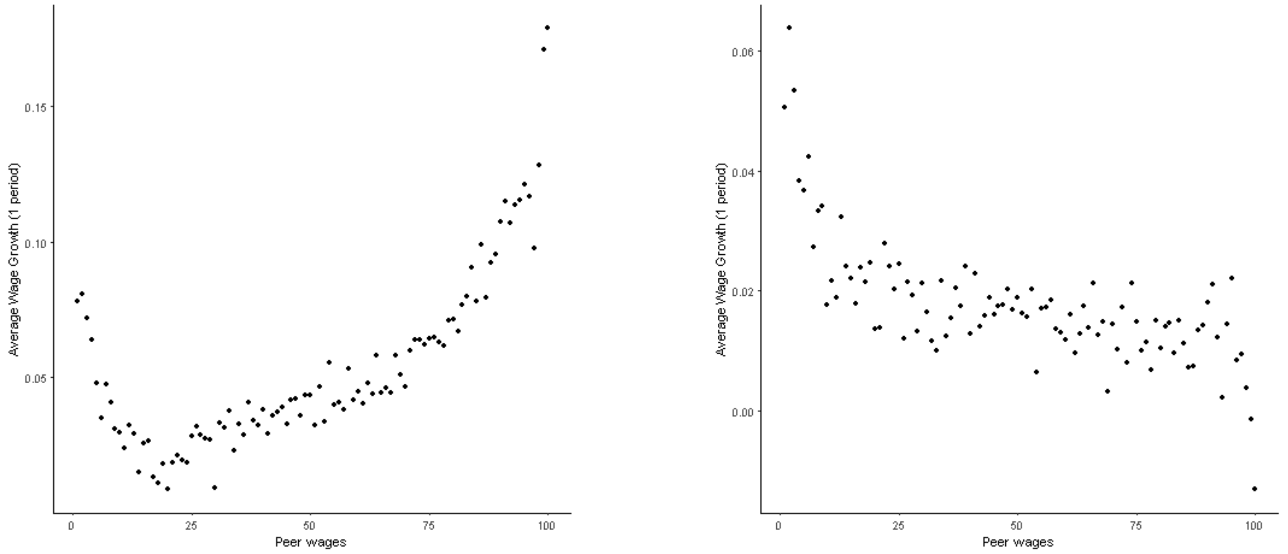
Results from Table 29 suggest important average effects of peer human capital on future wages while controlling for current wages. Here I explore the heterogeneity of peer wages on wage growth depending on the age of the worker. Figure 21 presents the differentiated effects of peer wages on the 1 period average wage growth for young (25 years at t) in Panel (a) and middle aged workers in Panel (b).[15] Both panels show that on average workers have wage increases. There are important differences on wage growth depending on age. Young workers have on average wage growths of 4.74% while middle aged workers have more moderate wage growths of 1.75%. They also present different learning patterns within age category. Young workers seem to benefit more the better coworkers they have whereas wage growth is rather flat for middle aged workers.[16] Figure 25 clearly shows that the average 1 period wage growth is decreasing in age. This is also in line with faster learning from peers from your employees.

[14] Estimates of 10 year ahead wages suggest that a worker is hurt by peer human capital ($\widehat{\phi}_{10} = -0.012$) opposed to the positive effects in the short run ($\widehat{\phi}_1 = 0.083$).

[15] I define one period ahead wage growth per worker $\Delta_{i,t}$ as the difference of log hourly wages. $\Delta_{i,t} = w_{i,t} - w_{i,t-1}$ where w is the log hourly wage. I then take averages per peer wage percentile.

[16] Figure 24 in the Appendix presents histograms of the wage gap for young and middle aged. The wage gap is the difference between the log hourly wage of the worker and the establishment average log hourly wage. The Figure shows that the wage gap of young workers is right skewed. They have on average lower than establishment mean wages. Table 36 shows that they are more benefited from peer learning than the average worker.

Figure 21 – Learning by Age



(a) Young

(b) Middle Age

Notes: Panels (a) and (b) present the average wage growth by age group. *Young* are employees with 25 years at t and *Middle Age* are the ones with 40 years at t. The y-axis is the average 1 period ahead wage growth (difference of log wages between t and $t-1$) and the the x-axis is the percentile of peer average log hourly wages.

3.3.3 Discussion

The main empirical challenge at the moment is to find exogenous variation that will help determine the importance of each of the channels. As a first step, I disentangle between both by estimating the above regressions for different subsets of workers trying to eliminate one channel at a time. Then I discuss

Network vs. Learning

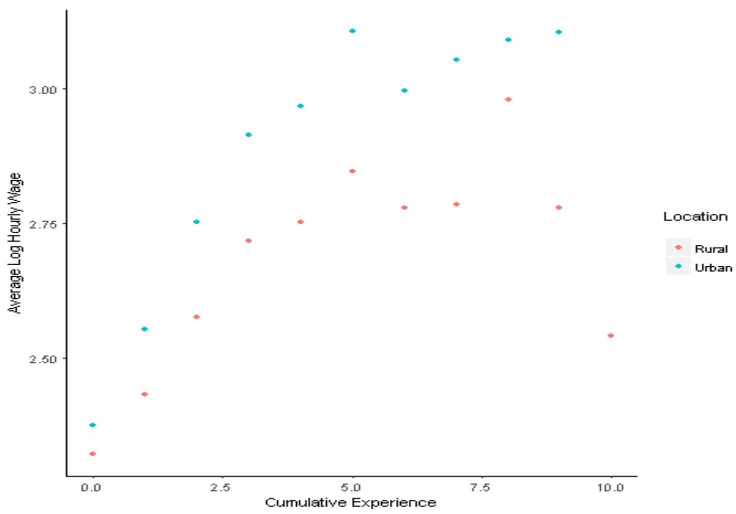
In order to identify the network effects excluding long lasting human capital accumulation effects through peer learning I estimate (3.1) but controlling for the average of the mean wage the previous 3 years of worker i, $\overline{w}_{\mathbf{J}(i,(t-3,t-1))}$. Results are in Table 35. Controlling for past average peer wages does not change the estimated effects of network (for both measures considered) into job mobility (the estimate is still $\widehat{\gamma}^{NC} = 0.4\%$) and wage gains.

Identifying learning effects excluding potential impacts of the network effect requires to focus on workers for whom the peer network is not relevant for their labor market outcomes. This is more likely to happen for the youngest worker cohorts who presumably did not have the time to build a coworker network. Table 36 presents the estimation results of (3.2) for the workers with less than 23 years. Learning from peers is related with wage gains that are twice as large the ones of the average gains. It seems indeed that the peer effects are very strong for workers who arguably have no coworker network. This evidence suggests that learning from peers is a strong mechanism through which workers are influenced by their workmates.

Location Differences

Here I explore differences in networks and earning profile depending on the location of the employees. Figure 22 shows the average log hourly wages as function of cumulative experience. Similarly to Martellini (2019) I find that wages in urban areas are higher and also present a steeper wage profile on experience.[17]

Figure 22 – Location Differences in Learning



Notes: The y-axis is the average log hourly wage of worker *i* at time *t* and the x-axis is the cumulative experience of the worker in year *t*. Details of variable construction are in the Appendix.

Table 30 presents percentage differences between urban and rural areas for both network measures. Urban areas perform worse in terms of the network measure capturing access to firm labor demand Ω_{it}^{LD}. The 10 biggest cities have networks that are almost 24% worse than the rest of the cities. This is potentially behind the negative results using the first measure. On the contrary, urban areas have on average better networks according to the count of number of connecting firms. There are also differences in the probability of job-to-job mobility from urban to rural areas that is 7.82% for the former and 1.88% for the latter.

[17] I define as urban areas the 10 biggest cities per year. Those are: Paris, Lyon, Marseille, Toulouse, Nice, Bordeaux, Montpellier, Rennes, Nantes, Lille and Strasbourg. I define the rest of the cities (*communes*) as rural locations.

Table 30 – Network: Location

	Urban-Rural Gain (%)
Ω_{it}^{LD}	-23.73
Ω_{it}^{NC}	16.77

Notes: I restrict to the subset with at least one connected firm. *Urban-Rural Gain* is the percentage gain of the average urban measures relative to the rural ones. In the case of Ω_{it}^{LD}, is $\mathbb{E}(\Omega_{it}^{LD}|urban) - \mathbb{E}(\Omega_{it}^{LD}|rural)$ over the average measure at the rural location ($\mathbb{E}(\Omega_{it}^{LD}|rural)$). Locations are taken as the ones of the worker that are not necessarily the sames as the ones of connected firms.

Finally, it is important to acknowledge that the current French administrative data is not the best suited to study coworker networks because one cannot identify the universe of workers in the available panel.[18]

3.4 Conclusion

This paper studies the effects of coworkers on mobility, location and wage growth prospects. I consider that coworkers affect each other by leading to better opportunities due to good networks or due to human capital spillovers among peers. I use French administrative data that only covers a panel of workers in a subsample. I do not find robust evidence favoring network effects among coworkers. On the contrary, learning from the peers improves wage profiles of the workers. In particular for the youngest cohorts. I disentangle between network and learning channels empirically by shutting down one at a time.

In the Appendix I present a structural model of learning from coworkers that I plan to use in future research to study how this mechanism determines team formation, worker location across the geography and firm wage inequality.

[18]It would be possible to track the main job per year (*poste principale*) of the workers using a different data source. Nevertheless, the data source using those main jobs might not be well suited neither as it requires minimum duration and/or earning restrictions that could contaminate the analysis. This is left for future research.

3.A Data Construction

This section provides data construction details of the different data sets: *DADS Panel EDP*, *DADS Postes* and *FICUS/FARE*.

3.A.1 DADS Panel EDP

The panel I use is from 2015 and covers about a 12th of the universe of French workers.[19] I observe the whole working history of sampled workers. I restrict the sample between 1980 and 2015 to compute worker experience but the analysis is done on employees between 2003 and 2015.[20] I restrict to employees between the ages of 20 and 60 working full time at the private sector.[21] I restrict to workers outside of France Telecom and La Poste (*PTT* equal to 0). I exclude observations with missing missing wages, firm identifier, working hours, location, duration, working type and start/end dates of the job spell.[22] Nominal wages are deflated with the CPI to 1998 constant prices.[23] I restrict to mainland France and workers with job spells of at least 5 days.[24] City codes are the concatenation of department (*dept*) and town codes (*comT*).[25]

Occupations and industries are defined as the variables using 2-digit precision. I construct cumulative experience by summing the duration of spells up to spell t. I convert those to yearly experience levels by defining the cumulative duration by 1820 hours per year following the full time definitions from the data.

3.A.2 DADS Postes

I restrict to workers living in mainland France and Corsica.[26] An establishment is the concatenation between the Siren and Siret identifiers. I take only spells with ordinary jobs, working full time and with positive hours worked, wages and duration.[27] Wages are first deflated using the CPI index as for the *DADS Panel EDP* and then throughout I use log hourly wages. I consider only establishments with 5 different spells in a given year. Establishment-year wage measures are constructed as: the average log hourly wage, 75th percentile of establishment log wages and 90th percentile.

3.A.3 FICUS/FARE

I filter firms with positive employment (*EFFSALM*) and with firm identifier information (*Siren* different to missing and zeros).

[19] The panel is *DADS Panel tous salariés 2015* which is a subsample based on the birthdates of the workers.

[20] Wage data before 1993 was estimated. The sample starts at 2003 due to a break in the coding of occupations that year.

[21] Employees at the private sector are filtered by taking the variable *sect* equal to *PRIV*. Full time workers are the ones with *ce* equal to *C*. This excludes workers working part time (*P*) at home (*D*) and unemployed (*A*).

[22] I take observations with positive net real wages (*SNR*), firm identifier (*sir*), hours (*nbheur*), location (*comR* and *comT*), duartion (*DP*), working type (*ce*) and start/end date (*DEBREMU* and *FINREMU* respectively) information.

[23] Source of CPI data https://www.insee.fr/fr/statistiques/serie/001643154

[24] I exclude observations with *dept* variable in 97, 98 and 99.

[25] I reclassify the city codes of Paris, Marseille and Lyon to be 75056, 13055 and 69123 respectively.

[26] I exclude workers with *regr* equal to 94 and 99.

[27] Ordinary jobs are the ones with typ_emploi equal to *O*. This exludes apprenticeships and internships. Full time spells are the ones with *CPFD* equal to *C*.

3.B Empirical Evidence

3.B.1 Network

Table 31 – Mobility and Network

	Dependent variable:			
	Probability J2J mobility			
	(1)	(2)	(3)	(4)
Ω_{it}^{NC}	0.022***	0.004***		0.008***
	(0.0001)	(0.0002)		(0.0002)
Ω_{it}^{LD}			−0.00000***	
			(0.00000)	
Controls	Y	Y	Y	Y
Person FE	Y	Y	Y	Y
Observations	7,415,371	1,892,681	2,071,871	683,757
R^2	0.229	0.470	0.470	0.066
Adj R^2	0.050	0.144	0.153	0.051

Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the probability of job-to-job mobility. Ω_{it}^{NC} is the count of connected firms and Ω_{it}^{LD} is the network measure capturing labor demand as defined in the main text. Columns (1), (2) and (4) present results with the former measure. Column (1) takes the whole sample (which assigns 0 to the workers with no network), Column (2) filters the observations having at least one connected firm, and Column (4) presents the results restricting to workers having suffered a mass layoff previously. Column (3) shows the results with Ω_{it}^{LD}. Controls are age, squared age, and occupation, industry-year and location fixed effects.

3.B.2 Learning

Figure 23 replicates the Figure in the main text but taking the distribution of the 90th percentile of wages across establishments.

Table 32 – Wage Growth and Network

	Dependent variable:			
	Probability wage gain			
	(1)	(2)	(3)	(4)
Ω_{it}^{NC}	-0.002^{***}	-0.001^{***}		
	(0.0002)	(0.0003)		
Ω_{it}^{LD}			-0.00000^{***}	
			(0.00000)	
Controls	Y	Y	Y	Y
Person FE	Y	Y	Y	Y
Observations	7,275,418	1,853,701	2,027,518	
R^2	0.202	0.393	0.384	
Adjusted R^2	-0.005	-0.062	-0.060	

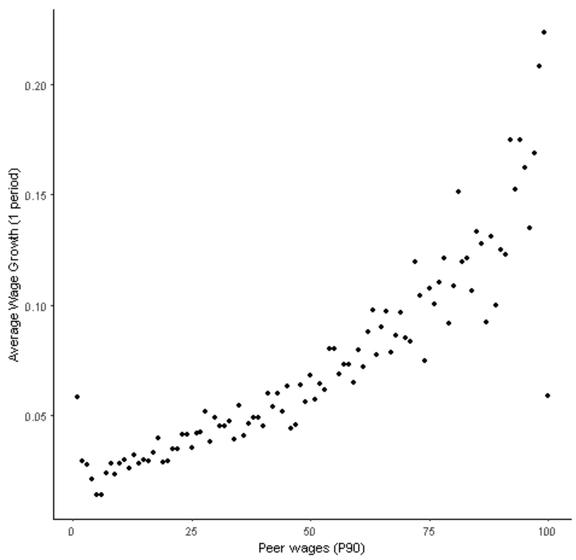
Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the probability of of having a wage gain. Ω_{it}^{NC} is the count of connected firms and Ω_{it}^{LD} is the network measure capturing labor demand as defined in the main text. Columns (1), (2) and (4) present results with the former measure. Column (1) takes the whole sample (which assigns 0 to the workers with no network), Column (2) filters the observations having at least one connected firm, and Column (4) presents the results restricting to workers having suffered a mass layoff previously. Column (3) shows the results with Ω_{it}^{LD}. Controls are age, squared age, and occupation, industry-year and location fixed effects.

Table 33 – Learning: Stayers

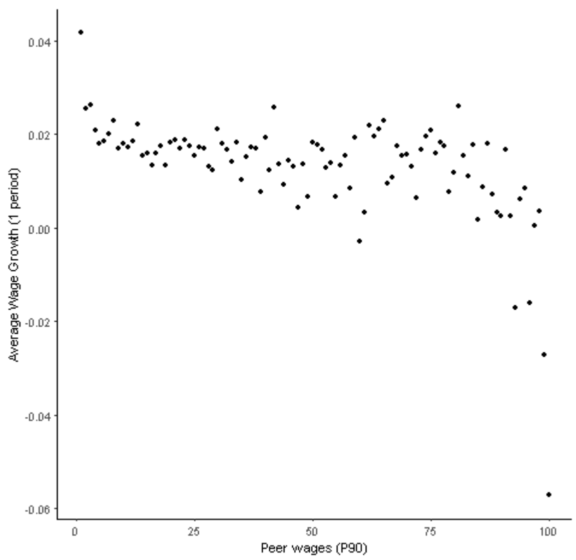
	Dependent variable:				
	w_{t+1}	w_{t+2}	w_{t+3}	w_{t+5}	w_{t+10}
	(1)	(2)	(3)	(4)	(5)
$\overline{w}_{J(i,t)}$	0.120***	0.132***	0.134***	0.135***	0.139***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.004)
Controls	Y	Y	Y	Y	Y
Person FE	N	N	N	N	N
Observations	3,024,596	2,405,801	1,879,507	1,089,557	94,299
R^2	0.827	0.810	0.801	0.792	0.781
Adjusted R^2	0.820	0.802	0.792	0.780	0.757

Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable w_{t+h} is the log hourly wage of worker i h spells ahead of time t, $\overline{w}_{J(i,t)}$ is the yearly average log wage per hour of full time employees at the establishment where worker i worked in year t. Controls are age, squared age, and occupation, gender and city-year fixed effects.

Figure 23 – Learning by Age. Percentile 90



(a) Young



(b) Middle Age

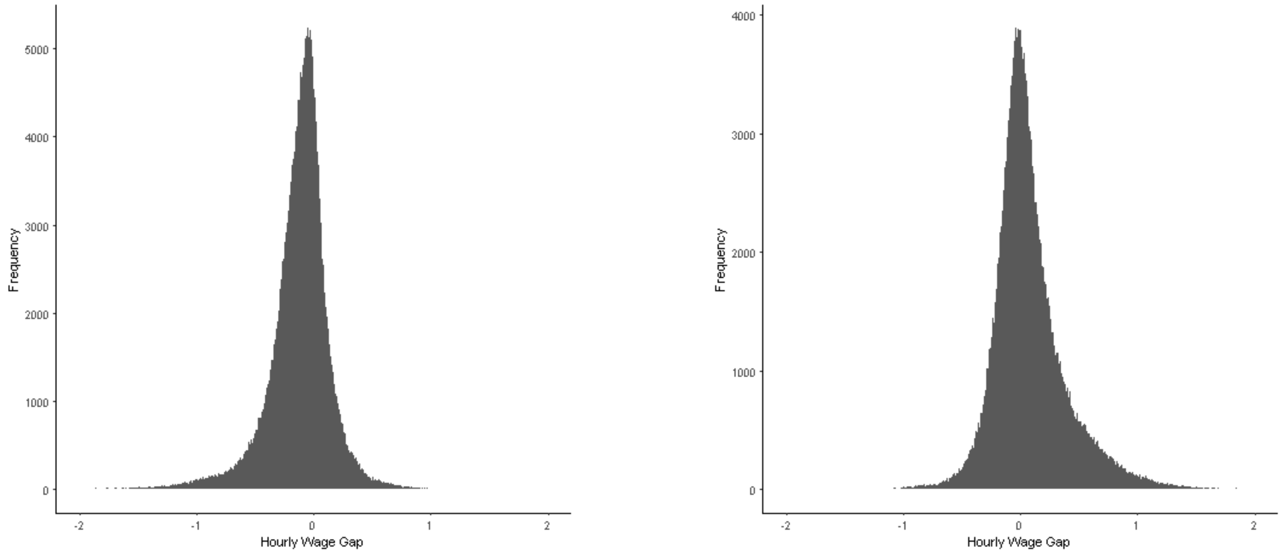
Notes: Panels (a) and (b) present the average wage growth by age group. *Young* are employees with 25 years at t and *Middle Age* are the ones with 40 years at t. The y-axis is the average 1 period ahead wage growth (difference of log wages between t and $t-1$) and the the x-axis is the percentile of the distribution of peer 90th percentile of the establishment.

Table 34 – Learning: Robustness

	Dependent variable:				
	w_{t+1}	w_{t+2}	w_{t+3}	w_{t+5}	w_{t+10}
	(1)	(2)	(3)	(4)	(5)
$\overline{w}_{J(i,t)}$	0.083***	0.054***	0.026***	0.008***	−0.012*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.007)
P75 $w_{J(i,t)}$	0.058***	0.038***	0.019***	0.006***	−0.005
	(0.001)	(0.001)	(0.001)	(0.001)	(0.005)
P90 $w_{J(i,t)}$	0.036***	0.025***	0.013***	0.004***	−0.004
	(0.0004)	(0.0004)	(0.0005)	(0.001)	(0.004)
Controls	Y	Y	Y	Y	Y
Person FE	Y	Y	Y	Y	Y
Observations	6,056,021	4,975,770	3,996,992	2,426,980	265,425

Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable w_{t+h} is the log hourly wage of worker i h spells ahead of time t, $\overline{w}_{J(i,t)}$ is the yearly average log wage per hour of full time employees at the establishment where worker i worked in year t. P75 $w_{J(i,t)}$ and P90 $w_{J(i,t)}$ are the 75th and 90th percentiles of log hourly wages at the establishment. Controls are age, squared age, and occupation, gender and city-year fixed effects.

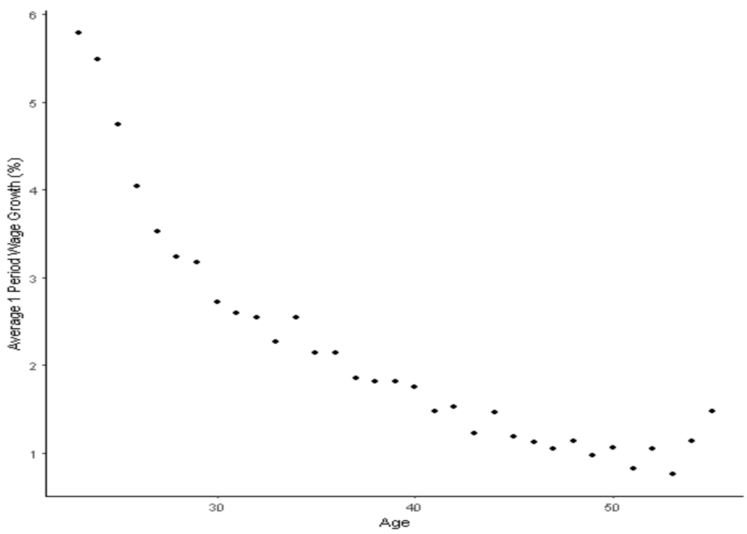
Figure 24 – Wage Gap by Age



(a) Young (b) Middle Age

Notes: Panels (a) and (b) present histograms of the wage gab by age group. *Young* are employees with 25 years at t and *Middle Age* are the ones with 40 years at t. The wage gap is defined as the difference between the log hourly wage of worker i at time t and the average log hourly wage at the establishment: $w_{i,t} - \overline{w}_{\mathrm{J}(i,t)}$ where w are log wages.

Figure 25 – Wage Growth by Age



Notes: The y-axis is the average hourly wage growth ($w_{i,t} - w_{i,t-1}$ where w are log wages) of worker i at time t and the x-axis is the age of the worker at time t.

3.B.3 Discussion

Table 35 shows the network results controlling for learning. Table 36 presents the results for the young cohorts of workers (aged 23 or below).

Table 35 – Mobility and Network: Controlling for Learning

	Dependent variable:		
	Probability J2J mobility		
	(1)	(2)	(3)
Ω_{it}^{NC}	0.022^{***}	0.004^{***}	
	(0.0001)	(0.0002)	
Ω_{it}^{LD}			-0.00000^{***}
			(0.00000)
Controls	Y	Y	Y
Person FE	Y	Y	Y
Observations	7,197,498	1,839,299	2,010,936
R^2	0.231	0.472	0.472
Adj R^2	0.051	0.144	0.153

Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable is the probability of job-to-job mobility. Ω_{it}^{NC} is the count of connected firms and Ω_{it}^{LD} is the network measure capturing labor demand as defined in the main text. Columns (1) and (2) present results with the former measure. Column (1) takes the whole sample (which assigns 0 to the workers with no network), Column (2) filters the observations having at least one connected firm. Column (3) shows the results with Ω_{it}^{LD}. Controls are age, squared age, and occupation, industry-year and location fixed effects. Controls now also include the average log hourly wage of the establishments where the worker was employed the previous 3 years.

Table 36 – Learning: Young

	Dependent variable:				
	w_{t+1}	w_{t+2}	w_{t+3}	w_{t+5}	w_{t+10}
	(1)	(2)	(3)	(4)	(5)
$\overline{w}_{\mathrm{J}(i,t)}$	0.151***	0.230***	0.283***	0.323***	0.340***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.019)
Controls	Y	Y	Y	Y	Y
Person FE	N	N	N	N	N
Observations	270,180	217,379	173,254	109,229	11,252
R^2	0.289	0.265	0.294	0.386	0.535
Adjusted R^2	0.128	0.084	0.106	0.194	0.168

Notes: *p<0.1; **p<0.05; ***p<0.01. The dependent variable w_{t+h} is the log hourly wage of worker i h spells ahead of time t, $\overline{w}_{\mathrm{J}(i,t)}$ is the yearly average log wage per hour of full time employees at the establishment where worker i worked in year t. Controls are age, squared age, and occupation, gender and city-year fixed effects.

3.C Model

Given the empirical evidence on the importance of learning I build a very preliminary model where workers' human capital accumulation is influenced by the average type of the coworkers. For the moment I abstract from coworker network effects.

Instantaneous utility is formed by wage earnings, mobility costs and a taste shock:

$$U(\omega_t, h_{it}, k_{it}, j) = w_t H(h_{it}, k_{it}, j) - C(k_{it}, j) + \varepsilon_{ijt}$$

where the market state ω_t is the economy wide wage rate w_t; $H(h_{it}, k_{it}, j)$ is the human capital accumulation function that depends on the initial human capital h_{it}, other state variables k_{it} and the chosen average peer type/unemployment status j; $C(k_{it}, j)$ is the mobility cost function that is only positive when the worker decides to change firms and not on the specific coworker type chosen and ε_{ijt} is a mobility shock.

Assumption 1: The market state ω_t evolves according to a Markov process that is unaffected by the decisions of an individual worker. That is, the conditional distribution of ω_{t+1} fulfills $T(\omega_{t+1}|\omega_t, j_{it} = j) = T(\omega_{t+1}|\omega_t)$ for all workers and all choices.

Following Scott et al. (2013) I simplify the notation by skipping the aggregate state common to all workers as a state by $V(\omega_t, h_{it}, k_{it}, \varepsilon) \equiv V_t(h_{it}, k_{it}, \varepsilon)$.

Assumption 2: Conditional on the aggregate and individual states ω_t, h_{it} and k_{it} the shocks on the instantaneous utility ε_{ijt} are identically and independently distributed across all i, j and t with a type 1 extreme value distribution.

There is a continuum of workers of mass 1. I assume worker i at time t is characterized by two states: the human capital at the beginning of the period h_{it} and other states k_{it} such as education and age. The human capital is assumed to be discrete with $\mathbf{H} = \{h_1, h_2, ..., h_N\}$ where $h_1 < h_2 < ... < h_N$. Each period a mass δ of workers dies and is replaced by the same mass of newborns that draw their initial human capital from the distribution Γ_0.

The choice set of the workers is to stay unemployed ($j = 0$), remain at the current firm ($j = 1$) or move ($j >= 2$). When moving to another firm, the worker also needs to choose the average peer type with whom she wants to match.

In order to be able to estimate the dynamic human capital gains from peer effects, I build a model of worker labor supply along the lines of Traiberman et al. (2017) and others by assuming extreme value type 1 (Gumbel) distributed shocks for the worker mobility decision. Worker's problem is:

$$V_t(h_{it}, k_{it}, \varepsilon) = \max_{j*} \{U(h_{it}, k_{it}, j*) + \varepsilon_{ij*t} + \beta\, \mathbb{E}[V_{t+1}(h'_{it}, k'_{it})|h_{it}, k_{it}]\}$$

where h_{it} is the human capital at the beginning of the period (that comes from the past choices), k_{it} are the rest of the worker states (age), $\varepsilon = \{\varepsilon_{ijt}\}$ is the vector of moving shocks for each of the possible coworker type choices j. $U(h_{it}, k_{it}, \overline{\theta}_j*)$ is the instantaneous utility and β is the discount factor. I assumed that the shocks ε are conditionally independent in order to take out ε from the next period's value function.

Denote by $v_t(h, k, j)$ the conditional value function representing the expected discounted returns for a worker conditional on having chosen action j and before the realization of the mobility shock ε_{ijt}:

$$v_t(h, k, j) = U(h, k, j) + \beta\, \mathbb{E}[V_{t+1}(h', k')|h, k]$$

where we have to take into account that the choice of today influences the continuation value by changing the instantaneous utility next period.

By the assumption of extreme value the conditional choice probability $p(j|h, k) \equiv Pr(j_{it} = j|h_{it} = h, k_{it} = k, \omega_t)$ is:

$$p(j|h, k) = \frac{exp(v_t(h, k, j))}{\sum_{j'} exp(v_t(h, k, j'))} \tag{3}$$

I borrow from Herkenhoff et al. (2018) and assume that the discrete human capital types evolve according to the worker's state and the chosen peers:

$$H(h_{it}, k_{it}, \overline{h}_j) = \begin{cases} h_{it} + 1 & \text{with probability } \Delta\, if\, h_{it} < N \\ h_{it} & \text{with probability } 1 - \Delta\, if\, h_{it} < N \\ h_{it} & \text{if } h_{it} = N \end{cases}$$

where $\Delta = \alpha_0^k + \alpha_1^{k,+} I(h_{it} > \overline{h}_j) + \alpha_1^{k,-} I(h_{it} > \overline{h}_j)$. $\alpha_1^{k,+}$ and $\alpha_1^{k,-}$ superscript k accounts for the fact that learning from coworkers can be different depending the age. I also consider the possibility that learning for depends on the human capital of the worker relative to the peers.

I assume that firms differ in productivity A which has exogenous distribution G. In order to introduce a notion to the firm I assume that they operate with a decreasing returns to scale technology. Production happens after the learning between coworkers takes place but firms do not internalize it. Firm's problem is:

$$\max_{\overline{h}_t} \{A_t \overline{h}_t^{\beta} - w_t \overline{h}_t\}$$

where $\overline{h}_t$ is simply the aggregate of human capital employed and will depend on the conditional choice probabilities (3). That is

$$\overline{h}_t = \sum_k \sum_h h_{it} p(j|h,k)\Lambda(k), \tag{4}$$

where $\Lambda(k)$ is the mass of workers with age k at time t. By assuming this linear form for the firm aggregate human capital in the production function, we are considering no complementarities in production and perfectly competitive labor markets. The first order conditions give:

$$\beta A_t \overline{h}_t^{\beta-1} = w_t \tag{5}$$

Denote by $\Gamma(h)$ the fraction of workers with human capital no greater than h. Let $N(\overline{h},h)$ denote the number of elements weakly less than h. Labor market clearing requires:

$$\Gamma(h) = \int N(\overline{h}(A),h)\, dG(A) \quad \forall h \tag{6}$$

where the left hand side is the labor supply of human capital h and the right hand side is the labor demand.

The distribution of human capital Γ evolves according to:

$$\Gamma_{t+1}(h) = \delta\Gamma_0(h) + (1-\delta)\sum_k \sum_x H(h|x,k,j)\, p(j|x,k)\, \Lambda(k)\, \Gamma_t(x) \tag{7}$$

where the the age distribution Λ is independent of human capital and evolves as:

$$\Lambda_{t+1}(k) = (1-\delta)\sum_k \Lambda_t(k-1) \tag{8}$$

where the starting age is normalized to 0. A fraction δ is newborn every period.

Equilibrium

A competitive equilibrium is a wage rate w_t, labor allocations $\{\overline{h}\}$, a coworker vector $\tilde{h}$ and a distribution of human capital such that:

1. Firms choose labor optimally (5),
2. Workers choose optimally (3),
3. Labor market clears (6),
4. The law of motion of human capital (7) is fulfilled.

Bibliography

Abowd, J., F. Kramarz, P. Lengermann, and S. Pérez-Duarte (2004): "Are good workers employed by good firms? A test of a simple assortative matching model for France and the United States," *Unpublished Manuscript*.

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

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

Allen, T., C. Arkolakis, and X. Li (2016): "On the existence and uniqueness of trade equilibria," *Unpublished manuscript*.

Andrews, M. J., L. Gill, T. Schank, and R. Upward (2008): "High wage workers and low wage firms: negative assortative matching or limited mobility bias?" *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171, 673–697.

Angrist, J. D. and J.-S. Pischke (2008): *Mostly harmless econometrics: An empiricist's companion*, Princeton university press.

Arbex, M., D. O'Dea, and D. Wiczer (2019): "Network search: climbing the job ladder faster," *International Economic Review*, 60, 693–720.

Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012): "Estimating spillovers using panel data, with an application to the classroom," *Quantitative Economics*, 3, 421–470.

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

Avron, H. and S. Toledo (2011): "Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix," *Journal of the ACM (JACM)*, 58, 8.

Azar, J., I. Marinescu, and M. I. Steinbaum (2017): "Labor market concentration," Tech. rep., National Bureau of Economic Research.

Azar, J. A., I. Marinescu, M. I. Steinbaum, and B. Taska (2018): "Concentration in US labor markets: Evidence from online vacancy data," Tech. rep., National Bureau of Economic Research.

Barkai, S. (2016): "Declining labor and capital shares," *Stigler Center for the Study of the Economy and the State New Working Paper Series*, 2.

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

Berger, D. W., K. F. Herkenhoff, and S. Mongey (2019): "Labor Market Power," Tech. rep., National Bureau of Economic Research.

Best, M. C., J. Hjort, and D. Szakonyi (2017): "Individuals and organizations as sources of state effectiveness," Tech. rep., National Bureau of Economic Research.

Binmore, K., A. Rubinstein, and A. Wolinsky (1986): "The Nash bargaining solution in economic modelling," *The RAND Journal of Economics*, 176–188.

Bloch, F., M. O. Jackson, and P. Tebaldi (2017): "Centrality measures in networks," .

Bonhomme, S., T. Lamadon, and E. Manresa (2015): "A distributional framework for matched employer employee data," *Unpublished manuscript, University of Chicago*.

Borovičková, K. and R. Shimer (2017): "High wage workers work for high wage firms," Tech. rep., National Bureau of Economic Research.

Brügemann, B., P. Gautier, and G. Menzio (2018): "Intra firm bargaining and Shapley values," *The Review of Economic Studies*, 86, 564–592.

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

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

Caldwell, S. and N. Harmon (2019): "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," .

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015): "The anatomy of French production hierarchies," *Journal of Political Economy*, 123, 809–852.

Calvó-Armengol, A. and M. O. Jackson (2007): "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of economic theory*, 132, 27–46.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and labor market inequality: Evidence and some theory," *Journal of Labor Economics*, 36, S13–S70.

Card, D., J. Heining, and P. Kline (2013): "Workplace heterogeneity and the rise of West German wage inequality," *The Quarterly journal of economics*, 128, 967–1015.

Cardoso, A. R., P. Guimaraes, P. Portugal, and H. Reis (2018): "The returns to schooling unveiled," .

Chandrasekhar, A. G., M. Morten, and A. Peter (2020): "Network-Based Hiring: Local Benefits; Global Costs," Tech. rep., National Bureau of Economic Research.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 126, 1593–1660.

Cingano, F. and A. Rosolia (2012): "People I know: job search and social networks," *Journal of Labor Economics*, 30, 291–332.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017): "Peer effects in the workplace," .

Costinot, A. and A. Rodríguez-Clare (2014): "Trade theory with numbers: Quantifying the consequences of globalization," in *Handbook of international economics*, Elsevier, vol. 4, 197–261.

Davidson, R. and J. G. MacKinnon (2006): "Bootstrap methods in econometrics," .

De Loecker, J. and J. Eeckhout (2017): "The rise of market power and the macroeconomic implications," Tech. rep., National Bureau of Economic Research.

——— (2018): "Global market power," Tech. rep., National Bureau of Economic Research.

Dhaene, G. and K. Jochmans (2015): "Split-panel jackknife estimation of fixed-effect models," *The Review of Economic Studies*, 82, 991–1030.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016): "Referral-based job search networks," *The Review of Economic Studies*, 83, 514–546.

Eaton, J. and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

Edmond, C., V. Midrigan, and D. Y. Xu (2018): "How costly are markups?" Tech. rep., National Bureau of Economic Research.

Falch, T. (2010): "The elasticity of labor supply at the establishment level," *Journal of Labor Economics*, 28, 237–266.

Gathmann, C., I. Helm, and U. Schönberg (2017): "Spillover effects of mass layoffs," *Journal of the European Economic Association*.

Gaure, S. (2014): "Correlation bias correction in two-way fixed-effects linear regression," *Stat*, 3, 379–390.

Glitz, A. (2013): "Coworker networks in the labour market," .

Gutiérrez, G. and T. Philippon (2016): "Investment-less growth: An empirical investigation," Tech. rep., National Bureau of Economic Research.

Herkenhoff, K., J. Lise, G. Menzio, and G. M. Phillips (2018): "Production and Learning in Teams," Tech. rep., National Bureau of Economic Research.

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and manufacturing TFP in China and India," *The Quarterly journal of economics*, 124, 1403–1448.

Hutchinson, M. (1989): "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines," *Communications in Statistics-Simulation and Computation*, 18, 1059–1076.

Iranzo, S., F. Schivardi, and E. Tosetti (2008): "Skill dispersion and firm productivity: An analysis with employer-employee matched data," *Journal of Labor Economics*, 26, 247–285.

Jarosch, G., E. Oberfield, and E. Rossi-Hansberg (2019a): "Learning from Coworkers," Tech. rep., National Bureau of Economic Research.

Jarosch, G., I. Sorkin, J. S. Nimczik, et al. (2019b): "Granular Search, Concentration and Wages," in *2019 Meeting Papers*, Society for Economic Dynamics, 1018.

Jochmans, K. (2018): "Heteroskedasticity-robust inference in linear regression models," *arXiv preprint arXiv:1809.06136*.

Jolivet, G., F. Postel-Vinay, and J.-M. Robin (2006): "The empirical content of the job search model: Labor mobility and wage distributions in Europe and the US," *Contributions to Economic Analysis*, 275, 269–308.

Kane, T. J. and D. O. Staiger (2008): "Estimating teacher impacts on student achievement: An experimental evaluation," Tech. rep., National Bureau of Economic Research.

Karabarbounis, L. and B. Neiman (2013): "The global decline of the labor share," *The Quarterly Journal of Economics*, 129, 61–103.

Kline, P., R. Saggio, and M. Sølvsten (2019): "Leave-out estimation of variance components," Tech. rep., National Bureau of Economic Research.

Kucheryavyy, K. (2012): "Continuity of a model with a nested CES utility function and Bertrand competition," *Economics Letters*, 117, 473–476.

Lamadon, T., M. Mogstad, and B. Setzler (2018): "Imperfect competition and rent sharing in the us labor market," *University of Chicago mimeo*.

Lipsius, B. (2018): "Labor Market Concentration Does Not Explain the Falling Labor Share," *Available at SSRN 3279007*.

Lopes de Melo, R. (2018): "Firm wage differentials and labor market sorting: Reconciling theory and evidence," *Journal of Political Economy*, 126, 000–000.

MacKinnon, J. G. and A. A. Smith Jr (1998): "Approximate bias correction in econometrics," *Journal of Econometrics*, 85, 205–230.

MacKinnon, J. G. and H. White (1985): "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *Journal of econometrics*, 29, 305–325.

Manning, A. (2011): "Imperfect competition in the labor market," in *Handbook of labor economics*, Elsevier, vol. 4, 973–1041.

Martellini, P. (2019): "The City-Size Wage Premium: Origins and Aggregate Implications," .

Mas, A. and E. Moretti (2009): "Peers at work," *American Economic Review*, 99, 112–45.

Morlacco, M. (2018): "Market Power in Input Markets: Theory and Evidence from French Manufacturing," .

Osborne, M. J. and A. Rubinstein (1990): *Bargaining and markets*, Academic press.

Redding, S. J. (2016): "Goods trade, factor mobility and welfare," *Journal of International Economics*, 101, 148–167.

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

Rigobon, R. (2003): "Identification through heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.

Saygin, P., A. Weber, and M. Weynandt (2014): "Coworkers, networks, and job search outcomes," .

Scott, P. T. et al. (2013): "Dynamic discrete choice estimation of agricultural land use," *Toulouse School of Economics Working Paper*, 526.

Staiger, D. O., J. Spetz, and C. S. Phibbs (2010): "Is there monopsony in the labor market? Evidence from a natural experiment," *Journal of Labor Economics*, 28, 211–236.

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

Traiberman, S. et al. (2017): "Occupations and Import Competition," in *2017 Meeting Papers*, Society for Economic Dynamics, 1237.

White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica: journal of the Econometric Society*, 817–838.