THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 1 Capitole

Présentée et soutenue par

Miguel ZERECERO ANTóN

Le 30 juin 2021

Essays on Labor Economics

Ecole doctorale : **TSE - Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche :
UMR 5314 -TSE-R

Thèse dirigée par
Christian HELLWIG

Jury

M. Chaney THOMAS, Rapporteur
M. Eeckhout JAN, Rapporteur
M. Christian HELLWIG, Directeur de thèse
M. Patrick FèVE, Président

Essays on Labor Economics

Ph.D. Thesis

Miguel Zerecero Antón[1]

Toulouse School of Economics

June 2021

[1]Address: 1, Esplanade de l'Université, 31080 Toulouse, France; email: miguel.zerecero@tse-fr.eu

Acknowledgements

I would like to thank my advisor Christian Hellwig for his invaluable guidance and help during my Ph.D. as well as for his support during the job market. I am also very grateful to Fabrice Collard and Patrick Fève for their generous time and help, even at late hours or by asking almost daily small econometrics questions. I am also indebted to other current and past members of the macro group at TSE, in particular, Tiziana Assenza, Martial Dupaigne, Eugenia Gonzalez-Aguado, Tim Lee, Robert Ulbricht, and Nicolas Werquin for building such a nurturing environment. I am grateful for their engagement at the workshops and seminars where most of my training happened, or by taking their time to discuss my research. I am also thankful to Nour Meddahi for his support along with every chapter of this thesis.

I am also thankful for the rest of the macro students, for their willingness to discuss research, and their disposition to help and encourage. I am particularly grateful to Miren Azkarate-Askasua with whom I wrote two chapters of my thesis. I will remember fondly our fun times in the small room at the Manufacture writing those papers, where I learned a lot by doing (and erring, so much erring). I am very fortunate for being her co-author. I also want to thank Rodrigo Arnabal, Hussein Bidawi, Jacopo Bregolin, Paloma Carrillo, Filippo Maria D'Arcangelo, Joana Duran, Alexandre Gaillard, Esther Goreichy, Alberto Grillo, Jeff Groesbeck, Julia Hoefer, Kunal Khairnar, Elia Lapenta, Gosia Majewska, Filip Mrowiec, Charles Pébereau, Kevin Remmy, Celia Ruiz, Fernando Stipanic, Marrit Teirlinck, Esli Trejo, Ling Zhou, and anyone who made this journey easier to travel. I am also especially thankful to Jacint Enrich, Jérôme Gonnot, and Willy Lefez for creating such a fun environment in our home and for helping me along the way. I could not have asked for better flatmates.

I am deeply grateful for my parents and siblings, for always being there when I needed to vent, and for always have time for giving a piece of advice or just to talk. I am very lucky to have your support.

Finally, I want to thank and acknowledge Mexico's National Council of Science and Technology (CONACYT) for their financial support when I was writing the last two chapters of this thesis.

Summary

This thesis contains three essays on the macroeconomic effects of labor markets with a special emphasis on the determinants of internal migration, spatial inequality, labor market power, and the determination of wages.

In the first chapter, I study a potential reason of why workers stay in economically distressed areas: people like to live close to what they call home. Using administrative data for France, I find: (i) the share of migrants who return to their birthplace is almost twice as large as the share of migrants who go to any other particular location; (ii) there is a negative relationship between labor flows and distance from the workers' birthplace; and (iii) workers accept a wage discount between 9 to 11 percent to live in their home location. To understand the implications of these findings, I build a dynamic quantitative migration model into which I introduce home bias, understood as a utility cost of living away from one's birthplace. I use the model to separately identify home bias and migration costs from the data. I find that differences in birth location lead to average welfare differences of up to 30 percent in consumption-equivalent terms, and explain 43 percent of the total dispersion in welfare. Finally, I show that a migration model without home bias overstates the migration response of agents. This underestimates the pass-through of local productivity to real wages and overestimates the efficiency costs associated with place-based policies.

In the second chapter, Miren Azkarate-Askasua 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 third chapter, Miren Azkarate-Askasua 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 and depending on the specification the correlation becomes positive after the correction.

Résumé

Ce travail de thèse est composé de trois chapitres traitant des effets macroéconomiques du marché du travail en mettant l'accent sur les déterminants de la migration interne, les inégalités spatiales, le pouvoir du marché du travail et la détermination des salaires.

Dans le premier chapitre, j'étudie une raison potentielle pour laquelle les travailleurs restent dans des zones économiquement en difficulté: les gens aiment vivre près de leur location d'origine. En utilisant des données administratives françaises, j'obtiens les résultats suivants: (i) la part de migrants qui retournent dans leur lieu de naissance est presque deux fois plus grande que la part de migrants qui se rendent dans une autre localité; (ii) il existe une relation négative entre les flux de main-d'œuvre et la distance par rapport au lieu de naissance des travailleurs; et (iii) les travailleurs acceptent une réduction de salaire de 9 à 11 pourcent pour vivre dans leur localité d'origine. Pour comprendre les implications de ces résultats, je construis un modèle de migration quantitative dynamique dans lequel j'introduis un biais d'origine, compris comme un coût en terme d'utilité de la vie loin de son lieu de naissance. J'utilise le modèle pour identifier séparément le biais domestique et les coûts de migration à partir des données. Je trouve que les différences de lieu de naissance entraînent des différences de bien-être moyen allant jusqu'à 30 pourcent en termes d'équivalence de consommation et expliquent 43 pourcent de la dispersion totale du bien-être. Enfin, je montre qu'un modèle de migration sans biais d'origine surestime la réponse migratoire des agents. Cela sous-estime la répercussion de la productivité locale sur les salaires réels et surestime les coûts d'efficacité associés aux politiques territoriales.

Dans le second chapitre, Miren Azkarate-Askasua 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 signi-

ficative du bien-être des salariés. La mobilité géographique est la clé pour réaliser les gains de la compétition.

Dans le dernier chapitre, Miren Azkarate-Askasua 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.

Table of Contents

List of Tables

List of Figures

Chapter 1

The Birthplace Premium

Miguel Zerecero[1]

Abstract

Why do people stay in economically distressed areas? In this paper, I explore a simple, yet overlooked hypothesis: people like to live close to what they call home. Using administrative data for France, I find: (i) the share of migrants who return to their birthplace is almost twice as large as the share of migrants who go to any other particular location; (ii) there is a negative relationship between labor flows and distance from the workers' birthplace; and (iii) workers accept a wage discount between 9 to 11 percent to live in their home location. To understand the implications of these findings, I build a dynamic quantitative migration model into which I introduce home bias, understood as a utility cost of living away from one's birthplace. I use the model to separately identify home bias and migration costs from the data. I find that differences in birth location lead to average welfare differences of up to 30 percent in consumption-equivalent terms, and explain 43 percent of the total dispersion in welfare. Finally, I show that a migration model without home bias overstates the migration response of agents. This underestimates the pass-through of local productivity to real wages and overestimates the efficiency costs associated with place-based policies.

JEL Codes: J61, R12, R13, R23

Keywords: Internal Migration, Home Bias, Spatial Inequality

[1]email: miguel.zerecero@tse-fr.eu. I am indebted to my advisor Christian Hellwig for his support throughout this project. I am also especially grateful to Fabrice Collard and Patrick Fève for their extensive suggestions. I also thank Tiziana Assenza, Miren Azkarate-Askasua, Matteo Bobba, Martial Dupaigne, Alexandre Gaillard, Eugenia Gonzalez-Aguado, Elia Lapenta, Tim Lee, Nour Meddahi, François Poinas and Nicolas Werquin for their comments and suggestions. I finally thank the participants at the TSE Macro workshop and researchers at Carlos III, CEMFI, Queen Mary, Royal Holloway, Stockholm School of Economics and UC Irvine for useful comments. I gratefully acknowledge financial support from TSE to get access to the administrative data provided by CASD. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'avenir' program (reference : ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD). All remaining errors are my own.

1.1 Introduction

Large groups of people tend to stay in less favorable areas within the same countries. It is puzzling that, even without legal impediments, they don't move to supposedly attractive locations. The literature has offered two main explanations. First, migration costs reduce mobility across regions, which limits workers' ability to arbitrage away differences in welfare.[2] Second, the observed variation in pecuniary measures, like real wages, might only reflect variation in local amenities. Thus, low-wage regions might only reflect a high level of amenities.[3]

In this paper, I focus on a different explanation for low mobility: people like to live close to their home. This home bias makes workers born in attractive regions better-off, as they don't have to compete with workers born in poorer regions who are reluctant to leave their home. Home bias can then generate significant average utility differences across space and birth cohorts. For example, considering the case of France, I find that the average worker born in an attractive area—like Paris, Nice, or Toulouse—has 5 to 7 percent more utility than the average French worker, measured in consumption terms. In contrast, the average worker born in Cantal, within the Massif Central region, or in Haute Marne in the North-East, has around 20 percent less utility than the average French worker. Thus, the difference between having a "good" and a "bad" birthplace can turn into a welfare difference of more than 30 percent, which is significant considering that France is a centralized and well-connected country.

In relative terms, these numbers imply that differences in birth location explain 43 percent of the overall welfare dispersion. With 53 percent of the welfare dispersion due to workers' idiosyncratic shocks, this means that differences in birth location account for almost all the rest of the variation. This result reflects the importance of home bias in shaping workers' location decisions which, combined with location-specific heterogeneity, makes the birthplace an important driver of expected lifetime utility. Ignoring the effect of home bias overstates the role of migration costs and the potential for policies to enhance mobility. It also overstates the costs of subsidizing poor locations, that may drive away workers from productive to unproductive regions.

I proceed in four steps. First, using administrative data for France, I document the prevalence of home bias in workers' migration decisions. The French data stand out as they register the birth location for all workers. This feature allows me to look at labor flows between two regions for workers who were born in different places, which is key for isolating the home bias from the effect of proximity in migration decisions. I find that labor flows are biased towards workers' home locations, even after controlling for proximity between origin and destination locations, and that workers who live in their home location have lower wages. Second, I build a general equilibrium dynamic Roy model of migration in which workers with heterogeneous preferences—defined by their birthplace—sort across locations with heterogeneous productivities and amenities. I use the structure of the model and the observed data on labor flows and wages to separately identify the standard migration costs from the home bias. Third, I use the estimated model to quantify the

[2]Bryan and Morten (2019) and Caliendo et al. (2019) have models with costly adjustment of labor across regions; Ahlfeldt et al. (2015) and Monte et al. (2018) propose a model where commuting is costly.

[3]Compensating variation in real wages because of amenities is a standard result in the traditional urban framework of Rosen (1979)-Roback (1982).

birthplace premium: the average utility a worker from a particular birthplace has in excess of the national average. Fourth, I illustrate the effect of ignoring home bias when modeling workers' mobility decisions.

I start by briefly describing the data in Section 1.2 and explaining how I define the different locations within France. The most disaggregated level of information for place of birth is the *département*. There are 94 *départements* in continental France with great variation in size and connectivity.[4] I aggregate them according to commuting flows, such that every location is a well integrated local labor market. I end up with 73 locations which still allows me for a disaggregated analysis of the home bias.

Section 1.3 provides empirical evidence of the home bias. I examine the labor flows across locations in France for the years 2002 to 2017. I find that the share of migrants who return to their birthplace is, on average, almost twice as large as the share of migrants who go to any other particular location. To distinguish between the effect of standard migration costs and home bias, I run a gravity-type regression, as used in the trade literature, and find that the labor flows to a particular destination is negatively related to distance from the workers' birthplace. This result holds while controlling for distance between origin and destination locations, that would capture normal migration frictions, as well as origin and destination fixed effects.

The biased labor flows suggest that workers dislike living away from their birthplace. This allows me to test whether idiosyncratic differences in wages are an important driver of workers' migration decisions. If workers select across locations based on differences in potential wages, and leaving the birthplace is costly, then workers who move away from their birthplace should have, on average, higher wages than those workers who stayed in their birth location. I find that for the vast majority of locations/periods of my sample the wages of workers living outside their birthplace are larger than the wages of workers living within their birthplace. This corresponds to an average 15 percent wage difference between the two groups. Thus, the evidence suggests that selection via wages is an important driver of the workers' location decision. I then estimate the average penalty workers face by living in their birthplace. I find that among workers who changed jobs between years, those who move back to their birthplace face a wage discount of 9 to 11 percent compared to going to another location.

In Section 1.4, I build a quantitative migration model in the spirit of Bryan and Morten (2019)—where differences in idiosyncratic productivities drive workers' migration decisions—but allowing for migration to be a dynamic decision, as in Caliendo et al. (2019). I add a fixed worker characteristic, birthplace, that biases the migration decision of workers towards their home. The static part of the model is a trade model à la Eaton and Kortum (2002) with housing, which works as a congestion mechanism. The combination of all these elements results in a dynamic discrete choice model—where workers with heterogeneous preferences defined by birthplace sort across heterogeneous locations based on idiosyncratic productivity shocks—with a static trade equilibrium determining output at each location.

The methodological challenge is to disentangle the role of home bias from standard migration costs along with identifying location-specific characteristics, like productivities and amenities,

[4] For continental France I mean the French *départements* that are in Europe, excluding the island of Corsica.

that are common in the trade and urban economics literature. Adding worker heterogeneity—like birthplace—allows for a richer analysis of phenomena, but it comes with a cost. A common feature in the discrete choice literature, especially when choices are persistent, is that a large probability mass is concentrated in a single alternative. Then, is usual to observe in the data a large fraction of alternatives where the number of people taking them is zero. Adding group heterogeneity, by conditioning in an extra dimension, increases the prevalence of zeros in the data. This represents a challenge when trying to bring together model and data. In my context, although the data consists of millions of observations, the number of workers migrating in a given year is around 4 percent of the total sample. Moreover, the number of origin-destination combinations per *each* group of workers with same birthplace is $73 \times 73 \approx 5,000$. These two elements make the data on observed combinations, conditional on birthplace, very sparse.[5]

As in Dingel and Tintelnot (2020), I address the "many-zeros" problem by assuming a discrete number of workers in the model. This assumption rationalizes the zeros in the data and guides the identification strategy in a transparent way. However, it poses challenges when solving the general equilibrium of the model.[6] Thus, I present two versions of the model: one with a discrete number of workers where the equilibrium needs not be in steady-state, and a more standard steady-state continuous-population model, which I use for computing general equilibrium counterfactuals.

In Section 1.5, I show how to identify and estimate the parameters of the model, using data on labor flows and wages. I show that, if migration costs are symmetric, they are non-parametrically identified from labor flows across locations.[7] I relax the sufficient identification conditions provided by Bryan and Morten (2019)—and the associated data requirements—such that the migration costs are identified from the location-pair fixed effects of a gravity Poisson regression on labor flows.[8] Bryan and Morten show that migration costs can be directly identified from the gross migration flows between two locations. In the context of my application, this requires to observe, for every pair of locations, an out-flow and an in-flow of labor for workers with the *same* birthplace and in the same year. In the data, less than 70 percent of the location pairs satisfy Bryan and Morten's conditions. With my weaker conditions, this number increases to more than 98 percent.

For tractability, I assume that the idiosyncratic productivity shocks are distributed Type 1 Extreme Value (or Gumbel). This assumption—ubiquitous in the discrete choice literature—delivers a closed form expression for the migration probability as a function of the expected utility and the migration costs.[9] Using the identified migration costs and count data on labor flows I estimate the underlying migration probabilities via maximum likelihood. I show that the solution to

[5]The sum of origin-destination combinations across workers with different birthplace is then $73^3 = 389,017$. I observe around 5% of the combinations each year.

[6]The lack of information about different alternatives might lead researchers to aggregate the alternatives into a smaller choice set, which makes it easier to combine model and data. This is a reasonable route for some applications. For example, Heise and Porzio (2019) analyze the effect of home bias for location decisions of East and West German workers. Germany stands out against other countries as it is obvious in how to group different locations in few regions for its analysis. For France though, is not obvious how to group locations into two, three or few more aggregate regions. Thus, aggregation could mask the effect of home bias in workers' migration decisions.

[7]By non-parametric I mean that I identify a single migration costs for every pair of locations.

[8]The gravity Poisson regression would be a *three-way* regression in the sens that it includes origin, destination and location-pair fixed effects.

[9]For a textbook treatment, see Train (2009).

this maximization problem is equivalent to solving for the 'source-country effects' of a balanced trade condition from a gravity-trade model.[10] I use the identified migration probabilities to impute model-consistent wages for those missing combinations in the data.

The result linking the maximization of the conditional likelihood and the gravity model complements the work of Dingel and Tintelnot (2020) on how to combine spatial quantitative models and sparse data on alternatives. Within my migration context, the system to solve is a collection of labor-movement equations, where the total labor at a destination is the sum of the probability of migrating to the destination—which is a function of the fixed effects—times the number of workers at origin locations. Thus, the fixed effects are estimated with the number of workers at every origin and destination in a given time and not the labor flows which are oftentimes unobserved. Fortunately, trade economists have already tackled the problem of how to efficiently solve these type of systems.[11] Thus, my result adds to the set of 'computational tricks' that allow for the feasible estimation of quantitative spatial models.

Next, I identify the home bias parameters using the information contained in the difference between the average wage of workers living outside their birthplace and the average wage of those returning to home. The idea is that the worker who returns home would accept a wage penalty, everything else equal. Similar to Artuç et al. (2010), I use the information from next period wages to control for the option value of future employment opportunities at each location which are embedded in the workers' continuation values. Similarly to the migration costs, I assume the home bias is symmetric across locations and birthplaces to non-parametrically identify them from the data.

I identify the remaining parameters, the distributions of productivities and amenities, following the standard approach in the quantitative spatial economics literature; see Redding and Rossi-Hansberg (2017). I identify the distribution of productivities by *inverting* the static part of the model such that the recovered distribution is consistent with the equilibrium and the observed wages. The amenities are recovered as a residual that explains the remaining variation in labor flows.

In Section 1.6 I compute counter-factuals to assess the welfare impact of birthplace preferences using the steady-state continuous-population version of the model.

As my main result, I compute the different birthplace premia and decompose welfare inequality where I distinguish between aggregate dispersion at the birthplace/location level and idiosyncratic dispersion, stemming from the individual-specific productivity shocks and geographic sorting. I find that individual heterogeneity and sorting explain 53% of the variance of individual welfare levels. Variance of between-birthplace average welfare explains 43% of the variance. The importance of home bias in determining where workers end up living—along with heterogeneity in attractiveness of locations—means that birthplace is a big determinant of expected lifetime utility.

[10] The term 'source-country effects' is borrowed from Eaton and Kortum (2002). In a gravity-type equation, let $X^{i,j}$ be the share of expenditure a country i spends in goods from country j. If $X^{i,j} = f(\mathcal{F}^j)$ is a function of some fixed effect $\mathcal{F}^j$ specific of the *source* country j, then all of these fixed-effects $\{\mathcal{F}^j\}$ are the 'source-country effects'.

[11] In particular, I borrow the algorithm proposed by Pérez-Cervantes (2014) which is well suited for a very large number of fixed effects and very easy to implement. Ahlfeldt, Redding, Sturm, and Wolf (2015) propose an alternative algorithm in the web appendix of their paper.

The main result shows that geography shapes long-run welfare inequality through birthplace. The reason is that home bias changes workers' location patterns in the long-run by making them gravitate around their home location. Thus, large differences across locations imply large welfare differences across workers with different birthplaces.[12] In contrast, without home bias, workers can arbitrage away the differences across locations, especially in the long-run. This makes initial geographic differences less important in shaping inequality.

Next, I compare the magnitudes of migration costs and home bias. To make migration costs, which are paid once, comparable to home bias, which corresponds to a flow utility costs, I rely on a compensating variation argument. I compute how much more consumption a migrant worker needs to have the same utility as a non-migrant worker. Similarly, I compute the compensating variation in consumption for a worker who lives outside her birthplace to have the same lifetime utility as a worker who lives in her birthplace. I find that the compensation for a migrant is 55.6 percent, while the compensation for a worker who lives outside her birthplace is 18.6 percent.

I then compare the effects of removing migration costs or home bias on output. Removing the home bias increases output by 11%, while removing migration costs raises output by more than 30%. In both cases, productivity gains are the result of better sorting of workers by idiosyncratic productivities, while gains from reallocation to more productive areas are minor and can even be negative.

In addition, I compare my model to one without home bias. I find that, while the estimated average migration cost is 10% larger, the average migration elasticity is 8% larger in the model without home bias, overstating the mobility response of agents. This in turn underestimates the average pass-through of productivity to real wages by 50% in the model without home bias, as the in-migration flow is larger which increases the price of housing.

In a similar vein, the model without home bias changes the predictions when evaluating place-based policies compared to my model with home bias. A common concern of such policies, is that, while aiming at some spatial redistribution of income, it also distorts the location decisions of workers of non-targeted locations. Thus, it can drive workers away from productive to unproductive locations, resulting in efficiency losses. However, if workers mobility is limited by their home bias, the associated efficiency costs to a place-based policy is limited. I impose a labor subsidy to each location, and compare the response on social welfare one-by-one in both models.[13] I find that the model without home bias has a misdiagnosis rate of 52%. This means that for more than half of the cases, the model without home bias predicts that subsidizing a particular location has the opposite effect on social welfare than a model with home bias.

All together, the different exercises teach us that home bias matters for the aggregate economy. By hindering the mobility of workers, home bias makes the birthplace an important determinant of overall welfare inequality. Neglecting its importance leads to over-stating the role of worker mobility as a force for welfare equalization.

[12] Consider the extreme case where home bias is prohibitive, and all workers live in their respective birthplace. Then, if geography would be the same, then there should be no dispersion of welfare across workers with different birthplaces.

[13] The social welfare would correspond to the sum of welfare across all agents in the economy, not just those that live in the subsidized location.

Literature This paper is related to several strands of the literature. First, it adds to the empirical evidence of the presence of a home bias in migration decisions. For example, Kennan and Walker (2011) find, for a sample of U.S. individuals, that half of the people who move return to their home location; Bryan and Morten (2019) find, for the case of Indonesia, that the share of people that migrate to a location from a particular birthplace is negatively correlated with distance; similarly, Heise and Porzio (2019) using data from Germany for the years 2009-2014, find that people born in East Germany are more attracted to live in East counties than individuals born in West Germany. My paper contributes to this literature by documenting a home bias effect for France. The presence of a strong home bias effect in France is not obvious a priori as: (i) it is a relatively small and well connected country, at least compared to the U.S. and Indonesia; (ii) it has been historically unified, in contrast to Germany; and (iii) it faces no linguistic or geographical barriers, which is the case of Indonesia.[14] Furthermore, the administrative data that I use allow for a clear separation of birthplace versus origin of the labor flow. This allows me to disentangle the effect of home bias versus the effect of proximity in driving the labor flows.

Second, the paper is related to the growing literature on the macroeconomic implications of worker sorting.[15] Akin to Bryan and Morten (2019), my paper bridges this literature on worker selection with the literature on the aggregate implications of workers' geographic mobility across heterogeneous locations.[16] Differently from them though, I combine selection and costly mobility in a dynamic framework to disentangle migration costs from the home bias. I also allow for costly trade across regions, where workers benefit from living *close* to a productive location. Without costly trade, all locations benefit equally from a productive location regardless of proximity.

Third, my work is related to the fast-growing quantitative spatial economics literature. I contribute to this literature by expanding the results of Dingel and Tintelnot (2020) on how to estimate these models without neglecting the sparsity of the data. Normally, quantitative spatial models are composed of agents making discrete choices from a large set of alternatives. It is usual for those models to assume a continuum of agents such that choice probabilities and the share of individuals taking that choice are (almost surely) equivalent. When the number of choices is large, say, the number of products or commuting patterns, these models encounter a 'many-zeros' problem, i.e., the observed data has many choices with no individuals taking them. This creates a disconnect between theory and data that is either ignored, or is addressed by ex-ante 'smoothing' the data, like in Almagro and Domınguez-Iino (2020).

In contrast to the previous literature, Dingel and Tintelnot (2020), propose a model with a discrete number of agents, which can rationalize the zeros in the data. They show that the estimation

[14] Indonesia is an archipelago that consists of 17,508 islands and there are more than 300 different native languages. Bahasa Indonesia is the official language, which is the mother tongue for only 7% of the population.

[15] Lagakos and Waugh (2013) and Young (2013) focus on the role of selection on unobservable skills to explain the rural-urban wage gap. Adão (2015) and Galle et al. (2017) present trade models where heterogeneous workers select across sectors. They use such frameworks to quantify the impact of trade on inequality and welfare. Young (2014) quantifies to what extent the differences in measured productivity between the manufacturing and service sector are due to worker selection. Hsieh et al. (2019), using a model of occupational choice due to heterogeneous skills, study how discrimination of minorities affected aggregate productivity in the U.S.

[16] For example, see Redding (2016), Diamond (2016), Monte et al. (2018), Caliendo et al. (2019), Caliendo et al. (2020), Schmutz and Sidibé (2019) and Monras (2020).

of such a model by means of maximum likelihood, which consists on estimating a non-linear model with a large number of fixed effects, is computationally feasible. They rely on a result from Guimaraes et al. (2003), who show that there is an equivalence relation between the likelihood function of the conditional logit and the Poisson regression.[17] Given the identification strategy I follow in my paper, I cannot exploit this result. Instead I show that the maximization of the conditional logit likelihood with one dimension of fixed effects is equivalent to solving the 'source-country effects' of a balance-trade condition in a gravity-type model.

The closest precedent to my paper are the works of Heise and Porzio (2019) and Zabek (2020). In addition to documenting a home bias effect when comparing East and West German workers, Heise and Porzio develop a general equilibrium job-ladder spatial model, where workers of heterogeneous productivity select across locations given their observed wage offers. They distinguish between traditional migration costs and the home bias. They calibrate and solve the model for two regions: East and West Germany. They find that spatial frictions are relatively small compared to other labor market frictions that prevent the reallocation of labor across firms. This creates modest output gains from removing migration costs.

Similar to Heise and Porzio (2019), I allow for a labor market friction that prevents workers to change jobs even within locations.[18] Different from them though, I show how to incorporate home bias and selection in an otherwise standard quantitative migration model which is suitable for the analysis of a geography consisting of a much larger number of locations.[19] In contrast to their results, I find that migration costs are actually important and removing them increases output by more than 30%, which dwarfs their change in output of 0.46%.[20] In both models, the wage premium to induce a worker to migrate must account for direct migration costs as well as changes in the option value of future employment opportunities. In my model changes in future employment opportunities are small relative to migration costs, since the probability of changing jobs is independent of one's birthplace and current residence. In their estimates, future wage offers depend on a worker's current residence and origin, suggesting that changes in future job prospects may constitute an important hidden cost of migration.

As in both this paper and in Heise and Porzio (2019), Zabek (2020) recognizes the importance of home bias to generate persistent differences in welfare across locations. He presents a Rosen-Roback model where workers of identical skills have stochastic preferences for staying at their home location. The distribution of these home preferences are the same conditional on birthplace, but in equilibrium, depressed places are going to endogenously retain workers who value their home

[17]Currently, there are several statistical packages that can handle the estimation of Poisson regressions involving a large number of fixed effects.

[18]In my model, I let migration to be a persistent choice, by understanding a movement across locations as a job-change, and assuming that an exogenous process determines with some probability if a worker has to change jobs between periods. Therefore, when a worker makes a migration decision, it takes into account that, whatever job she takes, it might last for long. Hence, initial differences in idiosyncratic productivities are magnified by the exogenous persistence of jobs, increasing the perceived variation of jobs across locations. This effect reduces the labor supply elasticity to a location.

[19]Furthermore, the French data that I use register birthplace, in contrast to Heise and Porzio, who assign a worker's home location to be the first location registered in the data. Also my data consists of more than 10 million observations per year which allows for a more disaggregated analysis of the home bias and migration costs.

[20]See Table 5 in Heise and Porzio (2019).

highly. This will generate lower real wages and smaller migration elasticities as the average inhabitant of a depressed place is more reluctant to leave. In contrast, in my model I don't have stochastic preferences but rather the home bias enters as a location-pair-specific utility cost. Nevertheless, it still generates smaller migration elasticities in depressed areas, as the proportion of natives would be endogenously larger in such places, making the average inhabitant also reluctant to go.[21] Another distinction is that the U.S. data he uses does not allow him to observe the place of previous residence, so he can't distinguish between home bias and standard migration costs. Finally, in his model, he allows for an endogenous evolution of the population birthplaces. This generates a spatial equilibirium force that would lead to eventual convergence in welfare, however it would be very persistence as it takes generations to change the home bias of workers. Given this very slow process of convergence, I abstract from such endogenous formation of workers' birthplaces in my model and focus on the evolution within a single generation.

1.2 Data

Most of my analysis relies on the *Déclaration Annuelle des Données, fichier Postes* (DADS Postes) data set for the years 2002-2010 and 2012-2017, which contains information about all non-agricultural workers in private and semi-public establishments in France. I don't include the year 2011 in my sample because there is no information about the birth department of workers for that year.[22] Appendix 1.E contains the details on the sample selection.

The unit of observation is a job, which is defined as a worker-establishment pair in a given year. This means that there might be more than one observation per worker every year. An establishment is a combination of firm/location. Therefore, by definition, a worker who moves across locations and does not commute to work in her old establishment would appear as having a new job.

The data have information on the workers' age, gender, wage, place of birth, residence and work. Starting in 2002, there is an indicator of which observation per worker is the main job (*Poste Principal*). A main-job is defined as the job with longest duration that a worker has in a given year. To keep one yearly observation per worker, I only use these main-job observations in the analysis. There is also information on the starting and ending dates of the job. While not being a panel, the data include information on the previous year's values for almost all of the variables. This allows me to recover migration patterns for people with different birthplace. It also allows me to identify which individuals changed jobs, even if they did not move. In Appendix 1.E.1 I explain in more detail how to identify these job *switchers* in the data.

The most disaggregated level of information for place of birth is the *département*. I constrain the analysis to continental France which excludes all the territories outside Europe, i.e. the *Départements et Régions d'Outre-Mer* (DROM), and the island of Corsica. In continental France there are 94 departements, which vary very much in size and connectivity among each other. For example,

[21] This composition mechanism is also present in Zabek's model. However, in his paper, he chooses to emphasize the mechanism that, conditional on being native, the preference to stay at the home location is on average higher for depressed places.

[22] For the interested reader, a similar dataset, *DADS Poste Principal* which is a sub-sample of *DADS Postes* does have the information for the year 2011. I currently don't have access to these data.

Figure 1 – Aggregation of *départements*. The locations that are aggregated are shaded in blue, while the old departemental borders are shown within the shaded area. In total I consider 73 locations for continental France.

there are 8 departements just in the super-dense region of Île-de-France, which has just about the same surface as the departement of Gironde—where Bordeaux is located.[23]

To make the geographical unit of analysis comparable, instead of using directly the departements, I aggregate a few departements according to their commuting patterns. Given data on departement of residence and of work for each worker, I can retrieve all the inter-departement commuting flows. I group two departements if two conditions are satisfied: first, the number of workers who commute from one departement to another is larger than 10% of the number of workers from the origin departement; second, both departements belong to the same *région* before the 2015 territorial reform.[24] After aggregating the different departements, I keep only the observations of workers who live and work within that same location.

In total I end up with 73 locations for continental France. Figure 1 shows the different locations I use in the analysis. The locations that are the union of different departements are shaded in blue. Within aggregated locations, the departemental borders are visible with finer lines. Most of the aggregated locations consist of two departements. The notable exceptions are the areas surrounding the cities of Paris, Lyon and Toulouse, which are, respectively, the first, third and fourth most populated cities of France.[25] The departement that has Marseille, which ranks as second in terms of populous cities, only aggregates with one neighboring departement.

My final sample consists of 202,521,533 job-worker observations distributed along 15 years and the different $73^3 = 389,017$ origin-destination-birthplace combinations.

[23] The surface of Île-de-France is $12,012.27$ km^2 while that of Gironde is $10,000.14$ km^2.

[24] There are 21 old *régions* in continental France. These would be similar to a State in the United States. In 2015 there was a territorial reform grouping some of these regions together. Currently there are 12 *régions* in continental France.

[25] The other exception would be the group formed by the Northeastern departements of Doubs, Haute Saône and Territoire de Belfort. The latter is, outside Île-de-France, the smallest departement in whole France and includes the relatively large city of Belfort, whose metropolitan area also includes a *commune*–Châllonvillars—that is in the departement of Haute Saône. Thus, the commuting flows between the two are large.

1.2.1 Basic terminology

Before describing the summary statistics let me introduce some terminology that I use in the rest of the paper. I say that a worker is a *native* if she lives in the same location where she was born. A *migrant* is a worker who just moved to a particular location in the current year, regardless of her birthplace. If in the next year the migrant stays in her current location, then she would stop being classified as a migrant. I call a *birth* cohort, or *birthplace* cohort, all the workers who were born in the same location. A *migration* cohort corresponds to all the workers with the same birthplace and with the same origin-destination locations in a given year. Thus, all those workers with the same birthplace who stay in the same location from one year to the next would constitute as well a migration cohort. Finally, I call a worker *switcher* if she changes jobs from one year to the next.

1.2.2 Summary Statistics

Table 1 presents worker and location level summary statistics for the final sample. The left panel shows some statistics about the number of workers per year/location in the sample. I observe over 13 million workers per year, but naturally the data at end of the sample—in 2017—are larger. The average number of workers per location-year is more than 180,000. However, as there are locations that are much larger—like Île-de-France or Lyon—the standard deviation is almost twice as large as the average number of workers per location. As the number of locations and birthplaces is the same, the average number of workers per birthplace-year is the same as the average for location-year. However, there is less heterogeneity across birth cohorts size than that of locations as the standard deviation is 5% smaller. This probably reflects the fact that some workers move out of their birthplace and concentrate in the most populous locations. There is a surprisingly high persistence in the relative number of workers of either locations or birth cohorts. The correlation between the number of workers in each location or with a particular birthplace for the first and last year of my sample—the years 2002 and 2017— is greater than 0.99.

The top-right panel in the table describes some details about different sub-groups of workers in the sample. The average proportion of workers who change jobs between years—or switchers—is 13 percent. Using the entire sample or only the switchers, I find that a similar proportion of around 65 percent of workers live within their birthplace. Only an average of 0.5 percent of the total sample migrates from year to year. When considering only switchers, the proportion of migrants increases to almost 4 percent. This is not surprising as each job is, by definition, linked to a location, so workers who migrate are necessarily switchers. Nonetheless, even for those workers who are changing jobs the proportion that migrate is still low. I also find that women have a smaller propensity to migrate, but not by much.

Regarding the age composition of the different groups in my sample, I find that, in general, switchers are younger, as shown in the bottom-right panel of the table. This can reflect that older workers find better, more stable jobs. In general, natives, non-natives and non-migrants have similar average age either for the whole sample or just focusing on switchers. Migrants have a similar age as those that don't migrate but change jobs. Finally, I find that those who return to their birthplace are on average older than those who leave it. This can indeed reflect that most workers start their

Table 1 – Summary statistics

	Value		All	Switchers
Number of workers		**Workers (%)**		
Per year	13,501,436	*Switching* jobs	13	–
Year 2002	11,052,111	Workers within birthplace	66	64
Year 2017	15,493,563	Workers Migrating	0.5	3.8
		Women Migrating	0.4	3.1
Average per Location/Year	184,951.2			
S.D. per Location/Year	339,745.6	**Age (years)**	40.58	35.04
S.D. per Birth Cohort/Year	325,787.1	Natives	40.07	34.06
		Non-Natives	41.57	36.86
Correlations, 2002-2017		Non-Migrants	40.61	35.08
Workers per Location	0.996	Migrants	–	34.25
Workers per Birthplace	0.998	Return Birthplace	–	36.21
		Leave Birthplace	–	30.96

Note: The left panel shows summary statistics regarding the number of workers in the sample. Average number of workers per location is the same as the average number of workers per birth cohort as the number of locations and birthplaces is the same. The first correlation is between the population vector living in each location in the years 2002 and 2017. The second correlation is the same but comparing size of birth cohorts. The right panel distinguishes, when possible, between the whole sample and using just the subset of workers who switch jobs. The top-right panel has summary statistics about the proportion of workers: (i) that change, or switch, jobs; (ii) that live within their birthplace; (iii) that migrate; and (iv) that are women. The bottom-right panel shows average ages for different sub-groups of the sample.

work life in their birthplace, so their first migration move has to be out of their birthplace.

1.3 Empirical Evidence on Home Bias and Selection

Using the labor flows and average wages, I document four empirical facts about the French labor market. These facts help to motivate the model I present in the next section.

Fact 1: Most workers live in their birthplace.

The average proportion of workers who live in their place of origin is 66%, as was already shown in Table 1. This could reflect just that workers tend to start their work life in their home location and later face strong migration costs. However, looking closer at the labor flows, systematic biases can be found, as Fact 2 below shows.

Fact 2: Labor flows are biased towards birthplace.

To establish Fact 2, I first show that the share of workers who returns towards the birthplace is larger, on average, than the share of workers migrating to any other location. Using workers with the same birthplace, I compute the number of workers who migrated between any two locations as a share of the total number of workers who migrated from the origin location. More formally, I

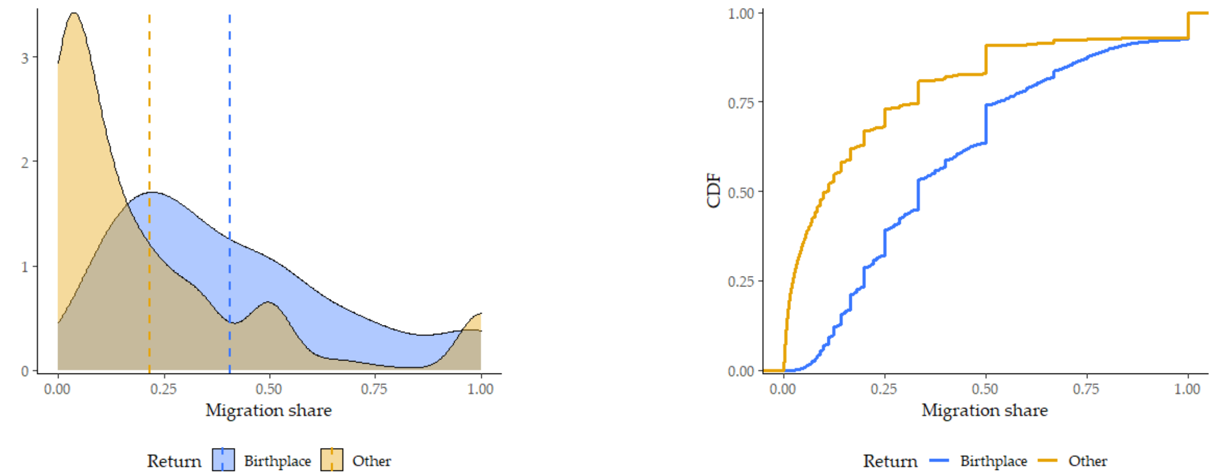


Figure 2 – Distribution of conditional migration shares. These are defined as $\tilde{s}^{i,j}_{b,t} = \frac{L^{i,j}_{t,b}}{\sum_{k \neq i} L^{i,k}_{t,b}}$, where $L^{i,j}_{t,b}$ is the number of workers who were born in location b and that moved from location i to j at year t. Both plots distinguish between the migration shares that returned to the workers birthplace versus all the other locations. The left panel plots the densities while the right panel plots the cumulative distribution function.

compute

$$\tilde{s}^{i,j}_{b,t} = \frac{L^{i,j}_{t,b}}{\sum_{k \neq i} L^{i,k}_{t,b}},$$

where $L^{i,j}_{t,b}$ is the labor flow, i.e., the number of workers who were born in location b and that moved from location i to j at year t.

Using these migration shares, I find that the share of migrants who return to their birthplace is, on average, almost twice as large as the share of migrants who go to any other particular location. For example, consider workers from Toulouse who live in Paris. Of those who are moving away from Paris, the share that moves back to Toulouse is, on average, twice as large as the share that goes to, say, Lyon.

The bias of migration shares towards workers birthplace becomes more evident if I look at the distributions instead of just the averages. I compare the distribution of migration shares $\tilde{s}^{i,b}_{t,b}$ for which $b = j$—where the destination is equal to the birthplace—with the distribution of all other migration shares, for which $b \neq j$. Without home bias, a worker's propensity to move to any other location should be independent of their birthplace, hence the two distributions of migration shares should look similar. The left panel of Figure 2 plots the densities of both distributions. The two distributions are very different: the distribution of return migration has a larger mean, median and mode, and is less skewed to the right. Moreover, as the right panel of Figure 2 shows, the distribution of shares associated with workers returning to their birthplace first-order stochastically dominates the distribution of migration shares going to alternative destinations.[26]

Although the share of workers who migrated back home is on average larger, this could just reflect that the origin locations were close to their home to begin with. Thus, the distribution

[26] As both figures show, some of the migration shares are equal to one. This means that for a particular year, the group of workers with the same birthplace that moved out of their current location all went to one particular destination. This is a reflection of the sparsity of the data that arises from conditioning migration shares by birthplace.

differences are only reflecting the effect of proximity, not home bias. To disentangle the effect of proximity from home bias, I estimate a gravity-type model directly over the labor flows. I find that labor flows are biased to the birthplace even if I control for traditional migration frictions, proxied by distance between origin and destination locations. In particular, I run the following Poisson regression

$$L_{t,b}^{i,j} = \exp\left(D_{t,j} + O_{t,i} + \mathbf{1}_{j\neq b}(\alpha_1 + \beta_1 \log(\mathrm{d}^{b,j})) + \mathbf{1}_{j\neq i}(\alpha_2 + \beta_2 \log(\mathrm{d}^{i,j})) + \varepsilon_{t,b}^{i,j}\right),$$

where $L_{t,b}^{i,j}$ is defined as above, the labor flow of workers born in b that move from location i to j at time t. The fixed effects $O_{t,i}$ and $D_{t,j}$ are, respectively, origin/year and destination/year specific and should control for any differences between the origin and destination that are constant across migration cohorts. This would include differences in size, amenities, cost of living, etc. The variable $\mathrm{d}^{i,j}$ denotes the distance between locations i and j, while $\mathbf{1}_{j\neq i}$ is an indicator function.

The model is in levels—instead of logs—to accommodate all the zero labor flows observed in the data. These zero flows are pervasive as the number of options per year is quite large and the percentage of people migrating every year is low.[27] If I were to estimate the model on log terms using only positive flows, I would lose a lot of information, potentially biasing the results. Thus, I estimate the previous model doing a Poisson regression.[28]

The first three columns of Table 2 show the results using different variables for distance.[29] As the table shows, there is a statistically significant negative relation between moving away from one's birthplace, as reflected by the estimated coefficient β_1. Both distance elasticities, β_1 and β_2 are estimated of similar magnitude. Although for some specifications the constant term α_1, associated to the dummy of living outside one's birthplace is estimated positive, this is only a reflection of the choice of unit of measurement for distance. The overall effect on the labor flows is always negative.[30]

What happens if, instead of using directly the labor flows, I use the workers who move as a share of the origin population, i.e., $L_{t,b}^{i,j}/\sum_k L_{t,b}^{i,k}$? The last three columns of Table 2 show the results of those regressions. With this specification, although the elasticity with respect to distance from birthplace β_1 is still negative, its magnitude is nowhere close to the elasticity of distance across origin and destination β_2. However, looking at the overall effect of living outside the birthplace, i.e. considering α_1, this is always negative and significant.[31]

The key takeaway from the gravity regressions is that, even after controlling for traditional migration frictions, the labor flows are negatively related to distance from the workers' birthplace. This result is robust to different specifications which are further explored in Appendix 1.H. I estimate both models using *département* as locations instead of the aggregated regions I used here.

[27] Recall that the number of options per year is equal to $73^3 = 389,017$.

[28] See Silva and Tenreyro (2006) regarding the advantages of using the Poisson regression over OLS with log terms for the estimation of gravity models.

[29] I use geodesic distance, driving distances and driving hours from Google Maps.

[30] In particular, the minimum value of log geodesic distance in kilometers in the sample is 3.82. The analogous for diving distance is 4.13. Thus, the maximum value of the total effect for a worker leaving her birthplace is always negative, i.e. i.e. $\max_{b,j}(\alpha_1 + \beta_1 \log(\mathrm{d}^{b,j})) < 0$.

[31] The reason why the estimates between specifications differ so much is because using flows versus shares changes the relative weights when solving for the score function of the Poisson likelihood. For more details, see Sotelo (2019).

Table 2 – Gravity regression

	Labor flows, $L^{i,j}_{t,b}$ PPML			Migration shares, $L^{i,j}_{t,b}/\sum_k L^{i,k}_{t,b}$ PPML		
	(1) Geodesic (km)	(2) Driving (km)	(3) Driving (hours)	(4) Geodesic (km)	(5) Driving (km)	(6) Driving (hours)
$\mathbf{1}(j \neq b)$	1.337***	1.947***	−3.122***	−0.112***	−0.109***	−0.127***
	(0.199)	(0.218)	(0.059)	(0.003)	(0.004)	(0.004)
$\mathbf{1}(j \neq b)\log(\mathrm{d}^{b,j})$	−1.105***	−1.157***	−1.267***	−0.004***	−0.004***	−0.005***
	(0.037)	(0.038)	(0.040)	(0.000)	(0.000)	(0.000)
$\mathbf{1}(j \neq n)$	1.099***	1.859***	−4.512***	0.403**	1.033***	−6.578***
	(0.206)	(0.204)	(0.036)	(0.130)	(0.132)	(0.025)
$\mathbf{1}(j \neq i)\log(\mathrm{d}^{i,j})$	−1.908***	−1.945***	−2.242***	−1.735***	−1.752***	−2.021***
	(0.045)	(0.042)	(0.049)	(0.027)	(0.026)	(0.028)
Adj. Pseudo R^2	0.964	0.965	0.948	0.789	0.789	0.789
Observations	5,835,255	5,835,255	5,835,255	5,835,255	5,835,255	5,835,255

Note: This table stores the results of two models, both estimated using Poisson Pseudo Maximum Likelihood (PPML). The first model uses the labor flows of workers with birthplace b that moved from location i to location j, $L^{i,j}_{b,t}$ as a dependent variable. The second model uses the migration shares $L^{i,j}_{t,b}/\sum_k L^{i,k}_{t,b}$. For each model I use three different distance measures: geodesic distance in hundreds of kilometers, driving distance in hundreds of kilometers, and driving time between locations in hours. I get driving distances and hours from Google Maps. Standard errors are in parenthesis. Significance levels: *p<0.1; **p<0.05; ***p<0.01

Fact 3: Workers select across locations via wages.

To establish Fact 3, I confront two different selection mechanisms across locations. Thus, I test the predictions if workers move primarily because of pecuniary reasons or instead they move for other reasons unrelated to income.

Consider a situation where there are costs of moving away from one's birthplace, as the evidence presented under Fact 2 suggests. Now, if selection is driven by differences in potential wages across different locations, and leaving the birthplace is costly, then we should expect that workers who move away from their birthplace to have, on average, higher wages than those workers who stayed in their birth location. Therefore, this selection mechanism gives the following prediction

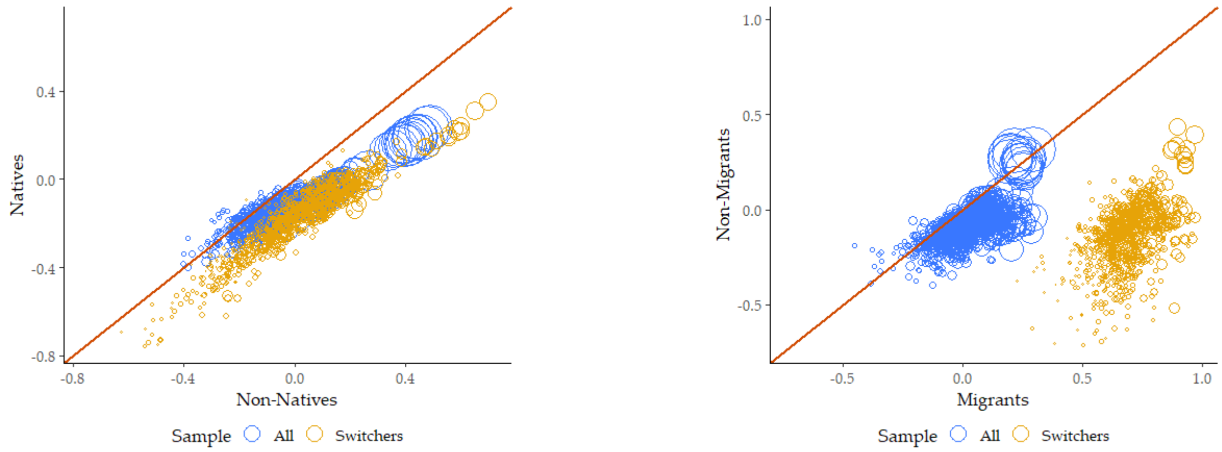
$$\mathbb{E}\left[\text{wage}|\text{No Native}\right] > \mathbb{E}\left[\text{wage}|\text{Native}\right],$$

meaning that the average wage of non-natives in a particular location should be larger than the average wage of natives. In contrast, if selection is driven by other elements orthogonal to wages, like, say, heterogeneous tastes for different locations, we should not observe a systematic difference between the wages of natives and non-natives.[32]

Figure 3a shows a plot where the y-axis corresponds to the mean of (log) wages in a particular location/period of natives after a normalization, whereas the x-axis does the same but for non-natives.[33] I include the 45 degree line to compare the relative magnitudes. Each of the blue circles in the Figure correspond to a location/year and the diameter of each circle is a function of the number of workers in such location. The graph shows that for almost all the locations/periods, the

[32] In reality, probably both mechanisms operate. However, using the prediction on average wages, I can see if the data rejects the selection-on-wages mechanism.

[33] I subtract the average wage of the entire sample used to each observation. This leaves the relative magnitudes unchanged.



(a) Selection Non-Natives

(b) Selection Migrants

Figure 3 – Selection via wages. The left panel compares the average (log) wages of non-native workers vs native workers. Wages from both groups are normalized by the average (log) wage of all the sample. The plot distinguishes two cases: when using the sample consisting of all workers and using the sample of workers who switched jobs. The plot in the right panel is analogous to the plot on the right, but compares (log) wages of migrants vs non-migrants.

average (log) wage of natives is lower than that of non-natives, consistent with the hypothesis that idiosyncratic differences in wages are an important driver of workers' migration decisions. Instead, if idiosyncratic differences not related to wages are the only thing that matters for migration, I would expect the points to gravitate around the 45 degree line.

I can restrict the sample to those workers who *switched* jobs from one year to the next. Using that sample, the selection mechanism appears to be stronger when comparing the wages of natives versus non-natives using all the workers. The orange circles in Figure 3a show this. Compared to the whole sample, the difference in the wages of non-natives versus natives is larger when using only the switchers. This is evident as the bulk of orange circles corresponding to job switchers are further down and to the right than the blue circles where I used all the workers.

If there are costs of migrating across locations, the same logic as above should apply with respect to wages of migrant versus non-migrant workers. The prediction would be that the average wage of migrants is larger than the average wage of those workers who stayed in the same location. The blue circles in Figure 3b shows the average wage of migrant versus non-migrant workers for every location/period after a normalization. The figure suggests that selection is less strong for year-to-year migration than when comparing natives vs non-natives, especially for large locations. The closer alignment to the 45 degree line can just reflect that some workers who were migrants in previous years and kept the same job are now classified as non-migrants. For example, if a worker migrated in a previous year because of a highly paid job and kept her job in subsequent years, she would appear as a non-migrant in the data, even though she clearly selected herself to that location via wages. On the other hand, migrant workers are, by definition, taking new jobs. Thus, a fair comparison would be to use those workers who changed jobs but stayed in the same location versus the workers who migrate into that location in the same year.

The selection mechanism via wages appears stronger when using workers who switched jobs from one year to another. Indeed, Figure 3b shows that the selection via wages appears to be stronger than when using all the workers. And not only is it stronger, the magnitude of the difference is very large: the horizontal distance between most of the circles and the 45 degree line is around 1. As I am comparing averages of log wages, this means that the wages of migrants are twice as large as those of non-migrants.[34]

The key takeaway for Fact 3 is that idiosyncratic differences in wages across locations are an important driver of workers' migration decisions. Also, that this selection mechanism appears stronger when using workers who change jobs between years, and that non-natives and migrants have higher average wages than natives and non-migrants, respectively.

Fact 4: Workers who Live in their Birthplace accept a Wage Penalty.

Facts 2 and 3 above show evidence of potential mobility frictions between a worker's birthplace and other locations, and that workers select into locations mainly via wages. Taken together, this suggests that workers who change jobs and move away from their birthplace should experience wage gains. In contrast, workers who change jobs but decide to stay in their birthplace or return to it, are likely to suffer a wage penalty.

To shed some light on these wage gains and penalties related to working within the birthplace, I estimate the following linear regression

$$\Delta \log w_{\iota;t,b}^{i,j} = \mathcal{P}_t^{i,j} + \mathbf{1}_{j=b}\beta_{In} + \mathbf{1}_{i=b} \times \mathbf{1}_{j\neq b}\beta_{Out} + \varepsilon_{\iota;t,b}^{i,j},$$

where $\Delta \log w_{\iota;t,b}^{i,j}$ is the year-to-year change in the log wage of worker ι who was born in b that moves from location i to j in t; $\mathcal{P}_t^{i,j}$ denotes an origin/destination pair fixed effect for year t that should absorb any constant differences across the two locations, as well as the compensation the worker needs for migrating; the dummy $\mathbf{1}_{j=b}$ indicates when a worker's destination j is her birthplace b; the interaction $\mathbf{1}_{i=b} \times \mathbf{1}_{j\neq b}$ indicates if a workers previous residence—or origin—i is the same as her birthplace b *and* that the destination j is different than the birthplace. This interaction captures all the workers who *leave* their birthplace in that period. The total gain from leaving the birthplace would be the composite of both effects, one that is from moving out from the birthplace plus not receiving the penalty of staying in the birthplace.

Table 3 shows the estimated wage gains of a worker who moves out of her birthplace and the penalty she incurs for staying/returning to it. The specification in the second column includes a quadratic polynomial in age and a gender dummy to account for possible differences in the composition of those workers who move back—or from—their birthplace.

The estimated penalty that workers entail to live in their home is between 4 and 8 percent. On the other hand, the expected wage gain a worker gets by moving away from her birthplace is between 9 and 11 percent. These results do not mean that in order for a worker to be indifferent between moving out of her birthplace, she needs to be compensated between 9 and 11 percent more

[34] In Appendix 1.H I make the same figures but using residual wages after running a regression for each year of log wages on a quadratic polynomial in age and a gender dummy. This controls for the differences in gender and age compositions across groups. Compared to the analysis using observed wages, the results are very similar and have the same implications. In particular, even after controlling for age and gender, the average wages of migrants are twice as large as non-migrants who changed jobs.

Table 3 – Birthplace penalty on wages

	Dependent variable: $\Delta \log w_{t,\iota}$	
	OLS	
	(1)	(2)
Destination = Birthplace	-0.042^{***}	-0.079^{***}
	(0.000)	(0.000)
Leaving Birthplace	0.072^{***}	0.008^{***}
	(0.002)	(0.002)
Origin/Dest./Year FE	✓	✓
Age and Gender Controls		✓
R^2	0.019	0.042
Observations	26,221,763	26,221,763

Note: The table shows the results of two linear regressions estimated using Ordinary Least Squares (OLS). The dependent variable is the time difference of the logarithm of the wage of an individual who switch jobs across years. Column 2 adds as controls a quadratic polynomial in age and a gender dummy. Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01

than her outside option. For this to be true, the outside option of the worker should be the wage she received in the previous period. However, the true outside option is not observed in the data as it will be the second best offer that a worker gets from a different location.

Taken together, Facts 1 to 4 suggest that workers prefer to live close to their home. This motivates the migration model I present in the next section, which I use to estimate this home bias. I then use the model to study the general equilibrium effects of the home bias in the aggregate economy and on the welfare distribution across birthplace cohorts.

1.4 A Migration Model with Home Bias

In this section, I first present a dynamic migration model with a finite number of workers. This assumption—which is not common in most macro-migration models—allows for a more transparent identification strategy later on. After this, I present the more standard steady-state continuous-population version of the discrete model, which I use later on to analyze the general equilibrium effects of the home bias.

My model is based on Caliendo et al. (2019) in that it combines a dynamic discrete choice problem for workers' migration decisions with a static trade equilibrium model determining output at each location. I add these new elements to their model: (i) I include home bias in workers' preferences; (ii) differences in idiosyncratic wages drive the migration decision of workers; and (iii) an exogenous process determines if a worker changes jobs, and therefore, whether she may migrate or not. As I show below, differences (ii) and (iii) combined have implications with respect to the labor supply elasticity.

I consider a discrete-time, infinite-horizon economy that consists of I locations, indexed by i, j and k. Each worker has a specific birthplace, indexed b. I assume there is a large, but discrete, number of workers who were born in location b, which is denoted by L_b, and I assume that it is constant across time.

Workers get utility from consuming a final good, assembled locally from a housing and non-housing good. Housing is in fixed supply. The non-housing good is assembled locally by a firm that uses tradable inputs, which are produced by intermediate firms from different locations.

In each location there are a finite number of fixed intermediate good firms produce a continuous mass of varieties, each of these produced according to a Cobb-Douglas technology that uses efficiency units of labor and housing as inputs. I assume that each firm-variety has different productivities and, following Eaton and Kortum (2002), I assume these are distributed Fréchet with a dispersion parameter equal to φ.[35] These firms trade across regions, subject to some iceberg costs, and non-housing good producers use the intermediate inputs to assemble the non-housing local good which is in turn used as an input by the final good producer.[36] The joint demand for housing by workers and firms generates a congestion force in the model: if a location attracts workers, this raises the price of housing and lowers the real wage.

Workers are forward-looking and have rational expectations. In every period, two things can happen: with some probability the worker keeps the same job and moves to the following period, or it becomes a job switcher, in which case the worker has to look for another job. If this is the case, then at the end of each period, workers observe a vector of location-specific idiosyncratic labor-augmenting productivity shocks for the next period. Given this information, workers optimally decide where to move in the following period subject to some migration costs. In addition to the migration costs, workers also pay a cost, in utility terms, from moving away from their birthplace.

Admittedly, the exogenous process that determines whether a worker has an opportunity to change jobs, and therefore migrate, is very simplistic. It can reflect several aspects of the labor market: separation rates and job finding rates, as well as on-the-job search. Regardless of how we interpret this exogenous process, it mainly captures that most workers do not take a migration decision in every period, and indeed just keep the same job.

Appendix 1.A contains the detailed derivations of the expressions in this section.

1.4.1 Workers

In period t, there is a discrete number $L^i_{t,b}$ of workers with birthplace b that live in each location $i \in \mathcal{I}$. Each worker ι supplies her efficiency units of labor, $\exp(\theta^i_{t-1,\iota})$ inelastically and receives a competitive efficiency wage w^i_t.

The worker uses her labor income to purchase and consume a local final good $C^i_{t,\iota}$ whose price is P^i_t. Formally, the worker's budget constraint is

$$P^i_t C^i_{t,\iota} = w^i_t \exp(\theta^i_{t-1,\iota}).$$

The final good is a composite of housing H^i_t and non-housing good Q^i_t which is assembled locally

[35]This assumption on the discrete number of firms allows me to accommodate a discrete number of workers and to keep the tractability that comes from assuming a Fréchet productivity distribution over a continuum of goods.

[36]The input output relation is as follows: Intermediate good $\rightarrow$ non-housing good $\rightarrow$ final good.

from tradable intermediates. These two goods are aggregated with a Cobb-Douglas technology

$$C_t^i = \left(Q_t^i\right)^{1-\alpha}\left(H^i\right)^{\alpha}.$$

Denote the housing and non-housing good prices as $P_{H,t}^i$ and $P_{Q,t}^i$. Then, the price index for the final good C_t^i is

$$P_t^i = \left(\frac{P_{Q,t}^i}{1-\alpha}\right)^{1-\alpha}\left(\frac{P_{H,t}^i}{\alpha}\right)^{\alpha}.$$

The flow utility that a worker ι, with birthplace b receives for living in location i at period t is

$$B^i + \log\left(C_{t,\iota}^i\right) - \kappa_b^i,$$

where B^i is a location specific amenity; $\kappa_b^i \geq 0$ is the utility cost of living away from one's birthplace, which I call the *home bias*: the larger κ_b^i is, the larger the preference of workers with birthplace b to stay home vis-a-vis location i. The home bias is common for all workers with birthplace b that live in location i.

At the beginning of each period, workers produce in their current location. Each of them then receives an independent shock that determines their immediate working situation: with probability ρ they stay in the same job and keep their same location-specific efficiency unit, and with probability $1-\rho$ they have to change jobs. If a worker has to change jobs, then she observes a vector of location specific idiosyncratic efficiency unit shocks $\mathbf{\Theta}_{t,\iota} \equiv \{\theta_{t,\iota}^k\}_{k\in\mathcal{I}}$. After observing the shocks, the worker optimally decides where to move, subject to some migration costs $\tau_t^{i,k} \geq 0$ measured in utility terms.

Workers discount the future at rate β. Given the assumptions on workers' behavior, I can write the lifetime utility of a worker with birthplace b living at location j recursively as:

$$\mathrm{v}_{t,b}^i(\theta_{t-1,\iota}^i, \mathbf{\Theta}_{t,\iota}) = B^i + \log\left(C_{t,\iota}\right) - \kappa_b^i + \beta\rho\mathbb{E}_t\left(\mathrm{v}_{t+1,b}^i(\theta_{t-1,\iota}^i, \mathbf{\Theta}_{t+1,\iota})\right) + \tag{1.1}$$

$$\beta(1-\rho)\max_k\left[\mathbb{E}_t\left(\mathrm{v}_{t+1,b}^k(\theta_{t,\iota}^k, \mathbf{\Theta}_{t+1,\iota})\right) - \tau^{i,k}\right]. \tag{1.2}$$

The sources of uncertainty in the model can be grouped in two: first, there is idiosyncratic uncertainty, i.e. the future realizations of the location specific efficiency unit shocks. Second, there is aggregate uncertainty. The sources of aggregate uncertainty can, in turn, be also grouped in two. First, location productivities might change from period to period given a known distribution. Second, given the discrete number of workers, labor supply at each location is stochastic. This last aspect differs from several macro-migration models with a continuum of agents. In such cases, this particular source of uncertainty would not be present. I summarize all the sources of aggregate uncertainty in a variable Z_t, which evolves according to the conditional distribution $F(Z_{t+1}|Z_t)$. Keep in mind though that in the steady-state continuous-population version of the model $Z_t = Z$, so the further characterization of its evolution is not necessary when solving that version of the model. I only include it to make clear that the identification strategy later on will not depend on the dynamics of Z_t.

I assume that the idiosyncratic efficiency shocks are distributed Gumbel with zero mean and variance equal to $\frac{\pi^2}{6}\delta^2$. This assumption, ubiquitous in the discrete choice literature, allows for a simple computation of the expectation of the maximum lifetime utility for next period. Let $V^i_{t,b} \equiv \mathbb{E}_{\Theta_t}\left(\mathrm{v}^i_{t,b}(\cdot) - \frac{\theta^i_{t-1,\iota}}{1-\beta\rho} \,\middle|\, Z_t \right)$ be the expected lifetime utility *net* of current discounted efficiency shocks $\theta^i_{t-1,\iota}/(1-\beta\rho)$, conditional on the aggregate shock vector Z_t. Then, given the assumption on the distribution of the idiosyncratic shocks, and substituting the budget constraint, I obtain

$$V^i_{t,b} = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho \overline{V}^i_{t+1,b} + \beta(1-\rho)\mathbb{E}_{\Theta_t}\left(\max_k \left[\overline{V}^k_{t+1,b} - \tau^{i,k} + \frac{\theta^k_{t,\iota}}{1-\beta\rho} \right] \right). \tag{1.3}$$

where $\overline{V}^k_{t+1,b} = \int V^k_b(Z_{t+1})dF(Z_{t+1}|Z_t)$ is the expected lifetime utility of moving to location k at period $t+1$. The scaled-up shock $\frac{\theta^k_{t,\iota}}{1-\beta\rho}$ is distributed Gumbel with mean zero but variance $\frac{\pi^2}{6}\lambda^2$, where $\lambda = \delta/(1-\beta\rho)$. Using the properties of the Gumbel distribution I can rewrite equation (1.3) as

$$V^i_{t,b} = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho \overline{V}^i_{t+1,b} + \beta(1-\rho)\lambda \log\left(\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda} \right). \tag{1.4}$$

The assumption on the distribution of the efficiency shocks allows me to compute a closed formed expression for the conditional migration probabilities. Conditional on changing jobs, the probability of a worker with birthplace b to move from location i to j, denoted $p^{i,j}_{t,b}$, is equal to

$$p^{i,j}_{t,b} = \frac{\exp(\overline{V}^j_{t+1,b} - \tau^{i,j})^{\frac{1}{\lambda}}}{\sum_{k\in\mathcal{N}} \exp(\overline{V}^k_{t+1,b} - \tau^{i,k})^{\frac{1}{\lambda}}}. \tag{1.5}$$

The parameter $\lambda \equiv \delta/(1-\beta\rho)$, which appears in expressions (1.4) and (1.5), represents the dispersion of the efficiency shocks after taking into account the probability of getting the same efficiency unit in the next period with probability ρ. Given the expression for the conditional probability of migrating (1.5), I interpret λ as the inverse labor supply elasticity. If the dispersion of shocks is smaller, jobs across locations are more alike, i.e. easier to substitute, which turns the labor supply more elastic.

When there is no persistence in the model, i.e. $\rho = 0$, the inverse supply elasticity is just the dispersion of the original efficiency shocks $\delta < \lambda$. But then, why is the persistence in the model making the labor supply more inelastic? When a worker is comparing different jobs across locations, she understands that with probability ρ she will keep the same job in the following period. Therefore, initial differences in efficiency units are magnified and their perceived variance increases. So the worker behaves *as if* the shocks she observes are distributed Gumbel with scale parameter $\lambda > \delta$. While other papers have considered exogenous persistence in the decision of workers, whether to migrate or change sector of employment, to the best of my knowledge, I am the first to link it to the (extensive margin) labor supply elasticity.[37] This is a consequence of workers selecting across locations via different job opportunities, as reflected in their efficiency shocks θ^j.

As there is a discrete number of workers in each location, the movement of labor from one location to another is a stochastic process governed by the above migration probabilities. Denote

[37] See section 5.3.2 in Caliendo et al. (2019) for an extension of their model where they add exogenous persistence in the migration decision. Also, Appendix 3 in the Online Appendix of Artuç et al. (2010) adds an extension to their sectoral choice model where some type of workers can't change sectors while others can. However, every worker has a probability to change type, so it is similar to a model with only exogenous persistence.

$\ell^{i,j}_{t,b}$ as the number of workers who migrate from i to j with birthplace b at the end of period t. Then, the distribution of labor in any location is equal to

$$L^{j}_{t+1,b} = \sum_{i \in \mathcal{I}} \ell^{i,j}_{t,b}.$$

To conclude the characterization of the dynamic sub-problem of the model, I show how the efficiency units per location evolve. The assumption on the distribution of the idiosyncratic shocks allows me to characterize analytically the expected amount of (idiosyncratic) efficiency units of a worker who, conditional on changing jobs, moved from location i to j. This is equal to

$$\mathbb{E}_{\iota}(\exp(\theta_{t,\iota})|i \to j) = \frac{\Gamma(1-\delta)}{\exp(\gamma\delta)}(p^{i,j}_{t,b})^{-\delta}, \tag{1.6}$$

where $\Gamma(\cdot)$ denotes the Gamma function and γ is the Euler-Mascheroni constant. The previous expression is intuitive: given the selection of individuals across locations, all infra-marginal workers have higher efficiency units than the marginal worker. Then, the more workers move into a particular location, the lower the average efficiency unit of that particular migration cohort.

Denote $h^{i,j}_{t,b}$ as the total amount of efficiency units of workers who have the opportunity to migrate and move from location i to j. Using (1.6), then

$$h^{i,j}_{t+1,b} = \frac{\Gamma(1-\delta)}{\exp(\gamma\delta)}(p^{i,j}_{t,b})^{-\delta}\ell^{i,j}_{t,b} + \chi^{i,j}_{t+1,b},$$

where $\chi^{i,j}_{t+1,b}$ is a zero mean expectation shock that captures deviations between the expected and realized efficiency units. Thus, the total amount of efficiency units per migration cohort is also a random variable as the labor flow $L^{i,j}_{t+1,b}$ and the expectation shock $\chi^{i,j}_{t+1,b}$ are stochastic variables.

Define the sum of total efficiency units of workers who did not switch jobs from one period to the next as $\tilde{N}^{j}_{t,b}$. Then, the evolution of the total amount of efficiency units of workers from birthplace b that live in location j is equal to

$$N^{j}_{t,b} = \tilde{N}^{j}_{t,b} + \sum_{i \in \mathcal{I}} h^{i,j}_{t,b}.$$

Finally, the total amount of efficiency units in location n is the sum of efficiency units across the different birth cohorts

$$N^{j}_{t} = \sum_{b} N^{j}_{t,b}.$$

The previous equations characterize the evolution of the total efficiency units supplied to each location j at every period t. Conditional on these allocations, I can now specify the static sub-problem of the model, and solve for the equilibrium efficiency wages at each time t such that labor markets clear in each location.

1.4.2 Production

The production side of the model is very similar to the one presented in the one-sector model of Caliendo et al. (2019) with the difference that the labor input is efficiency units. Another difference is that I assume balanced trade. This is because I lack data on trade flows across locations within France.[38]

In each location j I assume that there is a finite number of perfectly competitive intermediate firms each producing a continuum of varieties of intermediate goods. In order to produce a variety, the intermediate good firms use as inputs the total amount of efficiency units $\tilde{h}$ and housing $\tilde{H}$.[39] The total factor productivity is composed of two terms: a time-varying location specific component A_t^j, which is common for all varieties produced within the same location, and a variety specific component z^j, which is specific to variety z. This idiosyncratic productivity z^j is distributed Fréchet$(1, \varphi)$. Formally, the output of an intermediate producer with efficiency z^j for a given variety z is:

$$q_t^j\left(z^j\right) = z^j A_t^j \left(\tilde{H}^j\right)^{\eta} \left(\tilde{h}_t^j\right)^{1-\eta},$$

Intermediate firms pay the efficiency wage w_t^j for each effective unit of labor. The price of housing is P_H^j. Therefore, the unit price of an input bundle for the firm is

$$x^j = \left(\frac{w^j}{1-\eta}\right)^{1-\eta} \left(\frac{P_H^j}{\eta}\right)^{\eta}.$$

Cost minimization implies that the unit cost of an intermediate good z^j at time t is

$$\frac{x_t^j}{z^j A_t^j}.$$

Trade costs are represented by $\psi^{j,i}$. These are 'iceberg costs', meaning that, for one unit of any variety shipped from region i to j, it requires producing $\psi^{j,i} \geq 1$ units in location i. I assume that these costs are constant across periods. Competition in turn implies that the price paid for a particular variety z in location j is

$$\min_{i \in \mathcal{N}} \frac{\psi^{j,i} x_t^i}{z^i A_t^i}.$$

Local manufacturing goods in location j are produced by aggregating intermediate inputs from all the different locations in $\mathcal{N}$. Let Q_t^j be the quantity produced of local manufacturing goods in j and $\tilde{q}_t^j(z^j)$ the quantity demanded of an intermediate good of a given variety from the lowest-cost supplier. The production of local manufacturing goods is given by

$$Q_t^j = \left(\int \left(\tilde{q}_t^j(z^j)\right)^{\frac{\sigma-1}{\sigma}} d\xi(\mathbf{z})\right)^{\frac{\sigma}{\sigma-1}},$$

where $\xi(\mathbf{z}) = \exp\left(-\sum_{i\in\mathcal{N}} \left(z^i\right)^{-\varphi}\right)$ is the joint distribution over the vector $\mathbf{z} = (z^1, z^2, ..., z^I)$. Using the properties of the Fréchet distribution, the price of the consumption good at location j is

$$P_{T,t}^j = \overline{\Gamma} \left(\sum_{i\in\mathcal{I}} \left(\frac{x_t^i \psi^{j,i}}{A_t^i}\right)^{-\varphi}\right)^{-1/\varphi},$$

[38]This flows would have allowed me to compute the trade deficits for each location.

[39]I assume that the firm can split the efficiency units of a worker across the production of any variety

where $\overline{\Gamma}$ is just a constant term equal to $(\Gamma(1+(1-\sigma)/\varphi))^{1/(1-\sigma)}$ and, as it is standard, I assume that $1+\varphi > \sigma$.

The share of total expenditure in location j on goods from i is

$$\pi_t^{j,i} = \frac{\left(x_t^i \psi^{j,i}/A^i\right)^{-\varphi}}{\sum_{k\in\mathcal{N}}\left(x_t^k \psi^{j,k}/A^k\right)^{-\varphi}}.$$

Housing, as mentioned before is supplied inelastically, and is rented by both workers and intermediate firms in a perfect competition environment. I assume that owners of the housing stock consume just the local non-housing good Q_t^j.

1.4.3 Market clearing

In equilibrium, the sum of efficiency units and housing across all firms must be equal to the total supply in each location.

Let $P_{T,t}^j E_t^j$ be the total expenditure in location j on non-housing goods. Also, let $P_{T,t}^j Y_t^j$ be the total income of intermediate firms in location j. Then, non-housing goods market clearing implies

$$P_{T,t}^j Y_t^j = \sum_{i\in\mathcal{I}} \pi_t^{i,j} P_{T,t}^i E_t^i.$$

The labor market clearing condition implies

$$w_t^j N_t^j = (1-\eta) P_{T,t}^j Y_t^j.$$

while the market clearing condition for housing is

$$P_{H,t}^j H^j = \alpha w_t^j N_t^j + \eta P_{T,t}^j Y_t^j = \frac{\eta + \alpha(1-\eta)}{(1-\eta)} w_t^j N_t^j.$$

Finally, I assume trade is balanced, meaning

$$P_{T,t}^j Y_t^j = P_{T,t}^j E_t^j = \underbrace{(1-\alpha) w_t^j N_t^j}_{\text{Final demand workers}} + \underbrace{\alpha w_t^j N_t^j + \eta Y^j}_{\text{Final demand Housing owners}} = \frac{1}{1-\eta} w_t^j N_t^j.$$

Substituting into the non-housing goods market clearing condition

$$w_t^j N_t^j = \sum_{i\in\mathcal{I}} \pi_t^{i,j} w_t^i N_t^i.$$

1.4.4 Static equilibrium under symmetric costs

Let $W_t^j = w_t^j / P_{T,t}^j$ be the the efficiency wage deflated by the price of the local non-housing good in each location. Also, define $\tilde{A}^j = A^j \left(H^j\right)^\eta$ as a composite of both productivity and housing supply in location j. Then, if the trade costs are symmetric, i.e. $\psi^{i,j} = \psi^{j,i}$, the static equilibrium conditions can be collapsed into a single equation per location

$$\left(W^i\right)^{\tilde{\varphi}\varphi(1+\varphi)} \left(N^i\right)^{(1+\eta\varphi)(1-\tilde{\varphi}(1+\varphi))} = \sum_j \left(\psi^{j,i}\right)^{-\varphi} \left(\tilde{A}^i\right)^{\varphi} \left(\frac{\tilde{A}^j}{\tilde{A}^i}\right)^{\varphi\tilde{\varphi}(1+\varphi)} \left(W^j\right)^{\varphi(\tilde{\varphi}(1+\varphi)-1)} \left(N^j\right)^{1-\tilde{\varphi}(1+\varphi)},$$

where $\tilde{\varphi} = 1/(1+2\varphi)$. Appendix 1.A.3 contains the detailed derivations to get the expression above.

1.4.5 Steady-State continuous-population case

The model presented above with a finite number of workers per birthplace will guide the identification strategy in the next section. Solving such a model, however, is extremely challenging. To solve for the model, I consider a version of it where the economy fundamentals do not change and each birthplace cohort consists of a mass L_b of workers. These two assumptions render the model deterministic, in particular $V^i_{t,b} = \overline{V}^i_{t,b}$, while also putting the economy's aggregate variables on a steady state. Let

$$U^i_b = \exp\left(V^i_b\right), \quad \Omega^i_b = \left(\sum_k \exp\left(V^k_b - \tau^{i,k}\right)^{1/\lambda}\right)^{\lambda}, \quad \mathcal{B}^i = \exp\left(B^i\right)^{1/\delta}\left(H^j\right)^{\alpha/\delta},$$
$$T^{i,j} = \exp(\tau^{i,j})^{-1/\lambda}, \quad \text{and} \quad K^j_b = \exp(\kappa^j_b)^{-1/\delta}.$$

I can now summarize the steady-state continuous-population model. The static part of the equilibrium remains identical, which relates total efficiency units per location $\{N^i\}$ and deflated wages $\{W^i\}$

$$\left(W^i\right)^{\tilde{\varphi}\varphi(1+\varphi)}\left(N^i\right)^{(1+\eta\varphi)(1-\tilde{\varphi}(1+\varphi))} = \sum_k \tilde{\psi}^{k,i}\left(\tilde{A}^i\right)^{\varphi}\left(\frac{\tilde{A}^k}{\tilde{A}^i}\right)^{\varphi\tilde{\varphi}(1+\varphi)}\left(W^k\right)^{\varphi(\tilde{\varphi}(1+\varphi)-1)}\left(N^k\right)^{1-\tilde{\varphi}(1+\varphi)}. \tag{1.7}$$

The total efficiency units in a location

$$N^i = \sum_b N^i_b. \tag{1.8}$$

The rest of the equations characterize the total efficiency units in a location i per birthplace cohort b. The lifetime utility for a worker who was born in b and lives in location i is

$$\left(U^i_b\right)^{1/\lambda} = \mathcal{B}^i\left(W^i\right)^{\frac{1-\alpha}{\delta}}\left(N^i\right)^{-\alpha/\delta}K^i_b\left(\Omega^i_b\right)^{\frac{\beta(1-\rho)}{\delta}}. \tag{1.9}$$

The option value of living in location i is equal to

$$\left(\Omega^i_b\right)^{1/\lambda} = \sum_k T^{i,k}\left(U^k_b\right)^{1/\lambda}. \tag{1.10}$$

The evolution of the distribution of labor L^i_b is characterized by

$$L^i_b\left(U^i_b\right)^{-1/\lambda} = \sum_k T^{i,k}\left(\Omega^k_b\right)^{-1/\lambda}L^k_b. \tag{1.11}$$

The previous equation is scale invariant in $\{L^i_b\}$. The sum of total number of workers of a particular birthplace cohorts pins down the relative scale. Thus,

$$L_b = \sum_k L^k_b. \tag{1.12}$$

Finally, the total amount of efficiency units N^i_b is characterized as follows

$$N^i_b\left(U^i_b\right)^{\frac{\delta-1}{\lambda}} = \sum_k\left(T^{i,k}\right)^{1-\delta}\left(\Omega^k_b\right)^{\frac{\delta-1}{\lambda}}L^k_b. \tag{1.13}$$

Appendix 1.A.5 provides a detailed derivation of these expressions.

Table 4 – Parameter values

Parameter	Description	Value	Source
β	Discount factor	0.96	–
α	Share of housing consumption	0.3	Friggit (2013)
φ	Dispersion productivities	4.14	Simonovska and Waugh (2014)
η	Output elasticity	0.1	Gutierrez (2017)
$(\psi^{i,j})^{-\varphi}$	Trade Costs	—	Combes, Lafourcade, and Mayer (2005)
ρ	Prob of keeping job	0.867 (s.e. 2.4e^{-5})	1 - Proportion of Switchers

Definition 1 (Steady-State continuous-population competitive equilibrium)**.** *Given a distribution of birthplace cohorts* $\{L_b\}_{b\in\mathcal{I}}$*, the competitive equilibrium for the steady-state continuous-population economy is a vector of deflated wages,* $\{W^i\}_{i\ in\mathcal{I}}$*, total efficiency units per location* $\{N^i\}_{i\ in\mathcal{I}}$*, lifetime utilities* $\{U_b^i\}_{b,i\in\mathcal{I}}$*, option values* $\{\Omega_b^i\}_{b,i\in\mathcal{I}}$*, labor flows* $\{L_b^i\}_{b,i\in\mathcal{I}}$ *and efficiency units per birthplace cohort/location* $\{N_b^i\}_{b,i\in\mathcal{I}}$*, such that equations (1.7)-(1.13) are satisfied for all* $i, b \in \mathcal{I}$*.*

1.5 Identification and Estimation

The model presented in the previous section, entails a large number of parameters, as well as distributions of fundamentals, which need to be estimated or calibrated. In this section I explain how to identify and estimate the key parameters and the distributions of fundamentals.

Given that the static part of the equilibrium is fairly standard, I calibrate externally the parameters governing that part of the model, the trade costs and the discount factor, β. I choose values to match moments from other studies. For the discount factor β, I choose a value of 0.96 which is standard in the literature for annual frequencies. The trade elasticity φ is set to 4.14 which is the value proposed by Simonovska and Waugh (2014). The consumption elasticity with respect to housing α is set to 0.3, which is in line with survey studies on workers expenditures in France (Friggit (2013)). The output elasticity η is set to 0.1, in line with the profit share reported for France by Gutierrez (2017).[40] The internal trade costs, $(\psi^{i,j})^{-\varphi}$ are taken from Combes et al. (2005) who use data on commodity flows to estimate trade costs at the *département* level. Given that some of my locations are aggregates of different departements, I need to do some adjustments. I first compute all the trade costs across departements and then compute a population weighted average of these departemental trade costs to get the aggregate location trade cost. Regarding the persistence parameter ρ, in the data I can identify which workers changed main jobs between years. Appendix 1.E.1 explains how I do this. I estimate ρ using the average across years of the proportion of workers who stay in the same job between years. Table 4 summarizes the information of the parameters mentioned so far.

[40] The profit share is defined as total value added of non-financial corporations minus payments to labor and capital. As I don't have capital in the model, and given the Cobb-Douglass and perfect competition assumptions, the profit share would correspond to η in my model.

I use the structure of the model to identify the remaining parameters: the dispersion parameter δ, the mobility costs, $\{\tau^{i,j}, \kappa_b^j\}$, and the distribution of composite productivities and amenities $\{\tilde{A}^j, \mathcal{B}^j\}$.

I follow a sequential identification strategy which is inspired by Bryan and Morten (2019), Dingel and Tintelnot (2020) and Artuç et al. (2010). The merit of any identification strategy is related to its practical implementation. Thus, the steps in the identification sequence are not arbitrary, but are chosen such that the estimation procedure that follows is computationally feasible.

The main identification steps are as follows. First, I show how to use observed labor flows to identify the migration costs. I show how to relax the identification conditions provided by Bryan and Morten (2019), which in turn relaxes the data requirements. As I show later on, this will be important in the context of my application. Second, I show how to recover the dispersion parameter δ from the effect of migration costs on migrants' wages. Third, using the migration costs and labor flows, I show how to identify the underlying distribution of migration probabilities by means of maximum likelihood. I show that the maximization of such likelihood corresponds to solving a system of equations characterizing the balanced trade condition present in most gravity trade models. Trade economists have shown the existence and uniqueness of the solution of such systems and provided fast and efficient algorithms to find it.[41] Fourth, I show that efficiency wages are identified using average wages and the estimated migration probabilities. Fifth, I use average wage differentials across locations of the different migration cohorts to identify the home bias. The idea is that the wage of a worker outside home should be larger, all else equal, than the wage at home. I show how to control for all the other factors influencing the wage differential to isolate the effect of the home bias. Sixth, as in the trade literature, I show how to *invert* the static part of the model using observed wages to recover the underlying productivity distribution. Finally, as is standard in the urban economics literature, I identify the amenities as a residual that explains the remaining variation in labor flows.[42]

In what follows I explain with more detail each of the steps to identify the relevant parameters of the model.

1.5.1 (Scaled) Migration Costs $\tau^{i,j}/\lambda$

Given the logit structure of the migration probability, the conditional expectation of the labor flow between preiod t and $t+1$ $\ell_{t,b}^{i,j}$ can be rewritten as

$$\mathbb{E}_t(\ell_{t,b}^{i,j}) = p_{t,b}^{i,j} L_{t,b}^i = \exp\left(\mathcal{O}_{t,b}^i + \mathcal{D}_{t,b}^j - \tau^{i,j}/\lambda\right), \tag{1.14}$$

where $\mathcal{D}_{t,b}^j = \overline{V}_{t+1,b}/\lambda$ and $\mathcal{O}_{t,b}^i = -\log\left(\sum_k \exp(\overline{V}_{t+1,b}^k - \tau^{i,k})^{1/\lambda}\right) + \log L_{t,b}^i$. Then, conditioning on origin, destination and the location pair fixed effects, the conditional expectation of the labor flow is equal to the right hand side of (1.14). This moment condition is equivalent to the first order condition of a Poisson regression (or Poisson Pseudo Maximum Likelihood).

[41] For the existence and uniqueness results, see for example Ahlfeldt et al. (2015) and Allen et al. (2020a). For the algorithm, see Pérez-Cervantes (2014).

[42] For a discussion of the *inversion* of the model to recover fundamentals, as well as the identification of amenities as residuals, see Redding and Rossi-Hansberg (2017).

Identification of the migration costs by running a Poisson regression with fixed effects is not a priori obvious. For example, suppose there is an origin destination i with flows going to several destinations. Now, assume there is only one labor flow going to location j. Then, I could not separately identify, the destination fixed effect $D^{j}_{t,b}$ from the migration cost $\tau^{i,j}$.[43] In this section I show sufficient conditions for the identification of the migration costs when running a regression with fixed effects. First, I make the following assumption,

Assumption 1. *The migration costs are symmetric $\tau^{i,j} = \tau^{j,i}$ for all i, j in $\mathcal{I}$. Also, the cost of staying in the same location is zero, i.e. $\tau^{i,i} = 0$ for all i in $\mathcal{I}$.*

Now consider two locations, i and j. Then, for a particular birthplace cohort b at time t

$$\frac{p^{i,j}_{t,b}}{p^{i,i}_{t,b}} \frac{p^{j,i}_{t,b}}{p^{j,j}_{t,b}} = \exp\left(\mathcal{D}^{j}_{t,b} - \mathcal{D}^{i}_{t,b} - \tau^{i,j}/\lambda\right) \exp\left(\mathcal{D}^{i}_{t,b} - \mathcal{D}^{j}_{t,b} - \tau^{i,j}/\lambda\right) = \exp(-2\tau^{i,j}/\lambda). \tag{1.15}$$

Notice that I have used both the normalization and symmetry assumption to form this expression. The expression above means that if the data for a particular birthplace/year contains positive flows of workers going from i to j, workers going in the reverse direction, j to i, and workers staying within those two locations, then the (scaled) migration cost $\tau^{i,j}/\lambda$ is identified.

There is a simple intuition for why the product of these probability ratios identify the migration cost. First, the ratio $p^{i,j}_{t,b}/p^{i,i}_{t,b}$ controls for origin specific differences. The remaining variation is explained by the migration cost and differences in destination fixed effects. To control for the latter, I can use the same ratio but for the *reversed* flow $p^{j,i}_{t,b}/p^{j,j}_{t,b}$. Indeed, the variation in the reversed ratio accounts for the *reversed* difference of destination fixed effects and the migration cost. In the end, the larger the gross migration flow is, the smaller the migration cost. These are the conditions pointed out by Bryan and Morten (2019) which are summarized in the following proposition

Proposition 1. ***Bryan and Morten (2019)**. The (scaled) migration cost $\tau^{i,j}/\lambda < \infty$ is identified if $L^{i,j}_{t,b} > 0$, $L^{j,i}_{t,b} > 0$ and $L^{i,i}_{t,b} > 0$, $L^{j,j}_{t,b} > 0$, for some birth cohort b and period t.*

Proof. It follows from (1.14) and (1.15). □

These sufficient conditions for identification of the migration cost might be difficult to fulfill in my context. For example, I would need to observe, for a particular year, someone born in Toulouse migrating from Paris to Lyon, *and* someone born in Toulouse migrating from Lyon to Paris. In addition, I would need to observe, for the same year, someone from Toulouse staying in both Paris and Lyon. Maybe for this particular example, the data would easily fulfill the requirements for above's identification conditions, but these become increasingly hard to satisfy when comparing scarcely populated locations. Only 69.3% of the bilateral connections satisfy the conditions of Proposition 1. Thus, for 30.7% of the connections I would not be sure if I am actually identifying the migration costs when running a Poisson regression.

My objective is to relax the restrictions imposed by Proposition 1 to find a larger number of bilateral connections that are identified. In the rest of this subsection I provide an informal discus-

[43] This is an extreme example. However, finding other examples of data where I would fail to identify the migration costs are not hard to come up.

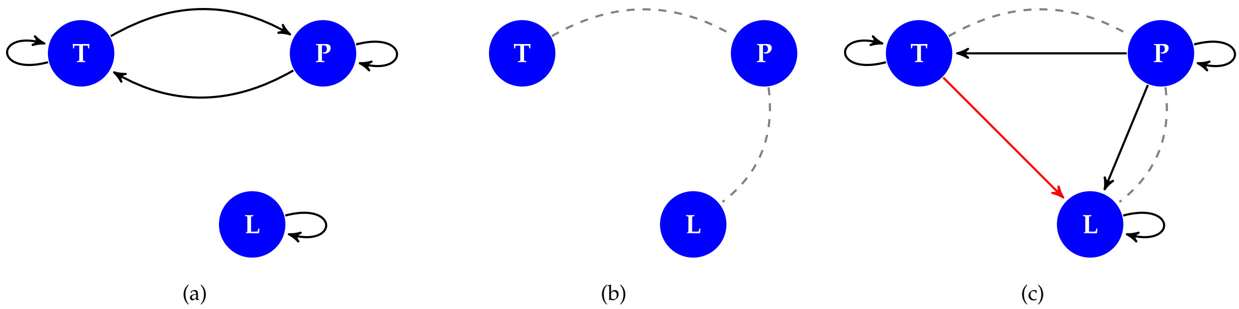


Figure 4 – Identification of Migration Costs in a Three Locations Example. The three locations are Toulouse (T), Paris (P) and Lyon (L). The left and right panel plot the graph representation of data where each (solid) edge represents some positive worker flow. The middle panel is an undirected graph where each (dashed) edge represents that the migration cost between the locations are *directly identified* (see main text).

sion of how to relax the data requirements for the identification of the migration costs and leave for Appendix 1.C.1 a more formal discussion of the details.

To keep things simple, suppose that there are only three locations in the data: Toulouse, Paris and Lyon. Suppose that for a particular period t and birth cohort b, I observe positive flows of migrants from Toulouse to Paris, and vice-versa, as well as workers who just stayed in each location. There are no outflows of workers from Lyon. The graph representation of such data is found in Figure 4a. From such data I can identify the (scaled) migration cost between Toulouse and Paris $\tau^{T,P}/\lambda$. Then I would say that the migration cost is *directly identified* from data for period t and birth cohort b.

I can do the same graph representation for different periods and birth cohorts. Suppose that for one of these periods and birth cohorts I can directly identify the migration cost from Paris to Lyon, $\tau^{P,L}/\lambda$. So using two different pairs of periods/birthplace I would have identified the migration costs between Toulouse to Paris and Lyon to Paris. This is represented in the graph in Figure 4b, where the edges as dashed lines represent that the migration costs between the connected locations are directly identified.

Now suppose there is a third pair of period/birth cohort data, t', b'. The following proposition summarizes sufficient conditions for identification of the migration cost from Toulouse to Lyon for the three location example, when the migration costs from Paris to Lyon and to Toulouse were already identified.

Proposition 2. ***Three locations.*** *Suppose that $\tau^{P,L}/\lambda$ and $\tau^{P,T}/\lambda$ are identified. Then, the (scaled) migration cost from Toulouse to Lyon $\tau^{T,L}/\lambda < \infty$ is identified from the labor flow data $\{L^{n,m}_{t,b}\}_{n,m\in\{T,P,L\}}$ if, for some birth cohort b and period t*

1. *There is a positive flow from Toulouse to Lyon, or viceversa.*
2. *There is an undirected path of labor flows from Toulouse to Lyon via Paris.*
3. *In all three locations there is a flow that stays.*

Proof. See Appendix 1.C.1 □

An example of period/birth cohort data fulfilling the identification conditions of Proposition 2, but not of Proposition 1, is represented in Figure 4c. The Figure also includes the dashed edges which represent the previously identified migration costs with data for other periods/birth cohorts. Differently from Figure 4a, there is only one flow going from Toulouse to Lyon, so the direct identification argument—the one from Proposition 1—is no longer valid to identify the migration cost between these two locations. The issue is that after normalizing the flow from Toulouse to Lyon by the flow that remains in Toulouse, the resulting expression

$$\frac{p_{t',b'}^{T,L}}{p_{t',b'}^{T,T}} = \exp\left(\mathcal{D}_{t',b'}^{L} - \mathcal{D}_{t',b'}^{T} - \tau^{T,L}/\lambda\right)$$

still contains the aggregate destination differences between the two locations. However, as Proposition 2 tells us that the data should be sufficient to identify the migration cost between Toulouse and Lyon. To see this, note that the destination dependent differences can be controlled by *pivoting* via Paris: use the difference between the probability of going to Toulouse from Paris and the probability of going to Lyon from Paris. Then, the remaining variation is

$$\frac{p_{t',b'}^{T,L}}{p_{t',b'}^{T,T}} \frac{p_{t',b'}^{P,T}}{p_{t',b'}^{P,L}} = \exp(-\tau^{T,L}/\lambda - \tau^{P,T}/\lambda + \tau^{P,L}/\lambda). \tag{1.16}$$

The ratio of labor flows going from Paris to Toulouse and to Lyon has information in the relative attractiveness of Toulouse versus Lyon, as well as the relative differences in migration costs. As both the migration costs of going from Paris to Toulouse and Lyon were already identified using data for other period/birth cohorts, then the migration cost between Toulouse and Lyon is also identified.

The example above is just one particular situation where the data fulfills the conditions of Proposition 2. However, note that in this example I identify the migration cost from Toulouse to Lyon under weaker conditions than those stated in Proposition 2, in particular the third condition: in equation (1.16) I did not use the labor flows that stayed in Lyon and in Paris. Similar case-by-case scenarios can be analyzed, but this becomes exponentially harder when the number of locations grows.[44] Therefore, for my context, I need identification conditions that are simple and easy to verify. Proposition 7 in Appendix 1.C.1 generalizes Proposition 2 and gives sufficient conditions for identification of a migration cost beyond the three location example.

As with Proposition 2, the more general Proposition 7 uses as a starting point some previously identified connections. It does not say how these have to be identified, though. Therefore, the identification argument is recursive: I can start with the directly identified migration costs and check which extra connections are identified. Then, I can use these new set of identified migration costs to find new ones, and so on. This recursive algorithm is explained with more detail in Appendix 1.C.1.

Although I can relax the data requirements for identification of the migration costs, the non-linear procedure that I use to estimate them might introduce some small-sample bias. I correct for

[44] Strictly speaking, the conditions are not weaker. I don't have to fulfill all the restrictions stated by the third condition of Proposition 3 because of the out-flows from Paris. This imposes a restriction in the direction of flows. However, the second condition of the Proposition says nothing about the direction of flows.

the bias by applying the split/panel jackknife estimation proposed by Dhaene and Jochmans (2015). The main idea is to split the panel in two and estimate for each half the migration costs. Then

$$\frac{\hat{\tau}_{BC}^{i,j}}{\lambda} = 2\frac{\hat{\tau}^{i,j}}{\lambda} - \frac{1}{2}\left(\frac{\hat{\tau}_1^{i,j}}{\lambda} + \frac{\hat{\tau}_2^{i,j}}{\lambda}\right),$$

is a bias-corrected estimate of the migration cost, where $\frac{\hat{\tau}^{i,j}}{\lambda}$ correspond to the estimate with the whole sample, and $\frac{\hat{\tau}_1^{i,j}}{\lambda}$ and $\frac{\hat{\tau}_2^{i,j}}{\lambda}$ correspond to the estimates for each half of the panel.[45]

Relaxing the data requirements for identification is even more important when doing the split-jacknife bias correction: compared to using the whole sample—where 69.3% of the connections satisfy the identification conditions of Proposition 1—only 47.9% satisfy the conditions for both sub-samples. In contrast, the connections satisfying the weaker conditions of Proposition 7 are 98.9% when using the entire sample, while 93.4% are satisfied in both sub-samples.

I parameterize those migration costs that are not identified in both sub-samples as a function of distance. Using the identified migration costs, I fit a linear model that depends on distance. I then use these estimates to impute the missing migration cost values. The function that parameterizes the (scaled) migration costs needs to fulfill an important property, besides the identification assumptions of normalization and symmetry. This property is that the change in a possible counterfactual scenario that corresponds to bringing all costs equal to zero should be invariant to the choice of unit of measurement for distance. Therefore, I parameterize the missing migration costs as:

$$\frac{\tau^{i,j}}{\lambda} = \begin{cases} 0 & \text{if } i = j \\ \mathcal{C}_\tau + \nu_\tau \log(\mathrm{d}^{i,j}) & \text{otherwise,} \end{cases}$$

where $\mathrm{d}^{i,j}$ is the distance between locations i and j, so ν_τ is just an elasticity. $\mathcal{C}_\tau$ is a constant that can be interpreted as a fixed cost of migrating, but that is linked with the choice of unit of measurement of distance. I choose geodesic distance in kilometers for parameterization of the migration costs, in order to fulfill the symmetry identification assumption. Figure 5 shows the estimated migration costs and the estimated values of $\mathcal{C}_\tau$ and ν_τ are, respectively, -0.474 (s.e. 0.319) and 1.427 (s.e. 0.054). As expected, the migration costs increase with distance. I leave their interpretation for section 1.6.2.

1.5.2 Wage dispersion parameter δ

In Appendix 1.A.2, I show that the expected log wage of a worker with birthplace b conditional on migrating from location i to j is

$$\mathbb{E}\left(\log\left(\mathrm{wage}_{t,b}^{i,j}\right)\right) = \log(w_t^j) - \delta \log(p_{t-1,b}^{i,j}). \tag{1.17}$$

Because of selection, the average wage of workers of a particular migration cohort is negatively related to the size of the migration cohort—which is related to the migration probability. When

[45] In a very simplified manner, what the correction is doing is to subtract an estimate of the bias. The idea is that the difference between the average of the difference between the estimates from one half of the sample and the entire sample is an estimate of the bias. By plugging in the negative of this average one can get the expression above.

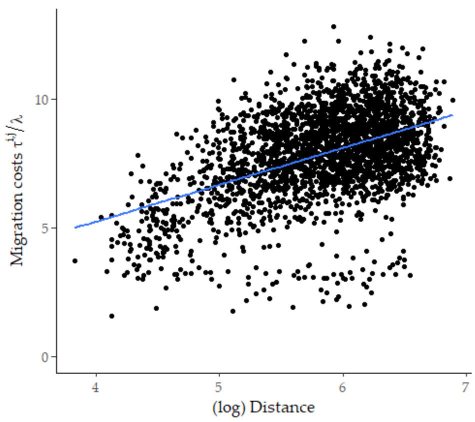


Figure 5 – Migration Costs vs Distance. The graph plots the migration costs vs (log) geodesic distance. Each point corresponds to a mobility cost and (log) geodesic distance of a pair of locations. The lines correspond to fitted a linear model. The slope corresponding is 1.43, s.e. 0.05 and the R^2 is 0.22.

more workers migrate, the efficiency of the marginal worker is smaller, reducing the average wage. The elasticity of average wage with respect to the migration probability is thus equal to $-\delta$.[46] Variation in migration costs imply variation in migration probabilities, which ultimately identifies δ. Substituting 1.5 into 1.17, I can write the previous expression as

$$\mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,b}\right)\right) = \tilde{\mathcal{O}}^{i}_{t,b} + \tilde{\mathcal{D}}^{j}_{t,b} + \delta\frac{\tau^{i,j}}{\lambda},$$

where $\tilde{\mathcal{O}}^{i}_{t,b}$ and $\tilde{\mathcal{D}}^{j}_{t,b}$ are origin/birthplace/period and destination/birthplace/period fixed effects. Note that the efficiency wage of the destination location is absorbed within the destination fixed effect. This expression reveals that the average compensation a worker needs in order to be willing to move from i to j is $\delta\frac{\tau^{i,j}}{\lambda}$. Then, I can run a regression with origin/birthplace and destination/birthplace fixed effects of individual (log) wages on the previously identified (scaled) migration costs to identify δ.[47]

I can control for differences in age and gender characteristics of individuals that should not

[46]The intuition of this elasticity is as follows: if efficiency shocks are more dispersed, i.e. higher δ, the gap in efficiency between marginal and average worker increases. Then, the average wage falls faster with the increase in cohort size, as the efficiency of the marginal migrant falls also at a faster rate.

[47]The idea of using the average wages to identify the dispersion parameter δ is similar in spirit to what Donaldson (2018) does to identify the trade elasticity. Donaldson collects data on different prices for a commodity, salt, as well as where production took place. Because of perfect competition and non-arbitrage, differences in prices between origin and destination should reflect the cost of trading across locations. Thus, high trade costs imply high prices. Donaldson uses the effect of trade costs on trade to recover the trade elasticity, which, in the context of his trade model à la Eaton and Kortum (2002) has a structural interpretation. In his model, buyers select where to import given differences in prices. The strength of this selection effect is driven by the trade elasticity, whose absolute value is negatively related to the dispersion of firms' efficiencies in each location. Higher values of the trade elasticity means that the relative efficiencies are more similar across goods, thus weakening the force of comparative advantage. Then, the effect of changes in trade costs over total imports is stronger when the comparative advantage motive is weaker. Similarly, in my model, migration locations are selected via wages, so high migration costs would imply high wages. I do the reverse as Donaldson as I use the labor flows—which would correspond to trade flows in his case—to infer (scaled) migration costs. Then I use the effect of migration costs on wages to infer the dispersion parameter of efficiency units. Analogous to his case, the dispersion parameter governs the strength of how workers pursue their comparative advantage in selecting migration destinations.

affect the migration decision, but might affect the wages in dimensions not captured by the model. Thus, before running the regression of migration costs on individual wages, for every year in my sample, I first run a regression of all wages on a quadratic polynomial in age and a gender dummy. I then take the residuals of those regressions as the main input for the remaining estimation steps.

Given the sequential identification strategy, the migration costs that I use to identify the dispersion parameter are measured with error. Even more so after the bias correction procedure, as it adds some variance as a cost for correcting the bias. This measurement error creates an attenuation bias on δ. To control for the bias, I instrument the migration costs. The estimated migration costs have a high correlation with geodesic distance, where pairs of locations that are further apart have on average a larger migration cost. For this reason, I instrument migration costs using geodesic distance and correct for the attenuation bias. [48] Details of the first stage regression are in Appendix 1.D.

After instrumenting, the estimated value of δ is 0.145 (s.e. 2e^{-4}), while the OLS estimate is lower, with an estimated value of 0.126 (s.e. 2e^{-4}).[49] My estimate is larger than the estimates found by Bryan and Morten (2019) for the U.S, 0.035, and Indonesia, 0.077, although their methodology only compares flows of natives versus non-natives.[50] Taking their estimate as a benchmark, this means that, in absence of a persistence force, i.e. $\rho = 0$, the migration elasticity in France $1/0.126 \approx 8$ is three and a half times smaller than that found for the U.S, $1/0.035 \approx 28$. This is in line with the idea that the U.S. has a much more mobile and dynamic labor market, although given the different models and identification steps, this should be taken with a grain of salt.

1.5.3 Conditional migration probabilities

Using the expression for the migration probabilities and the count data from workers migration decisions, I can write the conditional (log) likelihood function:

$$\log \mathcal{L} = \sum_t \sum_b \sum_{i,j} \ell_{t,b}^{i,j} \log \left(\frac{\exp\left(\mathcal{D}_{t+1,b}^{j} - \tau^{i,j}/\lambda\right)}{\sum_k \exp\left(\mathcal{D}_{t+1,b}^{k} - \tau^{i,k}/\lambda\right)} \right), \tag{1.18}$$

where $\mathcal{D}_{t+1,b}^{j} \equiv \overline{V}_{t+1,b}^{j}/\lambda$ are destination/birthplace/period specific fixed effects; and $\ell_{t,b}^{i,j}$ is the number of workers who changed jobs and moved from i to j with birthplace b at end of period t. It turns out that the direct maximization of the conditional (log) likelihood when the (scaled) migration costs are fixed is a highly tractable problem.

Proposition 3. *The values of the fixed effects $\mathcal{D}_{t+1,b}^{j}$, for all* j, b *and* t *that maximize the conditional (log) likelihood (1.18) are the same that solve the following system of equations*

$$\sum_i \ell_{t,b}^{i,j} = \sum_i \frac{\exp\left(\mathcal{D}_{t+1,b}^{j} - \tau^{i,j}/\lambda\right)}{\sum_k \exp\left(\mathcal{D}_{t+1,b}^{k} - \tau^{i,k}/\lambda\right)} \sum_h \ell_{t,b}^{i,h}, \quad \forall i,j \in \mathcal{I}. \tag{1.19}$$

[48] I consider as an instrument $\mathbf{1}(i \neq j)\log(\mathrm{d}^{i,j})$, so that the instrument is equal to zero for observations of workers who do not migrate. The correlation between the instrument and the migration costs is 0.93.

[49] Table 11 in Appendix 1.H contains the regression table of both the IV and OLS regressions.

[50] Bryan and Morten (2019) estimate the dispersion parameter by running a regression of average log wages against the (log) migration probabilities corresponding to equation (1.17).

Proof. See Appendix 1.C.2. □

The proof boils down to manipulating the first-order conditions of the maximization of the (log) likelihood.

The system above can be written more succinctly as $L_{t,b}^{j,\text{dest}} = \sum_i p_{t,b}^{i,j} L_{t,b}^{i,\text{orig}}$, where the inward labor flow is $L_{t,b}^{j,\text{dest}} = \sum_i \ell_{t,b}^{i,j}$ and the outward labor flow is $L_{t,b}^{i,\text{orig}} = \sum_h \ell_{t,b}^{i,h}$ for all locations i, j in $\mathcal{I}$. In other words, each of these expressions are just equal to a labor movement equation, where the sum of all the labor flows from a particular origin, $p_{t,b}^{i,j} L_{t,b}^{i,\text{orig}}$, have to be equal to the total labor observed in that destination, $L_{t,b}^{j,\text{dest}}$. Therefore the maximization of the likelihood corresponds to finding the fixed effects such that the migration probabilities satisfy such labor movement equations. The system is analogous to a balanced-trade equation that arises from gravity type-models. Trade economists have established the existence and uniqueness of the solution as well as developed efficient algorithms for computing it.[51]

The connection between the maximization of the conditional (log) likelihood and the labor flow equilibrium equation comes from relating two results: (i) there is a close relation between the maximization of the log likelihood and the PPML; (ii) estimation of gravity equations using PPML automatically satisfy the structural restrictions of gravity models.

Previous literature has pointed out, separately, these two connections. First, Guimaraes et al. (2003) show that solving *jointly* for the fixed effects $\mathcal{D}_{t+1,b}^{j}$ and migration costs $\tau^{i,j}/\lambda$ to maximize the conditional likelihood (1.18) is equivalent to doing a Poisson-Pseudo-Maximum Likelihood (PPML) estimation adding origin fixed effects, whose moment condition is equal to equation (1.14).[52] This moment condition is derived from the equilibrium expression of labor flows, which corresponds to a 'general gravity' framework.[53] Second, Fally (2015) shows that in a trade model where output and expenditures are consistent with the sum of outward and inward trade flows—which in my migration context is analogous to a consistent definition of inward and outward labor flows—the estimation of a 'general gravity' equation with origin and destination fixed effects using PPML automatically satisfies the structural restrictions imposed by the model.[54] Given that the maximization of the (log) likelihood and the PPML are closely related, it is therefore not surprising that the first order conditions of the maximization of the likelihood are as well closely related to the structural equations of a gravity model.

So why not estimate together the destination fixed effects and the (scaled) migration costs doing PPML? In doing so I would estimate, in one single step, the (scaled) migration costs and the underlying conditional migration probabilities. As pointed out by Dingel and Tintelnot (2020), doing the PPML instead of the maximization of the multinomial logistic log-likelihood is much more tractable as there are widely available algorithms which are extremely efficient, especially for high dimensional models like the one I consider here. The reason is that not all of the migration

[51]In Appendix 1.C.2 I use a general result from Allen et al. (2020a) and provide a simple proof for existence and uniqueness (up to a constant). I also describe the algorithm proposed by Pérez-Cervantes (2014) to find the solution.

[52]For a derivation of this result, see Appendix 1.G.

[53]Head and Mayer (2014) define as 'general gravity' for trade models when the trade flows can be written as $X^{i,j} = \exp(O^i + D^j - \vartheta \log(\widehat{\mathrm{d}}^{i,j}))$ where O^i and D^j are origin and destination specific fixed effects and ϑ is the trade elasticity.

[54]These restrictions are dubbed 'multilateral resistance' indexes by Anderson and Van Wincoop (2003).

costs are actually identified. Primarily because there are pairs of locations where no worker in the data migrated between the two in any year.

If the problem are the missing migration values I could, in principle, reverse the order of the identification steps I have followed so far. I could use the relationship between wages and migration costs to estimate them and impute values related to distance to those few that are missing. A slight difference is that the identified migration costs would have a different *scaling* factor. In particular the migration costs identified from the wages would be $\frac{\delta}{\lambda}\tau^{i,j}$. Using the migration costs estimates I could then estimate the underlying migration probability distribution and the dispersion parameter δ by doing a PPML estimation with origin and fixed effects.[55] A drawback of such an alternative is that the correction from the possible bias in the estimation of δ introduced by using the migration costs with measurement error is not trivial, especially from a computational point of view.

When taking the (scaled) migration costs $\tau^{i,j}/\lambda$ as given, I cannot benefit anymore from the computational advantages of the PPML estimation procedure when maximizing the conditional likelihood. In contrast to linear models, I cannot just re-define the endogenous variable of the Poisson regression and do the same estimation algorithm.[56] For example, when running a Poisson regression with origin and destination fixed effects whose left-hand-side variable is equal to $\ell^{i,j}_{t,b}\exp(\tau^{i,j}/\lambda)$, the estimated destination fixed effects would differ from those estimated by directly maximizing the conditional likelihood (1.18).

I use the fitted values that come from the maximization of the likelihood to compute estimates of the conditional migration probabilities. I use them in the next steps of my estimation strategy.

1.5.4 Efficiency wages w^j_t

Having identified both the dispersion parameter δ and the migration probabilities, I can identify the efficiency wage using the expression for the expected log wage of a worker (1.17). Passing $\delta\log(p^{i,j}_{t-1,b})$ to the left-hand-side, a simple average across migration cohorts with same destination would identify the efficiency wage.

1.5.5 Home Bias κ^j_b

To identify the home bias κ^j_b, I exploit the information contained in the expected log wages of the different migration cohorts. To ease notation, define the the expected log wage of a worker with birthplace b conditional on migrating from location i to j as

$$\omega^{i,j}_{t,b} \equiv \mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,b}\right)\right).$$

[55]Such an alternative identification strategy is developed formally in Appendix 1.G.

[56]Consider a linear model $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ to be estimated via Ordinary Least Squares. If $X_2\beta_2$ is fixed, I can redefine the endogenous variable as $y - X_2\beta_2$ and follow the same least squares algorithm to get an estimate of β_1.

The difference between the expected log wages of workers who move to j from location i with respect to the expected log wages of workers who return home b from i is

$$\begin{aligned}\omega_{t,b}^{i,j} - \omega_{t,b}^{i,b} &= \log(w_t^j) - \log(w_t^b) - \delta\left(\log(p_{t-1,b}^{i,j}) - \log(p_{t-1,b}^{i,b})\right)\\ &= \log(w_t^j) - \log(w_t^b) - \frac{\delta}{\lambda}\left(\overline{V}_{t,b}^{j} - \overline{V}_{t,b}^{b} - (\tau^{i,j} - \tau^{i,b})\right)\\ &= \log(w_t^j) - \log(w_t^b) - \frac{\delta}{\lambda}\left(V_{t,b}^{j} - V_{t,b}^{b} - (\tau^{i,j} - \tau^{i,b})\right) + \text{exp. error.}\end{aligned}$$

The last step exploits the rational expectations assumption: the migration decision at the end of period $t-1$ depends on the workers' expectation of the state of the world in period t, as reflected by the expected lifetime utility $\overline{V}_{t,b}^{j}$. Then, rational expectations imply that the difference between the expected utility $\overline{V}_{t,b}^{j}$ and $V_{t,b}^{j}$, which is the utility conditional on the realization of aggregate uncertainty in period t, is a mean zero expectation error.

The wage differentials reflect more than just aggregate differences between locations and migration costs. The utility differentials $V_{t,b}^{j} - V_{t,b}^{b}$ capture as well the effect of the home bias κ_b^j. I now show how to control for all the things that are not the home bias driving the wage differentials.

The lifetime utility $V_{t,b}^{j}$ of a location j is a function of: (i) a flow utility term that is constant across birthplace cohorts; (ii) the discounted expected utility of next period; and (iii), a birthplace specific option value for living in that particular location. Define

$$\zeta_{t,b}^{i,j} = \left(\omega_{t,b}^{i,j} - \omega_{t,b}^{i,b}\right) - \rho\beta\left(\omega_{t+1,b}^{i,j} - \omega_{t+1,b}^{i,b}\right) - (1-\rho\beta)\left(\omega_{t+1,b}^{j,i} - \omega_{t+1,b}^{b,i}\right).$$

Given that I have values for the discount factor β and the persistence parameter ρ, I can construct the sample analog of $\zeta_{t,b}^{i,j}$ using average wages for periods t and $t+1$. Substituting the equations for expected log wages (1.17), conditional migration probabilities (1.5), and lifetime utility (1.4), after some algebra I obtain that

$$\zeta_{t,b}^{i,j} = \mathcal{C}_t^j - \mathcal{C}_t^b + \left(\frac{(1-\beta)\delta}{\lambda}\right)(\tau^{i,j} - \tau^{i,b}) + \frac{\delta}{\lambda}\kappa_b^j + \text{exp. error,} \tag{1.20}$$

where $\mathcal{C}_t^j$ captures all the terms related to location j that are independent of birthplace.[57]

The expression above shows that the collection of wage differentials $\zeta_{t,b}^{i,j}$ depends only in aggregate differences across the two locations, relative differences in migration costs, and the home bias. The next-period wage differential $\omega_{t+1,b}^{i,j} - \omega_{t+1,b}^{i,b}$ controls for differences in next-period expected utilities. The third term in $\zeta_{t,b}^{i,j}$, the difference in average wages of the *reverse* migration cohorts for the next period, $\omega_{t+1,b}^{j,i} - \omega_{t+1,b}^{b,i}$, controls for differences in option values. The idea is that average wages of a migration cohort from a particular origin are informative about the option value of that origin location. The intuition is as follows. First, given selection of workers, the average log wage of a migration cohort is negatively related to their migration probability, as shown in equation (1.17). Now, consider the migration probability of going, lets say, from Toulouse to Paris versus the migration probability of going from Lyon to Paris. Assume, for the sake of the argument, that migration costs are the same. Then, as in both cases the destination is the same, the differences

[57]Formally, the constant term is defined as $\mathcal{C}_t^j = \log(w_t^j) - \rho\beta\log(w_{t+1}^j) - \frac{\delta}{\lambda}\mathcal{U}_t^j$, where $\mathcal{U}_t^j$ is the flow utility for living in location j net of the home bias.

in probabilities should reflect origin specific differences. If the probability of going to Paris from Toulouse is smaller than the probability of going to Paris from Lyon, it means that the alternative migration options attainable from Toulouse are relatively more attractive than those alternative options attainable from Lyon. In other words, the *option value* of being in Toulouse is higher than that of Lyon. Therefore, the difference between these two probabilities—and thus, of wages—are informative about the relative difference of option values between Toulouse and Lyon.[58]

The only thing left to control for are the aggregate differences $\mathcal{C}_t^j - \mathcal{C}_t^b$. So, similarly to the migration costs, to get rid of the aggregate differences I make the following symmetry assumption

Assumption 2. *The home bias is symmetric $\kappa_b^j = \kappa_j^b$ for all b, j in $\mathcal{I}$. Also, the cost of staying in the same location is zero, i.e. $\kappa_b^b = 0$ for all b in $\mathcal{I}$.*

Proposition 4 below shows that this assumption allows me to identify the home bias.

Proposition 4. *Let $\kappa_b^j = \kappa_j^b$ be symmetric, as defined by assumption 2. Then,*

$$\frac{\zeta_{t,b}^{i,j} + \zeta_{t,j}^{i,b}}{2\delta} = \frac{1}{\lambda}\kappa_b^j + exp.\ error. \tag{1.21}$$

Proof. It follows from equation (1.20). □

The Proposition shows how to exploit the wage information of workers with birthplace j, who mirror the behavior of those workers with birthplace b to identify the home bias. This means to use the information from the wage differentials by just interchanging the destination location with the birthplace location, i.e., to use the information in $\zeta_{t,j}^{i,b}$ to control for the remaining aggregate differences in $\zeta_{t,b}^{i,j}$. So the birthplace/destination pair fixed effects from a simple linear regression on $\frac{\zeta_{t,b}^{i,j} + \zeta_{t,j}^{i,b}}{2\delta}$ would identify the home bias scaled by $1/\lambda$.

I use the estimates of efficiency wages, the dispersion parameter δ, and the migration probabilities to complete the sample of average wages for those combinations that do not appear in the data. I impute values according to expression (1.17). If I were to impute all the average wages—instead of those that are just missing—it would be equivalent to use only the information contained in the estimated migration probabilities. I find the completion method a simple compromise to use both sources of information.[59]

Figure 6 plots the estimated home bias κ_b^j against distance. As it is clear, the relation is clearly positive, although it increases the variance the further the distance.[60] At first glance the birthplace costs are in general smaller than the migration costs. But keep in mind that the migration costs are paid only once, while home bias is present year after year. Thus, in present value the differences are less stark. I leave the discussion on how to interpret the migration costs and home bias costs for the next section.

[58] Similar to what I do with the next-period wages, Artuç et al. (2010) and Caliendo et al. (2019) use the next-period migration probabilities to control for the option value component of the expected utility. A slight difference is that in my case I also need to control for the persistence component in the expected utility.

[59] I could use both sources of information, the migration probabilities and the average wages and have an overidentified model and estimate it with GMM. I plan to do this in the future.

[60] Some of the estimated home bias are actually small and negative. These generally correspond to neighboring locations, like Deux-Sevres and Charente Maritime or Gironde and Pyrénées Atlantiques.

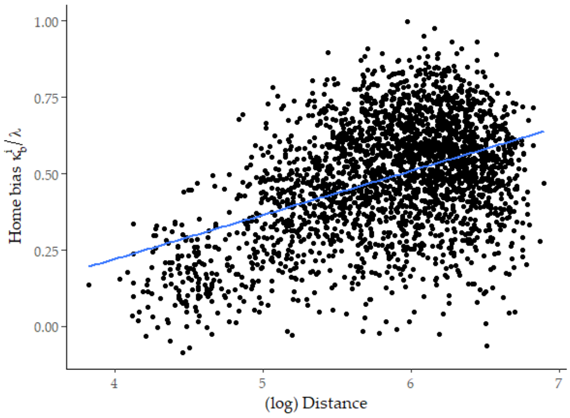


Figure 6 – Home Bias vs Distance. The graph plots the home bias vs (log) geodesic distance. Each point corresponds to a mobility cost and (log) geodesic distance of a pair of locations. The lines correspond to fitted a linear model. The slope corresponding is 0.14 s.e. 0.006 and the R^2 is 0.18.

At this stage I can perform a simple statistical test of the presence of home bias. Consider the null hypothesis $\mathcal{H}_0 : \kappa_b^j = 0$ for all $b, j \in \mathcal{I}$. The model with home bias nests the model without them, and then—under the null—all migration probabilities and lifetime utilities per location/period are the same across birthplaces. Under the null, the endogenous variables in equation (1.21) should all be equal to zero. Then, I can do all the previous steps of the estimation and do a joint significance test when estimating the home bias effects. The null hypothesis is rejected as the p-value associated to the F-stat is numerically indistinguishable from zero.[61] Therefore, I reject the hypothesis that there is no home bias.

I can also test whether I am overly complicating the model by estimating a home bias term for every location/birthplace combination instead of just a dummy that indicates whether a worker is outside her birthplace. In other words, I can test the null hypothesis $\mathcal{H}_0 : \kappa_b^j = \kappa$ for $b \neq j$. Again, I reject the null that home bias is constant across locations.[62]

1.5.6 Productivities A_t^j and Prices of Non-Housing Goods $P_{Q,t}^j$

In this section I explain how to *invert* the static part of the model to recover the underlying productivities that are consistent with the observed data. This will also allow me to recover price indices of non-housing goods, up to a constant.

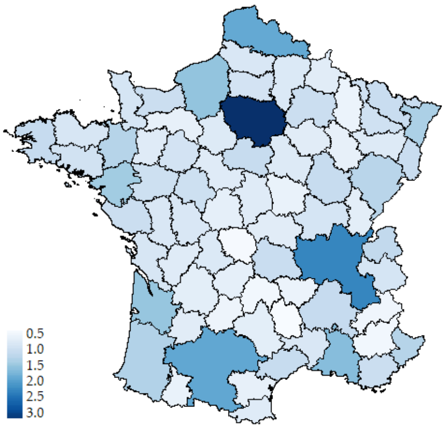
Combining the goods and labor market clearing conditions we get:

$$w_t^j N_t^j = \sum_{i \in \mathcal{I}} \frac{S^j \tilde{\psi}^{i,j}}{\sum_{k \in \mathcal{N}} S^k \tilde{\psi}^{i,k}} w_t^i N_t^i, \tag{1.22}$$

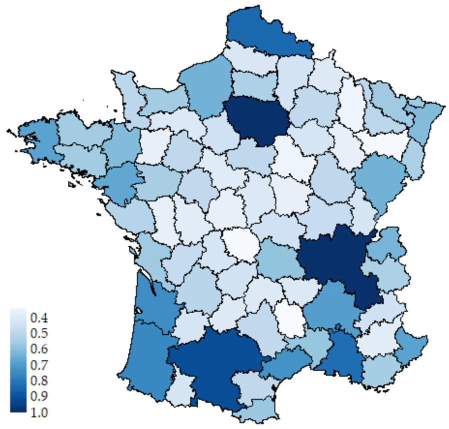
where $S^j \equiv \left(\frac{A_t^j}{x_t^j} \right)^{\varphi}$ is a *source* effect and $\tilde{\psi}^{i,j} = \left(\psi^{i,j}\right)^{-\varphi}$. Notice that these source effects also appear

[61]The F-stat is 8.03 with 2,628 and 2,560,620 degrees of freedom. The number 2,628 comes from the squared number of locations $73^2 = 2,701$ minus 73, as the constant term forces the normalization of a fixed effect per birth cohort.

[62]The associated p-value is again numerically indistinguishable from zero. The F-stat is 6.79 with degrees of freedom equal to 2,627 and 2,560,620.



(a) Productivities



(b) Amenities

Figure 7 – Estimated composite productivities and amenities. Both figures plot a composite value that includes also housing supply (see Section 1.5). The values of productivities are with respect to the national average. The values for amenities are with respect to the amenity in Île-de-France. The values for productivities correspond to the year 2017.

in the equation for the price index of the non-housing good $P^j_{Q,t}$.

As shown in Appendix 1.C.3, given the trade costs $\tilde{\psi}^{i,j}$ and the observed wage bills in the data, there is a unique solution, up to a constant, of the source effects. Using these soure effects along with the trade elasticity φ, I can identify the distribution of prices of non-housing goods (up to a constant).

Let, $\tilde{A}^j \equiv A^j \left(H^j\right)^\eta$ be a composite of productivity and housing supply. It summarizes how cheap is to produce something in location j by using an additional efficiency unit of labor. Substituting the price of the input bundle into the source effects and developing we get

$$\tilde{A}^j = \left(S^j\right)^{1/\varphi} \left(\frac{w^j}{1-\eta}\right)^{1-\eta} \left(\frac{w^j N^j}{\eta}\right)^{\eta}.$$

This means that given the estimated efficiency wages, the observed wage bills and the trade and output elasticities, along with the source effects, I can identify the distribution of the composite productivity/housing term $\tilde{A}^j$, up to a constant, which is all I need to solve the model. Figure 7a shows the map of composite productivities, where I have normalized the mean to be equal to one. As expected, the more productive regions are the most populated ones like Île-de-France, Lyon, Marseille, Toulouse, and Lille.

1.5.7 Amenities B^j

I conclude the identification section by explaining how I get the overall amenities from the residual variation in identified migration probabilities.

The Cobb-Douglas assumption on the technology of the final good plus market clearing imply that the price of housing is equal to $P^j_{H,t} \propto \frac{w^j_t N^j_t}{H^j}$. Substituting into the final good price index and then into the expression for lifetime utility (1.3) we get

$$V_{t,b}^{j} = \tilde{B}^{j} - \alpha \log\left(N_t^j\right) + (1-\alpha)\log\left(\frac{w_t^j}{P_{T,t}^j}\right) + \beta\rho V_{t+1,b}^{j}$$

$$+ \beta(1-\rho)\lambda \log\left(\sum_{k\in\mathcal{N}} \exp(\overline{V}_{t+1,b}^{k} - \tau^{i,j})^{\frac{1}{\lambda}}\right),$$

where $\tilde{B}^j \propto B^j + \alpha \log H^j$ is a composite of overall amenities and housing supply. As is clear from the expression above, the introduction of housing into the model works as a congestion force: the more efficiency units are in one location, the less attractive it becomes as the real wage decreases when the price of housing increases.

So I identify the composite amenity by exploiting the variation across migration probabilities, similar to what I did to identify the home bias. As will become clear, I can only identify the composite amenity up to a normalization, for which I pick a reference location x and put its corresponding value to be equal to zero. Then, the following ratio of probabilities is:

$$\log\left(\left(\frac{p_{t,b}^{i,j}}{p_{t,b}^{i,x}}\right)\left(\frac{p_{t+1,b}^{i,x}}{p_{t+1,b}^{i,j}}\right)^{\beta\rho}\left(\frac{p_{t+1,b}^{x,i}}{p_{t+1,b}^{j,i}}\right)^{\beta(1-\rho)}\right) = \frac{\tilde{B}^j}{\lambda} - \frac{\alpha}{\lambda}\log\left(\frac{N_{t+1}^j}{N_{t+1}^x}\right) + \frac{(1-\alpha)}{\lambda}\log\left(\frac{w_{t+1}^j}{P_{T,t+1}^j}\frac{P_{T,t+1}^x}{w_{t+1}^x}\right)$$

$$- \left(\frac{1-\beta}{\lambda}\right)(\tau^{i,j} - \tau^{i,x}) - \frac{1}{\lambda}(\kappa_b^j - \kappa_b^x) + \text{exp. error}.$$

Arranging all the terms to the left except for $\tilde{B}^j$, taking the averages for each location across the different periods would identify the composite of amenities and housing for each location.

Figure 7b show the map with the spatial distribution of the estimated composite amenities, where I have chosen Île-de-France as the reference location. To make it comparable to that of productivities—where all values are positive— I use the exponent of estimated amenities, $\exp(\tilde{B}^j)$. The map of amenities shows a similar pattern to the one of productivities: urban centers are more attractive. However, locations that are close to the coast, especially in the Southeast, close to the famous touristic regions of the Côte d'Azur, also have high values of amenities. Also the dispersion of productivities is almost twice that of amenities: the variance of the log of composite productivities $\tilde{A}^j$ is 0.13, while that of composite amenities $\tilde{B}^j$ is 0.07.

1.6 Model Solution and Counterfactual Analysis

I solve for the model in a steady-state and a continuous-population limit. As mentioned before, this renders the model deterministic and eases its solution. I choose a baseline year and solve for the model as if the productivities in the steady state are the same as those on the baseline year. I pick the year 2017 as the baseline.

With the solution of the model I first compute the birthplace premium: how much more welfare—in consumption terms—each birthplace cohort has compared to the national average. I then assess the importance of welfare differences due to birthplace in shaping overall welfare inequality. After, I compare the differences between the migration costs and the home bias. I show there is a direct correspondence between wage differentials of natives versus non-natives and the compensating variation in consumption a non-native needs to have the same utility as a native. This allows me to compare the steady-state of the model with the data. Finally, I compare my model to a model without home bias. I explore the implications of ignoring home bias for the response of real wages to a local productivity shock, and the costs associated with place-based policies.

Although solving this version of the model is computationally feasible, it is still challenging. To solve for the model I need to find the solution of a large system of non-linear equations. For example, there are 73^2 lifetime utilities to solve, one per each location/birthplace combination. However, by taking the total labor supply at each location as given, there is a sequential strategy to solve the rest of the variables very efficiently. I can show that part of the system are either contractions or can be represented as eigensystems with an eigenvalue equal to one. Solving these reduces to either iterating or finding the eigenvector associated with the unit eigenvalue. Both of these methods are computationally efficient. I explain the details of the solution algorithm in Appendix 1.B.

1.6.1 The Birthplace Premium and Decomposition of Welfare Inequality

I use the model in steady state to compute the differences in welfare across birthplace cohorts. This allows me to determine which workers are better off on average by the mere fact of being born in the right location.

Recall that the term V_b^i is equal to the expected utility of workers with birthplace b that live in location i *net* of current efficiency units. To recover the lifetime utility I need to sum again the current efficiency units and integrate across all the worker with birthplace b that live in i. Let the lifetime utility of an individual ι born in b, living in i, that migrated from j, and that has (log) efficiency θ^i be $\mathrm{v}_b^{j,i}(\theta_\iota^i) = V_b^i + \theta_\iota^i / (1-\beta\rho)$.[63] Then, the expected utility of workers living in i is

$$\tilde{V}_b^i = V_b^i - \frac{\lambda}{L_b^i} \sum_j \log\left(p_b^{j,i}\right) p_b^{j,i} L_b^j. \tag{1.23}$$

The second term to the right corresponds to an average selection term for workers living in i born in b. Appendix 1.A.7 contains the derivation of the expression above. Using these lifetime utilities per birthplace/location, I can compute the birthplace cohort average utility $\tilde{V}_b$ as well as the national average $\tilde{V}$ by

$$\tilde{V}_b = \sum_i \frac{L_b^i}{L_b} \tilde{V}_b^i, \quad \text{and} \quad \tilde{V} = \sum_b \frac{L_b}{L} \tilde{V}_b.$$

Definition 2. *The birthplace premium for birth cohort* $\boldsymbol{b}$*, denoted* $\boldsymbol{\varepsilon_b}$*, is defined as the average excess utility a worker born in* $\boldsymbol{b}$ *has compared to the national average, measured in consumption terms.*[64]

$$\tilde{V}_b + \frac{1}{1-\beta} \log(1-\varepsilon_b) = \tilde{V} \quad \Leftrightarrow \quad \varepsilon_b = 1 - \exp\left(\tilde{V} - \tilde{V}_b\right)^{1-\beta}.$$

When the birthplace premium is positive for a particular birth cohort b, it means that the welfare of that cohort is higher than the average French worker. Figure 8 shows the birthplace premium ε_b

[63] This expression comes from combining equations (1.1) and (1.3).

[64] Appendix 1.A.7 shows the detailed derivations to get the expression for the birthplace premium.

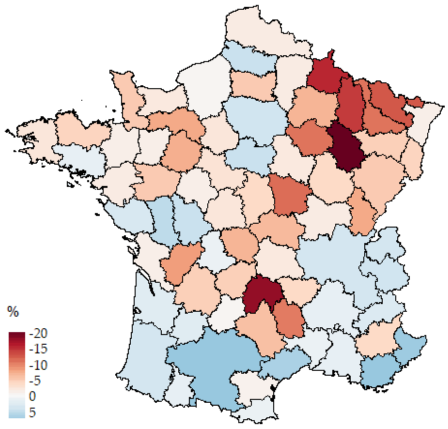


Figure 8 – Birthplace Premium. The map shows the different birthplace premia ζ_b for the different birth cohorts. The birthplace premium is the excess welfare, in consumption terms, that each birth cohort has on excess to the national average.

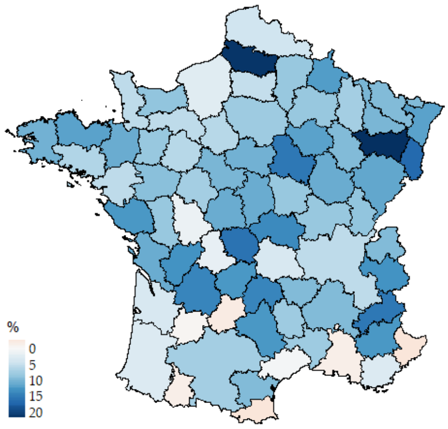
for the different cohorts. A location in the map represents a *birth* location and the color within a location represents the birthplace premium of the cohort born in such location. In absence of the home bias, these premia should all be equal to zero.

As is clear from the Figure, the inhabitants from the Île-de-France (Paris) region have a higher welfare, in consumption terms, compared to the national average. This is almost 5% larger than the national average real wage. In general, individuals born in locations that are overall attractive, as those in the South, or close to large agglomerations seem to be better off. The big winners are those born close to Toulouse in the South-West, or along the Côte d'Azur in the South-East, with a birthplace premium a little more than 7%. Some birth cohorts are doing very poorly in comparison. In the North-East, the cluster formed by Ardennes, Meuse, Meurthe-et-Moselle, Haute Marne, and Moselle have birthplace premia ranging from minus 10 to 20 percent. Another small cluster, towards the South-West in the Massif Central region, formed by the locations of Cantal and Lozère have birthplace premia of minus 17 and 10 percent, respectively.

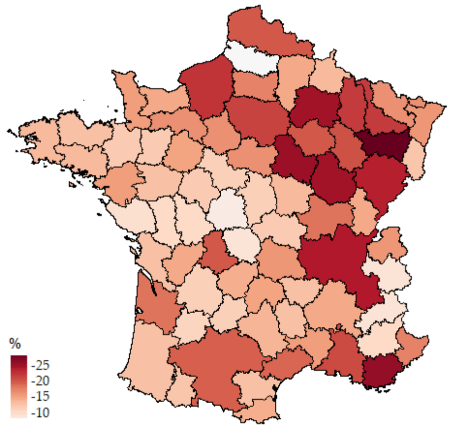
Almost all of the locations between the North-East and South-West clusters have negative birthplace premia. This region is known in France as the Empty Diagonal, which according to Wikipedia 'is a band of low-density population that stretches from the French department of the Landes in the southwest to the Meuse in the northeast.'[65] Looking back to the estimated amenities and productivities in Figures 7a and 7b, the locations in the Empty Diagonal are not attractive overall. The correlation between the birthplace premium and log composite productivities $\tilde{A}^i$ and amenities $\tilde{B}^i$ is 0.47 and 0.48, respectively. Thus, it is not surprising that the Empty Diagonal groups the big losers in terms of birthplace premium.

In the South, overall amenities are higher than in the North, and the productive and large population centers are more evenly distributed across space. Thus, even if someone born outside an attractive location within the South, it is probable that she lives in a productive location close to her birthplace, making her, on average better off. In addition, there are more options for relatively close, productive locations, for a worker born in the South than in the North. For those born in

[65] https://en.wikipedia.org/wiki/Empty_diagonal



(a) Workers Born in Toulouse



(b) Workers Born in Haute-Marne

Figure 9 – Excess utility Across Residence Location, Different Birthplace. The left panel shows the excess utility (compared to the national average) of the workers born in Toulouse that live in the different locations, measured in consumption terms. The right panel does the same but for workers born in Haute-Marne.

the North, Île-de-France is the only option if they want to live in a close-by productive location, while in the South and South East there is Lyon, Marseille and Toulouse that are relatively close to one another.[66] Therefore, the option value of being born in an unproductive region in the South is larger than in the North.

That a location b in the map shows a large birthplace premium does not mean that the inhabitants of region b have higher utility. Instead it means that those who were born in b have higher utility. Some of the workers might be living outside their birth location. However, those who are born in an attractive location would only move if the migration opportunity gives them more utility than in their birthplace. Thus, the average utility of a worker in any location is influenced by the workers outside option, which is their home location.

The influence of birthplace on average utility of workers regardless of residence location is illustrated in Figures 9a and 9b. The left panel shows the excess utility, measured in consumption terms, of the workers born in Toulouse, a location with high birthplace premium living in all the different locations. The right panel does the same but for workers born in Haute-Marne, which has a low birthplace premium. For both cohorts, there is heterogeneity in average utility across locations for workers with the same birthplace. However, the place of birth influences largely a workers' relative position with respect to the national average as the workers of Toulouse are better off relatively than those born in Haute-Marne, regardless of their residence location. Moreover, fixing residence location, the distribution of welfare across birth cohorts is very similar than the one portrait by Figure 8. For example, the correlation of excess utility for Toulouse residents with different birthplace, with the birthplace premium is 0.84; for the residents of Haute Marne the correlation is 0.92.

Welfare Decomposition. I now explore the relative importance of between-birthplace versus across-locations differences in shaping overall welfare inequality. I find that birthplace is a main driver

[66] There is also, Montpellier, Bordeaux and Nice for example.

of welfare inequality, and is almost as important as idiosyncratic differences and sorting across locations.

The dispersion of welfare is var $\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\right)$, where the variance is taken over all workers, who are indexed by ι. The overall dispersion can be decomposed as follows

$$\operatorname{var}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\right) = \underbrace{\operatorname{var}\left(\mathbb{E}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; j;b\right)\right)}_{\text{Between-birthplace/location}} + \sum_j\sum_b \frac{L_b^j}{L} \times \underbrace{\operatorname{var}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; j;b\right)}_{\text{Within-birthplace/location}}$$

$$= \operatorname{var}\left(\tilde{V}_b^j\right) + \sum_j\sum_b \frac{L_b^j}{L} \times \operatorname{var}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; j;b\right).$$

The first term is the variance across the average utility of workers born in b living in location j. The second term corresponds to the weighted average of the variance within each birthplace cohort and residence location. I further decompose each within-birthplace/location variance across the different migration cohorts

$$\operatorname{var}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; j;b\right) = \underbrace{\operatorname{var}\left(\mathbb{E}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; i\to j;b\right)\right)}_{\text{Between-migration cohort}} + \sum_j\sum_b \frac{L_b^j}{L} \times \underbrace{\operatorname{var}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; i\to j;b\right)}_{\text{Within-migration cohort}}$$

Conditional on a residence location j and birthplace b the only heterogeneity in utility comes from dispersion in the discounted (log) efficiency shocks $\theta_\iota^i/(1-\beta\rho)$. Conditional on a migration cohort $(i\to j;b)$ the (log) efficiency $\theta^i/(1-\beta\rho)$ is distributed Gumbel, with scale parameter λ and mean $-\lambda\log\left(p_b^{i,j}\right)$. Then,

$$\operatorname{var}\left(\mathrm{v}_b^{i,j}(\theta_\iota^i)\;\middle|\; j;b\right) = \underbrace{\operatorname{var}\left(\lambda\log\left(p_b^{i,j}\right)\;\middle|\; j;b\right)}_{\text{Selection}} + \underbrace{\frac{\pi^2}{6}\lambda^2}_{\text{Idiosyncratic}}.$$

The first term to the right is the variance across origins i of expected efficiency conditional on a birthplace b and residence j. It reflects the dispersion in average selection patterns for workers with different origin locations. The second term comes from the fact that conditional on a migration cohort, i.e. the conditioning on origin, destination and birthplace, the variance of efficiency wages is equal to the variance of the different (discounted) efficiency shocks, which are distributed Gumbel with scale parameter λ. The contribution of this term to total variance is fixed across different scenarios. Therefore, it constitutes a lower bound on overall welfare inequality.

I can decompose furthermore the variance across average utility of a birthplace/location $\tilde{V}_b^j$ as

$$\operatorname{var}\left(\tilde{V}_b^i\right) = \operatorname{var}_b\left(\tilde{V}_b\right) + \sum_b \frac{L_b}{L} \times \operatorname{var}\left(\tilde{V}_b^j\;\middle|\; b\right).$$

The first term to the right is the between-birthplace dispersion of average utilities per birth cohort $\tilde{V}_b$. This corresponds to the average differences between Figures 9a and 9b. The second term is the within-birthplace dispersion of utilities, weighted by the size of the birth cohort. This corresponds to the within heterogeneity across locations in Figures 9a and 9b.

Table 5 presents the results of the decomposition. I find for the baseline scenario that the between-birthplace component explains 43.2% of the dispersion in average location/specific welfare. The within-birthplace variation, that corresponds to differences across locations explains only

Table 5 – Decomposition of Welfare Inequality

		Between BP/Location (% Total Var)		*Within BP/Location (% Total Var)*		
	Total Var (/Baseline)	Between-BP	Within-BP	Selection	Idiosyncratic	Migration Rate (%)
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline	1	43.5	3.5	17	36	1.2
No Mig. Costs	1.61	78	0	0	22	11.8
High Home Bias	1.7	79	0	0	21	11.6
No Home Bias	≈1	0	8	56	36	3.2
High Mig. Costs	0.96	0	10	52	38	1.9
No Both	0.36	0	0	0	100	12.4
Without Geography	0.46	6	4	11	79	0.3
No Mig. Costs	≈1	64	0	0	36	12.2
High Home Bias	0.81	55	0	0	45	11.6
No Home Bias	0.91	0	3	58	39	1.6
High Mig. Costs	0.76	0	3	50	47	0.9
No Both	0.36	0	0	0	100	12.9

Note: The table shows the decomposition of variance of individual welfare for different scenarios. The different indentation means that some elements are changed in comparison to the immediate, less indented scenario. For example, the second row refers to the baseline scenario with no migration costs. The third row refers to the baseline scenario, no migration costs and high home bias. "Without Geography" means that all (composite) productivities and amenities are constant, trade costs are the same across locations and birthplace cohort sizes are equalized. The first column represents the total variance as a fraction of the variance in the baseline scenario. The second and third columns correspond to the percentage of total variance explained by the variance of average utility per birthplace/location $\tilde{V}_b^j$, while the fourth and fifth columns correspond to the within component. The fourth column explains how selection within a migration cohort drives inequality. The fifth column corresponds to idiosyncratic differences within each migration cohort.

3.4% of the total inequality. The within-birthplace/location explains most of the variation with 53.4% where selection contributes with 17% and the idiosyncratic component—conditional on a migration cohort—contributes with 36%.

The "Without Geography" row corresponds to a counterfactual scenario where I I shut down differences in (composite) productivities and amenities, as well as making trade costs the same across regions and equalizing the size of the birthplace cohorts.[67] When locations are more, the overall variance decreases by more than 50% and the fraction of total variance explained by the between-birthplace component is reduced from 43 to just 6 percent. The remaining differences are explained by the heterogeneity of the home bias of the different cohorts.[68] The within-birthplace/location selection term, which corresponds to column (4) is reduced from 17 to 11 percent. Homogeneous locations together with migration costs give little reason for workers to move around, as reflected by the migration rate, and therefore reducing the importance of selection.

Reducing migration costs but keeping the home bias increases the variance of welfare. As workers move more, but with different patterns across birthplace cohorts, then the average selection effects differences are magnified.[69] When increasing the home bias, the mobility patterns differ more, causing inequality to increase as well as the importance of the between-birthplace component.[70] Without home bias, the within-birthplace component explains only 8% of the total variance. Most of the variance is explained by the within-birthplace/location selection component: with the reduction in home bias, workers move more according to their comparative advantage, increasing welfare inequality. Increasing the migration costs mitigates this selection channel and increases the importance o heterogeneity across locations, as reflected by the increase from 8 to 10 percent in column (3).

Finally, when removing both the home bias and the migration costs, overall variance is explained entirely by the idiosyncratic shocks. In equilibrium, average welfare should be equal across all birthplace/locations. Without impediments to move around, the location decision is not determined by the origin location or birthplace of a worker. Then, workers would only choose where to live according to their idiosyncratic productivity. And while they indeed select across locations, the probability of going to any location is the same. Thus, the variance from the selection term is zero, leaving only the idiosyncratic component to explain the overall dispersion in welfare.

The small percentage of total dispersion explained by the within-birthplace component in the baseline scenario does not mean that heterogeneity of locations is unimportant in explaining welfare inequality. On the contrary, heterogeneity of locations is reflected in the between-birthplace component as illustrated by the "Without Geography" scenario, as it is the heterogeneity in locations—

[67] I allow for costly trade still. I take the average trade cost off the diagonal as a measure of cost between any two locations. For trade within each location I take the average of within diagonal trade costs. For the construction of trade costs in the baseline see Appendix 1.D

[68] The heterogeneity across birthplace cohorts of home bias itself increases the variance. However, it also changes the population composition across cities, even if they have the same fundamentals. This creates heterogeneity in real wages across locations which, combined with differences in employment distribution across the birthplace cohorts, adds to the heterogeneity.

[69] Also, there is more concentration towards productive areas, and while the differences in utility coming from differences in place of residence can be muted, the differences stemming from heterogeneous home bias can be magnified.

[70] To increase home bias in the third row, or migration costs in the fifth row, I take the off-diagonal value of the cost that enters the migration decision, which is an exponential function, and divide it by two.

along with the home bias—that determines the outside option of workers and influence their location decisions.

Home bias amplifies the role of geography in the long run welfare dispersion by making workers gravitate around their home location, preventing them from arbitraging away aggregate differences across locations. In contrast, migration costs prevent the short-run adjustment of labor and do not seem to matter much for the long run distribution of employment. Thus, geographic differences are better arbitraged away and dispersion is driven by the within-location, across-origin component.

1.6.2 Comparison of migration costs and home bias

In the model sketched in Section 1.4, both migration costs and home bias enter as utility costs. It is therefore tempting to compare their magnitudes for each location pair to determine their relative importance. Doing so requires adjusting the estimated magnitudes for the fact that migration costs are paid one time and home bias is a recurring cost.

I use a compensating variation argument to make both mobility costs comparable. First, for the migration costs, I look at how much larger the wage of a migrating individual needs to be in order to have the same utility as an individual that did not move. Similarly, I compute the compensating variation in wages such that a non-native individual has the same utility as a native.

If I compare two workers from the same birth cohort, one migrating from i to j and the other staying in j, then the wage of the migrating individual has to be larger for them to have the same utility. As shown in Appendix 1.A.6, this extra wage compensation in percentage terms is equal to

$$\xi_t^{i,j}(\tau) = \exp\left(\tau^{i,j}\right)^{(1-\beta\rho)} - 1.$$

The compensating variation, being a function of the migration cost is symmetrical. Because of this symmetry and the extreme value (Gumbel) assumption on the efficiency shocks, there is a simple correspondence with the data from observed wage differentials and the compensating differentials for migrants in the model.

Proposition 5. *For workers with same birthplace* b, *the compensating variation in wages a migrant needs to have the same utility as a non-migrant is identified from the following difference-in-differences in wages*

$$\log(1+\xi_t^{i,j}(\tau)) = \frac{1}{2}\left(\mathbb{E}\left(\log\left(\text{wage}_{t,b}^{i,j}\right)\right) - \mathbb{E}\left(\log\left(\text{wage}_{t,b}^{i,i}\right)\right) + \mathbb{E}\left(\log\left(\text{wage}_{t,b}^{j,i}\right)\right) - \mathbb{E}\left(\log\left(\text{wage}_{t,b}^{j,j}\right)\right)\right).$$

Proof. It follows from substituting the average utility (1.4) and the migration probability equation (1.5) into the expected wage equation (1.17). □

The Proposition follows the same logic as the identification of (scaled) migration costs when observing bilateral labor flows coming in both directions as stated by Proposition 1. It tells us that by comparing the wages of migrants versus those who stay in the same location we can back out the compensation.

In a similar way as with migrants, the extra wage compensation, in percentage terms, that a worker born in b who lives in location j needs in order to have the same utility as a native is

$$\xi_{t,b}^{j}(\kappa) = \exp\left(\kappa_b^j - \beta(1-\rho)(1-\beta\rho)\mathbb{E}_t \sum_{s=0}^{\infty} (\beta\rho)^{s-1}\left(\log\left(\Omega_{t+s,b}^{j}\right) - \log\left(\Omega_{t+s,j}^{j}\right)\right)\right) - 1. \quad (1.24)$$

The second term in the right hand side of the expression above correspond to a difference in the option values of living in location j between natives and non-natives. In the steady state, the expression above would be

$$\xi^{j}_{ss,b}(\kappa) = \exp\left(\kappa^{j}_{b} - \beta(1-\rho)\left(\log\left(\Omega^{j}_{b}\right) - \log\left(\Omega^{j}_{j}\right)\right)\right) - 1.$$

In contrast to the compensation to migrants, the compensation to non-natives is not symmetric. Because of the differences in option values changing the indices (b, j) would give different values. Thus, there is no direct correspondence of a compensating differential for an ordered pair (b, j) in the model and wage data. However, as Proposition 6 below shows, there is a correspondence for the compensating differentials that belong to the *unordered* pair (b, j).

Proposition 6. *For workers with same origin location* i*, the geometric average of the compensating differentials of a non-native worker living in* j *and a non-native worker living in* b *is identified from the following difference-in-differences of wages*

$$\frac{1}{2}\left(\log(1+\xi^{j}_{t,b}(\kappa)) + \log(1+\xi^{b}_{t,j}(\kappa))\right) = \frac{1}{2}\left(\begin{array}{c} \mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,b}\right)\right) - \mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,j}\right)\right) + \\ \mathbb{E}\left(\log\left(\text{wage}^{i,b}_{t,j}\right)\right) - \mathbb{E}\left(\log\left(\text{wage}^{i,b}_{t,b}\right)\right) \end{array}\right) \quad (1.25)$$

Proof. Same as in Proposition 5. □

Proposition 6 tell us that the wages can reveal a measure of compensation to non-natives for every pair of locations. The wages can give the average compensation for a non-native from Toulouse living in Lyon and the compensation to a non-native from Lyon living in Toulouse. To make it comparable to the previous compensating measures I say that the (geometric) average of the compensating differential of non-native for an unordered pair of locations (j, b) is

$$\tilde{\xi}^{j}_{t,b}(\kappa) = \sqrt{\log(1+\xi^{j}_{t,b}(\kappa))\log(1+\xi^{b}_{t,j}(\kappa))} - 1,$$

which is symmetrical. The right-hand side of (1.25) is then equal to $\log(1+\tilde{\xi}^{j}_{t,b}(\kappa))$.

Using the results of both Propositions, I compare the compensating differentials in the steady-state of the model with those observed in the data. An attractive feature of both Propositions is that they show there is a structural interpretation of the wage differentials without relying on any of the estimated parameters. In the model I can compute $\xi^{j}_{t,b}(\kappa)$ and $\xi^{b}_{t,j}(\kappa)$ without taking the geometric average. But to make it comparable with the compensations found in the data, I also compute the average compensation $\tilde{\xi}^{j}_{t,b}(\kappa)$ in the model.

In the model, I first compute the compensating differentials for every pair—either origin/destination or birthplace/destination. To compute the average compensating differential for migrants, I take a weighted average of each pair of compensating differentials, using the total migration flows between every pair as weights.[71] Similarly, for the average compensating differential of non-natives, I compute a weighted average using the birth cohort population in each location L^{j}_{b} as weights. When computing the compensating differential of migrants I exclude from the computation those

[71] In more detail, I compute $L^{i,j} = \sum_b L^{i,j}_{b}$ for weighting the migration compensating differential.

flows that remain in the same location. Similarly, I exclude the fraction of workers who stay in their birth location when computing the average compensating differential for non-natives.

In the data, using the results from Proposition 5, I first compute the compensating differentials for every pair using the difference in average log wages. I can either used just the observed wages or, similarly to when estimating the home bias, I can impute the model consistent average wage for those combinations that are missing by using the estimated efficiency wages and migration probabilities. Given that the compensating differentials for migrants (non-natives) are constant across birth cohorts (origin-destination cohorts), for every year, I can get an estimate by taking the following weighted average

$$\widehat{\xi}_t^{i,j}(\tau) = \sum_b \left(\exp(\overline{\log\left(\text{wage}_{t,b}^{i,j}\right)} - \overline{\log\left(\text{wage}_{t,b}^{j,j}\right)}) - 1 \right) \frac{L_{t,b}^{i,j}}{\sum_{b'} L_{t,b'}^{i,j}},$$

where $\overline{\log\left(\text{wage}_{t,b}^{i,j}\right)}$ is the sample average of log wages. I do the same for the compensating differential of non-natives. As in the model, to compute the average compensating differential, I compute a weighted average either using the total migration flow or the number of workers who live outside their birthplace. When using the sample with imputed wages, the number of workers correspond to the fitted values from the maximization of the likelihood function (1.18).[72]

The first column of Table 6 shows the results for the model. The average compensation that a non-native needs in order to have the same utility as a native is 18%. using the geometric average—which is comparable with the data— I found that is 12%. The average compensation for a migrant to have the same utility as a non-migrant is larger and equal to 55%. Compared to what I found using the data—that correspond to the second to fourth column in the table—, the compensating differential for migrants is smaller in the model, but is larger for the compensation to non-natives. By using the observed sample, the compensation for non-natives is 15%. However this is small when compared to the compensation for migrants which is more than 100%. This is similar when using the *Observed + Imputed* sample with observed and imputed wages. If I use only the imputed wages—as shown in the fourth column—the steady state model value and the data are more alike. This is expected as the migration costs were identified using information on the labor flows—which are related to the migration probabilities, which are then used to impute the model consistent wages. I conjecture that the estimated migration costs and the associated compensating differential in the model would be larger if I were to use the information in wages to estimate the migration costs. However, the compensations of non-natives are more alike in the model and the data when using only observed wages.

Compensating differentials for migrants are similar in magnitude to those previously estimated in the literature. For example, Kennan and Walker (2011) find that the average migration costs would correspond to an annual increase in the wage of between 36 to 76 percent.[73] Such migration costs might look implausible a priori, but when interpreted as average wage differences between

[72]If a wage is imputed for a combination it means that the associated observed labor flow is zero. Thus, I can't use them as weights as it will not change the outcome from a weighted average using only observed wages. That is why I use fitted values for labor flows instead, which are positive.

[73]In Kennan and Walker (2011), the estimated migration cost for the average mover is equal to 312, 146 dollars. Using a discount factor of 0.96, this corresponds to an increase of 15, 500 dollars per year for forty years. Given an estimated average wage of individuals in the

Table 6 – Compensating variation in wages

	Model		Data		
	Ordered pairs	*Geometric Average*	*Observed*	*Observed + Imputed*	*Just Imputed*
Non-Natives (%)	18.6	12	15	22	33
Migrants (%)	55.6		107	104	63

Note: The first row of table shows the average compensating variation in wages a non-native needs to have the same utility as a native. For the model I consider two cases: when using ordered and when using the geometric average. The second row shows the same but for a migrants to have the same utility as non-migrants. The first two columns show the values in the steady state of the model. The third to fifth columns show the values using different versions of the data. The third column, *Observed*, uses the direct observed wage differentials. The fourth column, *Observed + Imputed* completes the missing average wages in the original sample by imputing wages according to expression (1.17). The fifth column, *Just Imputed* only uses imputed wages.

migrants and non-migrants, these large costs should actually be expected. From Figure 3b, the difference between average log wages of migrants and non-migrants is around 1, i.e. wages of migrants are more than *double* that of non-migrants. Large migration costs are consistent with large observed wage differences between migrants and non-migrant workers.

The average compensating differentials computed above depend on the equilibrium allocation. First, they are weighted averages, thus they depend on how the labor flows—which are equilibrium objects—are determined. Second, for the compensation of non-natives, the differences in option values—also equilibrium objects—need to be taken into account. Thus, to make a comparison between the migration costs and the home bias that is invariant to the equilibrium allocation, I take a simple average of the ratio

$$\frac{\exp(\kappa_b^j) - 1}{\exp\left(\tau^{b,j}\right)^{(1-\beta\rho)} - 1} \tag{1.26}$$

across every pair of locations where $b \neq j$. The numerator is the compensation for a non-native *as if* the option values between natives and non-natives are equal. The denominator corresponds to the migrant's compensation, which is already invariant to the equilibrium.

I find that the average of the ratio (1.26) is 0.26. Therefore, for the average pair of locations, the magnitude of the home bias is 26% that of migration costs, measured in compensating differentials.

Even though home bias has a smaller magnitude than migration costs, it has a large impact on shaping the overall employment distribution. The reason is that home bias ties workers *permanently* to their preferred location.

1.6.3 The Effect on Output of Removing Home Bias vs Migration Costs

The main difference between the effect of migration costs and home bias on output is that, while both prevent workers from pursuing their comparative advantage, home bias affects the long run

50 percentile and with age 30 (42,850 dollars), this corresponds to an annual increase in the wage of 36%. Using the estimated average wage for individuals in the 50 percentile but of age 20 (20,166 dollars), this would correspond to an increase of 76%. There are similar estimates in the literature for changing sectors. For example, Artuç et al. (2010) estimate that the cost of changing sector is about ten times the average wage, which corresponds to an annual increase of 49% for forty years.

distribution of employment. The population size of a location is not only related to economic fundamentals—like productivity and amenities—but also to the size of the different birth cohorts. Moreover, home bias makes workers *gravitate* around their birthplace over time. In contrast, migration costs limit short run movements from the current location, so workers can do staggered movement towards the most productive areas.

I solve the model by shutting down the effect of each of the mobility costs. By easing the movement of workers, output can increase from two main factors. First, a selection effect: workers are able to select themselves to locations where they are relatively more productive. Second, a composition effect: if workers concentrate more in productive regions, then overall output increases.

I distinguish between the selection and composition effects as follows. Define total manufacturing output, $\mathcal{Y}$ as the sum of all real outputs per location, Y^i. Each of these local outputs can in turn be defined as $Y^i = A^i_L L^i$, where A^i_L is the labor productivity in location i. Labor productivity A^i_L is an endogenous object as it depends on the average efficiency units per location, which reflects how workers select into locations. Thus, total output can be though as a function of the distribution of labor productivities and workers

$$\mathcal{Y}(\mathbf{A}_L, \mathbf{L}) = \sum_i Y^i = \sum_i A^i_L L^i,$$

where $\mathbf{A}_L$ and $\mathbf{L}$ are vectors containing, respectively, all the labor productivities and the number of workers in each location. The sum of workers is normalized to 1, i.e., $\sum_i L^i = 1$. Thus, I can make the following decomposition

$$\mathcal{Y}(\mathbf{A}_L, \mathbf{L}) = \overline{A}_L + \sum_i (L^i - \overline{L})(A^i_L - \overline{A}_L) = \overline{A}_L + \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}).$$

The covariance term $\widetilde{\text{cov}}(\cdot)$ gives a measure of how concentrated is the population in the most productive locations.[74]

In models where labor productivity is exogenous, changes in aggregate output are driven entirely by the composition effect, i.e., by changes in the covariance term. However, as in my model labor productivities are endogenous, I will do something slightly different. For a given variable $\mathcal{X}$ in the baseline economy, define its value in the counterfactual as $\mathcal{X}'$. Then, the difference in total output between a counterfactual scenario and the baseline economy is

$$\begin{aligned}
\mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \mathcal{Y}(\mathbf{A}_L, \mathbf{L}) &= \mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \left(\overline{A}_L + \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L})\right) \\
&= \mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \left(\overline{A}_L + \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}')\right) + \left(\widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}') - \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L})\right) \\
&= \underbrace{\mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \mathcal{Y}(\mathbf{A}_L, \mathbf{L}')}_{\text{Change Selection}} + \underbrace{\left(\widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}') - \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L})\right)}_{\text{Change Composition}}. \qquad (1.27)
\end{aligned}$$

The first term in equation (1.27) corresponds to the selection effect as it leaves the allocation of labor across locations constant and changes the vector of labor productivities from $\mathbf{A}_L$ to $\mathbf{A}'_L$. The second

[74] Olley and Pakes (1996) propose this decomposition and use the change in the covariance term to evaluate the reallocation effect towards more productive plants following a deregulation reform in the telecommunication sector in the U.S.

Table 7 – Effects of Home Bias and Migration Costs on Output

Remove	Δ Output (%)	Δ Selection (%$\Delta\mathcal{Y}$)	Δ Composition (%$\Delta\mathcal{Y}$)
Home Bias	11	88	12
Migration Costs	35	106	-6
Both	37	115	-15

Note: The table shows the differences in output between the baseline economy and different counterfactuals, as well as its decomposition in selection and composition gains as a percentage of the change in output. First row, corresponds to a counterfactual where the home bias is removed. The second row corresponds to the case without migration costs. The third row removes both the home bias and the migration costs.

term captures the composition effect: how much of the output change be explained by changes in the labor allocation towards productive locations, while keeping productivities in each location constant.

Table 7 shows the decomposition of the output gains in the selection and composition terms. The left panel corresponds to the baseline model, estimated with home bias. The first row shows that output increases by 11% in the counterfactual where I remove the home bias. Of those output gains, the majority comes from selection with 88% and 12% from the composition effect. When I remove only migration costs, as represented in the second row, output increases more, with a gain of 35%. Moreover, the decomposition of the gains are very different from that where I remove the home bias. All of the gains come from better selection as they account for 106% of the gains in output. The third row of the table reports the output gains and its decomposition when removing both migration costs and the home bias. In that case output increases by 37% with all of the gains coming from selection. Therefore, the lion's share in output gains of removing both mobility costs comes from removing the migration costs.

What explains this difference in the source of output gains between removing either the home bias ot the migration costs? First, when there is no home bias, but workers still face the migration costs, workers can slowly reallocate towards more productive regions. Thus, in the new steady-state, more workers would end up in more productive areas. In contrast, when I only remove the migration costs, workers tend to gravitate around their birthplace, as changing locations do not affect the underlying mobility costs. Sure, workers would have more profitable opportunities close to their home location, which leads to a great increase in the selection component, but they would remain close to their birthplace. Thus, the reallocation of labor towards the most productive locations is limited.[75]

To sum up, removing the home bias or the migration costs has different long-run implications for the allocation of labor. In both cases, the main output gains come from the selection effect: by removing a mobility cost, workers can better pursue their comparative advantage. However,

[75] Additionally, because the birth cohort size is positively related to location attractiveness, then those born in attractive locations go to neighboring locations, which are, on average, less productive. In turn, this leads to less concentration of population in large productive areas as reflected by the negative covariance term.

removing home bias can lead to large reallocation of labor towards more productive regions, but the removal of migration costs allows people to sort better but mostly close to their home location.

1.6.4 How much do workers value living in their Home Location?

The model allows to compute how much consumption workers are willing to sacrifice in order to live in their home location. Consider a worker with birthplace b living away from her home in location i. Then, the difference in log efficiency units such that the worker is indifferent between staying in her current location or going back to her home location b is

$$\Delta_b^{i,b} = u^i - u^b + (1-\beta\rho)\tau^{i,b} + \beta(1-\rho)\left(\log(\Omega_b^i) - \log(\Omega_b^b)\right) - \kappa_b^i,$$

where u^i collects all the terms in the lifetime utility that depend only on location i and are independent of birthplace. The expression shows that in order to be indifferent, the worker needs to be compensated by the differences in aggregate utility and option values between the two locations, as well as the cost of migrating. In addition, the worker is also willing to forego some log efficiency units—to go back home, represented in the home bias term κ_b^i.

My objective is to compute how much of a pay cut workers are willing to accept to go back home, controlling for non-birthplace specific factors. The first column of Table 8 shows the average page cut a non-native needs to be indifferent about returning home when considering only the effect of κ_b^j. This is equal to 3.8%. When considering the differences in option values, which corresponds to the second column of the table, the pay cut decreases to 2.8%.

Even after imputing the adjustment for option values, there are non-birthplace specific factors affecting the number. For example, if a location has low migration costs towards all the other locations, which would be reflected in a high option value for *all* the birth cohorts. To control for these common differences in option values across birth cohorts, I compute

$$-\kappa_b^i + \beta(1-\rho)\left(\left(\log(\Omega_b^i) - \overline{\log(\Omega^i)}\right) - \left(\log(\Omega_b^b) - \overline{\log(\Omega^b)}\right)\right), \tag{1.28}$$

where $\overline{\log(\Omega^i)}$ is the across-birthplace average of option values for location i.[76] The idea of adding these average terms is that, without home bias, the birthplace specific option value Ω_b^i is equal to the average. In such case, the birthplace specific term in the expression above would be equal to zero. The third column of the Table shows that, after the adjustment for average differences in option values, the average pay cut non-natives are willing to take to go back home is 5.2%.

In similar lines, I can compute what is the pay raise a worker would need to be indifferent between leaving her birthplace or staying. The results are in the second row of Table 8. For my preferred specification, shown in the third column, the average wage gain that would have left indifferent the worker leaving outside her birthplace is 10.9%.

Thus, I can conclude that the average French worker who lives away from her birthplace is willing to accept a pay cut between 3 and 5 percent on her annual salary in order to get back home or needs a pay raise of at least 5 to 10 percent to leave it.

[76] I take a weighted average given the population of the different birth cohorts in each location. For each average, I exclude the corresponding migration cohort whose origin and destination is the same as their birthplace. That is, the average is $\overline{\log(\Omega^i)} = \sum \sum_{b \neq i} \log(\Omega_b^i) \frac{L_b^i}{\sum_{b' \neq i} L_{b'}^i}$.

Table 8 – Wage cut/raise to return/leave home

	Home Bias κ_b^j	Home Bias + Δ Opt. values	Home Bias + Δ Opt. values - Adjustment
Return (%)	-3.8	-2.8	-5.2
Leave (%)	4.5	7.3	10.9

Note: The first row shows the pay cut that the average worker who lives outside her home location is willing to take in order to go back to work at her home location. The first column refers to the pay cut when only the home bias terms κ_b^j are considered. The second column, in addition to the home bias, considers the changes in option values between the two locations. The third columns considers as well the differences in option values, but it makes an adjustment by subtracting the average difference on option values, as specified in equation (1.28). The second row does the analogous for the wage gain the average worker needs to remain indifferent between staying in her home and leaving.

1.6.5 Home Bias, Labor Mobility and the Pass Through of Productivity to Wages

In this section, I compare the general equilibrium outcomes of a more standard migration model without home bias with my model. I show that a model without home bias overstates the migration response to changes in the economy.

I first estimate the model without home bias. This means that I estimate the migration costs without conditioning on birthplace and I follow the same estimation steps thereafter.[77] Appendix 1.H shows a scatter plot comparing the migration costs of baseline versus the no-home-bias model. In general, the estimates for migration costs without the home bias are larger than in the baseline model. The estimated dispersion parameter is slightly smaller than in the baseline model, with a value of 0.135 (s.e. 1.8e-4) compared to 0.145 (s.e. 2e-4). The estimated distribution of composite amenities and productivities are very similar to the ones estimated for the baseline model. In particular, the correlation between the two estimates of amenities is 0.95 and of productivities 0.99.

In a model without home bias, the elasticity of the labor flow going from location i to j to a change in the efficiency wage in location j is

$$\tilde{\epsilon}^{i,j} = \frac{1}{\lambda}\left(1 - p^{i,j}\right),$$

where $p^{i,j}$ is the probability of going to j from i. In contrast, the same elasticity in my model is

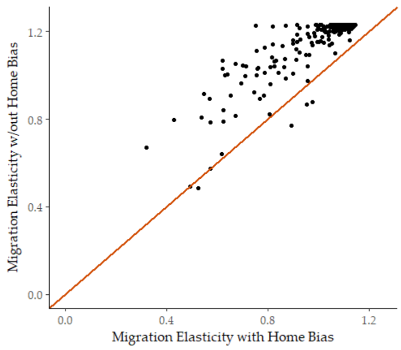
$$\epsilon^{i,j} = \frac{1}{\lambda}\left(1 - \sum_b \frac{L_b^{i,j}}{L^{i,j}} p_b^{i,j}\right),$$

where $L^{i,j} = \sum_b L_b^{i,j}$. Thus, the elasticity in the baseline model with home bias is a weighted average of the birth specific elasticities $\epsilon_b^{i,j} = \frac{1}{\lambda}\left(1 - p_b^{i,j}\right)$.

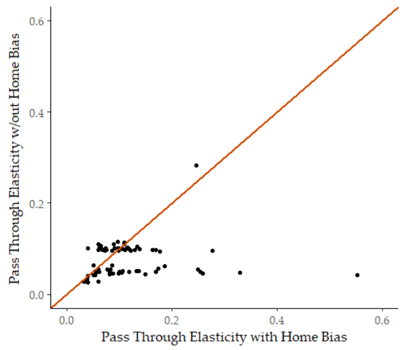
Figure 10a compares the different elasticities for the models with and without home bias in the steady state. Each dot corresponds to the migration elasticity for a pair of locations i, j. In almost all of the cases the elasticities are larger in the model without home bias. Thus, the predicted migration response to a change in any particular location would be higher, under-stating the long-term effect of a change in productivity on real wages.

In equilibrium, the pass-through of local productivity changes to real wages is counteracted by an increase in the price of housing, whose strength is governed mainly by two factors. First, by

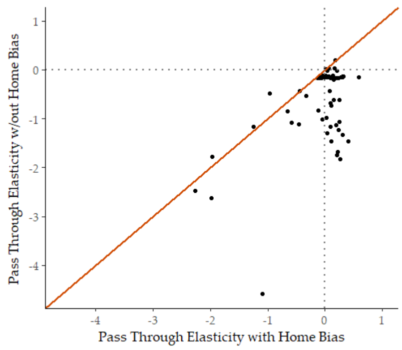
[77] I estimate the migration costs doing the same Poisson regression as in section 1.5.1 but the dependent variable is the total flow from i to j, i.e. $L^{i,j}$ and the origin and destination fixed effects do not depend on birthplace.



(a) Comparison Migration Elasticities



(b) Pass-Through to Efficiency Wages



(c) Pass-Through to Average Wages

Figure 10 – Comparison Model with and w/out Home Bias. The solid line in all three plots represents the 45° line. The top-left figure compares the migration elasticities in both models, and each dot represents a pair of locations i, j. The top-right panel compares the pass-through elasticity of productivity on real efficiency wages and each dot represents a location. The bottom panel compares the effect of pass-thorugh on real average wages.

how easy workers can substitute between housing and non-housing goods. Because of the Cobb-Douglas assumption, the elasticity of substitution between housing and non-housing is equal to one. Thus, in contrast, to the classic Rosen (1979)-Roback (1982) framework the local productivity gains are not fully appropriated by the land-owners as workers can substitute between housing and non-housing goods.[78] Second, the price of housing is affected by the mobility response of workers. If more workers go to a location after a productivity shock, then the price of housing increases.

Figure 10b compares the pass-through elasticity of local real efficiency wages from a local productivity shock for the models estimated with and without home-bias, where each dot represents a location. As expected the, pass-through is larger in the model with home-bias. The ratio of the average pass-through elasticity for the model with home bias over the same average for the model without home bias is 1.5. Thus the pass-through elasticity is on average 50% larger in a model with home bias, as the labor response is smaller.[79] However, the value of pass-trough is small, with an elasticity of 0.11. This is a consequence of the fixed housing supply. Although the housing supply elasticity appears to be very low for France (see Fack (2006)), including an elastic housing supply can be an interesting extension.

Figure 10c compares the pass-through elasticity of productivity shocks on average real wages. In contrast to the efficiency wages, the pass-through can be negative as an increase in productivity drives in less productive individuals lowering the average wage. However, in the model with home bias the majority of the pass-throughs are positive, in contrast to the mode without home bias, where almost all of the pass-throughs are negative. The average pass-through though, is negative in both models. In the model with home bias, the elasticity is -0.08, and the model without home bias the elasticity is -0.6.

A model without home bias overestimates the total migration response to changes in productivities across locations. Also, it would overstate the indirect effect on real wages to those locations non-affected by the productivity shock. These are indirectly affected by: (i) the out-migration of those locations towards the more productive region, and (ii) the decrease in the overall price of tradables.[80] Neglecting the home bias might lead to wrong conclusions when analyzing counterfactual scenarios.

1.6.6 The Effect of Home Bias in Place-Based Policies

The previous section shows that home bias matters for the mobility response of workers and the general equilibrium effects of productivity on real wages. Similarly, policy evaluations that neglect the home bias effect might give very different answers compared to an evaluation where home bias is taken into account.

[78] In the classic Rosen-Roback framework—at least the one presented in Moretti (2011)—workers are homogenous, housing is in fixed supply and there are no migration costs. The indirect utility of workers in location i is equal to $w^i - P_H^i$. This corresponds to a linear utility function in non-housing consumption where workers have to consume one unit of housing. In equilibrium, workers are indifferent between locations. Then, an increase in productivity in location i would increase the nominal wage w^i but also would increase in a same amount the price of housing P_H^i. Thus, the increase in productivity is fully capitalized by land owners.

[79] Appendix 1.H shows a plot comparing the employment response between the two models. The average employment elasticity to a local productivity shock is 30% less in a model with home bias versus a model without it.

[80] For an estimate of the indirect effects of productivity shocks in other locations for the U.S. see Hornbeck and Moretti (2020)

One of such policies are place-based policies: a subsidy to the inhabitants of a particular location financed by general taxes. In absence of productivity or amenity spillovers, such policies can be justified as a way to redistribute income across space. Because of the concavity in the flow utility of consumption, the total effect on overall social welfare—which is the sum of the utilities across all locations—might increase if redistribution reduces inequality. However, a common concern with such place-based policies is that, while aiming at some spatial redistribution of income, it also distorts the location decisions of workers of non-targeted locations. Thus, it can drive workers away from productive locations to poor locations, resulting in efficiency losses. Thus, the increase on social welfare that comes from redistribution might be trumped by the efficiency losses and a revenue neutral placed-based policy might reduce social welfare.

Which effects dominates in determining social welfare—redistribution or efficiency—depends ultimately in how strong is the migration response of workers. Therefore, a model without home bias—which overstates the mobility response of workers—would overstate as well the costs of a placed based policy.

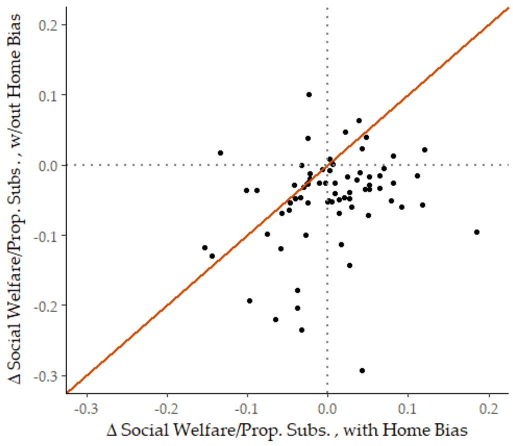
I compare the effects on social welfare of a 10% place-based labor subsidy between the model estimated with and without home bias.[81] In the exercise, I subsidize each location, one-by-one, in both models and compare the changes in social welfare. I find that the model without home bias—where the efficiency costs are more prevalent—finds almost always a negative effect in social welfare. In contrast, the model with home bias finds that in the majority of the cases, a placed based policy increases social welfare.[82]

Figure 11a plots, for every time I subsidize a location, the relation between the change in social welfare when there is a home bias and when there is not. I normalize each change in social welfare by the total subsidies spend as a proportion of output. Each dot corresponds to a subsidized location, with the x-axis measuring the percentage change in social welfare with home bias and the y-axis doing the same but for the scenario without home bias. The solid diagonal line corresponds to the 45 degree line. The Figure shows that for the majority of situations, social welfare is greater in the case with home bias. This indicates that the costs of placed-based policies are overstated in a model without home bias.

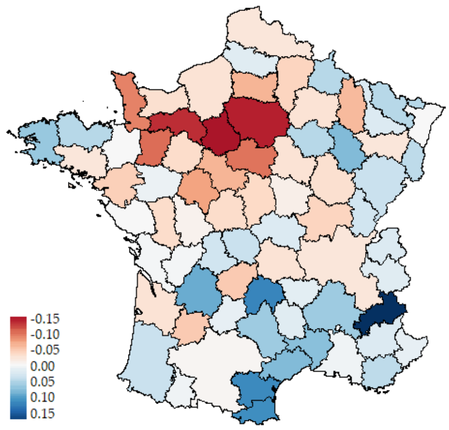
The Figure is also divided in four quadrants by a vertical and horizontal line at zero. This divide the situations when a policy increases welfare for both models. In the model without home bias only in 16% of the cases social welfare increases, while in the model with home bias this is 55%. The South-East quadrant is interesting as it collects 45% of the cases. This quadrant corresponds to the case where the policy increases welfare in the baseline scenario but the model without home bias predicts a reduction. The North-West quadrant shows the locations that a model without home bias predicts that social welfare increases if subsidizing those locations in contrast to the model with home bias, which predicts a reduction in social welfare.

[81] The introduction of tax policies affects the environment of the economy in two important ways. First, by collecting taxes and distributing subsidies from a centralized position, there would be locations that are net receivers of government transfers while other locations are net payers of taxes. In contrast to the baseline economy without policies, in the new situation trade is necessarily unbalanced.

[82] I am not taking a stance on whether these policies are the best for redistributing income. The aim of this section is to rather evaluate the costs.



(a) Comparison Between Models



(b) (Δ Social Welfare)/(Subsidies/Output)

Figure 11 – Response to Place-Based Subsidies. The left panel compares the social welfare gains between the baseline economy with home bias and a a model estimated without home bias, normalized by the subsidies as a proportion of output. Each dot corresponds to a subsidized location. The right panel presents a map shows the change in overall social welfare by subsidizing each location, normalized by the subsidies as a proportion of output. Locations in red mean that when subsidizing such locations, overall social welfare decreased.

The number of dots in both the North-West and South-East represent a measure of diagnosis of the model without home bias: it either wrongly predicts that social welfare increases—the North-West—or that social welfare is reduced–the South-East. In 52% of the cases, the diagnostic with a model without home bias is wrong.

The map on Figure 11b shows in red the locations where social welfare is reduced whenever they are subsidized in the model with home bias. It shows that subsidizing attractive and populous regions decreases overall social welfare. Although the output of manufacturing can increase by moving people to more productive regions, the regressive redistributive nature of subsidizing rich locations dominates, thus reducing the social welfare. Appendix 1.H plots the map of locations that reduce social welfare in a situation without home bias. As already implied by the South-West quadrant of Figure 11a, there is little intersection on which location reduces welfare between the baseline and counterfactual scenario.

The previous exercise shows that the negative effects of place based policies might not be that strong. While in the comparison I do not compute the combination of place-based policies that maximize social welfare, it can still be informative on the consequences of place-based policies in general.

The limited migration responses to the different policies in the presence of home bias should extend to more general settings. A proper analysis of how these preferences affect the design of optimal spatial tax policies, as in Gaubert et al. (2020), can be important to not overestimate, either the effects, or the costs, of implementing such policies.

1.7 Conclusion

In this paper I study the aggregate consequences of workers having a preference to live close to their home. To do so, I first show that the data support the presence of a home bias in workers migration decisions. I find that the labor flows are biased towards workers' birthplaces. I also find, via a gravity regression, that distance from one's birthplace is negatively related to the labor flow to a particular location. Additionally, I find that workers accept a wage discount for living in their home location.

After documenting the biased labor flows, I propose a framework to accommodate my empirical findings. I build a quantitative dynamic migration model embedded with home bias, understood as a cost from living away from one's home. I use data on wages and labor flows, along with the structure of the model, to identify and estimate the different parameters of the model. Among those, I show how to separately identify the migration costs and the home bias. I find that in the steady state of the model, the compensation a non-native needs in order to have the same welfare as a native is between 10 to 30 percent the compensation a migrant needs to have the same welfare as non-migrant.

Using the estimated model I solve for the steady state and compute the average welfare per birthplace cohort. I find that workers born within the attractive areas of Paris, Nice, or Toulouse, have 5 to 7 percent higher welfare, in consumption terms, than the average French worker.

The fact that the home bias effect is strong should affect how economists think about public policy programs that encourage mobility across regions by alleviating the pecuniary cost of moving. These types of policy can include subsidizing movers, making social security rights transferable (like in the European Union), etc. Policy makers should be cautious with the expectations on such programs, as the mere presence of ties to one's home could mitigate their effects. Understanding the precise nature of these preferences is important, in order to inform policy makers about the best policies to boost mobility and help people from economically distressed areas. Also, the presence of the home bias might help rationalize the existence of place-based policies.

1.A Derivations

In this appendix I derive the main equations of the model. I start by deriving the lifetime expected utility and the conditional migration probabilities. After that, I derive the expected wage per migration cohort. Then, I derive the expressions for comparing both the migration and home bias in terms of consumption terms to equalize the utility to a worker who did not move. Finally, I derive the welfare equations.

1.A.1 Lifetime Utility

A worker ι with birthplace b at location i with current efficiency of θ^i might change jobs with probability $1-\rho$. If it changes jobs, the worker observes a single offer per location. This translate into the worker observing a vector of log efficiency shocks Θ for the next period if it has the opportunity of changing jobs. If the worker does not change jobs it goes into the next period with the same efficiency unit. Without loss of generality, assume that the worker migrated to location i in the current period t. Then, the lifetime utility of worker ι is

$$\mathrm{v}^i_{t,b}(\theta^i_{t-1,\iota},\mathbf{\Theta}_{t,\iota}) = B^i + \log\left(C_{t,\iota}\right) - \kappa^i_b + \beta\rho\mathbb{E}_t\left(\mathrm{v}^i_{t+1,b}(\theta^i_{t-1,\iota},\mathbf{\Theta}_{t+1,\iota})\right) + \beta(1-\rho)\max_k\left[\mathbb{E}_t\left(\mathrm{v}^k_{t+1,b}(\theta^k_{t,\iota},\mathbf{\Theta}_{t+1,\iota})\right) - \tau^{i,k}\right],$$

subject to the per period/state budget constraint

$$P^i_t C_{t,\iota} = w^i_t \exp(\theta^i_{t-1,\iota}).$$

Using the budget constraint to substitute out consumption $C_{t,\iota}$ and iterating forward we get the following expression

$$\mathrm{v}^i_{t,b}(\theta^i_{t-1,\iota},\mathbf{\Theta}_{t,\iota}) = \mathbb{E}_t\left[\sum_{s=0}^{\infty}(\beta\rho)^s\left(\begin{array}{c} B^i + \log\left(\frac{w^i_{t+s}}{P^i_{t+s}}\right) + \theta_{t-1,\iota} - \kappa^i_b \\ +\beta(1-\rho)\max_k\left[\mathbb{E}_{t+s}\left(\mathrm{v}^k_{t+s+1,b}(\theta^k_{t+s,\iota},\mathbf{\Theta}_{t+s+1,\iota})\right) - \tau^{i,k}\right]\end{array}\right)\right].$$

Define as $\tilde{\mathrm{v}}^i_{t,b}(\mathbf{\Theta}_{t,\iota}) \equiv \mathrm{v}^i_{t,b}(\theta^i_{t-1,\iota},\mathbf{\Theta}_{t,\iota}) - \frac{\theta^i_{t-1,\iota}}{1-\beta\rho}$ as the lifetime utility of individual ι *net* of the present discounted value of efficiency units. Note that by subtracting $\theta^i_{t-1,\iota}/(1-\beta\rho)$, the net lifetime utility $\tilde{\mathrm{v}}^i_{t,b}(\mathbf{\Theta}_{t,\iota})$ is not longer a function of the state $\theta^i_{t-1,\iota}$. We can rewrite above's expression as

$$\tilde{\mathrm{v}}^i_{t,b}(\mathbf{\Theta}_{t,\iota}) = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho\mathbb{E}_t\left(\tilde{\mathrm{v}}^i_{t+1,b}(\mathbf{\Theta}_{t+1,\iota})\right) + \beta(1-\rho)\max_k\left[\mathbb{E}_t\left(\tilde{\mathrm{v}}^k_{t+1,b}(\mathbf{\Theta}_{t+1,\iota})\right) - \tau^{i,k} + \frac{\theta^k_{t,\iota}}{1-\beta\rho}\right],$$

There are two independent sources of uncertainty for individual ι: one concerns the idiosyncratic efficiency shocks, summarized at each period t in the vector $\mathbf{\Theta}_{t,\iota}$). The second source is aggregate uncertainty, related to possible changes in local productivity levels, and, because of the discrete nature of the model, uncertainty in the distribution of the aggregate labor supply even after conditioning on productivity levels. As in the main text, I summarize all of the aggregate uncertainty in the vector Z_t with a conditional distribution $F(Z_t|Z_{t-1})$. I assume that the vector of idiosyncratic shocks $\mathbf{\Theta}_{t,\iota}$) is independent of Z_s, for all $s = 1,2,..,t$. Let $V^i_{t,b} \equiv \mathbb{E}_{\mathbf{\Theta}_t}\left(\tilde{\mathrm{v}}^i_{t,b}(\mathbf{\Theta}_{t,\iota})\right)$ be the expected net lifetime utility conditional on the aggregate state Z_t, i.e. just taking the expectation over the log efficiency shock vector $\mathbf{\Theta}_t$. Also, I define $\overline{V}^i_{t+1,b} = \int V^i_b(Z_{t+1})dF(Z_{t+1}|Z_t)$ is the expected lifetime utility of living in location i at period $t+1$. Then, taking expectations over the vector of efficiency units conditional on the aggregate state we obtain

$$V^i_{t,b} = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho\overline{V}^i_{t+1,b} + \beta(1-\rho)\mathbb{E}_{\mathbf{\Theta}_t}\left(\max_k\left[\overline{V}^k_{t+1,b} - \tau^{i,k} + \frac{\theta^k_{t,\iota}}{1-\beta\rho}\right]\right).$$

I assume the idiosyncratic log efficiency shock θ^i is i.i.d over time and is distributed Gumbel (Type-I Extreme Value) with zero mean and variance equal to $\frac{\pi^2\delta^2}{6}$. Then, $\theta^i/(1-\beta\rho)$ is also distributed Gumbel with zero mean and variance $\frac{\pi^2\lambda^2}{6}$, where $\lambda \equiv \delta/(1-\beta\rho)$. So this means that by adding persistence to the model and letting agents to keep their efficiency shocks it is as if workers have realizations of shocks from a distribution with larger dispersion. Intuitively, what this is doing is that smaller differences on efficiency unit shocks across locations are magnified by the possibility, with probability ρ of keeping that same efficiency shock in the future. This make locations that a priori offer similar wages to be more differentiated in net present value. Therefore, the total labor supply response to differences in wages across locations is dampened.

In order to solve for the option value $\mathbb{E}_{\Theta_t}\left(\max_k\left[\overline{V}^k_{t+1,b} - \tau^{i,k} + \frac{\theta^k_{t,\iota}}{1-\beta\rho}\right]\right)$ and obtain the migration probabilities, first define the following distribution:

$$\begin{aligned}
G^{i,j}_{t,b}(v) &= \Pr\left(\overline{V}^j_{t+1,b} - \tau^{i,j} + \frac{\theta^j_{t,\iota}}{1-\beta\rho} < v\right) \\
&= \exp\left(-\exp\left(-\left(\frac{v - \overline{V}^j_{t+1,b} + \tau^{i,j}}{\lambda}\right) - \overline{\gamma}\right)\right),
\end{aligned}$$

where the second equality comes from the Gumbel distributional assumption on θ and $\overline{\gamma}$ is the Euler-Mascheroni constant.

To ease notation, define $\boldsymbol{u}^{i,j}_{t,b} \equiv \overline{V}^j_{t+1,b} - \tau^{i,j} + \frac{\theta^j_{t,\iota}}{1-\beta\rho}$. Fix $\boldsymbol{u}^{i,j}_{t,b} = v$. Then we have

$$\begin{aligned}
\Pr\left(v \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) &= \bigcap_{k\neq j} \Pr\left(\boldsymbol{u}^{i,k}_{t,b} \leq v\right) \\
&= \prod_{k\neq j} G^{i,k}_{t,b}(v) = \exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_{k\neq j}\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right) = G^{i,-j}_{t,b}(v).
\end{aligned}$$

Integrating $G^{i,-j}_{t,b}(v)$ over all possible values of v we get

$$\begin{aligned}
&\Pr\left(\boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) \\
&= \int_{-\infty}^{\infty} \exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_{k\neq j}\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right) dG^{i,j}_{t,b} \\
&= \int_{-\infty}^{\infty} \frac{1}{\lambda}\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\exp\left(\frac{\overline{V}^j_{t+1,b} - \tau^{i,j}}{\lambda}\right)\exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_{k}\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right) dv \\
&= \frac{\exp\left(\overline{V}^j_{t+1,b} - \tau^{i,j}\right)^{1/\lambda}}{\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}} \int_{-\infty}^{\infty} \frac{1}{\lambda}\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_k\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_{k}\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right) \\
&= \frac{\exp\left(\overline{V}^j_{t+1,b} - \tau^{i,j}\right)^{1/\lambda}}{\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}} \int_{-\infty}^{\infty} dG^{i}_{t,b} = \frac{\exp\left(\overline{V}^j_{t+1,b} - \tau^{i,j}\right)^{1/\lambda}}{\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}} = p^{i,j}_{t,b}.
\end{aligned}$$

Similarly, the probability of a worker of having at most utility v, conditional on having the possibility of changing jobs, is equal to

$$\Pr\left(\max_k \boldsymbol{u}^{i,k}_{t,b} \leq v\right) = \exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_k\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right) = G^{i}_{t,b}(v).$$

Taking the expectation associated with above's probability, we get

$$\begin{aligned}
\mathbb{E}\left(\max_k \boldsymbol{u}^{i,k}_{t,b}\right) &= \int_{-\infty}^{\infty} v\frac{1}{\lambda}\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_k\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma}\right)\sum_k\exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right) \\
&= \int_{-\infty}^{\infty} v\frac{1}{\lambda}\exp\left(-\frac{v}{\lambda} - \overline{\gamma} + \Lambda^i_{t,b}\right)\exp\left(-\exp\left(-\frac{v}{\lambda} - \overline{\gamma} + \Lambda^i_{t,b}\right)\right),
\end{aligned}$$

where

$$\Lambda^i_{t,b} \equiv \log\left(\sum_k \exp\left(\frac{\overline{V}^k_{t+1,b} - \tau^{i,k}}{\lambda}\right)\right).$$

Now, I change variables such that

$$x = \exp\left(-\frac{v}{\lambda} - \overline{\gamma} + \Lambda^i_{t,b}\right), \qquad dx = -\frac{1}{\lambda}\exp\left(-\frac{v}{\lambda} - \overline{\gamma} + \Lambda^i_{t,b}\right) dv.$$

Then,

$$\begin{aligned}
\mathbb{E}\left(\max_k \boldsymbol{u}^{i,k}_{t,b}\right) &= \int_0^\infty \left(-\overline{\gamma} - \log(x) + \Lambda^i_{t,b}\right)\lambda\exp(-x)dx \\
&= \lambda\left(-\overline{\gamma} + \Lambda^i_{t,b}\right)\int_0^\infty \exp(-x)dx - \lambda \underbrace{\int_0^\infty \log(x)\exp(-x)dx}_{=-\overline{\gamma}} \\
&= \lambda\Lambda^i_{t,b} = \lambda\log\left(\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}\right).
\end{aligned}$$

Substituting into the expression for expected lifetime utility we get the expression in the main text

$$V^i_{t,b} = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho\overline{V}^i_{t+1,b} + \beta(1-\rho)\lambda\log\left(\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}\right). \tag{29}$$

1.A.2 Expected efficiency units per migration cohort

To obtain the expected value of the efficiency unit of a worker, conditional on a particular migration decision, we first obtain the following probability function

$$\begin{aligned}
\Pr\left(\exp(\theta^j_{t,\iota}) < v \,\middle|\, \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j}\boldsymbol{u}^{i,k}_{t,b}\right) &= \Pr\left(\theta^j_{t,\iota} \leq \log(a) \,\middle|\, \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\ neqj}\boldsymbol{u}^{i,k}_{t,b}\right) \\
&= \frac{\Pr\left(\left(\theta^j_{t,\iota} < \log(v)\right)\cap\left(\boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j}\boldsymbol{u}^{i,k}_{t,b}\right)\right)}{\Pr\left(\boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j}\boldsymbol{u}^{i,k}_{t,b}\right)} \\
&= \frac{\Pr\left(\left(\boldsymbol{u}^{i,j}_{t,b} < \frac{\log(v)}{1-\beta\rho} + \overline{V}^j_{t+1,b} - \tau^{i,j}\right)\cap\left(\boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j}\boldsymbol{u}^{i,k}_{t,b}\right)\right)}{\Pr\left(\boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j}\boldsymbol{u}^{i,k}_{t,b}\right)}.
\end{aligned}$$

Note that the probability in the denominator is equal to the migration probability $p^{i,j}_{t,b}$. Then, above's expression is equal to

$$\begin{aligned}
\Pr\left(\exp(\theta^j_{t,\iota}) < v \;\middle|\; \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) &= \frac{1}{p^{i,j}_{t,b}} \int_{-\infty}^{\infty} \Pr\left(x < \frac{\log(v)}{1-\beta\rho} + \overline{V}^j_{t+1,b} - \tau^{i,j}\right) \prod_{k\neq j} G^{i,k}_{t,b}(x) dG^{i,j}_{t,b}(x) \\
&= \frac{p^{i,j}_{t,b}}{p^{i,j}_{t,b}} \int_{-\infty}^{\frac{\log(v)}{1-\beta\rho}+\overline{V}^j_{t+1,b}-\tau^{i,j}} dG^i_{t,b}(a) \\
&= \exp\left(-\exp\left(\frac{-\log(v)}{(1-\beta\rho)\lambda} - \left(\frac{\overline{V}^j_{t+1,b}-\tau^{i,j}}{\lambda}\right) - \overline{\gamma}\right) \sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}\right) \\
&= \exp\left(-v^{-1/\delta} \exp(-\overline{\gamma}) \frac{\sum_k \exp\left(\overline{V}^k_{t+1,b}-\tau^{i,k}\right)^{1/\lambda}}{\exp\left(\overline{V}^j_{t+1,b}-\tau^{i,j}\right)^{1/\lambda}}\right) \\
&= \exp\left(-v^{-1/\delta} \frac{\exp(-\overline{\gamma})}{p^{i,j}_{t,b}}\right) = \exp\left(-\left(\frac{v}{\exp(-\delta\overline{\gamma})\left(p^{i,j}_{t,b}\right)^{-\delta}}\right)^{-1/\delta}\right).
\end{aligned}$$

Then the distribution of efficiency units of workers who migrated from location i to j is distributed Fréchet with shape parameter $1/\delta$ and scale parameter $\exp(-\delta\overline{\gamma})\left(p^{i,j}_{t,b}\right)^{-\delta}$. [83] Given this distribution, the expected value is equal to:

$$\mathbb{E}\left(\exp(\theta^j_{t,\iota}) \;\middle|\; \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) = \frac{\Gamma(1-\delta)}{\exp(\gamma\delta)} (p^{i,j}_{t,b})^{-\delta}.$$

In a similar fashion, we can obtain the distribution of log efficiency units, conditional on a migration decision

$$\begin{aligned}
\Pr\left(\theta^j_{t,\iota} < v \;\middle|\; \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) &= \exp\left(-\exp\left(\frac{-v}{(1-\beta\rho)\lambda} - \left(\frac{\overline{V}^k_{t+1,b}-\tau^{i,k}}{\lambda}\right) - \overline{\gamma}\right) \sum_k \exp\left(\overline{V}^k_{t+1,b}-\tau^{i,k}\right)^{1/\lambda}\right) \\
&= \exp\left(-\exp\left(-\frac{v}{\delta} - \log(p^{i,j}_{t,b}) - \overline{\gamma}\right)\right).
\end{aligned}$$

Naturally, the log efficiency units conditional on a migration decision is now distributed Gumbel with scale parameter δ and location parameter $-\delta(\log(p^{i,j}_{t,b} + \overline{\gamma}))$.[84] The expected value of the log efficiency unit, conditional on a worker moving from location i to j is

$$\mathbb{E}\left(\theta^j_{t,\iota} \;\middle|\; \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) = -\delta \log(p^{i,j}_{t,b}).$$

I use this result later for the identification strategy, as the expected log wage of an individual with birthplace b that moved from location i to j is

$$\mathbb{E}\left(\log\left(\text{wage}^j_{t,\iota}\right) \;\middle|\; i \to j\right) = \mathbb{E}\left(\log\left(\text{wage}^j_{t,\iota}\right) \;\middle|\; \boldsymbol{u}^{i,j}_{t,b} \geq \max_{k\neq j} \boldsymbol{u}^{i,k}_{t,b}\right) = \log(w^j_t) - \delta \log(p^{i,j}_{t-1,b}).$$

1.A.3 Static equilibrium under symmetric trade costs and balanced trade

In what follows I will derive the system of equations that will solve for the vector of efficiency wages deflated by the price of tradables in each location given the labor supply distribution. As shown by Allen et al. (2020b),

[83] That the efficiency units is distributed Fréchet is expected as the underlying distribution of the log efficiency units was distributed Gumbel.

[84] Just note that the name for scale and location parameters in the Gumbel distribution will correspond to the shape and scale parameter, respectively, in the Fréchet distribution.

all the results follow under quasi-symmetric trade shocks. In the application trade costs are symmetric, so I don't do it under quasi-symmetric costs to ease notation.

First, given the Cobb-Douglas assumption on the production technology of the tradable good, the share of expenditure on each input is constant and equal to the output elasticity with respect to each input. This means that the total expenditures on housing by the intermediate firms in a location i is proportional to the wage bill

$$P_H^i H_P^i = \frac{\eta}{1-\eta} w^i N^i.$$

Additionally, the unit price of an input bundle for the firm is

$$x^i = \left(\frac{w^i}{1-\eta}\right)^{1-\eta} \left(\frac{P_H^i}{\eta}\right)^{\eta}.$$

Similarly, given the assumption on the utility of workers, the share of expenditures in housing as consumption is constant. Summing the expenditures from workers and firms we have that total expenditures on housing is equal to

$$P_H^i H^i = \frac{\eta + \alpha(1-\eta)}{(1-\eta)} w^i N^i.$$

We can conclude then that the price of housing is proportional to the ratio of total wage bill $w^i N^i$ and housing supply H^i

$$P_H^i \propto \frac{w^i N^i}{H^i},$$

while for the unit price of an input bundle we have

$$x^i \propto \left(w^i\right)^{1-\eta} \left(\frac{w^i N^i}{H^i}\right)^{\eta}. \tag{30}$$

Now passing to the trade part of the model. I assume that the housing owners in i, live in i and spend all their income on tradable goods. Given all the Cobb-Douglas assumptions, the total expenditures of people residing in location i is proportional to the total wage bill.

The share of total expenditure in market j on goods from market i is

$$\pi^{j,i} = \frac{\left(A^i/x^i\right)^{\varphi} \left(\psi^{j,i}\right)^{-\varphi}}{\sum_k \left(A^k/x^k\right)^{\varphi} \left(\psi^{j,k}\right)^{-\varphi}}.$$

We can also have the following expression for the price of tradables in location i

$$\left(P_T^j\right)^{-\varphi} = \boldsymbol{C}^{-\varphi} \sum_k \left(A^k/x^k\right)^{\varphi} \left(\psi^{j,k}\right)^{-\varphi},$$

where $\boldsymbol{C}$ is a constant. Substituting into above's expression we have

$$\pi^{j,i} = \left(A^i/x^i\right)^{\varphi} \left(\psi^{j,i}\right)^{-\varphi} \left(P_T^j\right)^{\varphi} \boldsymbol{C}^{-\varphi}.$$

Define $\tilde{A}^i = A^i \left(H^i\right)^{\eta}$ as a composite of both productivity and housing supply on location i. What this is saying is that, given the cost on efficiency units of labor w^i, the marginal cost can be reduced by having higher productivity A^i or a larger supply of housing. Also define $\tilde{\psi}^{j,i} = \left(\psi^{j,i}\right)^{-\varphi}$. Substituting (30) into the expression for $\pi^{j,i}$ we get

$$\pi^{j,i} = \left(\tilde{A}^i\right)^{\varphi} \left(w^i \left(N^i\right)^{\eta}\right)^{-\varphi} \tilde{\psi}^{j,i} \left(P_T^j\right)^{\varphi} \boldsymbol{C}^{-\varphi}.$$

Using the goods market clearing condition we have that income in location Y^i is equal to

$$P_T^i Y^i = \sum_j \left(\tilde{A}^i\right)^{\varphi} \left(w^i \left(N^i\right)^{\eta}\right)^{-\varphi} \tilde{\psi}^{j,i} \left(P_T^j\right)^{\varphi} \boldsymbol{C} P_T^j Y^j.$$

On the other hand, total expenditures E^i are equal to

$$P_T^i E^i = \sum_j \left(\tilde{A}^j\right)^{\varphi} \left(w^j \left(N^j\right)^{\eta}\right)^{-\varphi} \tilde{\psi}^{i,j} \left(P_T^i\right)^{\varphi} \boldsymbol{C} P_T^i Y^i.$$

Using the assumption that trade is balanced, i.e. $P_T^i Y_i = P_T^i E_i$ we have

$$\sum_j \left(\tilde{A}^j\right)^{\varphi} \left(w^j \left(N^j\right)^{\eta}\right)^{-\varphi} \tilde{\psi}^{i,j} \left(P_T^i\right)^{\varphi} Y^i = \sum_j \left(\tilde{A}^i\right)^{\varphi} \left(w^i \left(N^i\right)^{\eta}\right)^{-\varphi} \tilde{\psi}^{j,i} \left(P_T^j\right)^{\varphi} Y^j.$$

Define the origin and destination fixed effects as follow

$$\mathcal{F}_O^i \equiv \left(\tilde{A}^i\right)^{\varphi} \left(w^i \left(N^i\right)^{\eta}\right)^{-\varphi} \qquad \mathcal{F}_D^j \equiv \left(P_T^j\right)^{\varphi} P_T^j Y^j.$$

Then we can rewrite the balance trade condition as

$$\sum_j \mathcal{F}_O^j \mathcal{F}_D^i \tilde{\psi}^{i,j} = \sum_j \mathcal{F}_O^i \mathcal{F}_D^j \tilde{\psi}^{j,i}.$$

Under the assumption that trade costs are symmetric, i.e. $\tilde{\psi}^{j,i} = \tilde{\psi}^{i,j}$, Allen et al. (2020b), using the Perron-Frobenius theorem, show that the previous expression implies that the destination and origin fixed effects are equal up to a constant, meaning

$$\mathcal{F}_D^i \propto \mathcal{F}_O^i \quad \iff \quad \left(P_T^i\right)^{\varphi} P_T^i Y^i \propto \left(\tilde{A}^i\right)^{\varphi} \left(w^i \left(N^i\right)^{\eta}\right)^{-\varphi}. \tag{31}$$

Define the wage deflated by the price of tradables

$$W^i \equiv \frac{w^i}{P_T^i}. \tag{32}$$

Substituting the expression for the deflated wage into (31) and rearranging, we get an expression of the wage per efficiency unit as a function of the deflated wage, the labor supply and fundamentals, up to a constant

$$w^i \propto \left(\left(W^i\right)^{-\varphi} \left(\tilde{A}^i\right)^{-\varphi} \left(N^i\right)^{1+\eta\varphi}\right)^{-\tilde{\varphi}}, \tag{33}$$

where $\tilde{\varphi} \equiv \frac{1}{1+2\varphi}$ and I used the fact that $Y^i \propto w^i N^i$.

Coming back to the good markets clearing condition, and using the fact that total income is proportional to the wage bill, we get

$$w^i N^i = \sum_j \left(\tilde{A}^i\right)^{\varphi} \left(w^i \left(N^i\right)^{\eta}\right)^{-\varphi} \tilde{\psi}^{j,i} \left(P_T^j\right)^{\varphi} \boldsymbol{C}^{-\varphi} w^j N^j.$$

Substituting for P_T^j using (32) and rearranging we get

$$\left(w^i\right)^{1+\varphi} \left(N^i\right)^{1+\varphi\eta} \left(\tilde{A}^i\right)^{-\varphi} = \boldsymbol{C}^{-\varphi} \sum_j \tilde{\psi}^{j,i} \left(W^j\right)^{-\varphi} \left(w^j\right)^{1+\varphi} N^j.$$

Substituting (33) into above's expression and rearranging we obtain

$$\left(W^i\right)^{\tilde{\varphi}\varphi(1+\varphi)} \left(N^i\right)^{(1+\eta\varphi)(1-\tilde{\varphi}(1+\varphi))} = \sum_j \tilde{\psi}^{j,i} \left(\tilde{A}^i\right)^{\varphi} \left(\frac{\tilde{A}^j}{\tilde{A}^i}\right)^{\varphi\tilde{\varphi}(1+\varphi)} \left(W^j\right)^{\varphi(\tilde{\varphi}(1+\varphi)-1)} \left(N^j\right)^{1-\tilde{\varphi}(1+\varphi)},$$

where we have abstracted from the constant term $\boldsymbol{C}^{-\varphi}$, as we only care about the relative levels of the deflated wages W^i.[85] This is the expression in the main text.

[85] Alternatively, we can think we are solving for scaled wages $W^i \boldsymbol{C}$

1.A.4 Lifetime utility as a function of wages deflated by price of tradables

Given the assumption on the utility function we have that the price of the final good in location i is

$$P^i = \left(\frac{\eta + \alpha(1-\eta)}{(1-\eta)\alpha}\frac{w^i N^i}{H^i}\right)^{\alpha}\left(\frac{P_T^i}{(1-\alpha)}\right)^{1-\alpha}.$$

Substituting into (29) we get

$$V_{t,b}^i = B^i + \tilde{\boldsymbol{C}} + \alpha\log\left(H^i\right) - \alpha\log\left(N_t^i\right) + (1-\alpha)\log\left(\frac{w_t^i}{P_{T,t}^i}\right) - \kappa_b^i + \beta\rho\overline{V}_{t+1,b}^i + \beta(1-\rho)\lambda\log\left(\sum_k \exp\left(\overline{V}_{t+1,b}^k - \tau^{i,k}\right)^{1/\lambda}\right)$$

where $\tilde{\boldsymbol{C}}$ is a constant. Multiplying and dividing $\frac{w_t^i}{P_{T,t}^i}$ by $C^{-(1/\varphi)}$ and defining $\tilde{B}^i \equiv B^i + \tilde{\boldsymbol{C}} + \alpha\log\left(H^i\right) + \frac{1}{\varphi}\log\left(\boldsymbol{C}\right)$ we get the expression in the main text

$$V_{t,b}^i = \tilde{B}^i - \alpha\log\left(N_t^i\right) + (1-\alpha)\log\left(W_t^i\right) - \kappa_b^i + \beta\rho\overline{V}_{t+1,b}^i + \beta(1-\rho)\lambda\log\left(\sum_k \exp\left(\overline{V}_{t+1,b}^k - \tau^{i,k}\right)^{1/\lambda}\right). \tag{34}$$

1.A.5 Steady-state continuous-population economy

In the steady state, productivity levels stay constant. This, in addition to assuming a continuous population mass for each birthplace cohort yields the model deterministic. In particular, we have $V_{t,b}^i = \overline{V}_{t,b}^i$ and the share of workers migrating is equal to the probability. This implies as well that the total amount of efficiency units in location i for workers of birthplace b is

$$N_b^j = \frac{\Gamma(1-\delta)}{\exp(\gamma\delta)}\sum_i (p_b^{i,j})^{1-\delta} L_b^i.$$

Additionally, the la w of motion of labor is equal to

$$L_b^j = \rho L_b^j + (1-\rho)\sum_i p_b^{i,j} L_b^i \quad \Leftrightarrow \quad L_b^j = \sum_i p_b^{i,j} L_b^i.$$

Before setting the whole system of equations that describes the steady state equilibrium, let me define the following variables and parameters in order to simplify notation

$$U_b^i = \exp\left(V_b^i\right), \quad \Omega_b^i = \left(\sum_k \exp\left(V_b^k - \tau^{i,k}\right)^{1/\lambda}\right)^{\lambda}, \quad \mathcal{B}^i = \exp\left(\tilde{B}^i\right)^{1/\delta},$$
$$T^{i,j} = \exp(\tau^{i,j})^{-1/\lambda}, \quad K_b^j = \exp(\kappa_b^j)^{1/\delta}.$$

Then, the model on the steady state with a continuous population is summarized by the following system of equations

$$\left(W^i\right)^{\tilde{\varphi}\varphi(1+\varphi)}\left(N^i\right)^{(1+\eta\varphi)(1-\tilde{\varphi}(1+\varphi))} = \sum_k \tilde{\psi}^{k,i}\left(\tilde{A}^i\right)^{\varphi}\left(\frac{\tilde{A}^k}{\tilde{A}^i}\right)^{\varphi\tilde{\varphi}(1+\varphi)}\left(W^k\right)^{\varphi(\tilde{\varphi}(1+\varphi)-1)}\left(N^k\right)^{1-\tilde{\varphi}(1+\varphi)}, \tag{35}$$

$$\left(U_b^i\right)^{1/\lambda} = \mathcal{B}^i\left(W^i\right)^{\frac{1-\alpha}{\delta}}\left(N^i\right)^{-\alpha/\delta}K_b^i\left(\Omega_b^i\right)^{\frac{\beta(1-\rho)}{\delta}}, \tag{36}$$

$$\left(\Omega_b^i\right)^{1/\lambda} = \sum_k T^{i,k}\left(U_b^k\right)^{1/\lambda}, \tag{37}$$

$$L_b^i\left(U_b^i\right)^{-1/\lambda} = \sum_k T^{i,k}\left(\Omega_b^k\right)^{-1/\lambda}L_b^k, \tag{38}$$

$$N_b^i\left(U_b^i\right)^{\frac{\delta-1}{\lambda}} = \sum_k\left(T^{i,k}\right)^{1-\delta}\left(\Omega_b^k\right)^{\frac{\delta-1}{\lambda}}L_b^k, \tag{39}$$

$$N^i = \sum_b N_b^i, \tag{40}$$

$$L_b = \sum_k L_b^k. \tag{41}$$

Notice that I have not included the constant for average efficiency units $\frac{\Gamma(1-\delta)}{\exp(\gamma\delta)}$ as this will only affect the level of the deflated wages W^i and this won't affect neither the migration decisions, and therefore, the total supply of efficiency units per location.

1.A.6 Comparison of migration and home bias

In this section I derive the compensating variation in consumption such that migration and home bias would be canceled. This allows me to compare them.

First I derive the compensating variation in consumption a migrating worker needs to have to have the same utility as an individual that stayed in the same location. Consider two individuals, indexed $1, 2$, with birthplace b. One just moved to location j from location i while the other remained in the same location. We can then ask how much is the average difference in log efficiency units such that the migrating individual gets the same expected utility as the one that stays.[86] Formally, we need to find $\theta^j_{1,t-1} - \theta^j_{2,t-1}$ such that

$$\mathbb{E}_{\Theta_{t-1}}\left(\mathrm{v}^i_{t,b}(\theta^i_{t-1,1}, \mathbf{\Theta}_{t,1}) - \tau^{i,j} - \mathrm{v}^i_{t,b}(\theta^i_{t-1,2}, \mathbf{\Theta}_{t,2})\right) = 0.$$

This implies

$$\frac{1}{1-\beta\rho}\left(\theta^j_{1,t-1} - \theta^j_{2,t-1}\right) - \tau^{i,j} = 0 \quad\Longleftrightarrow\quad \theta^j_{1,t-1} - \theta^j_{2,t-1} = (1-\beta\rho)\tau^{i,j},$$

where all the location aggregate variables cancel each other and the difference in efficiency units are scaled up by $1/(1-\beta\rho)$ because there is the possibility for the worker of not changing jobs. Given the assumption of log utility in each period, the migrating worker needs to consume $\xi^{i,j}_{\tau} = \exp\left(\tau^{i,j}\right)^{(1-\beta\rho)} - 1$ percentage more than the staying worker in order to have the same utility.

Looking at the home bias, I can do a similar exercise by comparing two staying individuals, where the difference is that one is native to the location they are living, while the other was born somewhere else. There is a caveat, though. While the two individuals live in the same location, when comparing the differences in

[86] Recall that at the time of the migration decision, workers don't know the realization of efficiency units for the subsequent period

utility, all location specific terms will cancel. However, the option value for residing in a particular location is different for individuals with different birthplaces, so we need to adjust for that difference in option values.

Consider then, two individuals, again indexed $1,2$ with birthplace j and b, respectively. Both individuals stayed in their current location b. Assume they have the same utility and that the economy is in the steady state with a continuous population. Then, the difference on expected utilities is given by

$$\mathbb{E}_{\mathbf{\Theta}_{t-1}}\left(\mathrm{v}_{t,j}^{b}(\theta_{t-1,1}^{b},\mathbf{\Theta}_{t,1})-\mathrm{v}_{t,b}^{b}(\theta_{t-1,2}^{b},\mathbf{\Theta}_{t,2})\right)=\frac{1}{1-\beta\rho}\left(\theta_{t-1,1}^{b}-\theta_{t-1,2}^{b}-\kappa_{j}^{b}+\beta(1-\rho)\log\left(\Omega_{j}^{b}-\Omega_{b}^{b}\right)\right)=0$$
$$\Leftrightarrow\left(\theta_{t-1,1}^{b}-\theta_{t-1,2}^{b}\right)=\kappa_{j}^{b}-\beta(1-\rho)\log\left(\Omega_{j}^{b}-\Omega_{b}^{b}\right).$$

Similarly, the excess consumption, in percentage, that a non-native individual needs to have in order to have the same utility as a native would be given by

$$\xi_{\kappa,j}^{b}=\exp\left(\kappa_{j}^{b}-\beta(1-\rho)\log\left(\Omega_{j}^{b}-\Omega_{b}^{b}\right)\right)-1.$$

I can also consider a lower bound of excess consumption by not adjusting the differences in option values, i.e. $\tilde{\xi}_{\kappa,j}^{b}=\exp\left(\kappa_{j}^{b}\right)-1$. It is a lower bound as the option value of the native is, in general, larger than the non-native.[87] The benefit of this approach is that I do not require to assume the economy is in steady state. Additionally, I do not require to solve for the model in order to compute the lower bound as it is only a function of the estimated home bias.

1.A.7 Welfare derivations and Birthplace Premium

In this section I derive the expressions to compare welfare along the different counterfactual scenarios. I will assume only the steady state/continuous population case. This section follows closely Caliendo et al. (2019).

First, we can rewrite the expected lifetime utility net of current efficiency units V_{b}^{i} as

$$V_{b}^{i}=B^{i}+\log\left(C^{i}\right)-\kappa_{b}^{i}+\beta V_{b}^{i}+\beta(1-\rho)\lambda\log\left(\sum_{k}\exp\left(V_{b}^{k}-V_{b}^{i}-\tau^{i,k}\right)^{1/\lambda}\right),$$

where C^{i} is the real consumption that can be obtained with a unit of efficiency wage. Recall that the probability of choosing to stay within the same location, conditional on changing jobs, is equal to

$$p_{b}^{i,i}=\frac{\exp\left(V_{b}^{i}\right)^{1/\lambda}}{\sum_{k}\exp\left(V_{b}^{k}-\tau^{i,k}\right)^{1/\lambda}},$$

and therefore

$$\lambda\log\left(\sum_{k}\exp\left(V_{b}^{k}-V_{b}^{i}-\tau^{i,k}\right)^{1/\lambda}\right)=-\lambda\log p_{b}^{i,i}.$$

Substituting into the value function and rearranging, we get

$$V_{b}^{i}=\frac{1}{1-\beta}\left(B^{i}+\log\left(C^{i}\right)-\kappa_{b}^{i}-\beta(1-\rho)\lambda\log p_{b}^{i,i}\right).$$

However, V_{b}^{i} is the average lifetime-utility *net* of current efficiency units. To get the actual average welfare we need to take into account the heterogeneity in efficiency units. Recall that the migration decision is made at the end of every period and efficiency shocks are independent across period. Thus, I abstract from the vector of location specific efficiency shocks that govern the migration decision of the subsequent period.

[87] Exceptions might occur if for some combinations of locations the home bias is actually negative. This would mean that some natives have utility from *leaving* their birthplace. This would be extremely rare in the data.

In other words, I will look at the expected lifetime utility *after* a migration decision is made, but *before* the realization of the next period shocks. With some abuse of notation we can rewrite the expected lifetime utility of an individual with log efficiency θ^i that just decided to move from location j to i as

$$\mathrm{v}_b^i(\theta^i) = B^i + \log\left(\frac{w^i}{P^i}\right) - \kappa_b^i + \theta^i + \beta\rho\mathrm{v}_b^i(\theta^i) + \beta(1-\rho)\lambda\log\left(\sum_k \exp\left(V_b^k - \tau^{i,k}\right)^{1/\lambda}\right) = V_b^i + \frac{\theta^i}{1-\beta\rho}.$$

The previous expression adjusts the expected net lifetime utility V_b^i with two terms. The first one $\frac{\theta^i}{1-\beta\rho}$ corresponds to the net present value of log efficiency unit that a worker can get by moving to the current location. the second term just adds the migration cost as I am looking at the utility at the moment of the migration decision. The current workers who moved to location i from any other location constitute a fraction $(1-\rho)$ of the total workers with birthplace b that live in i. Now consider the workers who move to location i from location j. Their average utility is given by

$$\mathbb{E}\left(\mathrm{v}_b^i(\theta^i)\;\middle|\; j \to i\right) = V_b^i - \frac{\delta}{1-\beta\rho}\log p_b^{j,i} = V_b^i - \lambda \log p_b^{j,i}.$$

So the average utility is larger the smaller the migration flow $p_b^{j,i}$ as this would indicate only individuals with high efficiency units moved to i.

From the fraction of workers ρ that could not moved, there is a fraction $(1-\rho)$ that moved to i from j *two* periods ago. So their expected utility is the same as above. We can do the same reasoning for the previous periods. Aggregating all the workers who have moved from j to i in any period we have that their total utility is

$$(1-\rho)\sum_{s=0}^{\infty}\rho^s\left[V_b^i - \lambda\log p_b^{j,i}\right]p_b^{j,i}L_b^j = \left(V_b^i - \lambda\log p_b^{j,i}\right)p_b^{j,i}L_b^j.$$

The average utility of workers with birthplace b that live in location i is

$$\tilde{V}_b^i = \frac{1}{L_b^i}\sum_j\left(V_b^i - \lambda\log p_b^{j,i}\right)p_b^{j,i}L_b^j = V_b^i - \frac{\lambda}{L_b^i}\sum_j \log\left(p_b^{j,i}\right)p_b^{j,i}L_b^j.$$

So the average utility per birthplace cohort is

$$\tilde{V}_b = \sum_i \frac{L_b^i}{L_b}\tilde{V}_b^i,$$

and the average utility of the whole population is

$$\tilde{V} = \sum_b \frac{L_b}{L}\tilde{V}_b = \sum_b\sum_i \frac{L_b^i}{L}\tilde{V}_b^i.$$

The birthplace premium, denoted ζ_b, is defined as the compensating variation in consumption such that

$$\sum_i\left[\frac{1}{1-\beta}\left(B^i + \log\left(C^i\left(1-\varepsilon_b\right)\right) - \kappa_b^i - \beta(1-\rho)\lambda\log p_b^{i,i}\right)\frac{L_b^i}{L_b} - \frac{\lambda}{L_b}\sum_j \log\left(p_b^{j,i}\right)p_b^{j,i}L_b^j\right] = \tilde{V}$$

$$\Leftrightarrow \quad \log\left(1-\varepsilon_b\right) = (1-\beta)\left(\tilde{V} - \tilde{V}_b\right) \quad \Leftrightarrow \quad \varepsilon_b = 1 - \exp\left(\tilde{V} - \tilde{V}_b\right)^{1-\beta}.$$

1.B Solution algorithm

In this section I explain with more detail the algorithm to solve the model. Before doing so, I will present the following theorem, which is a special case of Theorem 1 in Allen et al. (2020a).

Theorem 1. *Consider the following system of $N \times K$ system of equations*

$$\prod_{h=1}^{K}\left(x_i^h\right)^{\beta_{kh}} = \sum_{j=1}^{K} K_{ij}^k \left[\prod_{h=1}^{H}\left(x_j^h\right)^{\gamma_{kh}}\right]$$

where $\{\beta_{kh}, \gamma_{kh}\}$ are known elasticities and $\left\{K_{ij}^k > 0\right\}$ are positive kernels related to bilateral frictions. Let $\mathbf{B} \equiv [\beta_{kh}]$ and $\mathbf{\Gamma} \equiv [\gamma_{kh}]$ be the $K \times K$ matrices of the known elasticities. Define $\mathbf{A} \equiv \mathbf{\Gamma}\mathbf{B}^{-1}$ and the absolute value (element by element) of $\mathbf{A}$ as $\mathbf{A}^p$. If the spectral radius of $\mathbf{A}^p$ (i.e. the absolute value of the largest eigenvalue, denoted $\tilde{\rho}(\cdot)$) is strictly smaller than one, i.e. $\tilde{\rho}\left(\mathbf{A}^p\right) < 1$, there exists a unique strictly positive solution to the above's system. Moreover, the unique solution can be computed by a simple iterative procedure. If $\tilde{\rho}\left(\mathbf{A}^p\right) = 1$, then the solution is unique up to a constant.

Theorem 1 gives us the conditions to apply a multidimensional contraction mapping to solve the model. I use this result to solve efficiently parts of the model and for parts of the identification, as detailed in section 1.C.2. Note than in the case of $K = 1$ the result is just an application of a standard contraction mapping.

I will describe how the sequence of the algorithm goes. Lets first assume a vector of total efficiency units per location $\{N^{i,(0)}\}_{i\in\mathcal{I}}$. Then, using Theorem 1, we can use the system of equations characterized by (35) to solve for the vector of deflated wages $\{W^i\}_{i\in\mathcal{I}}$, conditional on a vector of efficiency units supply per location $\{N^i\}_{i\in\mathcal{I}}$ using an iterative method as[88]

$$\left|\frac{\varphi(\tilde{\varphi}(1+\varphi)-1)}{\tilde{\varphi}\varphi(1+\varphi)}\right| = \frac{\varphi}{1+\varphi} < 1.$$

Substituting (36) into (37) we get

$$\left(\Omega_b^i\right)^{1/\lambda} = \sum_k T^{i,k}\mathcal{B}^k\left(W^k\right)^{\frac{1-\alpha}{\delta}}\left(N^k\right)^{-\alpha/\delta} K_b^k\left(\Omega_b^k\right)^{\frac{\beta(1-\rho)}{\delta}}.$$

Similarly, conditioning on wages and total efficiency units per location, we can use Theorem 1 to find the vector of option values $\{\Omega_b^i\}_{i\in\mathcal{I}}$ that solve the system above using an iterative method as

$$\beta < 1 \quad \text{and} \quad \rho \leq 1 \quad \implies \quad 0 < \frac{\lambda\beta(1-\rho)}{\delta} = \frac{\beta(1-\rho)}{1-\beta\rho} < 1.$$

Now, given a vector of total efficiency units per location, deflated wages and option values, we can characterize the overall welfare $\{U_b^i\}_{i\in\mathcal{I}}$. This in turn let us characterize all the migration probabilities (and thus shares as there is a continuum of workers) $p_b^{i,j}$. Define the migration matrix for workers with birthplace b as $\mathbf{P}_b$, where $(\mathbf{P}_b)_{(i,j)} = p_b^{i,j}$. Note that $\mathbf{P}_b$ is a stochastic matrix (i.e. the sum of each row is equal to 1). Define $\mathbf{L}_b$ as the vector of length I where the ith element is equal to L_b^i. Then, the system of equations characterized by 38 can be rewritten as

$$\mathbf{L}_b = \mathbf{P}_b'\mathbf{L}_b. \tag{42}$$

Note that system of equations (42) is an eigensystem and the vector $\mathbf{L}_b$ is equal, up to a constant, to the eigenvector associated to the unit eigenvalue. As $\mathbf{P}_b$ is a stochastic matrix, then the largest eigenvalue of $\mathbf{P}_b'$ is equal to one. By the Perron-Frobenius theorem, this eigenvalue is unique and there is a unique positive

[88] Recall $\tilde{\varphi} = 1/(1+2\varphi)$.

eigenvector that is associated to that eigenvalue. So to solve the previous system I find the eigenvector associated to the largest eigenvalue, which is a very fast procedure.[89] In order to pin down the level of employment for each birthplace, I use equation (41).

Finally, given the vector of labor allocation and the migration probabilities $p_b^{i,j}$, I can compute the total efficiency units per location $\{N^{i,(1)}\}_{i\in\mathcal{I}}$ using equations (36), (39) and (40). The solution is found when $\{N^{i,(0)}\}_{i\in\mathcal{I}} = \{N^{i,(1)}\}_{i\in\mathcal{I}}$.

The strategy to solve for the model is summarized in the algorithm below.

Algorithm 1 Model Solution

1: Initiate with a guess $\{N^{i,(0)}\}_{i\in\mathcal{I}}$.
2: Initiate $\{W^{i,(0)}\}_{i\in\mathcal{I}}$ and $d_W > \text{tol}_W$.
3: **while** $d_W > \text{tol}_W$ **do**
4: Get $W^{i,(1)}$ with

$$W^{i,(1)} = \left(\left(\frac{\left(\tilde{A}^i\right)^{\varphi}}{\left(N^{i,(0)}\right)^{(1+\eta\varphi)}} \right)^{1-\tilde{\varphi}(1+\varphi)} \sum_k \tilde{\psi}^{k,i} \left(\tilde{A}^k\right)^{\varphi\tilde{\varphi}(1+\varphi)} \left(W^{k,(0)}\right)^{\varphi(\tilde{\varphi}(1+\varphi)-1)} \left(N^{k,(0)}\right)^{1-\tilde{\varphi}(1+\varphi)} \right)^{1/\tilde{\varphi}\varphi(1+\varphi)}$$

5: $d_W = \left\| \{W^{i,(0)}\}_{i\in\mathcal{I}}, \{W^{i,(1)}\}_{i\in\mathcal{I}} \right\|_\infty$.
6: $W^{i,(1)} \to W^{i,(0)}$.
7: **end while**
8: **for** $b \in \mathcal{I}$ **do**
9: Initiate $\{\Omega_b^{i,(0)}\}_{i\in\mathcal{I}}$ and $d_\Omega > \text{tol}_\Omega$.
10: **while** $d_\Omega > \text{tol}_\Omega$ **do**
11: Get $\Omega_b^{i,(1)}$ with

$$\Omega_b^{i,(1)} = \left(\sum_k T^{i,k} \mathcal{B}^k \left(W^{k,(1)}\right)^{\frac{1-\alpha}{\delta}} \left(N^{k,(0)}\right)^{-\alpha/\delta} K_b^k \left(\Omega_b^{k,(0)}\right)^{\frac{\beta(1-\rho)}{\delta}} \right)^{\lambda}.$$

12: $d_\Omega = \left\| \{\Omega_b^{i,(0)}\}_{i\in\mathcal{I}}, \{\Omega_b^{i,(1)}\}_{i\in\mathcal{I}} \right\|_\infty$.
13: $\Omega_b^{i,(1)} \to \Omega_b^{i,(0)}$.
14: **end while**
15: Form matrix $\mathbf{P}_b$ with $(\mathbf{P}_b)_{(i,j)} = p_b^{i,j}$.
16: Find eigenvector v_b associated with the unit eigenvalue of $\mathbf{P}_b'$
17: Get $L_b^i = \frac{v_{b,(i)}}{\sum_k v_{b,(k)}} L_b$.
18: Get $N_b^i = \sum_k \left(p_b^{k,i}\right)^{1-\delta} L_b^k$.
19: **end for**
20: Get $N^{i,(1)} = \sum_b N_b^i$ for all $i \in \mathcal{I}$.
21: Check if $\{N^{i,(1)}\}_{i\in\mathcal{I}} \simeq \{N^{i,(0)}\}_{i\in\mathcal{I}}$. If not, go back to step 1 and update $\{N^{i,(0)}\}_{i\in\mathcal{I}}$

[89] Borrowed this idea from Eckert (2019).

1.C Identification details

In this section I discuss some of the details of the identification strategy. Some of the parameters I calibrate them externally, so in this section I take them as given. In particular these are the parameters concerning the static equilibrium part of the model plus the discount factor.

The observed migration share is the frequency estimator of the conditional migration probability. It is an unbiased estimator, so I can exploit this fact plus the closed form expression of the conditional migrating probabilities to form some moment conditions. These moment conditions would correspond to the first order conditions of a Poisson regression, or Poisson PseudoMaximum Likelihood (PPML).

The conditional migration probability can be rewritten as an expression that depends on a destination/period/birthplace and origin/period/birthplace specific fixed effects as well as the migration cost.

In this Appendix I provide sufficient conditions for identification of these migration costs. In the main text I give an intuitive explanation with an example. Here I give a more formal treatment on the matter.

I don't enter into details about the identification of the migration elasticity as it is already treated in the main text.

Using the closed form expression of the conditional migration probabilities to form the conditional likelihood function of observing the labor flows in the data. Taking the identified migration costs as given, I can estimate the location/birthplace/period specific expected utilities by treating them as fixed effects of the likelihood function. This allows me to obtain estimates of the conditional migration probabilities to be used later on the identification of the home bias.

In this Appendix I also show the details to obtain expression (1.19) of Proposition 3 from the first order conditions of the maximum likelihood problem. I prove that the solution for such system of equations is unique. While this system of equations is ubiquitous in the migration and trade literature and the uniqueness of the solution has been proved, at least from Ahlfeldt et al. (2015), I show the flexibility of Theorem 1 by offering and alternative and shorter proof. The details on the computational algorithm to solve for the system are left for Appendix 1.D.

I will abstract from the discussion of the persistence parameter and the distribution of fundamentals as they are already explained in the main text. I give more details on the identification of the migration costs, the migration elasticity and the home bias.

1.C.1 Migration Costs

Using labor flows $\ell^{i,j}_{t,b}$ or migration shares $s^{i,j}_{t,b}$ does not matter for the identification arguments. Then, I present the results of this section using the migration shares as there is less notation.

The following defines the graph for a particular year/birthplace

Definition 3. *Let $\mathcal{I}$ denote the set of all locations. Then, the graph for year t and birthplace b is an ordered pair $\mathcal{G}_{t,b} = (\mathcal{I}, \mathcal{E}_{t,b})$, where $\mathcal{E}_{t,b} = \{(i,j) | (i,j) \in \mathcal{I}^2 \text{ and } s^{i,j}_{t,b} > 0\}$.*

Note that in the above definition, I am allowing for the graph to have loops, meaning I allow for edges to have the same *destination* as the origin. Also note that $\mathcal{G}_{t,b}$ is defined as a *directed* graph, which means that each edge has an orientation. This is in contrast to *undirected* graphs where edges only show association between nodes.

It is useful to have a formal definition of the location pairs who fulfill the conditions of Proposition 1.

Definition 4. *For birthplace cohort b and period t, let data of labor flows from an unordered pair of locations (i,j) in $\mathcal{I}^2$ fulfill the conditions of Proposition 1. Then I say that (i,j) is **directly identified** on $\mathcal{G}_{t,b}$.*

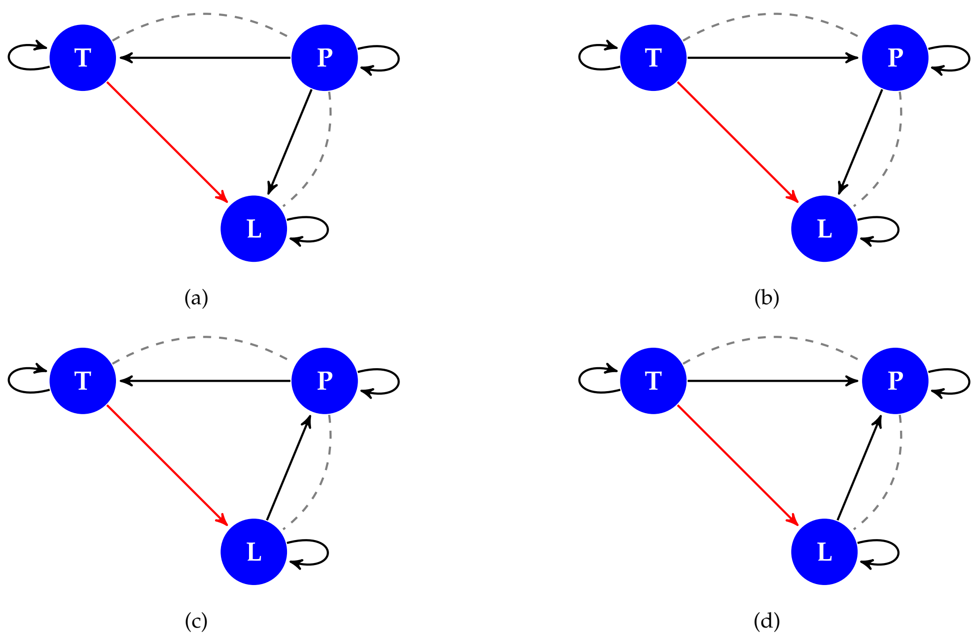


Figure 12 – Different cases for weakly connectivity between T and L through P.

I can define a graph collecting all the directly identified pairs for each birthplace cohort and year

Definition 5. *Denote all the set of directly identified pairs on $\mathcal{G}_{t,b}$ as $\mathcal{E}^d_{t,b}$. Then, the graph of directly identified pairs is the sub-graph $\mathcal{G}^d_{t,b} \subseteq \mathcal{G}_{t,b}$, $\mathcal{G}^d_{t,b} = (\mathcal{I}, \mathcal{E}^d_{t,b})$.*

In contrast to the graph $\mathcal{G}_{t,b}$, the sub-graph of directly identified pairs $\mathcal{G}^d_{t,b}$ is an *undirected* graph, i.e., edges have no orientation. Also it has no loops.

I can collect all the different directly identified pairs for every birthplace and year on a single graph.

Definition 6. *The graph of directly identified pairs is defined as $\mathcal{G}^d = \bigcup_{t,b} \mathcal{G}^d_{t,b}$.*

This graph summarizes all the pair combinations where, for some birthplace/year, the data satisfies the conditions of Proposition 1.

A path from i_1 to i_N on a graph is a sequence of nodes $(i_1, i_2, ..., i_N)$ where an edge connects every subsequent pair of nodes (i_n, i_{n+1}) and all nodes are distinct. Two nodes are *weakly connected* in $\mathcal{G}_{t,b}$ if there exists a path connecting i and j. Denote the set of all paths from i to j on graph $\mathcal{G}_{t,b}$ as $\mathcal{P}(i,j)$. Similarly, the set of all the paths from i to j on the graph of directly identified pairs $\mathcal{G}^d$ is defined as $\mathcal{P}^d(i,j)$. Now I can state the following result, which gives sufficient conditions for identification of the migration cost between two pairs of locations that are not directly identified.

Proposition 7. *The migration cost $\tau^{i,j}/\lambda < \infty$ is identified if, for some graph $\mathcal{G}_{t,b}$*

1. *$s^{i,j}_{t,b} > 0$ or $s^{j,i}_{t,b} > 0$, i.e. there is an edge connecting i and j in $\mathcal{G}_{t,b}$.*

2. *i and j are weakly connected in $\mathcal{G}_{t,b} \cap \mathcal{G}^d$, i.e. $\mathcal{P}_{t,b}(i,j) \cap \mathcal{P}^d(i,j) \neq \emptyset$.*

3. *For some $\mathcal{P}^d_{t,b} \in \{\mathcal{P}_{t,b}(i,j) \cap \mathcal{P}^d(i,j)\}$, $s^{k,k}_{t,b} > 0$ for all k in $\mathcal{P}^d_{t,b}$.*

Proof. I restrict the proof to the three location example as it is without loss of generality. The proof proceeds by looking at all the possible cases of directed graphs for three locations that fulfill the proposition's conditions. Consider then three locations, which are T, P and L. Let T-P and P-L be directly identified. Fixing the flow from T to L, there are four possible cases where T and L are weakly connected through P. There are all represented in Figures 12a to 12d. The reasoning when the flow is from L to T is symmetrical. I then check how each of the four cases identifies the migration cost between T and L. I abstract from period and birthplace subindices to keep notation simple.

(a) The first case is represented in Figure 12a. Then, $\frac{p^{T,L}}{p^{T,T}}\frac{p^{P,T}}{p^{P,L}} = \exp(-\tau^{T,L}/\lambda - \tau^{P,T}/\lambda + \tau^{P,L}\lambda)$.

(b) The second case is represented in Figure 12b. Then, $\frac{p^{T,L}}{p^{T,P}}\frac{p^{P,P}}{p^{P,L}} = \exp(-\tau^{T,L}/\lambda + \tau^{P,T}/\lambda + \tau^{P,L}\lambda)$.

(c) The third case is represented in Figure 12c. Then, $\frac{p^{T,L}}{p^{T,T}}\frac{p^{P,T}}{p^{P,P}}\frac{p^{L,P}}{p^{L,L}} = \exp(-\tau^{T,L}/\lambda - \tau^{P,T}/\lambda - \tau^{P,L}\lambda)$.

(d) The fourth case is represented in Figure 12d. Then, $\frac{p^{T,L}}{p^{T,P}}\frac{p^{L,P}}{p^{L,L}} = \exp(-\tau^{T,L}/\lambda + \tau^{P,T}/\lambda - \tau^{P,L}\lambda)$.

As the migration costs from P to T and P to L are identified, then in each of the previous cases, the migration cost from T to L is also identified. For the case where the flow is from L to T, is just the same four cases above, just interchanging L for T and viceversa. □

The argument for identification contained in Proposition 7 is recursive. Starting from the directly identified pairs, I can do one iteration with above's argument and check which pairs are identified. I can then use all the pairs that are identified and do another iteration to check with new pairs are identified and so on, and so on.

Algorithm to find identified pairs

Whether a particular graph fulfills the conditions spelled out in Proposition 7 is conceptually simple, the development of a recursive algorithm that uses the previously identified migration costs to show identification of other pairs might be somehow trickier–although not very difficult.

Before jumping into the description of the algorithm, let me present some basic concepts on graph theory and some notation. For any graph with N nodes, re-brand each node such that it corresponds to an integer from 1 to N. An adjacency matrix of a graph, is a matrix where the entry $(i, j) = 1$ if there is an edge connecting i to j. Undirected graphs have corresponding symmetric adjacency matrices. Normally, graphs have no loops, so the diagonal entries on the adjacency matrix are zero. For my context, I allow the graphs to have loops, so some diagonal entries would be equal to one. Also, denote as "$*$" the element-wise multiplication operator. Algorithm 2 presents the steps for finding recursively which pairs are identified according to Proposition 7. It consists of two parts. First, it finds all the directly identified pairs. Then, it proposes a recursive algorithm to find the identified pairs given the information of previously found pairs.

Algorithm 2 Find pairs that fulfill Identification Conditions

Let $\mathcal{A}_{t,b}$ be the corresponding adjacency matrix of graph $\mathcal{G}_{t,b}$.
$\mathcal{A}^{d}_{t,b} \leftarrow \mathcal{A}_{t,b}$, for all t, b.
for all t, b **do**
    if $\left(\mathcal{A}^{d}_{t,b}\right)_{(i,i)} = 0$ **then**
        $\left(\mathcal{A}^{d}_{t,b}\right)_{(i,j)} = 0$ for all j.
        $\left(\mathcal{A}^{d}_{t,b}\right)_{(j,i)} = 0$ for all j.
    end if
end for
$\mathcal{B}^{d}_{t,b} \leftarrow \mathcal{A}^{d}_{t,b} * \left(\mathcal{A}^{d}_{t,b}\right)'$, for all t, b.
$\tilde{M}_0 \leftarrow \sum_{t,b} \mathcal{B}^{d}_{t,b}$.
$\tilde{M}_0 \leftarrow \tilde{M}_0 > 0$.
$M_0 \leftarrow \tilde{M}_0$. diff$= 1$
while diff$\neq 0$ **do**
    for all t, b **do**
        $\mathcal{A}^{wc}_{t,b} \leftarrow \mathcal{A}^{d}_{t,b} + \left(\mathcal{A}^{d}_{t,b}\right)'$.
        $\mathcal{A}^{wc}_{t,b} \leftarrow \mathcal{A}^{wc}_{t,b} > 0$.
        $\mathcal{A}^{aux}_{t,b} \leftarrow \mathcal{A}^{wc}_{t,b} * M_0$.
        Form the path matrix of $\mathcal{A}^{aux}_{t,b}$, denoted $\mathcal{P}^{aux}_{t,b}$, where entry $(i, j) = 1$ if there is a path between i and j; equal to zero otherwise.
        $M_1 \leftarrow \mathcal{P}^{aux}_{t,b} * \mathcal{A}_{t,b} + M_0$.
        $M_1 \leftarrow M_1 > 0$.
        $M_0 \leftarrow M_1$.
    end for
    $\tilde{M}_1 \leftarrow M_0$.
    diff $= \left\|\tilde{M}_1, \tilde{M}_0\right\|_{\infty}$.
    $\tilde{M}_0 \leftarrow \tilde{M}_1$.
end while

1.C.2 Conditional migration probabilities

The first part of this section I give the proof for Proposition 3. From the maximization of the log-likelihood 1.18, the first order condition with respect to $\mathcal{D}^j_{t+1,b}$ is

$$\sum_i \ell^{i,j}_{t,b} \frac{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)}{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)} \left(\frac{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda) \sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda) - \left(\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)\right)^2}{\left(\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)\right)^2} \right)$$

$$- \sum_i \sum_{h \neq j} \ell^{i,h}_{t,b} \frac{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)}{\exp(\mathcal{D}^h_{t+1,b} - \tau^{i,h}/\lambda)} \left(\frac{\exp(\mathcal{D}^h_{t+1,b} - \tau^{i,h}/\lambda) \exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)}{\left(\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)\right)^2} \right) = 0,$$

$$\iff \quad \sum_i \ell^{i,j}_{t,b} \left(1 - \frac{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)}\right) - \sum_i \sum_{h \neq j} \ell^{i,h}_{t,b} \frac{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)} = 0$$

$$\iff \quad \sum_i \ell^{i,j}_{t,b} = \sum_i \frac{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)} \sum_h \ell^{i,h}_{t,b} \quad \iff \quad L^{j,\text{dest}}_{t,b} = \sum_i \frac{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)} L^{i,\text{orig}}_{t,b}.$$

Note that the expression above depends on the fixed effects of a particular period and birthplace class. Identification boils down to prove uniqueness of the system formed by the first order conditions of the fixed effects of particular year and birth cohort. To do so, I introduce the following useful Lemma

Lemma 1. *Consider a mapping of the form:*

$$A_n = \sum_{m \in \mathcal{N}} \frac{\omega_n K_{m,n}}{\sum_{n' \in \mathcal{N}} \omega_{n'} K_{m,n'}} B_m \quad \forall n \in \mathcal{N}.$$

For any strictly positive vectors $\{A_n\}$ $\{B_n\}$, where $\sum_n A_n = \sum_n B_n$ and any strictly positive matrix $\mathbf{K}$, where entries $(\mathbf{K})_{(m,n)} = K_{m,n}$, there exists a unique (to scale), strictly positive vector $\{\omega_n\}$.

Proof. The proof for the Lemma is just a corollary of the proofs of Lemmas A.6 and A.7 in the appendix of Ahlfeldt et al. (2015). However, to show the flexibility of Theorem 1, I propose an alternative, and I think easier, proof. To see this, first rewrite above's expression as

$$(\omega_n)^{-1} = \sum_{m \in \mathcal{N}} \frac{B_m}{A_n} K_{m,n} (z_m)^{-1}$$

$$z_m = \sum_{n' \in \mathcal{N}} K_{m,n'} \omega_{n'}.$$

Using the same notation as in Theorem 1, notice then that $\mathbf{A} = \Gamma \mathbf{B}^{-1}$ is

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then, the spectral radius of $\mathbf{A}^p$ is $\tilde{\rho}(\mathbf{A}^p) = 1$. Thus, by Theorem 1, there exists a unique up to scale solution to the system above. $\square$

The system of equations formed by the first order conditions falls into the class of systems where Lemma 1 applies. Thus, the fixed effects are identified up to a constant.

1.C.3 Prices of non-housing goods and Productivities/Housing

In order to get the prices of non-housing goods across locations, I need to solve for the source effects, defined as $S^i = \left(A^i/x^i\right)^{\varphi}$.

The system formed by equalizing total expenditures with total income takes the form of the equation in Lemma 1. Therefore, given the efficiency wages, the observed wage bill, the trade elasticity φ, the trade costs $\psi^{j,k}$ and the output elasticity η, I can identify, up to a constant, the source effects.

The price index of the non-housing goods can then be written in terms of the source effects as

$$P_T^j = \boldsymbol{C}^{-1}\left(\sum_k S^k \left(\psi^{j,k}\right)^{-\varphi}\right)^{-1/\varphi},$$

where $\boldsymbol{C}$ is a constant. The fact that the price of non-housing goods can be expressed directly as a function of the source effects is a consequence that the price (and the source effects) are directly a function of the price of an input bundle. This means that we could extend the model adding an arbitrary number of flexible inputs, like, say, capital, and the identified source effect would not change. However, the identified underlying fundamentals would differ.

Having identified the source effects, I can use the identified efficiency wage plus the observed wage bill to back out a composite of both productivity and housing supply in location i. Recall that the price of an input bundle is

$$x^i \propto \left(w^i\right)^{1-\eta}\left(\frac{w^i N^i}{H^i}\right)^{\eta}.$$

Thus, the source effect is equal to

$$S^i \propto \left(A^i\right)^{\varphi}\left(\left(w^i\right)^{1-\eta}\left(\frac{w^i N^i}{H^i}\right)^{\eta}\right)^{-\varphi} \quad \Longleftrightarrow \quad \tilde{A}^i \equiv A^i\left(H^i\right)^{\eta} \propto \left(S^i\right)^{1/\varphi}\left(w^i\right)^{1-\eta}\left(w^i N^i\right)^{\eta},$$

so the measure of competitiveness $\tilde{A}^i$ is identified up to a constant.

1.D Estimation details

1.D.1 Wage dispersion parameter δ

The wage of a worker can be influenced by idiosyncratic factors that are constant across locations. To control for these, I first run a regression, for each year, of the logarithm of wage on a quadratic polynomial on age and a gender dummy, for the entire sample –switchers and non-switchers alike. From this regression, I collect the residuals and keep only those of the workers who switch jobs across two subsequent years. I use the residual for the job switchers as my dependent variable to estimate δ.

Table 11, contains the estimates for both the OLS and IV estimates.

1.D.2 Conditional migration probabilities

Estimating via maximum likelihood the destination fixed effects on (1.18) boils down to solving different systems of equations for each birth cohort/period

$$L_{t,b}^{j,\text{dest}} = \sum_i \frac{\exp(\mathcal{D}_{t+1,b}^{j} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}_{t+1,b}^{k} - \tau^{i,k}/\lambda)} L_{t,b}^{i,\text{orig}}, \quad \text{for all } i,j,h \in \mathcal{I}. \tag{43}$$

Each of the systems correspond to 73 non-linear equations. While each of these non-linear systems are fairly large, the main complication is that the total number of systems to solve is way larger. In total there are $14 \times 73 = 1022$ systems to solve.

Although the separability of the problem is an advantage in computational terms, as it allows to easily parallelize the algorithm, the number of systems is still too large for traditional algorithms to be a feasible option. Trade economists have encountered this type of systems on gravity-type models, especially when they want to *invert* the model: compute fundamentals such that the observed data matches the equilibrium behavior of the model. Luckily, they have also developed tools for the quick and efficient solution of this type of systems. In particular, Pérez-Cervantes (2014) develops a stable and fast algorithm that deals with this type of systems. While each of the systems in my application is not very large, with only 73 unknowns, his algorithm is also well suited for high-dimensional systems of equations.[90]

Before, jumping to the description of the algorithm, note that (43) implies that, as long as there is someone in the destination location the estimated fixed effect has to be strictly positive. This would mean that the estimated conditional probability is also positive. To avoid problems with zeros in the estimated probabilities in the next steps of the estimation procedure, I impute a value of one worker, which stayed in the same location, for all those locations that did not have a worker of a particular birthplace in a given period. The number of observations modified corresponds to a little more than the 1% of the total number of locations/birthplace/periods.

Algorithm 3 describes the steps to solve for each system of equations for a particular birthplace b and period t. I don't include the time and birthplace subscripts to ease notation.

Algorithm 3 Pérez-Cervantes (2014) Solution Algorithm

1: Initiate with a guess $\{D^{i,(0)}\}_{i\in\mathcal{I}}$. A good guess is $\frac{L^{i,\text{dest}}}{\sum_j L^{j,\text{dest}}}$.
2: Pick adjustment parameter $\eta_{adj} < 1$.
3: Initiate $d_D > \text{tol}_D$.
4: **while** $d_D > \text{tol}_D$ **do**
5: Form the matrix Π, with entries $\Pi_{(i,j)} = \frac{\exp(\mathcal{D}^{j,(0)} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}^{k,(0)} - \tau^{i,k}/\lambda)}$.
6: Get

$$\mathbf{D}^{(1)} = \mathbf{D}^{(0)} + \eta_{adj}\left[L^{\text{dest}} - \Pi' L^{\text{orig}}\right],$$

where $\mathbf{D}^{(\text{iter})} = [D^{1,(\text{iter})}, ..., D^{I,(\text{iter})}]'$, iter $\in \{0,1\}$; $L^{\text{dest}} = [L^{1,\text{dest}}, ..., L^{I,\text{dest}}]'$ and $L^{\text{orig}} = [L^{1,\text{orig}}, ..., L^{I,\text{orig}}]'$.
7: **if** $\min_i \mathbf{D}^{(1)} < 0$ **then**
8: Adjust $\eta_{adj} \leftarrow c \times \eta_{adj}$, $c < 1$.
9: Go back to step 6.
10: **end if**
11: $d_D = \left\| \{D^{i,(0)}\}_{i\in\mathcal{I}}, \{D^{i,(1)}\}_{i\in\mathcal{I}} \right\|_\infty$.
12: $D^{i,(1)} \to D^{i,(0)}$.
13: **end while**

[90] In his application, Pérez-Cervantes (2014) computes the source effects, as in (1.22), for a single system with more than 3,000 equations.

1.D.3 Trade costs

Given the trade model sketched in the main text, the bilateral trade flow from i to j is

$$X^{i,j} = \left(x_t^i \psi^{j,i} / A^i\right)^{-\varphi} \left(P_t^j\right)^{\varphi} \overline{\Gamma}^{\varphi} E_t^j. \tag{44}$$

In logarithms

$$\log X^{i,j} = \mathcal{O}_t^i + \mathcal{D}_t^j - \varphi \log(\psi^{j,i}), \tag{45}$$

where the first two terms are origin and destination fixed effects per period. Thus, it is a linear expression that relates (log) bilateral trade with the trade costs. I can assume the iceberg costs are a function of some observables. A popular choice is distance between i and j and a dummy for contiguity

$$\psi^{j,i} = \beta_1 \mathbf{1}\left(\{i,j\} \text{ are neighbors}\right) + \beta_2 \log(\mathrm{d}^{i,j}).$$

Notice that running the regression of bilateral trade on a dummy of contiguity and (log) distance does not identify separately φ from β_1 and β_2. However, to get an estimate of the trade costs across locations is not important.

Combes, Lafourcade, and Mayer (2005), use data on commodity flows across *départements* and exactly this specification to estimate trade costs. See in particular the first column of Table 3 in their paper. They use great circle distance, which for the scale of France is almost identical to geodesic distance that I use in other parts of the paper. The only issue is how to compute distance within the departements. This is obtained by approximating each region as a disk upon all production is concentrated at the center and consumers are proportionally distributed throughout a given proportion of the total land-area of the region. They choose the proportion to be equal to 1/16, which they claim is a reasonable approximation of the concentration of population in France. All in all, the internal distance formula is given by $\mathrm{d}^{i,i} = 1/6\sqrt{\text{Area}/\pi} = 0.094\sqrt{\text{Area}}$, where the Area of each departement is measured in squared kilometers.,

They estimate an elasticity with respect to distance $-\varphi\beta_2$ of -1.76 and the dummy of contiguity $-\varphi\beta_1$ equal to 0.98.

While they are measuring trade flows across departements, here I am using a more aggregated location definition. To make the trade costs comparable at the location level I am using, I first obtain all the trade costs at the departement level and then take the population weighted average for each location. I use the year 2002 as it is the closest to their sample.

1.E Details on sample selection

In the data I have information for every job that a worker had. This means that if a worker had more than one job per year, there are more than one observations recorded for that single worker.

For every year, there is a variable that indicates if the job is the "main job" of the year or *poste principale* in French. Thus, from the dataset *DADS Postes* I select those jobs that:

1. Have a positive before tax wage, calculated from variable *sbrut*.
2. The job is the main job of the worker, indicated by the variable *pps*.
3. There are no missing values on current location of residence (*depr*), previous location of residence (*depr_1*) or birth departement (*dep_naiss*).

4. Neither selection of departement is outside of continental France. This means that neither the current and previous residence or individuals born outside continental France are considered. So I filter out all the observations where any of the three departement codes is higher than 95.

5. After 2010, domestic workers were included in the sample. I remove them from that year onward (industry code *ape_4* 970) to keep the different data comparable across years.

With these data, every year I run a regression of log wages on a quadratic polynomial in age and a gender dummy, whose variable is sexe. For future users of the dataset, is important to note that the variable is not encoded as a dummy as those individuals identified as women have a value equal to 2, and those identified as men a value equal to 1. I then store the residuals from such regressions and use them for the estimation of the model.

1.E.1 Identification of job switchers

The dataset also includes information on the job the worker had in the previous year. This includes date of termination and status as "main job" for the previous year. If a worker relation is not "terminated" then the dataset reports the value equivalent to the maximum number of working days.[91] Also the worker who enters the year with the same job (main job or not) as in the previous year would have the start date of current job equal to the minimum. I can use this information to identify those workers who changed "main jobs" across two years.

For a worker with a main job on a given year three options are possible

1. The job is the same main job as in the previous year.

2. The job started after January 1st.

3. The job started before January 1st but in the previous year.

For example, consider a worker who had a job from February to November of 2003. This job would be the "main job" for that worker in year 2003. Then the same worker starts a job in December 2003. When looking at this worker in the year 2004, we observe that the job he had on the previous year, which will correspond to the job she started in December 2003, was *not* the main job. So we can conclude that the worker changed "main jobs" from one year to the next.

If the worker started their current main job after January 1st of the current year, then we conclude she changed jobs. A job is linked to a location, therefore all the people that changed locations are classified as switchers as well.

Then, to identify those people that do not change "main jobs" across years they have to fulfill four conditions

1. Date of termination of main job in previous period is equal to the maximum.

2. The start date of current main job is equal to the minimum.

3. The previous year job is classified as that year's main job.

4. The worker stayed in the same location.

[91]The data stores dat in terms of number of days working from January 1st. The maximum number is set to be 360.

If the worker fulfills above's conditions, then it is classified as it she did not switch jobs.

There is a small risk of classifying incorrectly some of the workers. These would be the cases where workers indeed finished their previous job in the last date of the previous year and started their next job immediately after. Also notice that with this classification, some of the observations that are identified as job switchers could just be workers re-entering the labor force and who were not working in the previous year. This is would be the case if the tax authority has information in the departement of residence of the previous year. This is not a problem in the eyes of the model, as these workers can be thought of newcomers that have an opportunity to move after observing shocks for the different locations. They will affect very little the estimation of the persistence parameter ρ that tries to capture the strength of keeping one job rather than changes in the extensive margin of the total labor force.

1.F Birthplace as Proxy for Home

I use the birthplace as a proxy for the home location of individuals. This home location can be understood as a location where the workers grew up, made childhood friends etcetera. For those that moved at a very young age, the birthplace will be a wrong proxy for home.

Of course, there is no variable where individuals report which location they perceive as their home location so we have to rely on proxies. Another proxy for home could be the location where workers first appear in the sample, or the workers first job location. Unfortunately, the main data set is not a panel that would allow me to get the location of workers' first job. However, I can rely on a subsample of the data, the *Panel DADS* which is 1/12th of the original *DADS* data that I use in the main analysis.

Using the panel data, I can get the location of the first job of each worker in this subsample. I find that 66% of the workers first appear to work in their birthplace. So for those workers the proxy location for home would be the same.

Now, for workers that have a first location different than birthplace, it is not a priori obvious which one should be the proxy for home, as workers entering late into the job market won't necessary feel attached to that location if they just moved there. Thus, I look at those workers whose first job location is different than their birthplace and they got their first job when they were 18 years old or less. The idea is that those workers likely grew up in their first job locations, so by considering the birthplace as their home location I would be classifying wrongly their home location. The fraction of workers with age less or equal than 18 when having their first job that is not in their birthplace is only 21%.

To understand to what extent birthplace is a good proxy for home location, I compare where workers live given that they moved from their first job location. I focus, as I mentioned, on workers that had different first job location than birthplace at age 18 or younger.

Table 9 shows, of those workers whose first job location is different than their birthplace and moved from their first location job, the percentage that lives back in their first location job and in their birthplace at different ages. The first row shows that only 16% lives back in their first job location before turning 20. While this fraction is large, considering that there are 73 alternatives, it is still small compared with the fraction that works in their birthplace. Although the small fraction living back in their home location might just reflect the fact that workers only just moved out from the first job location to being with. What is surprising is that a large fraction goes to their birth location. Looking at older age groups, the percentage of workers that return to their first job location increases, although the fraction of birthplace is also large. All together what the table shows is that for some workers, the first job location is probably the best proxy for home location. However, even for some workers where a priori the first job location looks like the natural candidate to be the home

Table 9 – Location of workers: First job $\neq$ Birthplace at Age ≤ 18

	% of workers living in	
Age	*First job location*	*Birthplace*
< 20	16	45
$20 - 25$	31	35
$26 - 30$	28	28
$31 - 35$	26	25
36	25	22

Note: The data comes from the *Panel DADS* which is a 1/12th subsample of the *DADS* data used in the main analysis. I consider workers whose first job location was different than their birthplace when they were 18 or younger and that moved from their first job location. Each row corresponds to the percentage of workers that live either back again in their first job location or in their birthplace at different ages.

location, the birthplace is still attracting a large fraction. Thus, birthplace looks as good as a proxy as first job location for workers with different first job location than bithplace at age 18 or younger. Combined with the fact that a large fraction of workers, regardless of age, have their first job in their birthplace, males the birthplace a good proxy for home location, at least qhen considering the alternative proxies.

1.G Alternative Identification Strategy

In this section, I spell an alternative identification strategy for both the migration and home bias, as well as the dispersion parameter. The main difference with the identification strategy from the text is that I take a *reverse* approach. First, I use the average wages to identify the compensation workers need in order to migrate. These are related to the migration costs. Then I use these compensations along with the observed labor flows to identify the migration elasticity.[92]

I exploit the closed form expression of the conditional migrating probabilities to form the conditional likelihood function of observing the labor flows in the data. I estimate the migration elasticity via maximum likelihood using the previously identified migration costs. In order to control for the unobserved expected lifetime utility in the expression, I include birthplace/destination/time fixed effects in the maximum likelihood. Identification follows as the maximization of such likelihood is equivalent to estimate it via Poisson Pseudo Maximum Likelihood (PPML), i.e. a Poisson regression with the labor flows as dependent variables, as shown by Guimaraes, Figueirdo, and Woodward (2003). It is well known that the solution of the necessary first order conditions of the maximization problem of PPML is unique. The idea of using the PPML to estimate the original parameters from the multinomial likelihood function in the context of spatial models was introduced recently Dingel and Tintelnot (2020).

The main drawback from this strategy is that using previously estimated migration costs to identify the migration elasticity might introduce a bias. And in the case of the maximum likelihood, this bias is not easy to solve. In contrast, the identification of the migration elasticity in the main text corresponds to a linear

[92]This strategy is more closely related to that of Donaldson (2018) where he first use price differences to identify the trade costs. Second, he uses the variation on those trade costs and observed trade flows to identify the trade elasticity.

regression. The attenuation bias from using a regressor with measurement error is easy to solve, theoretically and computationally, with an instrument.

Estimating the migration elasticity via maximum likelihood allows me to use the observed zero migration flows in the data. On top of that, with the same estimation I can consistently estimate the underlying probabilities.[93] Then, I can use the probability estimates to identify the home bias. Using these estimates, I identify the home bias just as in the main text.

1.G.1 Migration costs

With a little abuse of notation, recall that the expected log wage of an individual with birthplace b that moved from location i to j is

$$\mathbb{E}\left(\log\left(\text{wage}_{t,b}^{i,j}\right)\right) = \log(w_t^j) - \delta \log(p_{t-1,b}^{i,j}).$$

Taking the difference with respect to the expected wage of an individual that stayed in the same location we get

$$\begin{aligned}\mathbb{E}\left(\log\left(\text{wage}_{t,b}^{i,j}\right) - \log\left(\text{wage}_{t,b}^{i,i}\right)\right) &= \log(w_t^j) - \log(w_t^i) - \delta\left(\log(p_{t-1,b}^{i,j}) - \log(p_{t-1,b}^{i,i})\right) \\ &= \log(w_t^j) - \log(w_t^i) - \delta\left(\overline{V}_{t,b}^j - \overline{V}_{t,b}^i\right) + (1-\beta\rho)\tau^{i,j},\end{aligned}$$

where I have used the assumption that $\tau^{i,i} = 0$. The expression above is intuitive. The differences between the average log wages of people that moved away from i to location j and the people that stayed in location i reflects overall differences in wage differences that are independent of mobility patterns, plus a component that reflects the compensation migrating individuals need to have to justify such a decision.

Assuming that migration costs are symmetric, i.e. $\tau^{i,j} = \tau^{j,i}$, I can use the reverse migration flow to control for overall differences on wages in the two locations that are independent to the migration costs. This means, using the difference on average wages of workers who went from j to i with workers who stayed in j. Formally

$$\mathbb{E}\left[\left(\log\left(\text{wage}_{t,b}^{i,j}\right) - \log\left(\text{wage}_{t,b}^{i,i}\right)\right) - \left(\log\left(\text{wage}_{t,b}^{j,j}\right) - \log\left(\text{wage}_{t,b}^{j,i}\right)\right)\right] = 2(1-\beta\rho)\tau^{i,j}.$$

Therefore, the expression above identifies non-parametrically the migration costs. The only dropback is the data requirements. We need simultaneously people from a particular birthplace doing a migrating pattern plus the reverse. That is why I assume that $(1-\beta\rho)\tau^{i,j}$ is a linear function of distance.

1.G.2 Wage dispersion parameter

Recall the expression for the probability of a worker with birthplace b of moving from location i to j for a worker with birthplace b, conditional on changing jobs

$$p_{t,b}^{i,j} = \frac{\exp\left(\overline{V}_{t+1,b}^j - \tau^{i,j}\right)^{1/\lambda}}{\sum_k \exp\left(\overline{V}_{t+1,b}^k - \tau^{i,k}\right)^{1/\lambda}}.$$

I can then form the following conditional log-likelihood function

$$\log\mathcal{L}\left(\delta, \{\mathcal{D}_{t+1,b}^j\}\right) = \sum_t\sum_b\sum_{i,j} \ell_{t,b}^{i,j} \log p_{t,b}^{i,j} = \sum_t\sum_b\sum_{i,j} \ell_{t,b}^{i,j} \log\left(\frac{\exp\left(\mathcal{D}_{t+1,b}^j - (1-\beta\rho)\tau^{i,j}/\delta\right)}{\sum_k \exp\left(\mathcal{D}_{t+1,b}^k - (1-\beta\rho)\tau^{i,k}/\delta\right)}\right),$$

[93] Even with the inclusion of fixed effects, consistency follows as the number of fixed effects to be estimated grows at a rate of I^2 = location $\times$ birthplace, while the number of observations grow at rate I^3 = origin $\times$ destination $\times$ birthplace.

where $\mathcal{D}^{j}_{t+1,b} \equiv \overline{V}^{j}_{t+1,b}/\lambda$ are destination/birthplace/period specific fixed effects; and $\ell^{i,j}_{t,b}$ is the number of workers who move from i to j with birthplace b at period t, conditional on changing jobs.

By using the identified scaled up migration costs $(1-\beta\rho)\tau^{i,j}$ and controlling for the destination/period/birthplace fixed effects I can identify the migration elasticity δ.

The conditional log-likelihood is a highly non-linear object and the identification and consistency estimation of δ might be concerned. However, I can show, just as Guimaraes et al. (2003), that the conditional likelihood and a Poisson regression extended with origin fixed effects would imply solving for the same optimization problem. Identification follows because a Poisson regression has a unique solution, if such solution exists.[94]

To see the equivalence between the likelihood and the Poisson regression, consider $\ell^{i,j}_{t,b}$ to be independently distributed with

$$\mathbb{E}\left(\ell^{i,j}_{t,b}\right) = \tilde{\mu}^{i,j}_{t,b} = \exp\left(\mathcal{O}^{i}_{t+1,b} + \mathcal{D}^{j}_{t+1,b} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right).$$

The log-likelihood would then be written as

$$\begin{aligned}\log \mathcal{L}_P &= \sum_t \sum_b \sum_{i,j} \left(-\tilde{\mu}^{i,j}_{t,b} + \ell^{i,j}_{t,b} \log \tilde{\mu}^{i,j}_{t,b} - \log \ell^{i,j}_{t,b}!\right) \\ &= \sum_t \sum_b \sum_{i,j} \left(-\exp\left(\mathcal{O}^{i}_{t+1,b} + \mathcal{D}^{j}_{t+1,b} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right) + \ell^{i,j}_{t,b}\left(\mathcal{O}^{i}_{t+1,b} + \mathcal{D}^{j}_{t+1,b} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right) - \log \ell^{i,j}_{t,b}!\right).\end{aligned}$$

Taking the first order conditions with respect to $\mathcal{O}^{i}_{t+1,b}$ we obtain

$$\frac{\partial \log \mathcal{L}_P}{\partial \mathcal{O}^{i}_{t+1,b}} = \sum_j \left(\ell^{i,j}_{t,b} - \exp\left(\mathcal{O}^{i}_{t+1,b} + \mathcal{D}^{j}_{t+1,b} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right)\right) = 0$$

and therefore

$$\exp\left(\mathcal{O}^{i}_{t+1,b}\right) = \frac{\ell^{i}_{t,b}}{\sum_j \exp\left(\mathcal{D}^{j}_{t+1,b} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right)},$$

where $\ell^{i}_{t,b} = \sum_j \ell^{i,j}_{t,b}$. Concentrating out the origin fixed effects $\mathcal{O}^{i}_{t+1,b}$ we get

$$\begin{aligned}\log \mathcal{L}_{Pc} &= \sum_t \sum_b \sum_{i,j} \ell^{i,j}_{t,b} \log p^{i,j}_{t,b} + \sum_t \sum_b \sum_{i,j} \ell^{i,j}_{t,b} \log \ell^{i}_{t,b} - \sum_t \sum_b \sum_i \ell^{i}_{t,b} \\ &= \sum_t \sum_b \sum_{i,j} \ell^{i,j}_{t,b} \log p^{i,j}_{t,b} + \boldsymbol{C}_{\mathcal{L}_P},\end{aligned}$$

where $\boldsymbol{C}_{\mathcal{L}_P}$ is a constant that does not depend on the parameters. Therefore, the concentrated log-likelihood of the Poisson regression is equal, up to a constant to the original likelihood. Hence, the maximization problem will yield the exactly same estimate. There are currently very efficient packages that allow for a fast estimation of a Poisson regression with a large number of fixed effects.[95]

Consistency of the estimators follows from results on consistency on non-linear panels with two way fixed effects, as explained by Weidner and Zylkin (2020). Moreover, the fixed effects are also consistently estimated. Heuristically, the reason for this is that, while increasing the sample size by increasing the number of locations increases the number of origin and destination fixed effects to be estimated at a rate I^2 (location

[94]In principle the inly risk of identification here would be that the solution does not exists. This can happen if all the observations with zero observations, i.e. $\ell^{i,j}_{t,b} = 0$ are collinear. For example, if for a certain year there are zero workers from a particular birthplace/origin in a destination. This entails less than 1 % of the combinations in my sample. In order to avoid such problems I would just assume there is one worker who just stayed within the same destination when I have zero workers.

[95]I use the package *fixest* for R.

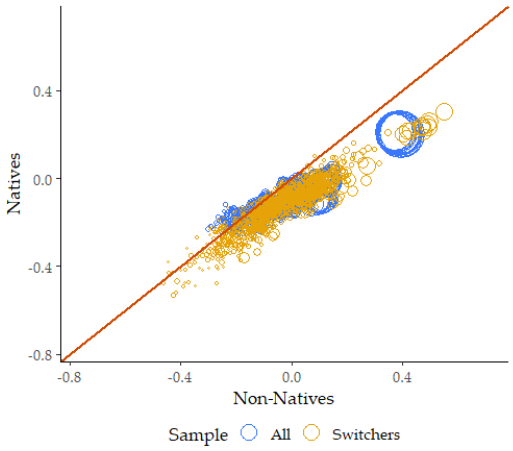
$\times$ birthplace), the sample size grows at rate N^3. This also means that we could consistently estimate the migration elasticity using a single birthplace cohort. Additionally, this also means that I can obtain consistent estimates of the underlying distribution of migration probabilities $\{p_{t,b}^{i,j}\}$, which I can then use to identify the home bias.

1.H Additional Figures and Tables

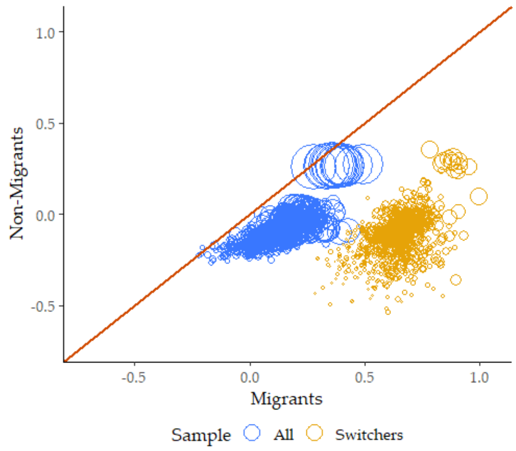
Table 10 – Gravity regression using *départements*

	(log) Labor flows, $\log L_{t,b}^{i,j}$			(log) Migration shares, $\log\left(L_{t,b}^{i,j} / \sum_k L_{t,b}^{i,k}\right)$		
	PPML			*PPML*		
	(1)	(2)	(3)	(4)	(5)	(6)
	Geodesic (km)	Driving (km)	Driving (hours)	Geodesic (km)	Driving (km)	Driving (hours)
$\mathbf{1}(j \neq b)$	2.318***	2.921***	−2.625***	-0.127***	-0.124***	−0.145***
	(0.154)	(0.179)	(0.031)	(0.003)	(0.218)	(0.059)
$\mathbf{1}(j \neq b)\log(\mathrm{d}^{b,j})$	−1.289***	−1.323***	−1.687***	−0.004***	−0.005***	−0.005***
	(0.030)	(0.033)	(0.029)	(0.000)	(0.000)	(0.000)
$\mathbf{1}(j \neq n)$	−2.554***	−2.052***	−7.452***	-1.871***	-1.353***	−7.050***
	(0.124)	(0.137)	(0.029)	(0.128)	(0.139)	(0.03)
$\mathbf{1}(j \neq i)\log(\mathrm{d}^{i,j})$	−1.299***	−1.320***	−1.837***	−1.349***	−1.368***	−1.773***
	(0.025)	(0.026)	(0.029)	(0.024)	(0.025)	(0.026)
Adj. Pseudo R^2	0.963	0.964	0.963	0.798	0.798	0.798
Observations	12,458,760	12,458,760	12,458,760	12,458,760	12,458,760	12,458,760

Note: This table stores the results of two models, both estimated using Poisson Pseudo Maximum Likelihood (PPML). The first model uses the labor flows of workers with birthplace b that moved from location i to location j, $L_{b,t}^{i,j}$ as a dependent variable. The second model uses the migration shares $L_{t,b}^{i,j} / \sum_k L_{t,b}^{i,k}$. For each model I use three different distance measures: geodesic distance in hundreds of kilometers, driving distance in hundreds of kilometers, and driving time between locations in hours. I get driving distances and hours from Google Maps. Standard errors are in parenthesis. Significance levels: *p<0.1; **p<0.05; ***p<0.01



(a) Selection Non-Natives



(b) Selection Migrants

Figure 13 – Selection via wages after controlling for age and gender. I first run a regression in each year of the logarithm of the wage on a quadratic polynomial of age and a gender dummy for all the workforce and collect the residuals. The left panel compares the average residual (log) wages of non-native workers vs native workers. Wages from both groups are normalized by the average residual (log) wage of all the sample. The plot distinguishes two cases: when using the sample consisting of all workers and using the sample of workers who switched jobs. The plot in the right panel is analogous to the plot in the right, but compares residual (log) wages of migrants vs non-migrants.

Table 11 – Estimates scale parameter δ

	Dependent variable: residual log wage$_t$	
	OLS	*IV*
	(1)	(2)
δ	0.126***	0.145***
	(0.0002)	(0.0002)
Origin/Dest./Year FE	✓	✓
Adj. R^2	0.056	0.056
Observations	26,237,598	26,237,598

Note: The table shows the results of two linear regressions estimated using Ordinary Least Squares (OLS) and Instrumental Variables (IV). I first run a regression in each year of the logarithm of the wage on a quadratic polynomial of age and a gender dummy for all the workforce and collect the residuals. The dependent variable is the residual of an individual who switch jobs across years. Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01

Table 12 – First stage regression $\tau^{i,j}/\lambda$

	Dependent variable: $\log \tau^{i,j}/\lambda$
	OLS
	(1)
$\mathbf{1}(j \neq i)\log(\mathrm{d}^{i,j})$	0.965***
	(0.0057)
Origin/Dest./Year FE	✓
Adj. R^2	0.901
R^2-Within	0.862
Observations	26,237,598

Note: The table shows the results of the first stage regression when instrumenting (scaled) migration costs $\tau^{i,j}/\lambda$. Standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01

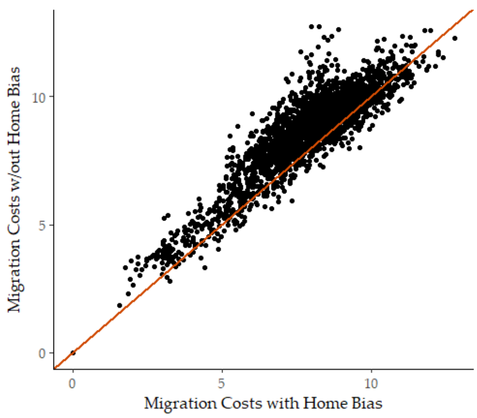


Figure 14 – Comparison Migration Costs. The plot compares the migration costs estimated in a model with home bias and a model without.

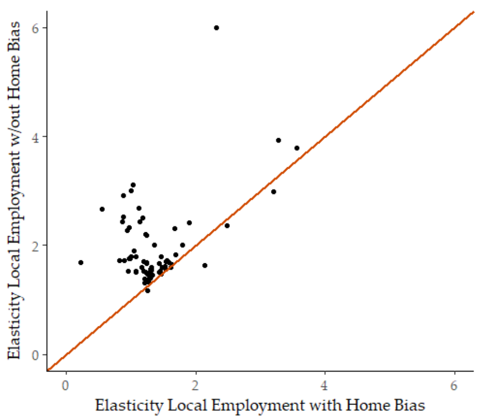


Figure 15 – Local Employment Response to a Productivity Shock. The plot compares the local employment elasticities to a productivity shock for a model estimated with home bias and a model without.

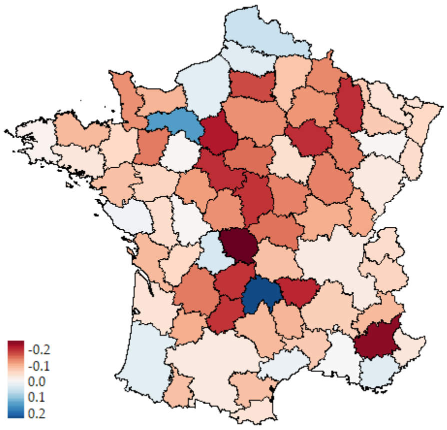


Figure 16 – Response to Place Based Policies, no Home Bias. The map shows the change in overall social welfare by subsidizing each location, normalized by the subsidies as a proportion of output, in an economy without home bias. Locations in red mean that when subsidizing such locations, overall social welfare decreased.

Chapter 2

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 a negative effect of employment concentration on wages thanks to mass layoff shocks to competitors. Second, we develop a multi-sector 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. We show that observing wage and employment data is enough to compute counterfactuals. Third, we evaluate the efficiency and welfare losses from labor market distortions. Eliminating employer and union labor market power increases output by 1.6 percent 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, Wage setting, Misallocation

[1]Azkarate-Askasua: miren.askasua@gmail.com; Zerecero: miguel.zerecero@tse-fr.eu. 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, Ramon Marimon, 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, ECB, Tilburg University, ITAM Business, Stockholm University, EUI, CEMFI, Banco de España, CREi, UPF, Mannheim University and Goethe University for useful comments. We gratefully acknowledge financial support from TSE to get access to the administrative data provided by CASD. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'avenir' program (reference : ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD). Zerecero acknowledges the financial support from CONACYT (reference: 329103/383874). First version: October 1, 2018. All errors are our own.

2.1 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 average 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 inverse 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

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

[3] The 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 2.3.4 provides more detail on the French institutional setting.

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

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 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 inverse labor supply curve which potentially 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 important element of the model is collective wage bargaining. We assume wages are set at the establishment-occupation level between establishments and unions acting symmetrically. Both sides internalize how rents are generated and bargain with zero as outside option.

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 generates 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 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

[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.

[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.

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. 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. We call these establishments 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 the labor supply and demand, we identify the elasticities 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.

In a second step we estimate the *within* local labor market labor supply elasticities by directly estimating the labor supply for each establishment. 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 industry 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, productivities and amenities, at the establishment-occupation level that rationalizes the observed data on wages and employment. Using establishment wages, we back out amenities to match employment shares. However, given that we do not observe physical output, we can only identify the *revenue* productivity, which is a

[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, σ_v are different across sub-samples.

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 separately from these prices 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 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 individual physical productivity part which is assumed fixed in the counterfactuals.

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 competitive and workers have no bargaining power leading to wages that are equal to the marginal revenue product of labor. 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 expected 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 but a bit lower than the estimates of Berger et al. (2019) for the U.S. Interestingly, we find that removing the bargaining process would marginally reduce output compare to the baseline. Thus, given the presence of employers labor market power, unions seem to have no negative efficiency effects. They do, however, have a large redistributional effect as the labor share without unions is reduced by almost 10 percentage points.

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, in the counterfactual with no distortions, 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.

[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.

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 (Azar et al., 2017; Benmelech et al., 2018; Azar et al., 2018; Schubert et al., 2020; Dodini et al., 2020; Marinescu et al., 2020). The concentration critically relates to the definition of a *local* labor market which most of the papers consider as rigid entities based on combinations of location-industry or location-geography identifiers. There have been some advances in considering the endogeneity of local labor markets either based on labor flows (Nimczik, 2017), on commuting (Manning and Petrongolo, 2017), on skill composition (Macaluso, 2017; Dodini et al., 2020) or broadly on workers' outside options (Schubert et al., 2020) inferred from labor flows. Contrary to Schubert et al. (2020) and Dodini et al. (2020) who study the effects of market concentration incorporating outside options to the notion of local labor markets, we take a more traditional approach and define them based on location-industry-occupation identifiers.

The cited empirical papers 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-affected establishments. This allows us to identify the effect of the employment share at the local labor market, our proxy of the strength of employer labor market power, on wages. Recently and independently to our work, Dodini et al. (2020) use involuntary displacements such as mass layoffs and plant closures in an event study of individual labor market outcomes. They compare differential labor market outcomes of displaced workers depending on the concentration in their geographical area. The main difference is that in our reduced form we use *competitors'* mass layoff shocks as exogenous variation to the structure of the local labor markets and measure the differential wages of establishments within a firm in affected and non-affected local labor markets.

Dodini et al. (2020) study how outside options shape labor market concentration but there is mixed recent evidence on the effects of outside options on wages. Jäger et al. (2020) find that wages do not respond to the outside option of unemployment in Austria while recently Caldwell and Danieli (2018) and Caldwell and Harmon (2019) find employment outside options do influence wages, and Hafner (2020) finds wage and employment effects on French local labor markets opening to cross border commuting. We view our work as complement to these empirical papers as our structural model neatly incorporates the effect of outside options within the local labor market into wages while allowing for labor market power from unions.

This paper also contributes to structural work on employer labor market power. We depart from the traditional monopsony power framework (e.g. Manning, 2011; Card et al., 2018; Lamadon et al., 2018) by having heterogeneous markdowns arising from market structure and by extending it to allow for wage bargaining. The paper is complementary to Jarosch et al. (2019) in the sense that they consider employer labor market power in a search framework. We contribute to those papers by incorporating unions. Marinescu et al. (2020) find negative effects of local labor market concentration on wages for new hires in France that are mitigated in more unionized industries like Benmelech et al. (2018) for the U.S. These findings are in line with our structural model and we find that allowing for collective bargaining is key to match certain empirical regularities.

In contemporaneous and independent work, Berger et al. (2019) build a structural model with oligopsonistic competition in local 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.

This 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.

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. Recently Hershbein et al. (2020) and Wong (2019) disentangle between output and labor market power using, respectively, a production function approach for the U.S. and France. They both find the presence of employer labor market power even when controlling for production function heterogeneity and output 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 et al., 2020; De Loecker and Eeckhout, 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) with the exception of Hershbein et al. (2020). The focus of this paper is therefore not on labor share trends but on the effects of 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 firm 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. Dube et al. (2018) estimate a labor supply elasticity to firm level wage policies that are between 3 and 4 which are close to our within local labor market supply elasticities and the average elasticities in the meta-analysis of Sokolova and Sorensen (2018). On the contrary, the median estimate in the meta-analysis and the estimates in Webber (2015) are near 1 and therefore close to our across local labor market supply elasticity.

The rest of the paper is organized as follows. Section 2.2 introduces the data. Section 2.3 shows the stylized facts and our empirical strategy. Section 2.4 introduces the model. Section 2.5 discusses details about identification and estimation of the model. Section 2.6 discusses the results from counterfactual exercises. Section 2.7 presents extensions of the model and Section 2.8 concludes.

2.2 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 data set 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.[12]

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 2.F.3.

[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] We use interchangeably 3-digit industry or sub-industry throughout the text.

Table 13 – Establishment-Occupation Summary Statistics

	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
All Sample					
L_{iot}	11.1	1.1	2.3	6.2	59.5
$w_{iot}L_{iot}$	367.2	31.6	71.8	196.6	2,379.5
w_{iot}	34.0	20.9	27.4	39.5	117.1
$s_{io\|m}$	0.20	0.01	0.05	0.24	0.30
(a) Monolocation					
L_{iot}	7.4	1.0	2.1	5.1	29.7
$w_{iot}L_{iot}$	216.7	29.7	64.5	159.6	925.2
w_{iot}	32.8	20.3	26.6	38.5	35.5
$s_{io\|m}$	0.18	0.01	0.04	0.19	0.29
(b) Multilocation					
L_{iot}	26.6	1.3	4.1	15.1	120.3
$w_{iot}L_{iot}$	1,004.7	45.7	139.3	533.0	5,052.4
w_{iot}	39.0	23.6	30.7	43.7	257.7
$s_{io\|m}$	0.29	0.02	0.11	0.48	0.35

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. Number of observations for All Sample is 4,151,892. For the Monolocation sample is 3,359,236; and for the Multilocation sample is 792,656. 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.

2.2.1 Summary Statistics

Table 13 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 the same occupation in different locations, which we denote as multilocation occupations. The micro evidence in the next Section focuses on multilocation firm-occupations.[13] Panels (a) and (b) of Table 13 have the summary statistics of occupations belonging to monolocation and multilocation firms. Occupations in firms with establishments at multiple locations have larger average number of employees of 27 versus 7 average employees for firms with establishments at a single location. In both groups, the distribution of employment is concentrated in few large employers, as both medians are smaller than the means. 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. Table 14 contains summary statistics of sub-industries for the year 2007, which is the year we use

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

Table 14 – Sub-industry Summary Statistics.

Variable	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
N_h	2,840	493	1,261	2,639	4,530.5
L_h	30,466	7,559	15,070	50,036	33,899.3
$\overline{w}_h$	34.6	29.6	33.0	37.531	6.9
LS_h	0.52	0.48	0.53	0.58	0.10
KS_h	0.26	0.17	0.23	0.32	0.13

Notes: There are 97 3-digit industries, or sub-industries, in the sample. 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 get the capital shares following Barkai (2016). All the nominal variables are in thousands of constant 2015 euros.

for our counterfactuals. The average 3-digit industry labor share is 52% and the share of capital is 26%.[14] 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%.

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 29 in Appendix 2.F.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 occupation o. We take the standard approach of defining local labor markets as rigid entities and abstract from flexible labor markets as in Nimczik (2017) for Austria, or how easy is to change to similar occupations, as considered by Macaluso (2017) and Schubert et al. (2020) using rich mobility data coming from resumes in the U.S.

Table 15 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.[15] 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 2.F.1.

[14] We follow Barkai (2016) to compute the capital share.

[15] 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.

Table 15 – Local Labor Market Summary Statistics. Baseline Year

Variable	Mean	Pctl(25)	Median	Pctl(75)	St. Dev.
N_m	4.76	1	2	4	14.4
L_m	51.0	2.8	9.4	34.9	196.2
$\overline{w}_m$	36.6	24.3	30.2	42.5	36.1
$\widehat{w}_m$	36.2	24.1	30.0	42.2	25.6
HHI($s_{io\|m}$)	0.67	0.38	0.68	1.00	0.32
HHI($s^w_{io\|m}$)	0.68	0.39	0.70	1.00	0.32

Notes: There are 57,940 local labor markets in the year 2007. 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.

2.3 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 the link between wages and concentration. We later present evidence on the institutional framework of French labor market and the importance of wage bargaining.

2.3.1 Labor Market Concentration and the Labor Share

We start by establishing the relationship between aggregate concentration measures 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 at time t, HHI_{mt}, is the sum of the squared employment shares of the plants present in m at a given year. Labor share at the 3-digit industry level, LS_{ht}, is the ratio of the wage bill over value added at time t. 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 therefore build a sub-industry concentration index $\overline{HHI}_{ht}$ by taking the employment weighted mean of HHI_{mt} across different local labor markets.[16]

We run the following linear regression:

$$\log(LS_{h,t}) = \delta_{b,t} + \beta \log(\overline{HHI}_{h,t}) + \varepsilon_{h,t}. \tag{2.1}$$

Table 16 presents the results. In general, the results indicate that more concentrated sub-

[16]The HHI index at market m and year t is: $HHI_{mt} = \sum_{i \in \mathcal{I}_m} s^2_{io|m}$ 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 16 – Concentration and Labor Share

	Dependent variable: $\log(LS_{h,t})$		
	(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 (2.1). 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

industries have a lower labor share. 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. Column (3) shows that the negative relation between employment concentration and the labor share is robust to controlling for industry and industry-year fixed effects.

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 15). Estimates in column (3) with industry-year fixed effects are very similar. The low estimated effects imply that wages, and therefore the labor share, are not very responsive to differentiated levels of concentration. Nevertheless, one cannot interpret that they rule out employer labor market power because in a setting where all the firms acted as pure monopsonists facing an equal labor supply elasticity, wages (and the labor share) would be insensitive to concentration as all establishments would have the same markdown.

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.

2.3.2 Labor Market Concentration and Wages

This section explores the relationship between employer labor market power and wages at the establishment level. 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. In what follows, we focus on multi-location occupations where the effects are estimated using residual variation across local labor markets within a firm-occupation-year.[17]

The baseline specification is:

$$\log(w_{io,t}) = \beta\, s_{io|m,t} + \psi_{\mathbf{J}(i),o,t} + \delta_{\mathbf{N}(i),t} + \varepsilon_{io,t}\,, \tag{2.2}$$

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 also 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's and occupation o's 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 17 shows the results. The first two columns recover estimates of the specification (2.2) 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 (2.2) is problematic and expected to be biased towards zero. We indeed find that both OLS estimates are very close

[17] Recall that a multi-location occupation of a firm is an occupation that is present in several establishments across the geography.

Table 17 – 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

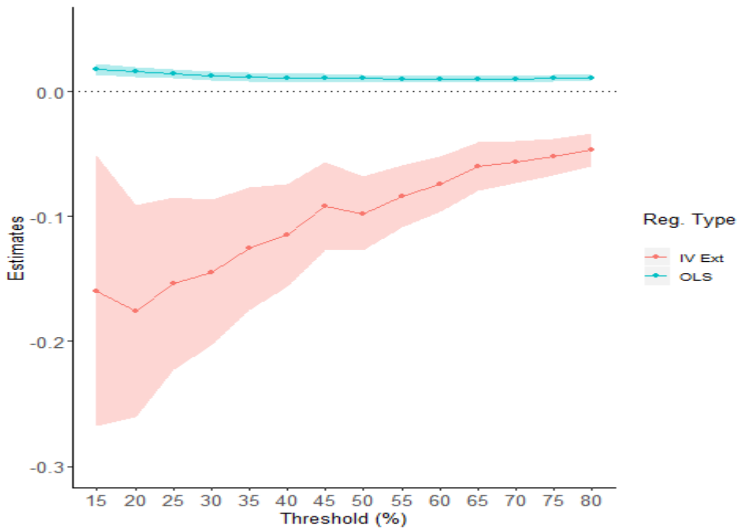
to zero and positive. Columns (2) and (4) present the results once we instrument for the employment share. Both specifications (with CZ and CZ-year fixed effects) give the same point estimates. These estimates imply that an increase of one percentage point (p.p. henceforth) of the local labor market 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 link between labor market concentration and wages. We now instrument the 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 to a competitor is exogenous to the residual within firm-occupation variation across local labor markets that identifies the effect. The main specification is an instrumental variable regression where we compare establishment-occupations of firms that had exogenous increases in concentration due to the competitors' shock against establishment-occupations that were not exposed to the competitors' shock. Here we provide some detail of the construction of the instruments that is complemented in Appendix 2.G.

We first need to identify the firms suffering 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

Figure 17 – 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 (2.2) with commuting zone fixed effects. Results with commuting zone-year fixed effects are in Section 2.3.3.

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 firm identifiers keeping the number of competitors at the local market constant. Our instrument is a proxy to capture the impact of a firm's large employment shock into the competitors.[18] We restrict the sample to non affected firm-occupations with establishments in local labor markets with and without a competitor suffering a mass layoff.

The choice of κ leads to a trade-off as 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.[19]

Results with commuting zone fixed effects are in Figure 17. OLS estimates of β from (2.2) are in blue slightly above zero and IV estimates are in red.[20] Both are plotted with 95% confidence intervals.[21] The employment share being endogenous, the OLS estimates are biased towards zero and are in line with the column (3) of Table 17. 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 estimated effect. We estimate an elasticity of 0.17 with $\kappa = 20\%$ (a loss of 80% of employment). 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.[22] For the more standard threshold of $\kappa = 70\%$ (reduction of 30%

[18] See Appendix 2.G for a graphical illustration of the identification.

[19] A standard value in the literature is κ =70% (e.g. Hellerstein et al., 2019; Dodini et al., 2020). That is a 30% loss of employment.

[20] We are restricting to firm-occupations 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.

[21] Details of the point estimates and confidence bands are in Appendix 2.G.

[22] This computation is done taking the employment share differences between the percentile 75 and 25 from Table 13 for the median

employment) the elasticity is almost divided by 4 to 0.06 which implies a twice as big reduction as with lagged concentration measures in Table 17. As we increase the threshold the estimated coefficient converges to the OLS estimate and the variance is reduced.

2.3.3 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 25 in Appendix 2.G.2 shows a robustness check where the new instrument is not binary anymore 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 26 in Appendix 2.G.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 instead of 3-digit industries as in the baseline specification.[23]

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.

2.3.4 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.[24] According to administrative data, the unionization rate in France was 9% in 2014 which 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%).[25]

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 18 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 affiliation rates.[26] This is in stark contrast to the

wage. The analogous computation with the average wage gives a wage reduction of roughly 1300 euros.

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

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

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

[26]The source of collective bargaining agreements is the OECD as for unionization rates.

Table 18 – Union Density and Collective Bargaining Coverage

Country	Union Density	Coverage	Country	Union Density	Coverage
Western Europe			**Southern Europe**		
Austria	27.7	98.0	Italy	36.4	80.0
France	9.0	98.5	Spain	16.8	80.2
Germany	17.7	57.8	**Americas**		
Netherlands	18.1	85.9	Canada	29.3	30.4
Switzerland	16.1	49.2	Chile	15.3	19.3
Northern Europe			United States	10.7	12.3
Finland	67.6	89.3	**Asia & Oceania**		
Ireland	26.3	33.5	Australia	15.1	59.9
Norway	49.7	67.0	Japan	17.5	16.9
United Kingdom	25.0	27.5	Korea	10.0	11.9
			Turkey	6.9	6.6

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 data comes from the OECD and the sources are administrative data except for Australia, Ireland and United States which are based on survey data. The regions are defined according to the United Nations M49 area codes.

U.S. collective bargaining agreements that only apply to union members and therefore coverage is very similar to the unionization rate.

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.[27] 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.[28]

Legal requirements regarding union representation depend on firm or plant size. The first requirements start when the establishment reaches 10 employees and there is an important tightening of duties when reaching the threshold of 50 employees.[29] 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) (See Table 1 of Naouas and Romans, 2014).[30]

Theoretically, workers organize into unions to extract rents from the firm through bargaining. Bargaining can happen at different levels in France and here we want to inform the modeling decisions in the next section by quantifying bargaining differences depending on industries or occupations. We build a proxy of rents at the firm level and then compare how the correlation of wages with rents is differentiated depending on the industries and occupations. In particular we

[27]Several collective agreements can coexist at a given establishment.
[28]Source DARES.
[29]The Appendix of Caliendo et al. (2015) provide a comprehensive summary of size related legal requirements in France.
[30]The prevalence of wage bargaining was 44% for establishments with 11 employees or more.

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 with respect to rents 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 2.H.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.

2.4 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, 3-digit 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.

Capital and output markets are competitive. Establishments are owned by absent entrepreneurs who rent the capital and collect the profits. We assume the economy is a small open economy and capital is specific for each industry. Thus, the industry specific rental rates of capital R_b are exogenous.[31]

[31] As it is a small open economy, it is not important whether the entrepreneurs own capital or not.

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. We assume that establishments and unions are symmetric. Both have zero threat points and internalize how the marginal cost changes when moving along the labor supply curve. Null outside options for workers are not common in the literature but the assumption is in line with new evidence of insensitivity of wages to the value of nonemployment (Jäger et al., 2020).[32]

Having a discrete set of establishments per local labor market means that when bargaining, both parties 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.

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

Production

The final good Y 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:

$$Y = \prod_{b \in \mathcal{B}} Y_b^{\theta_b}, \tag{2.3}$$

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:

$$P_b Y_b = \theta_b P Y. \tag{2.4}$$

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. Y_b is the aggregate of output of all the firms in that sector:

$$Y_b = \sum_{i \in \mathcal{I}_b} y_i, \tag{2.5}$$

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

[32]We show in Appendix 2.A.4 that the same equilibrium wages arise with a different bargaining protocol where employer labor market power is incorporated through workers' outside options.

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:

$$y_i = \sum_{o=1}^{O} y_{io} = \sum_{o=1}^{O} \widetilde{A}_{io} K_{io}^{\alpha_b} L_{io}^{\beta_b}. \tag{2.6}$$

The choice of this particular production function is motivated by tractability and empirical reasons. The linearity of the aggregation within establishments allows for the separability of different local labor markets.[33] 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, unless we were to assume establishment-specific output elasticities. Appendix 2.I lays out the model and proofs with a Cobb-Douglas production function.

The separability of local labor markets also requires 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 2.4.4.

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

$$y_{io} = F_b^{\alpha_b(1+\varepsilon_b\delta)} A_{io} L_{io}^{1-\delta}, \quad A_{io} \equiv \widetilde{A}_{io}^{\frac{1}{1-\alpha_b}} \left(\frac{\alpha_b}{R_b}\right)^{\frac{\alpha_b}{1-\alpha_b}}, \tag{2.7}$$

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.[34] Details of these derivations are in Appendix 2.A. From now on we work with the production function after substituting out the capital.

Labor Supply

We now present worker preferences that give rise to upward sloping establishment-occupation specific inverse labor supplies. A worker k derives utility by consuming the final good c_k 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:

$$\mathcal{U}_{kio} = c_k z_{kio} u_{km}. \tag{2.8}$$

[33] The solution and characterization of the model are in Section 2.4.4.

[34] $F_b = P_b^{\frac{1}{\chi_b}}$, $\chi_b = (1-\alpha_b)(1+\varepsilon_b\delta)$ 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 two independent Fréchet distributions:

$$P(z) = e^{-T_{io}z^{-\varepsilon_b}}, \quad T_{io} > 0, \varepsilon_b > 1 \tag{2.9}$$

$$P(u) = e^{-u^{-\eta}}, \quad \eta > 1, \tag{2.10}$$

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 one 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 taste shocks. We name the parameters ε_b and η as the *within* and *across* labor market elasticities of labor supply. 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 Frisch elasticity 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.[35]

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

$$\mathcal{U}_{kio} = w_{io} z_{kio} u_{km}, \tag{2.11}$$

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 2.A. The probability $\Pi_{io} = s_{io|m} \times s_m$ writes as:

$$\Pi_{io} = \frac{T_{io} w_{io}^{\varepsilon_b}}{\sum_{j \in I_m} T_{jo} w_{jo}^{\varepsilon_b}} \times \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\sum_{m' \in \mathcal{M}} \Phi_{m'}^{\eta/\varepsilon_{b'}} \Gamma_{b'}^{\eta}}, \tag{2.12}$$

where $\Phi_m = \sum_{j \in I_m} T_j w_{jo}^{\varepsilon_b}$ is a local labor market aggregate, and the economy wide constant Φ is $\Phi = \sum_{m \in \mathcal{M}} \Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}$. Γ_b is just an industry-specific constant. In equilibrium, the first fraction is equal to $s_{io|m}$ and the second term in (2.12) is s_m.

[35]One can view these taste shocks as mobility costs in a static model that could be present when changing jobs across the geography, industry and occupations.

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

$$L_{io}(w_{io}) = \frac{T_{io} w_{io}^{\varepsilon_b}}{\Phi_m} \frac{\Phi_m^{\eta/\varepsilon_b} \Gamma_b^{\eta}}{\Phi} L = \Pi_{io} L. \tag{2.13}$$

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

2.4.1 Absence of Bargaining

To ease the exposition of our baseline model, in this section we characterize equilibrium wages in the absence of bargaining. Given the labor supply curves with finite elasticities, establishments post wages taking into account the (inverse) labor supply curves (2.13) 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).[36] 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:

$$w_{io}^{MP} = \frac{e_{io}}{e_{io}+1} \beta_b A_{io} L_{io}^{-\delta} P_b^{\frac{1}{1-\alpha_b}}, \tag{2.14}$$

where $e_{io} = \varepsilon_b (1 - s_{io|m}) + \eta \, s_{io|m}$ 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. The fraction $\frac{e_{io}}{e_{io}+1}$ in equation (2.14) is the markdown and it is defined as:

$$\mu(s_{io|m}) = \frac{\varepsilon_b (1 - s_{io|m}) + \eta \, s_{io|m}}{\varepsilon_b (1 - s_{io|m}) + \eta \, s_{io|m} + 1}. \tag{2.15}$$

In the absence of bargaining, the wedge between the marginal revenue product of labor and the wages boils down to a markdown (2.15).[37] We denote this object in short notation as μ_{io}.

As long as workers have less elastic labor supplies across local labor markets than across establishments within a given local labor market (i.e. as long as $\eta < \varepsilon_b$), the markdown (2.15) is a decreasing function of the employment share $s_{io|m}$. Once an establishment is big with respect to the nearby competitors, it internalizes that it is facing a more inelastic labor supply of workers willing to stay and applies a markdown further away from 1. In the limit where ε_b and η tend to infinity, establishments face an infinitely elastic labor supply and the labor market would be perfectly competitive with a markdown $\mu(s_{io|m}) = 1$.

[36]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.

[37]Appendix 2.A derives this expression.

Heterogeneous markdowns distort relative wages across establishment-occupations and therefore the labor allocations. By 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 2.4.4.[38]

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 longer true if we were to allow an endogenous leisure or labor force participation decision. Counterfactually increasing wages would increase total labor supply L and therefore total output.[39]

2.4.2 Bargaining

We now introduce the bargaining between employers and unions.[40] 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.

We assume that workers and establishments are symmetric in the bargaining protocol: first, both parties enter the bargaining with a null outside option and, second, internalize how they generate rents as they move along the labor supple curve. The former implies that if bargaining were to fail, workers could not earn any income and establishments could not produce. The zero outside option for the workers is in line with recent evidence of lack of response of wages to changes in unemployment benefits (Jäger et al., 2020). The second assumption, where unions also internalize how the marginal cost changes when introducing an additional worker, is behind the idea that unions will be bargaining to extract part of the generated rents.

The bargained equilibrium wage is the solution to a reduced form Nash bargaining where the union's bargaining power is φ_b and the one of the establishment is $1 - \varphi_b$. Appendix 2.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:

$$w_{io} = \underbrace{\left[(1 - \varphi_b)\, \mu_{io} + \varphi_b \frac{1}{1 - \delta} \right]}_{\text{Wedge } \lambda(\mu_{io}, \varphi_b)} \times \underbrace{\beta_b A_{io} L_{io}^{-\delta} P_b^{\frac{1}{1-\alpha_b}}}_{\text{MRPL}}. \tag{2.16}$$

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} that would be present under

[38] Appendix 2.H provides an illustration of the distributional and efficiency consequences.

[39] The constant $\mu = \frac{\eta}{\eta+1}$ 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.

[40] We use interchangeably workers or unions.

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 when $\delta > 0$ we have that $\frac{1}{1-\delta} > 1$. 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 $\frac{1}{1-\delta} > 1$.

If the within local labor market elasticity $\varepsilon_b < \eta$, then the labor supply elasticity e_{io} is decreasing in the local labor market employment share. Hence, even if unions bargain over the wages, one would observe a negative relationship between employment shares $s_{io|m}$ and wages w_{io} as long as they don't extract all the quasi-rents i.e. as long as $\varphi_b < 1$. A desirable feature of the model is that it nests the oligopsonistic competition only and bargaining only models as special cases. The former is equivalent to the limit where the 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 we take the alternative bargaining from Appendix 2.A.4 and worker's outside option is the competitive wage. The wedge in that case is equal to: $1 - \varphi_b + \varphi_b \frac{1}{1-\delta} = 1 + \varphi_b \frac{\delta}{1-\delta}$. The bargained wages incorporate a markup over the marginal product and become $w^B = (1 + \varphi_b \frac{\delta}{1-\delta}) \times MRPL$.

2.4.3 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 output Y, such that equations (2.3)-(2.13) and (2.16) are satisfied $\forall\ io \in \mathcal{I}_m$, $m \in \mathcal{M}$ and $b \in \mathcal{B}$.

2.4.4 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 together with the assumption on the production function 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 industry b 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 inverse labor supply (2.13) into (2.16) and simplifying we obtain:

$$w_{io} = \left(\beta_b \lambda(\mu_{io}, \varphi_b) \frac{A_{io}}{\left(T_{io}\Gamma_b^{\eta}\right)^{\delta}} \right)^{\frac{1}{1+\varepsilon_b \delta}} \Phi_m^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi}{L}\right)^{\nu_b} F_b, \tag{2.17}$$

where $\nu_b \equiv \frac{\delta}{1+\varepsilon_b \delta}$ is 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 (2.17), their relative wages are:

$$\frac{w_{io}}{w_{jo}} = \left(\frac{\lambda(\mu_{io}, \varphi_b)}{\lambda(\mu_{jo}, \varphi_b)}\right)^{\frac{1}{1+\varepsilon_b\delta}} \left(\frac{A_{io}}{A_{jo}} \frac{T_{jo}^{\delta}}{T_{io}^{\delta}}\right)^{\frac{1}{1+\varepsilon_b\delta}}. \tag{2.18}$$

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 inverse labor supply (2.13). 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 (2.17) can be separated in two terms. First, a sub-market m constant $(\Phi_m^{(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi}{L}\right)^{\nu_b} F_b)$; and second, an establishment-occupation specific component. We denote this second term as:

$$\widetilde{w}_{io} = \left(\beta_b \lambda(\mu_{io}, \varphi_b) \frac{A_{io}}{\left(T_{io}\Gamma_b^{\eta}\right)^{\delta}}\right)^{\frac{1}{1+\varepsilon_b\delta}}, \tag{2.19}$$

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 (2.17) in partial equilibrium:

Proposition 8. *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 (2.17).*

Proof. See Appendix. □

Proposition 8 tells us that if we take the aggregate terms as constants, then the solution for the system (2.17) 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. The system (2.19) has a unique solution as we can use Proposition 8 with $\Phi_m = \Phi = L = F_b = 1$.

We now turn to the second step of the equilibrium characterization. Given the solutions to the establishment-occupation components we build industry level labor supplies and productivity measures. We can then write the model at the industry b level.

Aggregating the individual labor supplies (2.13), the industry labor supply is $L_b = \frac{\Phi_b(\mathbf{w}_b)}{\Phi} L$, where $\Phi_b(\mathbf{w}_b) = \sum_{m \in \mathcal{M}_b} \Phi_m(\mathbf{w}_m)^{\eta/\varepsilon_b}$. $\mathbf{w}_b$ and $\mathbf{w}_m$ are, respectively, vectors of wages for industry b and local labor market m. Then, Phi_b is just a function aggregating wages at the industry b level. At this stage we do not know the wages for industry b $\mathbf{w}_b$, but rather the normalized wages $\widetilde{\mathbf{w}}_b$ (see

Proposition 8). However, the following holds $\Phi_b(\mathbf{w}_b) = \Phi_b(\widetilde{\mathbf{w}}_b)F_b^{\psi_b\eta}\left(\frac{\Phi}{L}\right)^{1-\delta}$. Thus, the industry b labor supply L_b is

$$L_b(\mathbf{F}) = \frac{\Phi_b(\mathbf{w}_b)\Gamma_b^{\eta}}{\Phi(\mathbf{w}_b)}L = \frac{F_b^{\psi_b\eta}\widetilde{\Phi}_b\Gamma_b^{\eta}}{\sum_{b'\in\mathcal{B}}F_{b'}^{\psi_b\eta}\widetilde{\Phi}_{b'}\Gamma_{b'}^{\eta}}L, \quad \psi_b = \frac{1+\varepsilon_b\delta}{1+\eta\delta}. \tag{2.20}$$

where $\widetilde{\Phi}_b = \Phi_b(\widetilde{\mathbf{w}}_b)$ and $\mathbf{F} = \{F_b\}_{b\in\mathcal{B}}$ is a vector of transformed prices. So given the solution for all the normalized wages, the industry labor supplies L_b are just a function of the transformed prices $\mathbf{F}$.

Starting from the establishment-occupation output (2.7) we aggregate up to industry output:

$$Y_b = F_b^{\alpha_b(1+\varepsilon_b\delta)}A_bL_b^{1-\delta}, \quad A_b = \sum_{io\in\mathcal{I}_b}A_{io}s_{io|m}^{1-\delta}s_{m|b}^{1-\delta}, \tag{2.21}$$

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 (2.4) to find industry prices. Using the final good production function (2.3) and the intermediate good demand (2.4), we obtain

$$F_b^{1+\varepsilon_b\delta}A_bL_b(\mathbf{F})^{1-\delta} = \theta_b\prod_{b'\in\mathcal{B}}\left(F_{b'}^{\alpha_{b'}(1+\varepsilon_{b'}\delta)}A_{b'}L_{b'}(\mathbf{F})^{1-\delta}\right)^{\theta_{b'}}. \tag{2.22}$$

Steps to get to this expression are in Appendix 2.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.[41] By 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.[42]

Given the solution for normalized wages, we can think of industry productivity A_b and the aggregator for 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 9. *For any set of parameters* $\{\beta_b, \theta_b \quad s.t. \quad 0 \le \beta_b, \theta_b < 1, \forall\, b \in \mathcal{B}\}$, $0 \le \delta \le 1$, $\{\psi_b \equiv \frac{1+\varepsilon_b\delta}{1+\eta\delta}\}_{b\in\mathcal{B}}$, *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 system formed by* (2.22) *and it's characterized by:*

$$F_b = X_bC^{\frac{1}{\psi_b(1+\eta)}}, \tag{2.23}$$

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

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

Proof. See Appendix. □

[41] Recall that B is the number of different 2-digit industries.

[42] This will be useful in the extension with endogenous labor force participation in Section 2.7.

Proposition 9 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 8 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 8 and 9 we can therefore conclude that there exists a unique solution to the model for any set of valid parameters and vectors of productivities and amenities.

2.5 Identification and Estimation

In this section, we describe the identification strategy, the estimation procedure and present the results. We have two types of parameters: (i) related to the labor supply and bargaining, and (ii) related to technology. The labor supply and bargaining parameters are the within and across local labor market elasticities ($\{\varepsilon_b\}_{b=1}^B$ and η respectively) and the workers' bargaining powers ($\{\varphi_b\}_{b=1}^B$). The technology parameters are: the inverse elasticity of the labor demand (δ), the industry output elasticities ($\{\alpha_b\}_{b=1}^B, \{\beta_b\}_{b=1}^B$) and the elasticities of intermediate goods in the final good production function ($\{\theta_b\}_{b=1}^B, \{\beta_b\}_{b=1}^B$). Given our restriction δ, we only need to estimate either the capital elasticities $\{\alpha_b\}_{b=1}^B$ or the labor ones $\{\beta_b\}_{b=1}^B$.

We lay out a recursive identification strategy in three steps. First, we identify the across local labor market labor supply elasticity η and the inverse elasticity of labor demand δ by exploiting differences in the variance-covariance matrix of structural shocks across occupations. 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 output elasticities of capital $\{\alpha_b\}_{b=1}^B$ to match industry capital shares. Then, we calibrate the union bargaining powers $\{\varphi_b\}_{b=1}^B$ to match the industry labor shares, and we calibrate yearly industry elasticities with respect to intermediate good in the final good production function to match 2-digit industry output in the data.

We start the identification of global parameters by taking 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.

Full monopsonists being the only players in the local labor market, the markdown part of the labor wedge is constant and equal to $\mu(s=1) = \frac{\eta}{\eta+1}$. Their labor demand is:

$$w_{io} = \left[(1-\varphi_b)\,\frac{\eta}{\eta+1} + \varphi_b\,\frac{1}{1-\delta}\right]\beta_b P_b^{\frac{1}{1-\alpha_b}} A_{io} L_{io}^{-\delta}, \tag{2.24}$$

and the labor supply they face is:

$$L_{io} = \frac{T_{io}^{\eta/\varepsilon_b} w_{io}^{\eta} \Gamma_b^{\eta}}{\Phi} L. \tag{2.25}$$

Similar labor supply and demand systems can be formed for each occupation. This system suffers from standard identification issues when we have simultaneous equations as independent identification 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. First, we assume that the variances of the shifters differ across occupations, i.e., we assume heteroskedasticity. To gain intuition of how identification through heteroskedasticity works, consider a simple demand and supply system. Now assume we start increasing the variance of the supply shifter relative to the variance of the demand. Then, the new observed scatter describing price-quantity pairs will be more tilted towards the demand. This variation allows to identify the parameters. In our preferred specification, we group the occupations into two categories, white collar (top management and clerical) and blue collar (supervisor and operational), and assume that the covariance between the demand and supply shifters (productivity and amenity respectively) are constant within each of the two categories. This assumption reflects the idea that amenities such as working hours, repetitiveness of the tasks or more general working environments are similarly related to productivity within our two categories.

Taking logarithms and demeaning by substracting the industry b average per year, the system formed by (2.24) and (2.25) for occupation o is:

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

We estimate the variance covariance matrix of the left-hand side, employment and wages per occupation, from the data. The restriction we impose 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 for each category we obtain a system of two equations that do not depend on the within local labor market labor supply elasticity ε_b anymore and depends only on η and δ. More details about this identification argument are in Appendix 2.D.

The second step is devoted to the estimation of the within local labor market labor supply elasticities ε_b. Those are estimated exploiting the labor supply equation of non full monopsonists. The labor supply they face (2.13) in logs is:

$$\ln(L_{iot}) = \varepsilon_b \ln(w_{iot}) + f_{mt} + \ln(T_{iot}),$$

where f_{mt} is a local labor market times year 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 violating the exclusion restriction. We therefore instrument for the wages using a proxy $\widehat{A}$ of firm productivity.

$$\widehat{A}_t = \frac{P_{bt} Y_{jt}}{\sum_{\mathbf{J}(i),t} L_{iot}^{1-\delta}},$$

where Y_{jt} is value added at the firm j in year t and $\mathbf{J}(i), t$ denotes the set of establishments belonging to firm j. The first estimation step did not require independence of the structural shocks. In order to minimize the potential of endogeneity bias coming from the correlation between amenities and productivities, we use a lagged instrument instead of the contemporaneous one.

In the final step we calibrate the capital elasticities, the union bargaining powers, we non-parametrically recover amenities and productivities, and we calibrate the elasticities of the final good production function. We start the final step by calibrating the capital elasticities to target the average industry capital shares. We follow Barkai (2016) to construct the industry interest rates or required rates $\{R_{bt}\}_{b=1}^{B}$ per year to build yearly capital shares.[43] From the first order condition for capital, the industry b capital share of output is:[44]

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

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 $\frac{\beta_b}{1-\alpha_b} = 1 - \delta$.

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:

$$LS_{io} = \frac{w_{io} L_{io}}{P_b y_{io}} = \beta_b \lambda(\mu_{io}, \varphi_b), \tag{2.26}$$

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.[45] 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 3-digit industry value added measures.[46] Amenities and revenue productivities are non parametrically identified to match wages and labor allocations in equilibrium.

We recover non-parametrically establishment-occupation TFPRs (revenue TFPs) using the wage first order conditions. We observe employment and nominal wages at the establishment-occupation level from the data. Equation (2.16) in nominal terms is:

$$P_t w_{iot} = \beta_b \lambda(\mu_{iot}, \varphi_b) \; P_t F_{bt}^{1+\varepsilon_b \delta} A_{iot} L_{iot}^{-\delta}, \tag{2.27}$$

where $P_t w_{iot}$ and L_{iot} are observed and $\beta_b \lambda(\mu_{iot}, \varphi_b)$ depends on the estimated parameters and observed employment shares. Equation (2.27) makes clear that given the observed nominal wages

[43] Details are in Appendix 2.F.4.

[44] This is derived in Appendix 2.A.

[45] The model aggregation of the labor share is in Appendix 2.A.5 and the industry markdown is characterized in equation (41).

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

Table 19 – 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

and employment, one can only back out transformed TFPRs, $Z_{iot} = P_t F_{bt}^{1+\varepsilon_b \delta} A_{iot}$, that are a function of the establishment-occupation physical productivity A_{iot} and prices $P_t F_{bt}^{1+\varepsilon_b \delta}$.[47] Details of how we back out amenities T_{iot} to ensure that we match employment are in Appendix 2.D.3.

We calibrate the elasticities of the final good production function $\{\theta_b\}_{b/in\mathcal{B}}$ for every year of the sample such that the industry expenditure shares are equal to the shares of industry value added in the data. Table 27 in Appendix 2.E 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.

2.5.1 Estimation Results

Table 19 shows the estimation results of the main parameters. The most important parameters of the estimation are arguably the 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.[48] The across local labor market elasticity being lower than the within ones ($\widehat{\eta} < \widehat{\varepsilon}_b \quad \forall b$), workers are more elastic within than across local labor markets. This implies that the markdown μ_{io} is decreasing in the employment share and therefore 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 2.3.

Our estimates for the labor supply elasticities are within the ball park of the recent estimates from Berger et al. (2019) for the US. 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.[49]

[47] Revenue Total Factor Productivities are defined as $P_t P_{bt} A_{iot}$. With some abuse of notation, we name the transformed revenue total factor productivities $P F_{bt}^{1+\varepsilon_b \delta} A_{iot}$ as TFPRs. Given that one cannot observe industry prices P_{bt}, backing out productivities A_{iot} from the data would require carry out some normalizations to get rid of industry prices and be able to compute counterfactuals.

[48] Table 26 in Appendix 2.E provides all of the industry estimates.

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

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 doing 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 establishment 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. 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.[50] 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 $\{\alpha_b\}_{b\in\mathcal{B}}$ allow us to recover the values for the output elasticities with respect to labor, $\{\beta_b\}_{b\in\mathcal{B}}$, as $\beta_b = (1-\alpha_b)(1-\delta)$. These elasticities go from 0.56 for *Transport* to the 0.85 for *Shoe and leather production*.

2.5.2 Estimation Fit

Using the point estimates we first check the fit on non-targeted labor shares at the sub-industry level and total value added. We then replicate the empirical exercises of Section 2.3 with model generated variables and compare them with the results obtained with actual data.

Not-targeted Moments

Figure 18 depicts the fit of the model and non-targeted data. In panel (a) we have 3-digit 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 a 2-digit industry. We see that most of the dots are aligned around the 45 degree line.

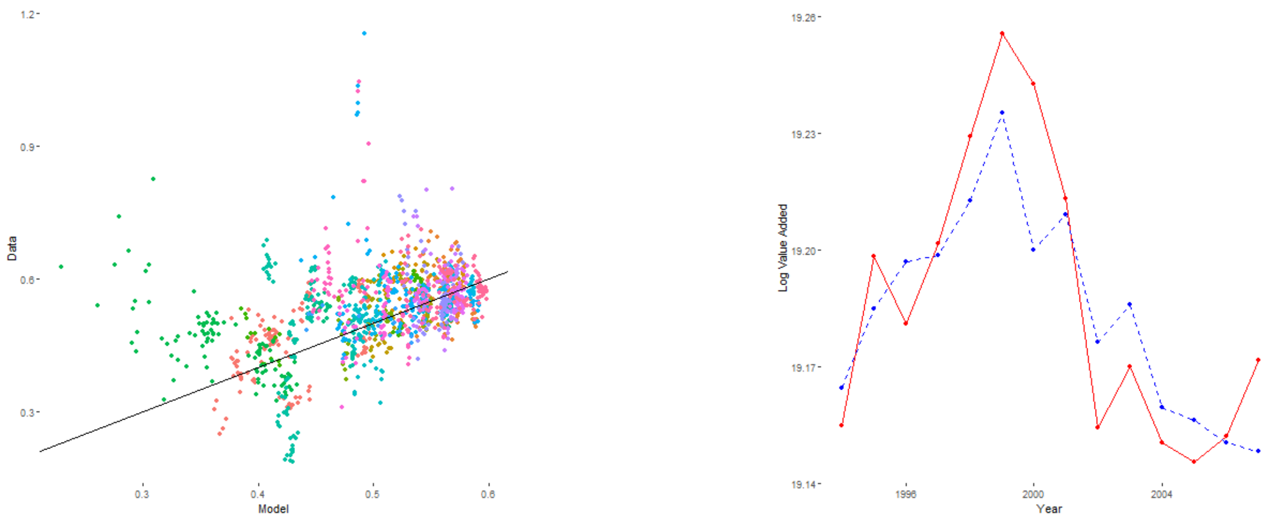
Panel (b) of Figure 18 shows the model matches the evolution of aggregate value added. This in fact might not be surprising as there is a very strong relationship between establishment's production and wage bill in the model and in the data. Since the model exactly matches the establishment's wages and labor allocations, it also has a good fit of the value added.

Going Back to the Empirical Evidence

We further validate the model by replicating the empirical evidence of Section 2.3 linking micro-level concentration to wages and industry concentration to the industry labor shares. We first simulate the model to explore the quantitative relation between employment shares and wages at

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

Figure 18 – Model Fit Non Targeted Moments



(a) Sub-industry Labor Share

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

the establishment level. We then compare how well the model matches the regression coefficients linking aggregate measures of employment concentration and the labor shares.

In the micro-level empirical evidence from Section 2.3 we measure the effects of concentration on wages by using shocks to local competitors to capture an exogenous change in the relative position within a local labor market of an establishment. Thus, our aim in the simulation and empirical exercise presented below is to induce such exogenous changes to the local labor market of establishments. Using the identified amenities, T_{io}, and TFPRs, Z_{io}, for the year 2007 as baseline, we simulate *proportional* changes in establishment-occupations' productivities and solve again for the equilibrium wages within each local labor market. Then, we explore the link of employment shares to log wages according to the following linear model:

$$\log(w_{io}) = f_b + \beta s_{io|m} + u_{io},$$

where $\log(w_{io})$ is the logarithm of simulated normalized wages, f_b is an industry fixed effect, $s_{io|m}$ is the equilibrium employment share of establishment-occupation io in the local labor market m, and u_{io} is an error term. As we are only solving for the equilibrium wages within each local labor market, we are implicitly normalizing local labor market, industry and economy wide constants that would be incorporated after solving for the general equilibrium of the model.[51] We include of the industry fixed effect f_b because we use revenue productivities as a fundamental in the simulation.[52]

To replicate the exogenous change in the local labor market structure, we instrument the employment share with the weighted average of the productivity proportion changes of each establishment-occupation's competitors, where the weights are the employment shares in the baseline scenario.

[51] We consider normalized counterfactual wages similar to the baseline ones in equation (2.19), but with revenue productivities Z_{io} instead of productivities A_{io}.

[52] Recall that revenue productivities are a function of both an industry level price and the productivity at the baseline equilibrium. The fixed effect would capture such industry prices.

More clearly, the instrument for the employment share of io in local labor market m is:

$$\sum_{(io)' \in \{m \setminus io\}} \frac{Z'_{(io)'}}{Z_{(io)'}} \frac{L_{(io)'}}{\sum_{(io)'' \in \{m \setminus io\}} L_{(io)''}},$$

where $Z'_{(io)'}$ is the simulated revenue productivity for establishment-occupation $(io)'$, and $L_{(io)'}$ is its employment in the baseline, i.e. before the simulation.

The estimated coefficient is -0.203 with a standard error of 0.035. The point estimate is a little below to the one presented on Figure 17 but well within the confidence intervals. We take this as evidence that the model is able to replicate the intensity of the link between employment concentration and wages at the establishment occupation level.[53]

We now turn to the model validation using the aggregate empirical evidence of Section 2.3. Here we generate model consistent value added and labor shares taking the market structure and fundamentals (amenities and productivities) from the data. Table 20 presents in the first 2 columns the empirical evidence of Table 16 with fixed effects (columns (2) and (3)) and the rest are devoted to compare two alternative models. Model results present the same regressions as the ones for the data for two competing models: one with oligopsonistic competition only $LS^{M,MP}_{h,t}$ (columns (3) and (4)) and 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. On the contrary, 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. These results support the mechanism of our structural model where union bargaining power and employer labor market power are relevant.

2.6 Counterfactuals

In this section we evaluate efficiency and welfare effects of the labor wedges coming from labor market power. We start by showing that counterfactuals can be computed observing establishment Revenue Total Factor Productivities (TFPRs) instead of the underlying productivities, which are unobserved. 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. That is, we characterize the competitive equilibrium allocation where wages are equal to the marginal revenue product. We also consider other counterfactual situations where labor wedges remain and are equal to the bargaining only or oligopsonistic competition only cases. These aim to disentangle the relative distortions coming from each side of labor market power. We perform the main counterfactuals for 2007, the last year of our sample.

[53] The first stage of the instrumental variable is highly significant (p-value<0.001) and the point estimate is negative. This is an expected result, as for a larger average productivity of the competitors the employment share should diminish. The OLS estimate is 0.851 with standard error 0.005. This is expected as all the variation comes from productivity changes, so the only thing shifting, leaving market structure constant, is the establishments demand. Thus the supply elasticity would be identified.

Table 20 – 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 16 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

Our baseline counterfactuals assume free mobility of labor. We perform three additional counterfactuals modifying 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 location-industry-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 industry level. Compared to the previous case, in this last counterfactual, 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 the de-industrialization process over time.

2.6.1 Counterfactuals using Revenue Productivities

This section shows that is possible to compute the counterfactuals in general equilibrium by using 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 on misallocation 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 perform general equilibrium counterfactuals that also take into account productivity differences across industries.[54] We show that we can: (i) carry out counterfactuals in general

[54] 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

equilibrium by writing the model in relative terms from a baseline scenario; and (ii) compute the movement of production factors across industries.

Our approach is to write counterfactual industry prices relative to the baseline and to fix the transformed revenue productivities.[55] Using the definition of the transformed revenue productivities Z_{io}, the equation (2.27) for nominal wage 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 a hat the relative variables (e.g. $\widehat{F}_b = \frac{F_b'}{F_b}$). Writing the model as deviations from a baseline scenario has been dubbed 'exact-hat-algebra' by Costinot and Rodríguez-Clare (2014). We can then rewrite the revenue productivity of a counterfactual as:

$$Z_{io}' = P'(F_b')^{1+\varepsilon_b\delta} A_{io} = \widehat{P}\widehat{F}_b^{1+\varepsilon_b\delta} Z_{io}.$$

The counterfactual revenue productivity is a function of the relative (transformed) price $\widehat{F}_b$ and the observed revenue productivity Z_{io}. Denoting by λ_{io}' the counterfactual wedge, the counterfactual real wages are:

$$\begin{aligned} w_{io}' &= \beta_b \lambda_{io}' Z_{io}' L_{io}'^{\,-\delta} \frac{1}{P'} \\ &= \beta_b \lambda_{io}' Z_{io} \frac{\widehat{F}_b^{1+\varepsilon_b\delta}}{P} L_{io}'^{\,-\delta}, \end{aligned} \tag{2.28}$$

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

The system (2.17) in the counterfactual writes as:

$$w_{io}' = \left(\beta_b \lambda_{io}' \frac{Z_{io}}{(T_{io}\Gamma_b^{\eta})^{\delta}}\right)^{\frac{1}{1+\varepsilon_b\delta}} \frac{\widehat{F}_b}{P^{\frac{1}{1+\varepsilon_b\delta}}} \Phi_m'^{\,(1-\eta/\varepsilon_b)\nu_b} \left(\frac{\Phi'}{L'}\right)^{\nu_b}, \tag{2.29}$$

where 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 establishment-occupation components are enough to compute the employment shares at the local labor market level, $s_{io|mo}'$, and at the industry level, $s_{m|b}'$. Following the same steps as in the baseline, the industry level system of equations is analogous to (2.22) but with relative variables and solving for relative industry prices we can compute the industry employment L_b'.[56] Propositions 8 and 9 apply and therefore the solution for the relative counterfactuals exists and is unique.

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 cannot 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.

[55] Solving the counterfactuals in levels as stated in Section 2.4 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 Total Factor Productivity, TFP) is constant across industries and get rid of industry relative prices by normalizing the minimum TFP per industry.

[56] Appendix 2.A provides the steps for the computation of the relative counterfactuals.

2.6.2 Main Counterfactuals

We consider four different counterfactual situations to evaluate the efficiency and welfare effects of labor market power in general, and of each of the sides of labor market power in particular. First, the main counterfactual characterizes the competitive equilibrium where labor wedges disappear and establishments and workers acts as price takers leading to the equalization of the wages and the marginal revenue products. Second, we present the counterfactual characterizing the limit case of our framework where there is only bargaining. Third, we have the limit case where employer labor market power is the only one present, and finally, a situation where unions collect all the profits.

Table 21 shows results of different counterfactuals under the free mobility assumption. The first column present labor shares in the baseline or 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 21. Eliminating labor wedges coming from employer and union labor market power increases aggregate output by 1.62%. Setting wages equal to the marginal revenue product induces efficiency gains that translate into output gains.

The second 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 none of the sides would internalize movements along the labor supply but bargain over the wages. The labor wedges become $\lambda(1, \varphi_b) = 1 + \varphi_b \frac{\delta}{1-\delta}$. It is important to note that the assumed institutional framework for the unions with heterogeneity only across industries makes labor wedges almost constant. This reduced heterogeneity of labor wedges (only different across industries) almost eliminates the allocative distortions and is behind the result of almost attaining the output gains of the main counterfactual.

Comparing now to the third counterfactual with employer labor market power, we see that output is reduced by 0.21% with respect to the baseline. Union bargaining power therefore attenuates the labor market distortions in our calibrated model. This reduction does not incorporate extensive margin responses of total employment as it is fixed. The mechanism behind the reduction in output is that labor wedges would be slightly more heterogeneous than in the baseline and distortions are amplified. 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.

In respect of the distributional effects or 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 Λ_b are sufficient statistics to compute the aggregate labor share which is a value added weighted sum of industry labor shares. Using the demand of the final good producer (2.4), the aggregate labor share is:[57]

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

Aggregate labor share is equal in all the variations of the main counterfactual without labor wedges that we present later. This comes from Λ_b being equal to one for all industries b.

[57] The derivation of the theoretical labor share is in Appendix 2.A.5.

Table 21 – Counterfactuals: Efficiency and Distribution

	LS (%)	Gains (%)		
		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
Bargain $\lambda(1, \varphi_b) = 1 + \varphi_b \frac{\delta}{1-\delta}$	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 *Bargain* is the standad bargaining framework where the workers' outside options are the competitive wages and they don't internalize movements along the labor supply. *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.

Column (1) of Table 21 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 reduced by 19% in the oligopsonistic case when the wedges become $\lambda(\mu, 0) = \mu_{io}$. Increases in the aggregate wage do not imply that wage inequality is reduced. Figure 29 in Appendix 2.H shows that the demeaned wage distributions are very similar on the baseline and the price taking counterfactuals (in Panel (a) and (b) respectively). This Figure highlights that even in the absence of labor wedges, wages across establishments are not equalized. This result is due to the idiosyncratic preferences of workers for different establishments.

Aggregate wage changes translate into welfare differences. We study the median expected welfare for workers. This median expected utility is:[58]

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

Column (4) of Table 21 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.

[58] 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.

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.

We perform three additional counterfactuals to locate the output gains in an environment with mobility costs. They differ in restrictions imposed on mobility where we allow mobility to happen only within industry, industry-occupation and local labor market. Table 22 compares the free mobility case with the restricted mobility cases. Comparing the output gains in 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.

Table 22 – Counterfactuals: Limited Mobility

			Contribution (%)		
	ΔY (%)	Δ Prod (%)	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

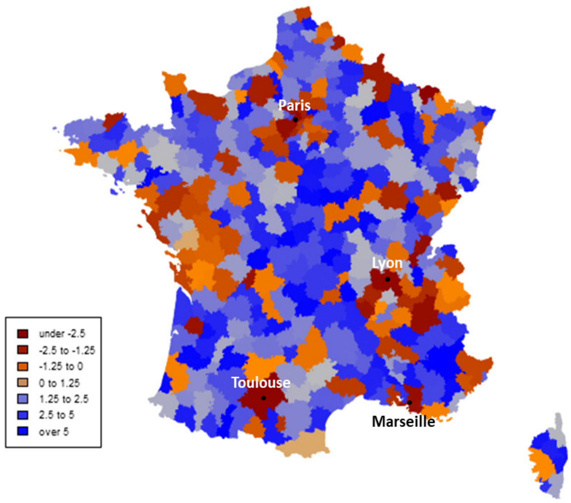
Notes: All the table presents results in percentages. First column presents the ΔY is the change of aggregate output with respect to the baseline, Δ *Prod* is the change in aggregate productivity from decomposition (2.30). Last three columns present the contribution of each of the elements of the decomposition (2.30) 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.

These results underscore the importance of free mobility of labor across locations as the main driver for output gains. Figure 19 shows the percentage change of manufacturing employment in the free mobility case. Each block is the aggregation of local labor markets to the commuting zone. 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 relative wage and employment gains.[59]

Turning now to the source of the output gains, we can use the aggregate production function and the relative industry output from Appendix 2.A (equation (43)), and decompose the logarithm

[59] 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.

Figure 19 – 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.

of the relative final output into three terms:

$$\ln \widehat{Y} = \underbrace{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{F}_b^{\alpha_b(1+\varepsilon_b \delta)}}_{\Delta \text{ GE}} + \underbrace{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{Z}_b}_{\Delta \text{ Productivity}} + \underbrace{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{L}_b^{1-\delta}}_{\Delta \text{ Labor}} . \tag{2.30}$$

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 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.

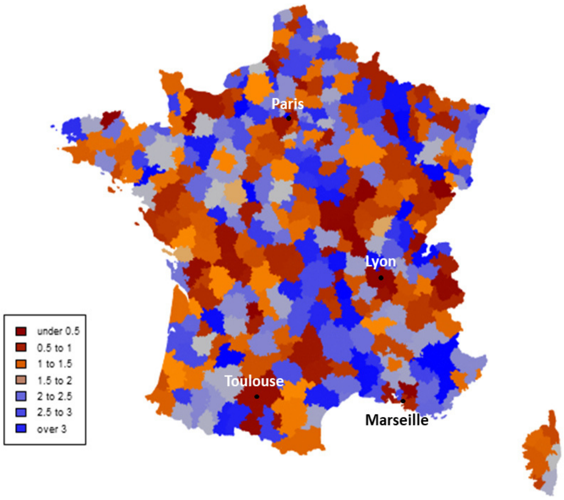
Columns (3) to (5) of Table 22 show the decomposition of relative changes of output.[60] The main source of output gains come from productivity. Industry productivity is an employment weighted sum of establishment-occupation productivities (that are unchanged). The original source of productivity and output gains is therefore 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 but mobility restrictions below the industry level prevent the reallocation towards productive establishments and reduce the productivity gains.

Figure 20 shows geographical differences of productivity gains in the free mobility case. The Figure is similar to Figure 19 in the sense that most significant gains of the counterfactual productivity happen outside urban areas. As a result, the largest gains relative to the baseline in wages and employment are in commuting zones without big cities.

[60] 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 $\frac{\sum_{b \in \mathcal{B}} \theta_b \ln \widehat{Z}_b}{\ln \widehat{Y}}$. Each row from columns 3 to 5 sums up to 1.

Figure 20 – 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 2.6.1, keeping fixed the baseline revenue productivities, any change in the counterfactual comes from changes in productivities. Counterfactuals are performed in 2007.

2.6.3 Pass Through

The efficiency losses are the consequence of distortions in the pass through of productivity to wages. The structural wage equation (2.29) relates our recovered measure of productivity Z_{io} to equilibrium wages. Taking logs, equilibrium wage in the baseline economy is:

$$\log w_{iot} = \frac{1}{1+\varepsilon_b\delta}\left(\log Z_{iot} - \delta \log T_{iot} + \log \lambda(\mu_{iot}, \varphi_b)\right) + f_{mt}, \tag{2.31}$$

where f_{mt} is a local labor market times year constant. 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:

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

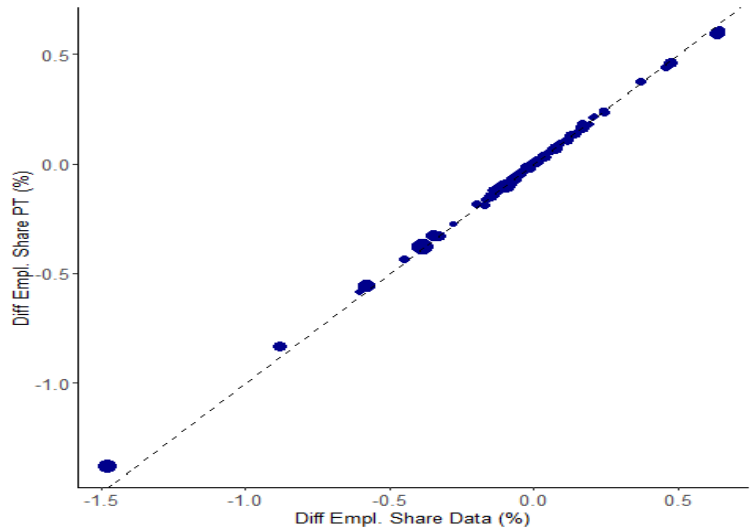
where ϵ_s^λ and ϵ_Z^s denote respectively the elasticity of the wedge λ_{io} with respect to the employment share s and the elasticity of the employment share s with respect to the transformed TFPR Z. The equation above emphasizes the origin of potential distortions coming from labor market power. When the wedge λ 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 33 in Appendix 2.J 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

Figure 21 – De-industrialization differences



Notes: The x-axis shows the percentage differences of commuting zone employment shares out of manufacturing 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 year is 1994 and the final one is 2007.

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.

2.6.4 Mobility and Wage Gap

Figure 19 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 urban-rural mobility over time and the urban-rural wage gap.

Mobility over time

We compare the urban-rural mobility process observed in the data to the one from yearly counterfactuals without labor market power. In the data, the de-industrialization or the reduction of manufacturing employment occurred primarily in cities leading to the gain in relative importance of rural areas within manufacturing. Figure 21 compares the relative employment shares observed in the data to the one in a counterfactual without labor wedges for each commuting zone.

First, we performed the main counterfactual where there are no labor wedges because establishments and unions act as price takers (PT) for the initial year 1994. Then 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. To compare the mobility over time, we compute the differences over time of the commuting zone employment shares in the data ($\Delta^D = S^D_{07} - S^D_{94}$) and in the counterfactual ($\Delta^M = S^{PT}_{07} - S^{PT}_{94}$). Figure 21 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 of the commuting zone. The counterfactual urban-rural mobility is very similar to the process observed in the data which is mostly guided by exogenous productivity and firm location decisions and not by labor market distortions.

The line generated by the largest population commuting zones in Figure 21 is slightly flatter than

Table 23 – 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.

the 45 degree line. Cities would loose their relative importance a bit slower in the counterfactual. A potential reason is the closure of manufacturing firms in the largest cities that became more concentrated over time leading to distortions that are closer to the ones present in rural areas.

Wage Gap

Table 23 presents wage levels and the urban/rural wage gap.[61] Both, urban and rural areas, experience important wage gains in the counterfactual. Gains being bigger outside cities the wage gap is reduced from 36% to 23% in the counterfactual. This reveals that labor market distortions account for more than a third of the urban/rural wage gap.

2.7 Extensions

We made important assumptions in the main counterfactual: workers were perfectly mobile, total labor supply was fixed and there were no agglomeration externalities. In this section, we propose extensions to relax the last two assumptions. First, we allow for an endogenous labor participation decision. Second, we introduce agglomeration forces in the local labor markets.

2.7.1 Endogenous Participation

We briefly present the extension with endogenous labor force participation decisions. We assume workers can decide between working and staying at home. In the latter case, they earn 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 as 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.

[61] 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.

Table 24 – Counterfactual: Endogenous Participation

	ΔY (%)	Δ Prod (%)	Δ L (%)	Contribution (%)		
				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
Bargain $\lambda(1, \varphi_b) = 1 + \varphi_b \frac{\delta}{1-\delta}$	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 (2.30) and ΔL is the counterfactual change in total employment. Last three columns present the contribution of each of the elements of the decomposition (2.30) 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. *Bargain* is the standad bargaining framework where the workers' outside options are the competitive wages and they don't internalize movements along the labor supply. *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.

Defining out-of-the-labor-force, from now on OTLF, as a new 3-digit 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 both, the employed and the OTLF workers. The proportion of OTLF workers in each local market identifies the home production amenity times the wage $T_{uo} w_{uo}^{\varepsilon_b}$.[62] This wage is fixed in the counterfactuals while the real wages of firms change depending on the counterfactual wedges.

Table 24 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 to the output gain decomposition in Table 22, around 30% of the gains come from the increased total employment. Labor force increases 1% in the main counterfactual without wedges. This extensive margin of 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 by -0.75%. Despite featuring high wage gains, the increase in total employment is minor in the counterfactual because we assume that workers have idiosyncratic shocks to stay OTLF.

[62] Details on the theoretical model with endogenous participation are in Appendix 2.B.

Table 25 – 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 (2.30). Last three columns present the contribution of each of the elements of the decomposition (2.30) 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.

2.7.2 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:

$$Pw_{io} = \beta_b \lambda(\mu_{io}, \varphi_b) Z_{io} L_{io}^{-\delta} L_m^{\gamma}. \tag{2.32}$$

Similarly to the baseline counterfactual, we back out the transformed TFPRs Z_{io} to perfectly match observed establishment-occupation wages w_{io} under the assumption of agglomeration externalities. 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.[63]

Table 25 summarizes the counterfactual results for different values of γ. All the counterfactuals in Table 25 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 (2.32). 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 (No Agglomeration). Output gains are monotonic in the importance of agglomeration externalities.

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

2.8 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 also present reduced form evidence on the heterogeneity of union bargaining power across industries within manufacturing.

We lay out a quantitative general equilibrium model that links structural labor wedges to employment shares and union 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 separately identify parameters leading to employer and union labor market power by implementing a recursive estimation strategy. We first estimate global parameters by imposing restrictions on structural shocks, second we identify the industry specific labor supply elasticities using the structural equation of the labor supply and then we calibrate the rest of parameters to match industry moments.

We evaluate the efficiency and welfare costs of employer and union labor market power. We find that removing structural labor wedges increases output by 1.62%. Gains are amplified 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 establishments. Removing labor market distortions also leads to significant labor share and wage gains. These results imply that the employer labor market power is more important than the one of unions on the labor wedge for manufacturing in France.

The counterfactual suggests that there is missing employment in French rural areas due to employer labor market power. Eliminating these distortions would not only increase wages but also the efficiency of manufacturing. The potential insights for policy are clear. Our calibrated model suggests that unions counteract employer labor market power but promoting unions would not completely overcome the distributional and efficiency effects. On the contrary, 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. Unfortunately, the implementation of this policy would be very cumbersome as it would require taking into account the structure of the labor market and the fundamentals of establishments. Alternatively, the efficiency gains could be partially achieved by attracting employment to remote locations. Place based policies aimed to improve the amenities of rural areas would possibly trigger the employment gains necessary to fulfill the output gains from removing employer labor market power.

2.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 establishment (inverse) 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.

2.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:

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

2.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 (2.13) 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.[64] Getting rid of the time index t, the first order conditions of this problem are:

$$\begin{aligned} w_{io} &= \beta_b \frac{e_{io}}{e_{io}+1} P_b \widetilde{A}_{io} K_{io}^{\alpha_b} L_{io}^{\beta_b - 1}, \\ R_b &= \alpha_b P_b \widetilde{A}_{io} K_{io}^{\alpha_b - 1} L_{io}^{\beta_b}. \end{aligned} \tag{33}$$

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

[64] The construction details of the rental rate of capital or the required rate are in Appendix 2.F.4.

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:

$$y_{io} = \left(\frac{\alpha_b}{R_b}\right)^{\frac{\alpha_b}{1-\alpha_b}} \widetilde{A}_{io}^{\frac{1}{1-\alpha_b}} L_{io}^{\frac{\beta_b}{1-\alpha_b}} P_b^{\frac{\alpha_b}{1-\alpha_b}}. \tag{34}$$

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:

$$y_{io} = P_b^{\frac{\alpha_b}{1-\alpha_b}} A_{io} L_{io}^{1-\delta}. \tag{35}$$

2.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:

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

2.A.4 Bargaining Details

We provide derivations under the baseline bargaining protocol where employers and unions have zero outside options and internalize movements along the labor supply curve. An alternative bargaining protocol leading to the same equilibrium condition for the wages is at the end.

Each establishment has different occupation profit functions $(1-\alpha_b)P_bF(L_{io}) - w^u_{io}L_{io}$, where the optimal capital decision has been taken. We assume that workers and establishments are symmetric both having null threat points and internalizing the generation of rents as they move along the labor supply curve.

During the bargaining establishments and unions choose wages to maximize:

$$\max_{w^u_{io}} [w^u_{io}L_{io}(w^u_{io})]^{\varphi_b} \left[(1-\alpha_b)P_bF(L_{io}(w^u_{io})) - w^u_{io}L_{io}(w^u_{io})\right]^{1-\varphi_b},$$

where we made explicit the fact that both parties internalize how labor supply is a function of equilibrium wages. φ_b is 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, $(1-\alpha_b)F(L_{io})$ is the output of the establishment-occupation after substituting for the optimal decision of capital. The first order conditions of the above maximization problem are:

$$\varphi_b \frac{(1-\alpha_b)P_bF(L_{io}) - w^u_{io}L_{io}}{w^u_{io}L_{io}} \left[L_{io} + w^u_{io}\frac{\partial L_{io}}{\partial w^u_{io}}\right] + (1-\varphi_b)\left[(1-\alpha_b)P_b\frac{\partial F(L_{io})}{\partial L_{io}} - L_{io} - w^u_{io}\frac{\partial L_{io}}{\partial w^u_{io}}\right] = 0.$$

Using the definition of the perceived labor supply elasticity $e_{io} = \dfrac{\partial L_{io}}{\partial w_{io}}\dfrac{w_{io}}{L_{io}}$ and rearranging the first order condition:

$$w^u = \varphi_b(1-\alpha_b)P_b\frac{F(L_{io})}{L_{io}} + (1-\varphi_b)(1-\alpha_b)P_b\frac{\partial F(L_{io})}{\partial L_{io}}\frac{e_{io}}{e_{io}+1},$$

where $\mu(s_{io}) \equiv \dfrac{e_{io}}{e_{io}+1}$ is the markdown that establishments would set under oligopsonistic competition.

In the case of a Cobb-Douglas production function, the marginal revenue product of labor is proportional to the labor productivity, i.e. $(1-\alpha_b)p\frac{\partial F(L_{io})}{\partial L_{io}} = \beta_b\frac{pF(L_{io})}{L_{io}}$, 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:

$$w^u = \underbrace{(1-\alpha_b)P_b\frac{\partial F(L_{io})}{\partial L_{io}}}_{MRPL_{io}}\left[(1-\varphi_b)\frac{e_{io}}{e_{io}+1} + \varphi_b\frac{1}{1-\delta}\right],$$

where we recovered the expression from the main text.

The alternative bargaining assumption leading to the same equilibrium wages is that employers and unions bargain over the wages without internalizing movements along the labor supply and workers' outside options are the oligopsonistic competition wages w^M_{io} under the allocation with the given equilibrium wages. This alternative protocol is quite ad-hoc as the employer labor market power is embedded in the workers outside options. The bargaining problem would be:

$$\max_{w^u_{io}} \left[w^u_{io}L_{io} - w^M_{io}\right]^{\varphi_b} \left[(1-\alpha_b)P_bF(L_{io}) - w^u_{io}L_{io}\right]^{1-\varphi_b}.$$

2.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 (2.17),

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

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

$$\Phi_m = \widetilde{\Phi}_m^{\psi_b} F_b^{\varepsilon_b \psi_b} \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \varepsilon_b}, \quad \widetilde{\Phi}_m = \sum_{i \in I_m o} T_{io} \widetilde{w}_{io}^{\varepsilon_b}, \quad \psi_b \equiv \frac{1+\varepsilon_b \delta}{1+\eta\delta} \geq 1, \tag{36}$$

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,

$$w_{io} = \widetilde{w}_{io} \widetilde{\Phi}_m^{\frac{\psi_b - 1}{\varepsilon_b}} F_b^{\psi_b} \left(\frac{\Phi}{L}\right)^{\frac{\delta}{1+\eta\delta}}. \tag{37}$$

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 (36),

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

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 (2.5):

$$Y_b = F_b^{\alpha_b(1+\varepsilon_b\delta)} A_b L_b^{1-\delta}, \quad A_b = \sum_{m \in \mathcal{M}_b} \sum_{io \in \mathcal{I}} A_{io} s_{io|m}^{1-\delta} s_{m|b}^{1-\delta}, \tag{38}$$

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

$$L_b = \frac{\Phi_b \Gamma_b^\eta}{\sum_{b' \in \mathcal{B}} \Phi_{b'} \Gamma_{b'}^\eta} L = \frac{F_b^{\psi_b \eta} \widetilde{\Phi}_b \Gamma_b^\eta}{\widetilde{\Phi}} L, \quad \widetilde{\Phi} = \sum_{b' \in \mathcal{B}} F_{b'}^{\psi_b \eta} \widetilde{\Phi}_{b'} \Gamma_{b'}^\eta. \tag{39}$$

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 (36) 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 (2.4) 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 (2.23) 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 (2.16) to the industry level,

$$w_b L_b = \beta_b \Lambda_b P_b Y_b, \tag{40}$$

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 (34) and (2.21), the industry markdown Λ_b yields the following expression:

$$\Lambda_b = \frac{\sum_{io \in \mathcal{I}_b} \lambda_{io} A_{io} s_{io|m}^{1-\delta} s_{m|b}^{1-\delta}}{A_b}. \tag{41}$$

Industry and aggregate labor shares are:

$$LS_b = \beta_b \Lambda_b, \quad LS = \frac{\sum_{b \in \mathcal{B}} w_b L_b}{\sum_{b \in \mathcal{B}} P_b Y_b}. \tag{42}$$

Substituting (40) 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.

2.A.6 Hat Algebra

From the main text, we get that the counterfactual wage w'_{io} from (2.29) can be written as: $w'_{io} = \omega_{io} \frac{\widehat{F}_b}{P^{\frac{1}{1+\varepsilon_b \delta}}} \Phi'^{(1-\eta/\varepsilon_b)\nu_b}_m \left(\frac{\Phi'}{L'}\right)$ 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,

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

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

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

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

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:

$$\widehat{Y}_b = \widehat{F}_b^{\alpha_b(1+\varepsilon_b\delta)} \widehat{Z}_b \widehat{L}_b^{1-\delta}, \tag{43}$$

where $\widehat{Z}_b = \frac{Z_b(s')}{Z_b(s)}$. Using L'_b and equation (2.4) we get,

$$\widehat{F}_b^{\psi_b(1+\eta)} \widehat{Z}_b \left(\frac{\widetilde{\Phi}'_b \Gamma_b^\eta}{L_b}\right)^{1-\delta} = \prod_{b' \in \mathcal{B}} \left(\widehat{F}_{b'}^{\alpha_b(1+\varepsilon_b\delta)+(1-\delta)\psi_b\eta}\right)^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \widehat{Z}_{b'}^{\theta_{b'}} \prod_{b' \in \mathcal{B}} \left(\frac{\widetilde{\Phi}'_{b'} \Gamma_{b'}^\eta}{L_{b'}}\right)^{(1-\delta)\theta_{b'}}. \tag{44}$$

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 (2.23) in Proposition 9 with hat variables:

$$\widehat{F}_b = \widehat{X}_b \widehat{C}^{\frac{1}{\psi_b(1+\eta)}}, \tag{45}$$

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

Fixed Labor

In the case where employment is fixed at the industry level b, the counterfactual wage (2.29) 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 (44). 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 8 and 9 therefore also apply in the relative counterfactuals with fixed labor at the industry level b (or at a lower level).

2.B Extensions

2.B.1 Endogenous Participation

We showed in the proof of Proposition 9 that the solution of transformed prices **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 2.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 (36),

$$\Phi_{b,e} = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \sum_{m \in \mathcal{M}_b} \widetilde{\Phi}_m^{\psi_b \eta/\varepsilon_b} F_b^{\psi_b \eta} \Gamma_b^{\eta} = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \widetilde{\Phi}_{b,e} F_b^{\psi_b \eta} \tag{46}$$

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

and,

$$\Phi_e = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \sum_{b \in \mathcal{B}} \widetilde{\Phi}_{b,e} F_b^{\psi_b \eta} \Gamma_b^{\eta} = \left(\frac{\Phi}{L}\right)^{\psi_b \nu_b \eta} \widetilde{\Phi}_e \tag{47}$$

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

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 (47),

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

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.

2.B.2 Agglomeration

Plugging the labor supply into (2.32), 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,

$$\Phi_m = \widetilde{\Phi}_m^{\widetilde{\psi_b}} P^{-\frac{\widetilde{\psi_b}\varepsilon_b}{1+\varepsilon_b\delta}} \left(\frac{\Phi}{L}\right)^{\widetilde{\psi_b}\widetilde{\nu_b}\varepsilon_b}, \quad \widetilde{\psi_b} \equiv \frac{1+\varepsilon_b\delta}{1+\eta(\delta-\gamma)}. \tag{48}$$

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:

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

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

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:

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

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:

$$\widehat{F}_b^{1+\varepsilon_b\delta}\widehat{Z}_b\left(\frac{L'_b(\widehat{\mathbf{F}})}{L_b}\right)^{1-\delta+\gamma} = \prod_{b'\in\mathcal{B}}\widehat{F}_{b'}^{\alpha_{b'}(1+\varepsilon_b\delta)}\widehat{Z}_{b'}\left(\frac{L'_{b'}(\widehat{\mathbf{F}})}{L_{b'}}\right)^{1-\delta+\gamma}$$

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

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 9, uniqueness of the equilibrium with agglomeration forces only needs the additional requirement of $\gamma \neq \frac{1}{\eta} + \delta$.

2.C Proofs

Proof of Proposition 8.

Existence. We follow closely the proof by Kucheryavyy (2012). Define the right hand side of (2.17) 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:

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

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 (2.19). □

Uniqueness. First we introduce the following Theorem and Corollary that we will use later to establish uniqueness in our proofs. These are transcribed from Allen et al. (2016) as they are not present any more in the current version of the paper Allen et al. (2020a):

Theorem 2. *Consider $g : \mathbb{R}^n_{++} \times \mathbb{R}^m_{++}$ for some $n \in \{1, ..., N\}$ and $m \in \{1, ..., M\}$ such that:*

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

(ii) gross-substitution property: $\frac{\partial g_i}{\partial x_j} > 0$ for all $i \neq j$,

(iii) 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. We proceed by contradiction. Suppose there are two different up-to-scale, solutions, x^1, x^2, such that $f(x^1) = f(x^2) = 0$ i.e. $g(x^1, y^0) = g(x^2, y^0) = 0$. Without loss of generality, suppose there exists some $t > 1$ such that $tx^1_j \geq x^2_j$ for all $j \in \{1, ..., n\}$ and the equality holds for at least one $j = \bar{j}$. Then the inequality must strictly hold since x^1 and x^2 are different up-to-scale. Condition (iii) $\frac{\partial g_i}{\partial y_k} \geq 0$, for all i, k implies $g(tx^1, y^0) \leq g(tx^1, ty^0) = 0$ where $g(tx^1, ty^0) = 0$ is from condition (i) (and also $g(tx^2, ty^0) = 0$ because x^1 and x^2 are solutions). However, condition (ii) implies $g_j(tx^1, y^0) > g_j(x^2, y^0) = 0$, thus a contradiction. □

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. Define $m(x, y)$ as a vector function where $m_i(x, y) = \sum_{j=1}^{\nu_f} y^{\bar{\alpha} - \alpha_j} g^j_i(x) - \sum_{k=1}^{\nu_g} y^{\bar{\alpha} - \beta_k} h^k_i(x)$. Obviously, $m(x, y)$ is of homogenous degree $\bar{\alpha}$ and $\frac{\partial m_i}{\partial y} \geq 0$. Also we have $f(x) = m(x, y^0)$ where $y^0 = 1$, thus the above theorem applies.

Furthermore, if $f_i(x)$ is not homogeneous of some degree because $\alpha_j \neq \beta_k$, there is at most one solution. Suppose not, tx^1 and x^1 are the solutions, then $f_i(x^1) > t^{-min(\beta_k)} f_i(tx^1) = 0$, also a contradiction. □

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 m 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.[65] 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 $g(\mathbf{w}) = 0$. □

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

Proof of Proposition 9.

Developing equation (2.22) we get

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

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

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:

$$\mathbf{f} = \mathbf{C} + \mathbf{D}'\mathbf{f}. \tag{49}$$

A solution to the system (49) 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:

$$\sum_b \mathbf{d}_b = 1 \Leftrightarrow \sum_b \left(\alpha_b(1+\eta\delta) + \eta(1-\delta) \right) \theta_b = 1+\eta$$

$$\Leftrightarrow \sum_b \alpha_b \theta_b (1+\eta\delta) = 1+\eta-\eta(1-\delta) \Leftrightarrow \sum_b \alpha_b \theta_b = \frac{1+\eta-\eta(1-\delta)}{1+\eta\delta} \Leftrightarrow \sum_b \alpha_b \theta_b = 1.$$

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 $0 \leq \beta_b, \theta_b < 1 \ \forall b \in \mathcal{B}$ and $0 \leq \delta \leq 1$.

In order to obtain the closed form solution, rewrite (2.22) 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 9. □

2.D Identification Details

2.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 (2.13) 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 (2.16) 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:

$$\begin{aligned}\overline{\ln(L_{io})} &= \eta \overline{\ln(w_{io})} + \frac{\eta}{\varepsilon_b} \overline{\ln(T_{io})}, \\ \overline{\ln(w_{io})} &= -\delta \overline{\ln(L_{io})} + \overline{\ln(A_{io})}.\end{aligned} \tag{50}$$

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 (50) we can not identify the elasticities.[66]

In order to identify the elasticities using the labor supply and demand equations in (50), 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 variances of the transformed productivities and amenities for each of the four occupational categories.[67]

[66] 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.

[67] 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.

We can rewrite the system (50) in the following way:

$$\frac{\eta}{\varepsilon_b}\overline{\ln(T_{io})} = \overline{\ln(L_{io})} - \eta\,\overline{\ln(w_{io})},$$
$$\overline{\ln(A_{io})} = \delta\,\overline{\ln(L_{io})} + \overline{\ln(w_{io})}. \tag{51}$$

Denote the covariance matrix of the structural errors for occupation o in category S (meaning the left hand side of system (51)) by Ψ_{oS}. Denote the covariance matrix between employment and wages of the full monopsonists by $\widehat{V}_{oS}$. The covariance of system (51) 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 2.D.

The system (51) 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:

$$\widetilde{\delta}^2[\omega_{11,1}\omega_{12,2} - \omega_{11,2}\omega_{12,1}] - \widetilde{\delta}[\omega_{11,1}\omega_{22,2} - \omega_{11,2}\omega_{22,1}] + \omega_{12,1}\omega_{22,2} - \omega_{12,2}\omega_{22,1} = 0. \tag{52}$$

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 (50) 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 (52).

2.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:

$$\mathbb{E}_t \left[LS^M_{bt}(\varphi_b) - LS^D_{bt} \right] = 0. \tag{53}$$

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 (53) with $\varphi_b \in [0, 1]$ this has to be unique.[68]

2.D.3 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 (2.13), 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:

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

2.E Additional Estimation Results

Table 27 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.

[68] 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 26 – 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.

2.F Data Details

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

2.F.1 Additional Summary Statistics

Table 30 – 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.

Table 27 – 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).

2.F.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.[69] Before 2000 we take table sources in euros and from 2001 onward we use data from consolidated economic units.[70] After excluding firms without firm identifier the raw data sample contains about 29 million firms from which about 2.8 million are manufacturing firms.[71] Manufacturing sector (sector code equal to *D*) constitutes on average 10% of the observations, 19.2% of value added and 27.2% of employment.

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

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

[70] 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.

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

Table 28 – 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.

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.[72]

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

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

Table 29 – 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.

zone level.[73] 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 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.[74]

2.F.3 Variable Construction

Ficus:

- Value added: value added net of taxes (*VACBF*). We restrict to firms with strictly positive value added.[75]
- 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*).[76]

[73] 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.

[74] We get rid of wage per capita outliers by truncating the sample at the 0.5% below and 99.5%.

[75] We follow the advise of the French statistical instiute (INSEEE) in using net value added to perform comparisons across industries.

[76] For firms declaring at the BIC-BRN regime (*TYPIMPO*= 1) we only take *SALTRAI* .

- 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.[77] 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.[78]

2.F.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 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:

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

[77] Occupations with first digit 1 and 7 are excluded. They constituted less than 0.05% of the matched sample.

[78] City codes are the concatenation of department (*DEP*) and city (*COM*).

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

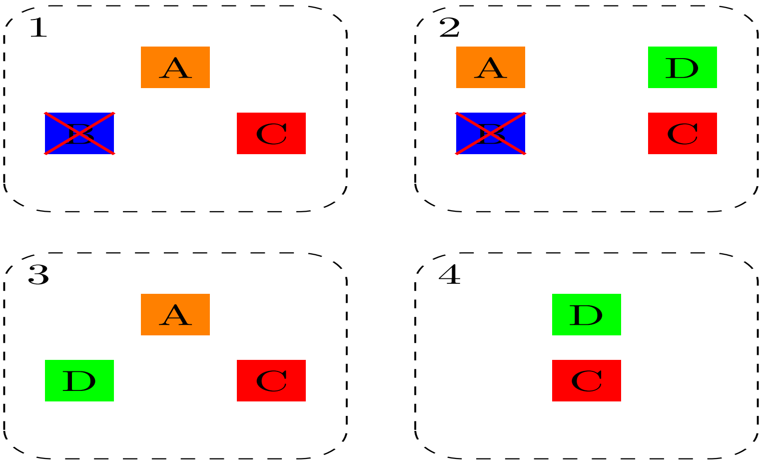
$$\sum_s R_{sb} P^K_{sb} K_{sb} = \underbrace{\sum_s \frac{P^K_{sb} K_{sb}}{\sum_{s'} P^K_{s'b} K_{s'b}} R_{sb}}_{R_b} \underbrace{\sum_s P^K_{sb} K_{sb}}_{P^{Kb} K_b},$$

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

2.G Empirical Evidence

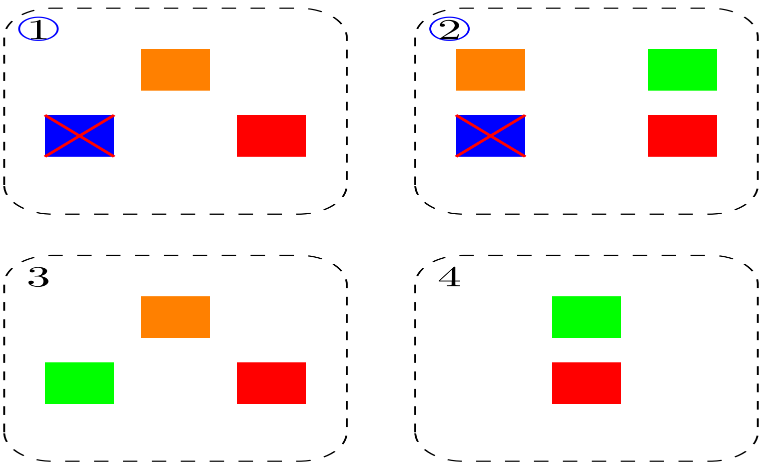
Example of an economy with four local labor markets and four firms identified by color. Each firm is multi-location with plants at different local labor markets. The blue firm is affected by a mass layoff at the national level (in all the local markets where it is present). Natural experiment on $s_{io|m}$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

Figure 22 – Local Labor Markets with and without shock



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

Figure 23 – Treated Establishments



The first order condition for wages is:

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

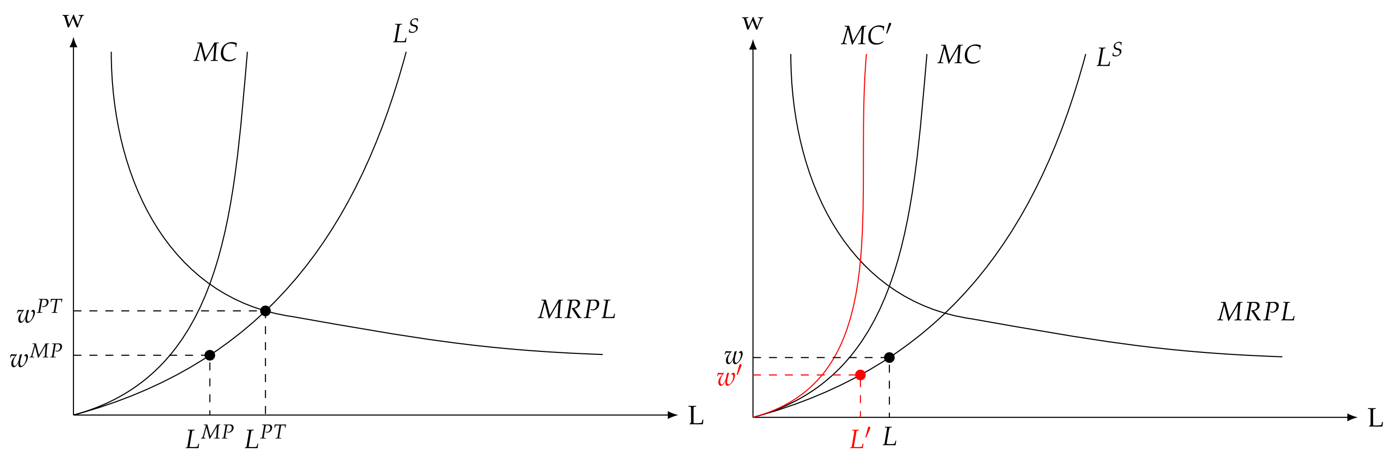
where the right hand side is the marginal cost ($\frac{\partial w_{io} L_{io}}{\partial L_{io}}$) when internalizing movements along the labor supply curve. Noting that $\frac{\partial w_{io}}{\partial L_{io}} \frac{L_{io}}{w_{io}} = \frac{1}{e_{io}}$ is the inverse of the labor supply elasticity e_{io}, 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 24 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 24 gives an intuition of our instrument.

Figure 24 – Instrument



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

(b) Instrument

2.G.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.[79] The first stage is:

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

[79]A firm j at occupation o is hit by a negative shock if $\mathbb{1}\{L_{io,t}/L_{io,t-1} < \kappa\, \forall i \text{ where } \mathbf{J}(i), t = j\} = 1$. 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:

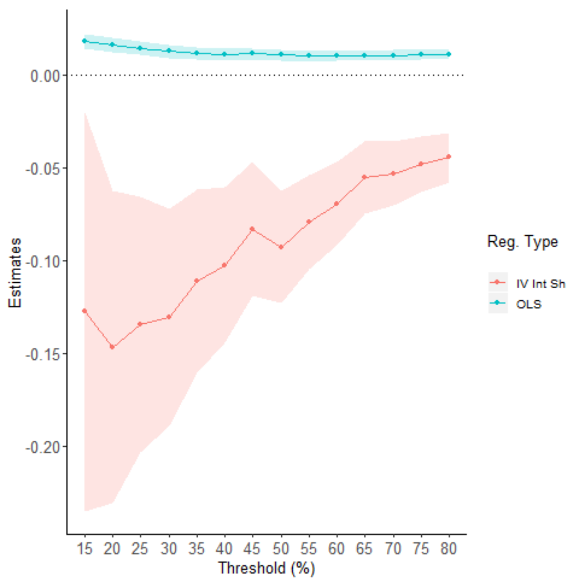
$$\log(w_{io,t}) = \psi_{\mathbf{J}(i),o,t} + \delta_{\mathbf{N}(i)} + \alpha\, \widehat{s_{io|m,t}} + u_{io,t} \tag{54}$$

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.[80]

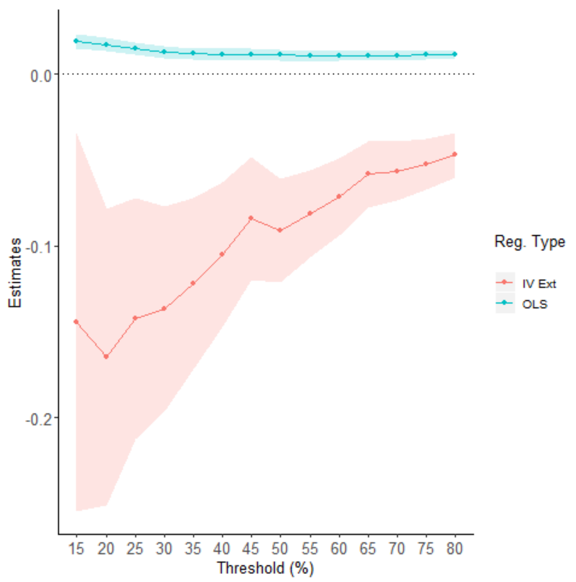
2.G.2 Robustness Checks

Figure 25 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 25 – Robustness



(a) Instrument: Intensive Share



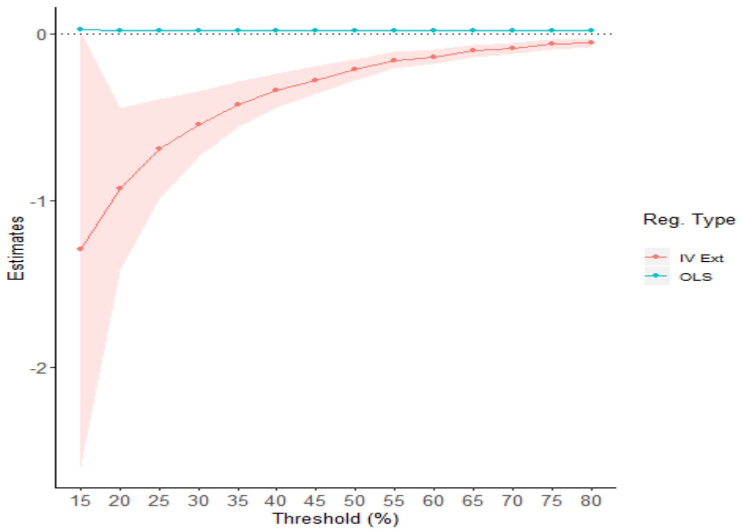
(b) CZ-year fixed effects

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.

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

Figure 26 – Robustness. Local Labor Market at 2-digit Industry



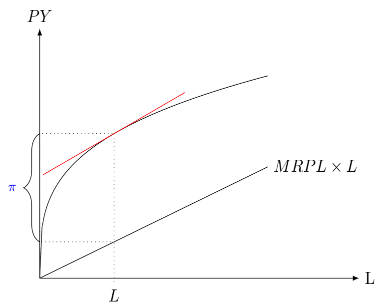
Notes: This figure presents the point estimates and 95% confidence bands of the OLS and IV exercises on the y-axis. The x-axis presents different thresholds κ that define a mass layoff shock. We focus on non-affected competitors (not suffering a mass layoff shock). The instrument is the presence of a 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 24 is that the local labor market is at 2-digit rather than 3-digit industry.

2.H Distributional and Efficiency Consequences

Here we illustrate the distributional and efficiency effects when the labor wedge λ is below one. Figure 27 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 where employer monopsony power dominates. Wages are below the marginal revenue product because the wedge λ is below one. This generates additional profits for the firm, reducing wage bill payments and therefore the labor share.

Figure 27 – Distributional Consequences

(a) Perfect Competition



(b) Labor Market Power

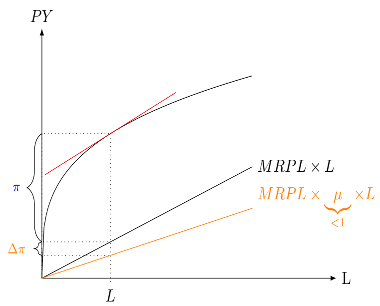


Figure 28 shows the efficiency consequences due to the misallocation of resources. The left panel shows two firms with the same labor wedge. For simplicity we assume that all firms and local labor markets have the same amenities so workers being indifferent, all establishments will have the same wage in equilibrium. With homogeneous wedges, the marginal revenue products are equalized across establishments. In particular, if firm B is more productive we have in equilibrium $L_B > L_A$. On the right panel we show an example with heterogeneous wedges. 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, $\mu_B < \mu_A$ and therefore $\lambda_B < \lambda_A$. 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 $L_A > L_B$. Shifting employment from A to B, from low to high marginal revenue product firms, there could be efficiency gains.

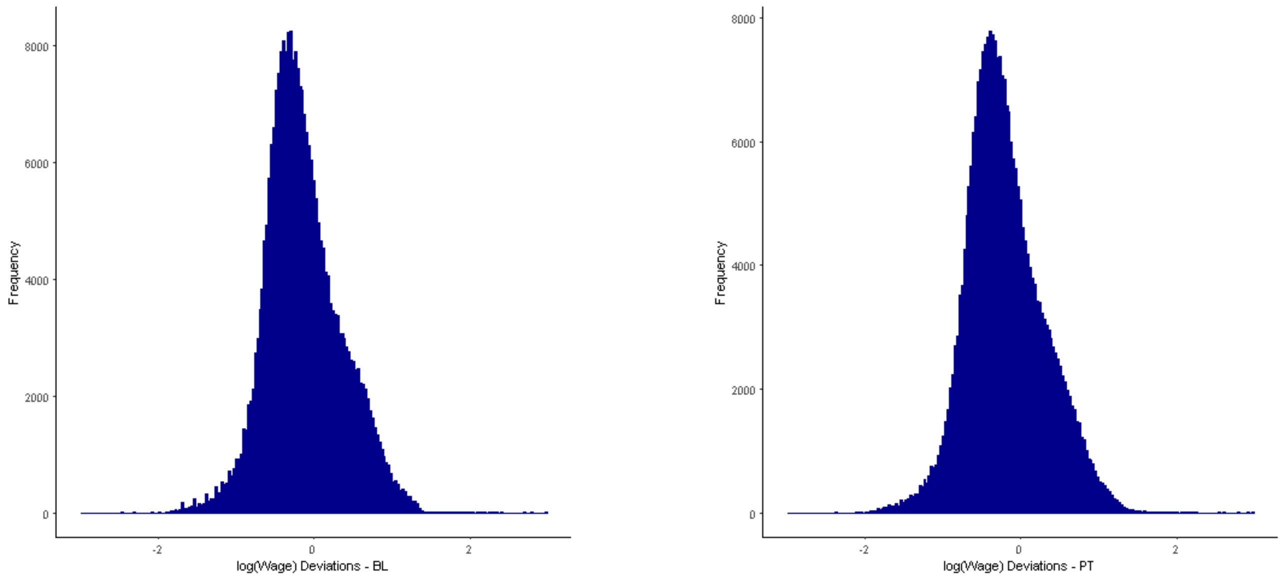
Figure 28 – Efficiency Consequences

(a) Homogeneous Wedges

(b) Heterogeneous Wedges

The next Figure shows the wage inequality in the baseline scenario and in the counterfactual without labor wedges when workers and firms act as price takers. Wages are demeaned.

Figure 29 – Wage Distribution



(a) Baseline Demeaned Wage Distribution

(b) Counterfactual (PT) Demeaned Wage Distribution

2.H.1 Union

Tables 31 and 32 present respectively the rent sharing elasticities for industries and occupations.

Table 31 – 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 32 – 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

2.I 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:

$$y_i = \widetilde{A}_i K_i^{\alpha_b} H_i^{\beta_b} = \widetilde{A}_i K_i^{\alpha_b} \left(\prod_{o \in \mathcal{O}} L_{io}^{\gamma_o} \right)^{\beta_b}, \quad \sum_o \gamma_o = 1, \quad \alpha_b + \beta_b \leq 1. \tag{55}$$

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 (2.13) 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,

$$w_{io} = \left[\lambda(\mu_{io}, \varphi_b) \frac{\gamma_o}{T_{io} \Gamma_b^{\eta}} (\beta_b A_i)^{\frac{1+\varepsilon_b}{1+\varepsilon_b \delta}} (\Upsilon (T_i \Gamma)^{1/\varepsilon_b})^{\frac{\varepsilon_b(1-\delta)}{1+\varepsilon_b \delta}} \left(\prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \left(\prod_{o' \in \mathcal{O}} \Phi_{m'}^{(\eta/\varepsilon_b - 1)\gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \Phi_m^{1-\eta/\varepsilon_b} \right]^{\frac{1}{1+\varepsilon_b}} \tag{56}$$

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

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:

$$w_{io} = \left[\lambda(\mu_{io}, \varphi_b) c_{io} \left(\prod_{o' \in \mathcal{O}} \lambda(\mu_{io'}, \varphi_b)^{\varepsilon_b \gamma'_o} \right)^{\frac{1-\delta}{1+\varepsilon_b \delta}} \frac{\Phi_m^{1-\eta/\varepsilon_b}}{\prod_{o' \in \mathcal{O}} \Phi_{m'}^{(1-\eta/\varepsilon_b)\gamma'_o}} \right]^{\frac{1}{1+\varepsilon_b}} C_l F_b. \tag{57}$$

The last system is equivalent to the one in (56) 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:

$$\widetilde{w}_{io} \equiv \left[\lambda(\mu_{io},\varphi_b)c_{io}\left(\prod_{o'\in\mathcal{O}}\lambda(\mu_{io'},\varphi_b)^{\varepsilon_b\gamma_o'}\right)^{\frac{1-\delta}{1+\varepsilon_b\delta}} \frac{\widetilde{\Phi}_m^{1-\eta/\varepsilon_b}}{\prod_{o'\in\mathcal{O}}\widetilde{\Phi}_{m'}^{(1-\eta/\varepsilon_b)\gamma_o'}}\right]^{\frac{1}{1+\varepsilon_b}}. \tag{58}$$

$\widetilde{w}_{io}$ 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 (58) 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 (58), define $\widehat{w}_{io}$ as:

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

$$w_{io} = \widehat{w}_{io}\left[\frac{\widetilde{\Phi}_m^{1-\eta/\varepsilon_b}}{\mathcal{L}_l}\right]^{\frac{1}{1+\varepsilon_b}} C_l F_b = \widehat{w}_{io} z_l = \widetilde{w}_{io} C_l F_b. \tag{59}$$

We can show that the system formed by (59) has a solution and is unique.

Proposition 10. *For given parameters $0 \le \alpha_b, \beta_b < 1$, $1 < \eta < \varepsilon_b$, $0 \le \delta \le 1$, 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* (59).

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 (59) exists and is unique. $\square$

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 2.A.5 of the paper.

2.J Pass Through

Table 33 – 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 (2.31) 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 3

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. The direct computation for a bias correction is not feasible when the number of covariates is large. We propose a bootstrap method for correcting this bias that accommodates different assumptions on the structure of the error term including general heteroscedasticity and serial correlation. Our approach is suited to correct variance decompositions and 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 this to other methods in the literature. Using administrative data for France, we apply our method by carrying out 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 and depending on the specification that the correlation becomes positive after the correction.

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 Fabrice Collard, Thomas Crossley, Patrick Fève, Simen Gaure, Silvia Goncalves, Cristina Gualdani, Koen Jochmans, Elia Lapenta, Tim Lee, Thierry Magnac, Nour Meddahi, Jean-Marc Robin and Uta Schönberg for helpful comments. We gratefully acknowledge financial support from TSE to get access to the administrative data provided by CASD. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'avenir' program (reference : ANR-10-EQPX-17 – Centre d'accès sécurisé aux données – CASD). Zerecero acknowledges the financial support from CONACYT (reference: 329103/383874). All errors are our own.

3.1 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 the 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 main application of the method is the correction of variance decompositions of multi-way fixed effects. The method is very similar to MacKinnon and Smith Jr (1998) when the bias is flat or independent of the initial estimates, but is more efficient than one that follows their approach directly. Compared to other methods in the literature, our estimator also has the benefit of being fast and easy to implement 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 heteroscedastic and when there is serial correlation or clustering of the errors. Our procedure consists of a wild bootstrap under the assumption of diagonal covariance matrix, and a

[2]By large matrix we mean a matrix with dimension in the order of hundreds of thousands or millions. Instead of computing the inverse directly, researchers usually rewrite the object of interest as a system of linear equations that can be solved by preconditioned conjugate gradient methods. A particular example of such systems are the normal equations in OLS regressions.

wild block bootstrap (Cameron et al., 2008) for those that are non-diagonal that is valid for unrestricted dependence of the error terms within group and heteroscedasticity. The method is flexible in the definition of the group and therefore allows for example clustering of the errors depending on the geographical area or serial correlation within the worker-firm match.

We apply our method to French administrative data and perform a variance decomposition of an estimated AKM type model. Consistent with the 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 correlation at the connected set passes from -0.10 to -4.1e-04 under the assumption of serial correlation of the error terms within the match.[3] Abowd et al. (2004), also using 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 8% 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. (2020). 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.[4] He 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*.[5]

Kline et al. (2020) (KSS henceforth) follow a similar approach to Gaure (2014) in estimating the small sample bias of second order moments. They compute the trace term leading to the bias and estimate the covariance matrix 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 differs 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 to be faster and more flexible. Monte Carlo

[3]In a more restricted sample fulfilling the requirements to apply the method of Borovičková and Shimer (2017), the estimated correlation passes from -0.06 to 0.08 after the correction.

[4]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).

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

simulations show that our correction is between 30% to 60% faster than KSS, has similar accuracy and is suited to perform a full variance decomposition while their method focuses on the two main fixed effects. Nevertheless, KSS go one step further and propose how we can 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 estimating a bias correction for each interested quadratic form, as it requires solving a large system of linear equations in each iteration that are particular to each quadratic form. 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 almost feasible 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. Regardless of the number of moments to correct, we need to solve one system of linear equations per bootstrap, while with the Gaure and KSS methods one needs to solve as many systems of equations per iteration as needed corrections.[6] They implement corrections of the second order moments of the two leading fixed effects while we can directly perform a full variance decomposition, which is therefore suited for corrections on multi-way fixed effect regressions.

Following the work of MacKinnon and Smith Jr (1998) bootstrap methods have been used to 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 with a special focus on two-way fixed effects and to compare it to other methods 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.[7] The first restriction implies that the sample used for BS is about half of the original sample of private firms.[8] Suggestive of the potential sample selection issues is that the plug-in estimate of the correlation between worker and firm fixed effects is -0.10 under the original data whereas is -0.06 under the connected set generated from BS data. Both approaches now yield different estimates. They estimate a correlation between worker and firm types of 0.56 while we estimate correlations of 0.09 or 0.16 depending on the specification. We estimate our corrections of

[6] 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 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*.

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

[8] The original data of private firms has 5.8 million observations while after filtering of two job and worker restrictions the sample has only 2.5 million observations.

the variance components at the connected set originated from the BS data restrictions.

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. (2019). Recently Jochmans and Weidner (2019) characterize the dependence between the bias of nonlinear functionals on the parameters and the network structure of the data allowing for more than second order moments. Exploring these dependencies is out of the scope of this paper.

3.2 The Bias

For clarity of exposition we layout the source of the bias characterized by Andrews et al. (2008). Consider the following linear model:[9]

$$Y = X\beta + u, \tag{3.1}$$

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

$$\widehat{\varphi}_{PI} = \beta' A \beta + u' Q' A Q u + 2u' Q' A \beta. \tag{3.2}$$

Using the general formula for the expectation of quadratic forms and the exclusion restriction $\mathbb{E}(u|X) = 0$ we obtain,[10]

$$\mathbb{E}\left(\widehat{\varphi}_{PI}|X\right) = \beta' A \beta + \text{trace}\left(Q' A Q \mathbb{V}(u|X)\right) = \varphi + \delta, \tag{3.3}$$

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 almost feasible bias correction term $\widehat{\delta}$ is defined as,

$$\widehat{\delta} \equiv \text{trace}\left(Q' A Q \widehat{\mathbb{V}}(u|X)\right), \tag{3.4}$$

[9] We somewhat follow the notation in Kline et al. (2020) for the interested reader to compare the papers.

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

where $\widehat{\mathbb{V}}(u|X)$ is an estimator of the conditional variance of the error term $\mathbb{V}(u|X)$. The almost feasible 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 $\mathbb{V}(u|X)$.[11] Therefore we need an unbiased estimate of the conditional variance to be able to compute the almost feasible bias correction $\widehat{\delta}$.[12] We can define then the following unbiased estimate of φ as:[13]

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

Unfortunately, when the number of covariates is very large, computing the almost feasible 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.[14] 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.

3.2.1 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 (3.1) 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 $\beta' = [\beta_1' \, \beta_2']$.

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}.[15] 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

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

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

[12] 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 almost feasible bias correction as $\widehat{\delta} = \widehat{\sigma}_u^2 \times \text{trace}\left(A(X'X)^{-1}\right)$.

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

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

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

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.

3.3 Bootstrap Correction

The bootstrap correction estimates the almost feasible bias correction (3.4) by replicating the bias structure of the plug-in estimates (3.2). In this section we present the bootstrap correction and discuss different implementations depending on the choice of the covariance matrix estimate.

Suppose that we have the residuals of our original regression $\widehat{u} = Y - X\widehat{\beta}$. Using these residuals we can construct an estimate of the covariance matrix, $\widehat{\mathbb{V}}(u|X)$, which we discuss in Section 3.3.1. We generate a new dependent variable for the bootstrap Y^* as:

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

where v^* is a vector containing the bootstrapped residuals. This is equivalent to performing a traditional bootstrap as in MacKinnon and Smith Jr (1998), while setting $\widehat{\beta} = \mathbf{0}$. The construction of 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 $\widehat{\mathbb{V}}(u|X)$. The following proposition states the main result of the paper and all the proofs are left to Appendix 3.A:

Proposition 11. *Suppose the regression model (3.1) 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*

$$\widehat{\delta}_b = \mathbb{E}_{v^*}\left(\beta_j^{*\prime} A \beta_j^* | X, u\right)$$

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

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

The intuition behind our bias estimator is that in every bootstrap iteration we are replicating the source of the bias, which is the noise embedded in the estimated parameters. MacKinnon and Smith Jr (1998) propose a similar bootstrap to correct for flat biases like the one we want to eliminate.[16] MacKinnon and Smith Jr (1998) propose building the bootstrapped dependent variable by using the original estimate of β, $Y^* = X\widehat{\beta} + v^*$. In our application we would next compute the quadratic objects $\beta_{j,MS}^{*\prime} A \beta_{j,MS}^*$ and their correction would be: $\widehat{\delta}_{b,MS} = \mathbb{E}_{v^*}\left(\beta_{j,MS}^{*\prime} A \beta_{j,MS}^* | X, u\right) - \widehat{\beta}' A \widehat{\beta}$. 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 (3.2) we have that in bootstrap j: $\beta_{j,MS}^{*\prime} A \beta_{j,MS}^* = \widehat{\beta}' A \widehat{\beta} + (v_j^*)' Q' A Q v_j^* + 2 v_j^{*\prime} Q' A \widehat{\beta}$. When the covariance matrix of the errors is diagonal, it can be shown that the covariance of the last two terms conditional on X and u is equal to 0. Thus we have that the conditional variance of their estimator in the bootstrap

[16] 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 (3.3) is independent of $\widehat{\beta}$.

is: $\mathbb{V}(\widehat{\delta}_{b,MS}|X,u) = \frac{1}{n^*}\mathbb{V}((v^*)'Q'AQv^*|X,u) + \frac{4}{n^*}\mathbb{V}(v^{*\prime}Q'A\widehat{\beta}|X,u)$ where we used the fact that $\widehat{\beta}$ is not a random variable once we condition on X and u to eliminate the first term, and the independence of the bootstrap errors v^* across j to enter the variance into the sum. We therefore have that the conditional variance of their estimator:

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

is higher than ours, attributable to the presence of the last term similarly to equation (3.2). Both methods are unbiased and consistent but ours is more efficient as we show in Section 3.3.2.

The computational burden of our method comes from estimating β_j^* for each bootstrap. The advantage of our method is twofold. First, we can correct several moments simultaneously, without increasing the computational time. 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 the result of a large number of fixed effects. This is the case in most applications.

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 almost feasible bias correction term $\widehat{\delta}$. Therefore, the bootstrap procedure has to be consistent with the underlying assumption on the structure of the error term.

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 11:

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.

3.3.1 Choice of covariance matrix estimate

We divide the discussion in this section into 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. Second, we discuss estimators of non-diagonal covariance matrix, in particular, when we have clustering or serial correlation.

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 11 suggests an algorithm to make simultaneous M

corrections.[17] Let $\widehat{\psi}_i$ be the estimate of the variance for the ith observation error term. Algorithm 4 in Appendix 3.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.[18]

The White (1980) estimator is biased:[19]

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

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 include:

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

In the homoscedastic case we can use the well known unbiased estimate $HC_1\widehat{\psi} = n/(n-k)\sum_{i=1}^n \widehat{u}_i^2$.[20] Alternativelly, under the homoscedastic case we could replace steps (3) and (4) of Algorithm 4 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.[21]

In the general hereoscedasticity case, MacKinnon and White (1985) acknowledge the existence of a bias in all three but denote HC_2 as an almost unbiased estimate of the variance. These heteroscedasticity robust are inconsistent when the model has many covariates (Cattaneo et al., 2018) as is usually the case with multi-way fixed effects. Recently, Kline et al. (2020) and Jochmans (2018) have proposed the following unbiased estimator of the ith conditional variance:

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

In practice, estimating $\widehat{\psi}_i$ with HC_U some observations have a negative estimated variance and that prevents us from taking the square root in step (4) of the Algorithm 4.[22] 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,

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

[17] 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.

[18] 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.

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

[20] The origine of the bias is again a trace term that under homoscedasticity is equal to $n-k$. For a textbook explanation see Proposition 1.2. in Hayashi (2000).

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

[22] 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 HC_2. We could alternatively 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,

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

The use of HC_2 or HC_M requires the computation of the leverage h_{ii} for each observation. Moreover we need them to be smaller than 1 for every observation. In the following we describe how we ensure that the leverages are below the unity by computing the leave-one-out connected set, how we estimate them and finally propose a diagnosis and an adjustment for our leverage estimates.

Leave-one-out connected set. Two-way fixed effect models are only estimated at the connected set. In typical applications on the labor market or teacher evaluations, firm (school) fixed effects are only identified within the connected set that is generated by moving workers (teachers). Movers therefore determine the connected set of firms (schools) whose fixed effect can be identified. The need to have $h_{ii} < 1$ for all i requires that no single observation is necessary to estimate a particular fixed effect. That is, eliminating any observation the set of fixed effects in the connected set needs to remain the same. We achieve this by first pruning the data to get the leave-one-out connected set without critical movers identifying a given firm fixed effect and eliminating unique observations. The pruning is analogous to Kline et al. (2020) and we leave the details for the Appendix.

Estimation of leverage. The direct computation of the leverage, by using the diagonal of the projection matrix H, is computationally infeasible when the number of covariates is large.[23]

We propose a way to estimate the leverage of each observation that is similar to our bias estimator. We simulate repeatedly random variables and use the fitted values of the projection into X to estimate the leverage. The procedure starts by generating 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 11, 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.[24] Estimating the leverage this way entails two different sets of bootstrap, one for the leverage and the second one for the estimation of the bias correction.[25]

Alternatively, an exact computation of the leverage is possible by using the definition of fitted values $\widehat{Y} = HY$ and a regression intensive procedure. We have that the leverage of observation i is

[23] $H \equiv X\left(X'X\right)^{-1}X'$ is a function of the inverse of a very large matrix X.

[24] This is exactly the way Kline et al. (2020) estimate the leverage in their paper. However, they directly solve for the normal equations of the regression using the preconditioned conjugate gradient method.

[25] One is usually interested in estimating corrections for at least three moments that involve solving two systems of linear equations (See Gaure, 2014; Kline et al., 2020).

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 $\widehat{Y}'$. Thus the leverage, that is a 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 in what would involve n regressions. When the data set is large, this is clearly not plausible and we leave the exact computation for the problematic ones identified in the following diagnostic.

Diagnostic and adjustment. We detect problematic leverage estimations by checking that they are under unity and above an observation specific lower bound. Let $\tilde{X} = X\mathbf{1}$, meaning $\tilde{X}$ 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 $\tilde{h}_{ii} = \tilde{x}_i^2 / \sum_{i=1}^{n} \tilde{x}_i^2$. We diagnose our leverage estimates by comparing them to $\tilde{h}_{ii}$. Those that are underestimated or are above 1 can thus 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 $\frac{1}{1-\widehat{h}_{ii}}\left(1 - \frac{1}{(1-\widehat{h}_{ii})^2}\frac{\widehat{\text{var}}(\widehat{\mathbf{y}}_i^2)}{n_h}\right)$, where the last term corrects for a non-linear bias with $\widehat{\text{var}}(\widehat{h}_{ii})$ being a sample variance of the different estimates of the squared fitted values.

Algorithm 7 in Appendix 3.C takes as inputs the covariates X and gives output a combination of actual and estimated leverages, as well as the variance $\widehat{\text{var}}(\widehat{h}_{ii})$ 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 no longer diagonal. We restrict our attention to dependence of the errors only within a given group. The variance covariance matrix is block diagonal as there are non zero elements around the diagonal corresponding to the dependence of the errors within the group g.[26] 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 4 to a wild block bootstrap as proposed by Cameron et al. (2008). This consists of a wild bootstrap that takes into account the group or cluster dependence of the data. Differently to the sieve bootstrap

[26] 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 (3.1) is:

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

(e.g. Davidson and MacKinnon, 2006) it has the benefit of accommodating any structure of the dependence within group and also heteroscedasticity of the true innovations.

Following Roodman et al. (2019) we estimate the variance of observation i, $\widehat{\psi}_i$, with a variant of HC_1 from the previous section that takes into account the number of groups G: $\widehat{\psi}_i = \frac{G}{G-1}\frac{n}{n-k}\widehat{u}_i^2$.[27] Algorithm 5 in the Appendix describes the procedure for our bias estimator that keeps the dependence structure through a wild block bootstrap. It takes as inputs X, $\{\widehat{\psi}_i\}_{i=1}^N$ and the different matrices $\{A_m\}_{m=1}^M$.

3.3.2 Simple example

Monte Carlo simulations illustrate the effectiveness of our bias correction method. The model design is the same as in (3.1) with homoscedastic errors and sample size $n = 500$. The number of covariates is $k_1 = k_2 = 200$. We keep this number relatively low to compute the direct correction. We have 10,000 simulations in total. In each simulation, conditioning on X, we draw new error terms to form the dependent variable. We estimate $\widehat{\beta}$ and compute the almost feasible bias correction terms. After the estimation, we perform $n^* = 100$ bootstraps and use them to compute the estimation of the bootstrap correction terms.[28]

Figures 30 and 31 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.

Table 34 presents the mean and variance of the differences of our bootstrap method $\widehat{\delta}_b$ and the bootstrap following MacKinnon and Smith Jr (1998) $\widehat{\delta}_{b,MS}$ with respect to the direct correction $\widehat{\delta}$.[29] The mean differences of our method 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. The alternative bootstrap correction $\widehat{\delta}_{b,MS}$ in Columns 3 and 4 performs worse in terms of mean differences and also variances.

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

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

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

[27] Introducing covariance estimation refinements proposed by Bell and McCaffrey (2002) and further developed in Imbens and Kolesar (2016) are left for future research. They extend the HC_2 corrections to the cases where the covariance matrix is not diagonal by correcting by the leverage of the group.

[28] We use the covariance estimator HC_1 and therefore skip the part of computing that involves the leave-one-out connected set.

[29] AS previously stated, they propose to generate the bootstrap dependent variable as $Y^* = X\widehat{\beta} + v^*$. Their correction is: $\widehat{\delta}_{b,MS} = \frac{1}{n^*}\sum_{j=1}^{n^*}\left(\beta^{*\prime}_{j,MS}A\beta^*_{j,MS}\right) - \widehat{\beta}'A\widehat{\beta}$, where the last term is the plug-in estimate.

Table 34 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. As expected, our bootstrap has lower MSE than the alternative directly following MacKinnon and Smith Jr (1998).

3.3.3 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 the choice of the number of bootstraps. Our estimator $\widehat{\delta}_b$ is a sample mean estimate of the almost feasible bias correction term $\widehat{\delta}$. Using results from probability theory we can exploit the information given by Chebyshev's inequality.

In Proposition 11 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 11) and $\mathbb{V}(\widehat{\delta}_b|X,u) = \frac{1}{n^*}\eta^2$.[30] Then, by Chebyshev's inequality we have

$$\mathbf{P}\left(\left|\widehat{\delta}_b - \widehat{\delta}\right| \geq k\frac{\eta}{\sqrt{n^*}} \;\middle|\; X,u\right) \leq \frac{1}{k^2}.$$

Next one can choose the number of bootstraps n^* such that the distance between the bootstrap estimate $\widehat{\delta}_b$ and the almost feasible 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

$$\frac{1}{k^2} = \alpha, \quad \frac{k}{\sqrt{n^*}} = \lambda.$$

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 almost feasible 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.

3.4 Comparison of Methods

In this section we first compare our method to Gaure (2014), Kline et al. (2020) and Borovičková and Shimer (2017).[31] The closest methods to ours are the ones by Gaure (2014) and Kline et al. (2020). All three aim to compute the trace term in equation (3.3). Yet, Borovičková and Shimer (2017) propose

[30] 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.

[31] The codes to implement our method are here.

a method to compute the correlation of theoretically different worker and firm types. Second, we present the 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 error structures 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 implemented his bias correction method on the R package *lfe* only under the assumption of homoscedastic errors. Moreover, KSS and ourselves provide corrections under serial correlation or clustering of the errors. Second, our method is the only one implementing corrections on the full variance decomposition correcting 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 and our approach is therefore suited for multi-way fixed effect models or full variance decompositions in two-way fixed effect models. KSS and Gaure on the contrary need to compute new sets of normal equations per additional correction and only implement corrections of the two leading fixed effects. Our method is also more flexible than Gaure and KSS in the type of corrections by allowing for yearly corrections and different types of dependencies on the error structures. Finally, 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 3.3.1, KSS estimate the covariance matrix by HC_U and our baseline application is with HC_2 even if we explore other covariance matrix estimates.

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:

$$w_{it} = \theta_i + \psi_{J(i,t)} + q_{it}\gamma + \varepsilon_{it}, \tag{3.5}$$

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 (3.5) can be estimated by OLS where the person/firm fixed effect estimators have the same structure as the one in Section 3.2. Thus the second order moments exhibit a similar bias and the implementation of the correction is analogous.

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 to 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.[32]

[32]We refer to an old version of the Borovičková and Shimer from 2017 where they provide a way to translate the variances and

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:[33]

$$\sigma_\theta = \frac{\sigma_\lambda - \rho\sigma_\mu}{1-\rho^2}, \quad \sigma_\psi = \frac{\sigma_\mu - \rho\sigma_\lambda}{1-\rho^2}.$$

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 paid) 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 is mean zero conditional on the types (and other covariates) with the underlying assumption of exogenous mobility.

3.4.1 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 3.3.1.

We compare all the methods under conditional homoscedasticity of the errors. Results are in Table 35. All the methods improve the initial bias of the plug-in estimate. The least accurate method is BS reducing by 69% the MSE of the naive estimates whereas the other three methods reduce it by 98%.[34] 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 35 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 the correlation of the AKM types $\rho_{\theta,\psi}$. Their original types are, on the contrary, not suited to perform a variance decomposition. We find that the MSE taking their types are orders of magnitude greater.[35] Gaure, KSS and our method are very similar in terms of MSE, Gaure being slightly more accurate than the other two.[36] Figure 33 shows the distribution of the bias of the firm variances for the naive estimate ($\sigma^2_{\psi,PI} - \sigma^2_\psi$) 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. Our method is the one performing fastest among the three closest competitors (Gaure, KSS and our method) as it has the lowest computing time.[37]

covariances of their worker and firm types to the ones in AKM. In the latest version, they slightly changed their estimator and no longer provide this link.

[33] See Proposition 1 in Borovičková and Shimer (2017).

[34] We wrote the code for BS following Borovičková and Shimer (2017) and converting the data to the match level.

[35] The scaled MSE (MSE $\times 10^2$) of $\sigma^2_\lambda, \sigma^2_\mu$ and $\sigma_{\lambda,\mu}$ are respectively 57.4, 81.9 and 4.00.

[36] 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 run our corrections in Matlab with tolerance of 1e-5.

[37] KSS and our method do not incorporate the simplifications that come from having homoscedastic errors. In particular, under homoscedasticity of the errors HC_1 is an unbiased estimate of the variance and one could skip the pruning of the data.

Table 36 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 85% the MSE of the plug-in estimate in the low mobility case.[38] Our method is slightly more accurate for both mobility cases, it also outperforms KSS in terms of time. Figure 34 shows the distribution of the bias of the plug-in estimate, KSS and our method. Both corrections have similar distributions but the bootstrap method has smaller variance for reasons shown in Section 3.3.1. Table 37 compares the different covariance matrix estimators applicable with our method. All the estimators have similar MSE but HC_2 outperforms the rest.

Tables 38 and 39 present 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 and we allow the true innovation to be homoscedastic and heteroscedastic. We compare the plug-in estimate to our bootstrap with the wild block bootstrap method from Algorithm 5 *Boot*, *Boot Av Match* where observations are averaged to the match, and the KSS correction methods where the observations are also averaged to the match *KSS Av Match*. The best performing correction method is *Boot Av Match* both in terms of time and MSE under homoscedasticity (Table 38) and heteroscedasticity (Table 39). Our method using a wild block bootstrap *Boot* also improves the MSE of the plug-in estimates but has higher MSE than *Boot Av Match* because of a higher bias of the person fixed effects component. *KSS Av Match* improve the MSE of the naive estimates but perform worse than our method with and without taking match averages and under homoscedasticity and heteroscedasticity.

3.5 Application

In the application we use a panel data from the French statistical agency (INSEE) from 2002 to 2014.[39] Our dependent variable is (log) gross daily wage of full time employees with ages between 20 and 60 working at private firms.

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 guide our choice of number of bootstraps, we perform some simulations with a fixed set of covariates with low mobility and simulate a thousand samples by simulating the error. With each dataset we perform corrections from one to 300 bootstraps as in the Monte Carlo simulations of Section 3.4. Figure 32 shows the mean squared error between the true covariance between person and firm fixed effects and the corrected one for different number of bootstraps.[40] The figure shows that with the first 25 bootstraps the MSE reduces significantly and around 150 it flattens. This suggests that few bootstraps are enough to gain accuracy.[41]

[38]Table 1 in Kline et al. (2020) shows that their connected set is similar to our low mobility scenario with 2.7 movers per firm and average firm size of 12.

[39]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 on certain days. The sample size was increased in 2002 so we took this as the starting year.

[40]For all the samples we take the corrections obtained with different bootstraps and take the mean squared error against the true moment.

[41]Throughout the application corrections we run corrections with 300 simulation to estimate the leverage and 1000 bootstraps to

Table 40 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 under the assumption of serial correlation within match. 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 a 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 in the case of 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 36 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 both are similar 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 and explanations for this phenomenon are outside the scope of this paper.

3.5.1 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 to ensure that annual wage observations are independent conditional on type, which might not be the case especially for workers who do not switch firms. In order to accommodate for the extra covariates in the BS method, we first run a linear regression of log wage versus q_{it} (age and education interacted by year effects) and take the residual. 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. We estimate the bootstrap corrected moments at the connected set or leave-one-out-corrected set of the BS final sample.

Table 41 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 to the Monte Carlo simulations that satisfied the assumptions for both methods, estimates differ by a large amount when using French labor market data. The bootstrap corrected estimated correlation is 0.16 (0.09) under HC_2 (HC_1) covariance matrix estimation, well below the estimated one using BS method, 0.55.[42] Looking at each of the components of the corre-

estimate the corrections of second order moments.

[42]The BS estimates are obtained by using the formulas of Section 5.2. in Borovičková and Shimer (2017).

lation, both variances are larger and the covariance is smaller when using the bootstrap corrected method instead of BS method.

There are different reasons why BS estimates might differ from ours. To begin with, the types defined by BS are fundamentally different from the ones defined in the AKM model. They are related only 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 is 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 violate both the AKM and BS identification assumptions.

Results in Table 41 under our method also differ from the ones previously reported in Table 40. Table 42 presents some summary statistics of the original data differentiated by being in the final BS data or not.[43] The Table shows that the requirements to use the Borovičková and Shimer (2017) method are quite demanding as only 43% of the original observations are included in their final sample. Furthermore, Table 42 shows that their data requirements lead to a sample with similar average wage but almost 5 years younger on average and slightly more educated. The applied user might be worried by sample selection when using the BS method to estimate worker and firm correlation as Lentz et al. (2018) document that most of the worker-firm sorting happens early in the career which would lead to higher correlations for younger workers.

3.6 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. The application to French labor market data shows that the correction increases the correlation between firm and worker fixed effects. Depending on the sample and on the specification, our bias correction method changes the sign of that correlation and in all cases it changes the relative importance of the different components in explaining the variance of log wages.

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 to estimate the model several times. The correction can thus be applied easily to any study running an AKM type regression or multi-way fixed effects regressions. The method is faster than Kline et al. (2020) and similarly accurate. Besides the speed, the other biggest advantage of our approach is its flexibility because it allows for yearly corrections or an increase in the number of moments to correct without increasing the computational costs.

[43] The original data constitutes of almost 5.9 observations that translate into a connected set of 5.1 million observations as in Table 40.

Comparisons to other models through Monte Carlo simulations show that, in terms of accuracy, our method is comparable in accuracy to Gaure (2014) and Kline et al. (2020). Our method is broader than the implementation of Gaure (2014) given that he only offers corrections under homoscedasticity of the errors. Compared to Kline et al. (2020) our approach is faster, is more flexible in the type of corrections it can perform and can 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.

3.A Proofs

Proposition 11.

Proof. First, note that for any bootstrap estimate of the quadratic form $\beta_j^{*\prime} A\beta_j^*$ we have that

$$\beta_j^{*\prime} A\beta_j^* = (v_j^*)' Q' AQ v_j^*.$$

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

$$\mathbb{E}_{v^*}\left(\beta_j^{*\prime} A\beta_j^* \,\middle|\, X, u\right) = \text{trace}\left(Q' AQ\mathbb{V}(v_j^*|X)\right).$$

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

$$\mathbb{E}(\widehat{\delta}_b|X) = \mathbb{E}_u\left(\mathbb{E}_{v^*}(\widehat{\delta}_b|X, u) \,\middle|\, X\right) = \mathbb{E}_u(\widehat{\delta}|X) = \delta.$$

□

3.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.[44]

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. (2020).[45] 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 a total sample size of roughly 22000 observations.

Worker and firm fixed effects are random draws from normal distributions. We assume that there is sorting depending on the permanent types, which leads to non negative correlations between worker and

[44] We thank Simen Gaure for sharing with us a piece of code that we used as a base for the simulations.

[45] See Table 1 in Kline et al. (2020) where each worker is observed twice.

firm fixed effects while fulfilling exogenous mobility. That is, 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 an 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 (3.5) without other covariates.

3.C Algorithms

Here we detail the implementation algorithms of our method. Algorithm 4 and 5 describe respectively the estimation of the bias correction for diagonal and non diagonal covariance matrices. Algorithm 6 describes how to prune the data to ensure that the maximum leverage is below 1 and Algorithm 7 details how to estimate the leverage.

Algorithm 4 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 $\sqrt{\widehat{\psi}_i} \times r_i^*$.
  Compute β^* as the estimate of a regression of v^* on X.
  Compute $\widehat{\delta}_{aux,m}^{(j)} = (\beta^*)' A_m \beta^*$ for all $m \in \{1, ..., M\}$.
end for
Compute $\widehat{\delta}_{b,m} = \dfrac{\sum_{j=1}^{n^*} \widehat{\delta}_{aux,m}^{(j)}}{n^*}$ for all $m \in \{1, ..., M\}$.

Algorithm 5 Estimate $\{\widehat{\delta}_{b,m}\}_{m=1}^{M}$ when covariance matrix is non diagonal

Let $\mathbb{G} = \{1, ..., G\}$ be the set of groups g each with length n_g.
for $j = 1, ..., n^*$ **do**
  Simulate a vector r_g^* of length G of mutually independent Rademacher entries. All the observations withing the group will have the same Rademacher entry.
  Generate a vector of residuals v^* of length n whose ith entry belonging to group g is equal to $\sqrt{\widehat{\psi}_i} \times r_g^*$.
  Compute β^* as the estimate of a regression of v^* on X.
  Compute $\widehat{\delta}_{aux,m}^{(j)} = (\beta^*)' A_m \beta^*$ for all $m \in \{1, ..., M\}$.
end for
Compute $\widehat{\delta}_{b,m} = \dfrac{\sum_{j=1}^{n^*} \widehat{\delta}_{aux,m}^{(j)}}{n^*}$ for all $m \in \{1, ..., M\}$.

Algorithm 6 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 7 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

3.D Tables and Figures

Table 34 – Results of simple Monte Carlo simulations

	$\widehat{\delta} - \widehat{\delta}_b$		$\widehat{\delta} - \widehat{\delta}_{b,MS}$		Mean Squared Error				True
	Mean	Variance	Mean	Variance	Naive	Ideal	Boot	Boot MS	Moment
$\widehat{\text{var}}(X_1\beta_1)$	0.33×10^{-3}	0.00156	-0.79×10^{-3}	0.07945	45.53	22.94	22.94	23.04	243.19
$\widehat{\text{var}}(X_2\beta_2)$	0.36×10^{-3}	0.00156	-4.25×10^{-3}	0.15211	67.59	44.75	44.76	44.98	458.83
$\widehat{\text{cov}}\,(X_1\beta_1, X_2\beta_2)$	-0.33×10^{-3}	0.00138	0.93×10^{-3}	0.04417	22.54	12.50	12.51	12.57	-9.09

Notes: The first two columns represent, respectively, the mean and the variance of the difference between the almost feasible correction and the bootstrap correction. Columns 3 and 4 are analogous to the bootstrap following MacKinnon and Smith Jr (1998). Columns 5 to 9 compute the MSE between the estimated moments and the true ones and Column 10 presents the true moments in the simulation.

Table 35 – Monte Carlo simulations. Homoscedastic errors.

	Time	Mean Squared Error (MSE×10^2) $\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		6.43813	0.28682	0.11545	2.28013
BS	0.3	1.95426	0.11898	0.02482	0.69935
Gaure	1.2	0.04790	0.10116	0.01338	0.05415
Boot	1.1	0.04792	0.10328	0.01369	0.05496
KSS	2.7	0.04842	0.10288	0.01382	0.05504

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* is our method with HC_2 covariance matrix estimator, and *KSS* is the Kline et al. (2020) 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 36 – Monte Carlo simulations. Heteroscedastic errors.

Mov/firm	Model	Time	Mean Squared Error (MSE×10^2) $\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Low Mobility						
3	Plug-in		19.53008	1.12075	7.41524	9.35536
3	Boot	1.4	0.17457	3.38480	0.38844	1.31594
3	KSS	2.0	0.20498	3.46213	0.41224	1.35979
Mid Mobility						
5	Plug-in		10.44445	0.41531	1.90528	4.25501
5	Boot	1.3	0.09246	1.00829	0.13824	0.41299
5	KSS	2.3	0.09735	1.01536	0.13689	0.41653

Notes: *Plug-in* is the naive plug-in estimator, *Boot* refers to our method with HC_2 covariance matrix estimator, and *KSS* is the Kline et al. (2020) 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 37 – Comparison of variance estimations.

	Mean Squared Error (MSE$\times 10^2$)			
Model	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in	19.896	5.997	5.034	10.309
Boot HC_0	2.236	1.245	0.792	1.424
Boot HC_1	0.428	0.716	0.291	0.478
Boot HC_2	0.173	0.622	0.171	0.322

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. $\hat{\sigma}^2_\theta$, $\hat{\sigma}^2_\psi$ and $\hat{\sigma}_{\theta,\psi}$ present respectively the mean squared errors of the 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). Simulated data exhibits low mobility like in the top panel of Table 36 and all the estimations are done at the leave-one-out.

Table 38 – Monte Carlo simulations. Serial correlation with homoscedasticity.

		Mean Squared Error (MSE$\times 10^2$)			
	Time	$\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		91.43846	1.52054	0.52576	31.16158
Boot	0.5	9.14621	0.23902	0.04491	3.14338
Boot Av Match	0.3	5.58738	0.38655	0.23283	2.06892
KSS Av Match	2.3	18.62253	0.36095	0.04201	6.34183

Notes: *Plug-in* is the naive plug-in estimator, *Boot Av Match* refers to our method with HC_1 covariance estimator where the observations are averaged to the match and wages are transformed to average match wage. *Boot* refers to our method with a wild block bootstrap where each match defines a block. In both, *Boot* and *Boot Av Match* we skip the pruning of the data. *KSS Av Match* is the Kline et al. (2020) method where the observations are averaged to the match. 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) multiplied by 100 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 39 – Monte Carlo simulations. Serial correlation with heteroscedasticity.

	Time	Mean Squared Error (MSE$\times 10^2$) $\hat{\sigma}^2_\theta$	$\hat{\sigma}^2_\psi$	$\hat{\sigma}_{\theta,\psi}$	Average
Plug-in		91.03966	1.53866	0.53512	31.03782
Boot	0.5	9.04595	0.28665	0.04516	3.12592
Boot Av Match	0.3	5.46460	0.43006	0.22812	2.04093
KSS Av Match	2.3	18.45137	0.39646	0.04425	6.29736

Notes: *Plug-in* is the naive plug-in estimator, *Boot Av Match* refers to our method with HC_1 covariance estimator where the observations are averaged to the match and wages are transformed to average match wage. *Boot* refers to our method with a wild block bootstrap where each match defines a block. In both, *Boot* and *Boot Av Match* we skip the pruning of the data. *KSS Av Match* is the Kline et al. (2020) method where the observations are averaged to the match. 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) multiplied by 100 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 40 – Application. Plug-in vs corrected decomposition.

	Plugin		Boot Serial	
	Component	Exp. Sh.	Component	Exp. Sh.
$Var(y)$	0.216	1.00	0.216	1.00
$Var(\widehat{\theta}_i)$	0.163	0.75	0.135	0.62
$Var(\widehat{\psi}_j)$	0.049	0.23	0.032	0.15
$Var(\mathbf{q}\widehat{\gamma})$	0.008	0.03	0.007	0.03
$2Cov(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.033	-0.15	-0.006	-0.03
$2Cov(\widehat{\theta}_i, \mathbf{q}\widehat{\gamma})$	-0.000	-0.00	-0.000	-0.00
$2Cov(\widehat{\psi}_j, \mathbf{q}\widehat{\gamma})$	-0.000	-0.00	-0.000	-0.00
$Var(\widehat{\epsilon})$	0.030	0.14	0.049	0.23
$Corr(\widehat{\theta}_i, \widehat{\psi}_j)$	-0.100	-	-0.000	-
Obs.	5108399	5108399	5108399	5108399

Notes: *Plug-in* refers to the uncorrected estimates of each of the variance components and *Boot Serial* refers to the estimates after our bootstrapped correction using a wild block bootstrap. $Var(y)$ 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 and *Obs.* is the number of observations.

Table 41 – Application. Comparison of the Methods.

	BS	Plugin	Boot HC_1	Boot HC_2
$\widehat{\sigma}^2_\theta$	0.061	0.095	0.066	0.063
$\widehat{\sigma}^2_\psi$	0.005	0.038	0.024	0.019
$\widehat{\sigma}_{\theta,\psi}$	0.010	-0.004	0.003	0.005
$\widehat{\rho}_{\theta,\psi}$	0.558	-0.064	0.087	0.156
Obs.	945356	942235	942235	931925

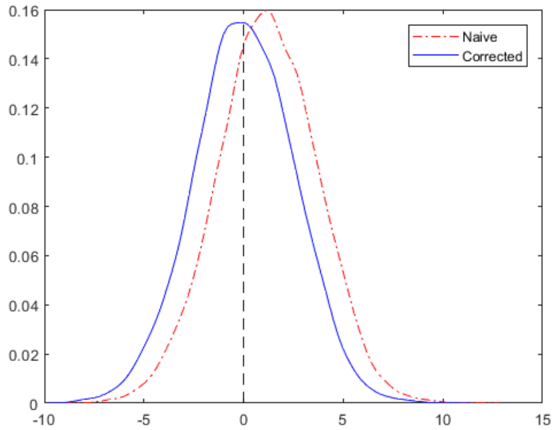
Notes: The results of *BS* correspond to the AKM worker and firm types of Borovičková and Shimer. *Plugin* are the plug-in estimates at the connected set originated from BS data, *Boot* HC_1 are the results of our method under diagonal covariance matrix estimator HC_1 at the connected set originated in the BS data, *Boot* HC_2 are the results of our method under diagonal covariance matrix estimator HC_2 at the leave-one-out connected set in the BS data. $\widehat{\sigma}^2_\theta$ and $\widehat{\sigma}^2_\psi$ are respectively the estimates of the variance of worker and firm fixed effects. $\widehat{\sigma}_{\psi,\theta}$ is the covariance, $\widehat{\rho}_{\psi,\theta}$ the correlation between worker and firm fixed effects and *Obs.* is the number of observations.

Table 42 – Application. Summary Statistics.

BS Data	Obs.	Mean Wage	Mean Age	Mean Education
No	3311804	4.39	41.43	4.56
Yes	2541773	4.37	36.94	4.95

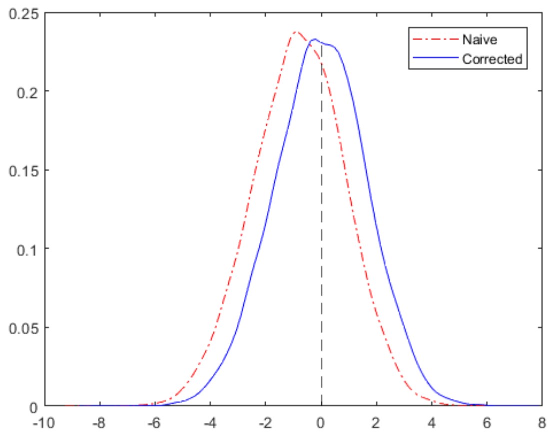
Notes: *BS Data* is an indicator if the observation belongs to the final sample of Borovičková and Shimer (2017), *Obs.* is the number of observations before taking match level averages in the original data and before computing the connected set, *Mean Wage* is the average log daily wage, *Mean Age* is the average age in years and *Mean Education* is the average education where education is a discrete variable between 1 (no education) and 8 (university degree).

Figure 30 – Density of $\widehat{\sigma}^2_{1,PI} - \sigma^2_1$ and $\widehat{\sigma}^2_{1,b} - \sigma^2_1$



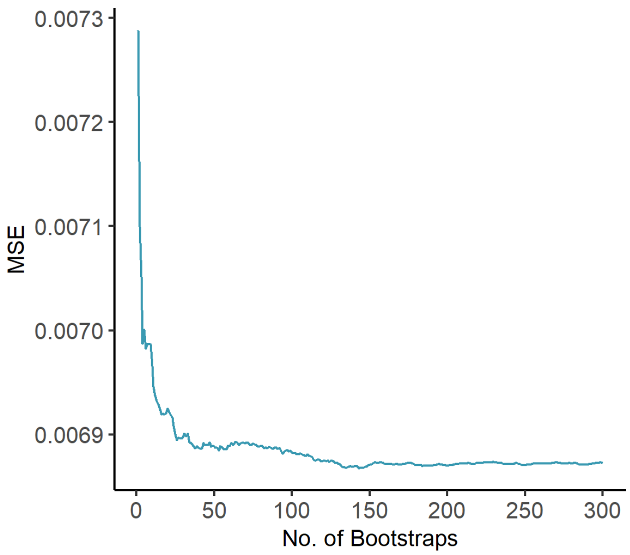
Notes: This figure presents the distributions of the differences between the true variance σ^2_1 and both, the naive plug-in estimated variance $\widehat{\sigma}^2_{1,PI}$ and the bias corrected estimated variance $\widehat{\sigma}^2_{1,b}$. The distribution of the difference between the true moment and the bias corrected estimated covariance is centered at zero.

Figure 31 – Density of $\widehat{\sigma}_{12,PI} - \sigma_{12}$ and $\widehat{\sigma}_{12,b} - \sigma_{12}$



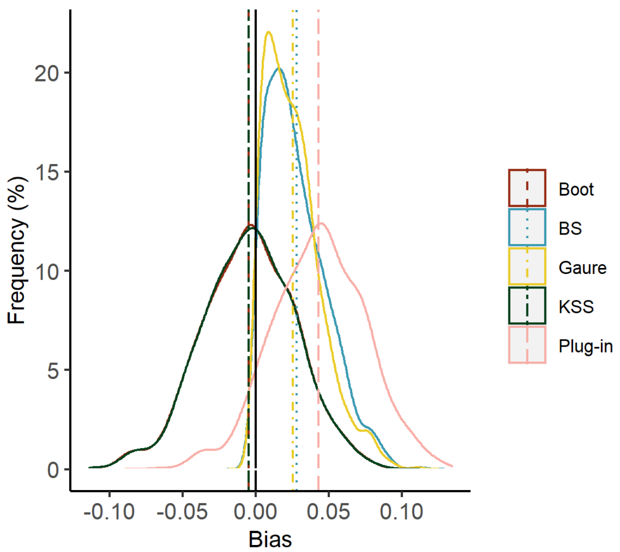
Notes: This figure presents the distributions of the differences between the true covariance σ_{12} and both, the naive plug-in estimated covariance $\widehat{\sigma}_{12,PI}$ and the bias corrected estimated covariance $\widehat{\sigma}_{12,b}$. The distribution of the difference between the true moment and the bias corrected estimated covariance is centered at zero.

Figure 32 – MSE of corrected $\widehat{\sigma}(\theta, \psi)$ by number of bootstraps.



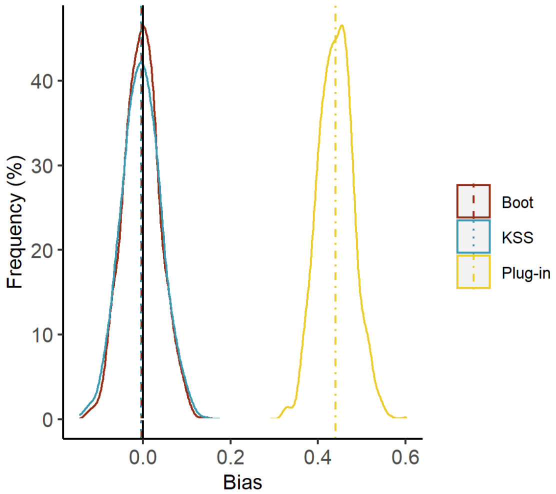
Notes: This figure presents the mean squared error (MSE) of the covariance between worker-firm fixed effects $\widehat{\sigma}(\theta, \psi)$ across 1000 homoscedastic error simulations. The bootstrap correction assumes a diagonal covariance matrix and we use the HC_1 covariance matrix estimator.

Figure 33 – Model Comparison: Homoscedastic Errors.

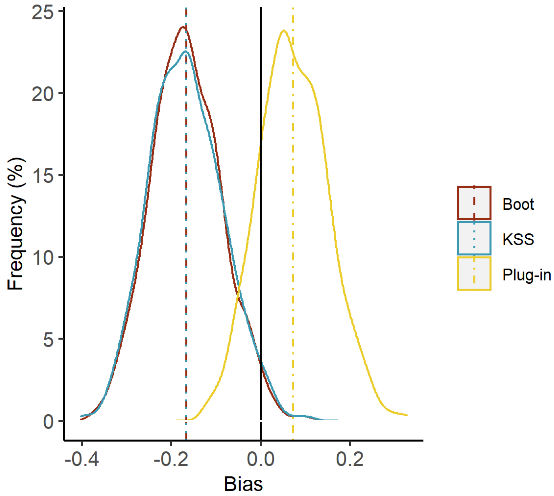


Notes: This figure presents the distributions of the bias of $\widehat{\sigma}^2_{\psi}$ for the naive plug-in estimate and the corrected moments for the different methods. Simulated errors are homoscedastic and labor mobility is high.

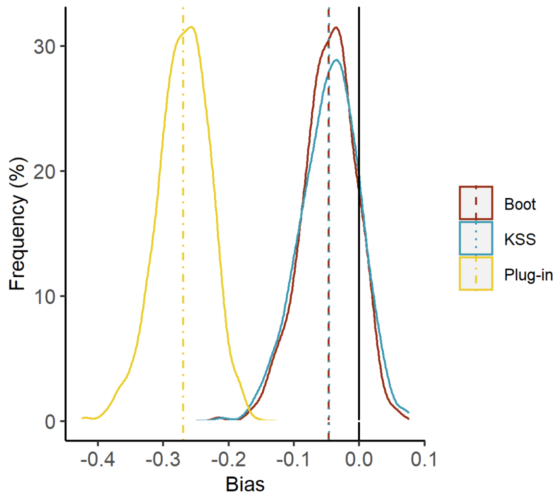
Figure 34 – Model Comparison: Heteroscedastic Errors.



(a) Bias of $\widehat{\sigma}^2_\theta$



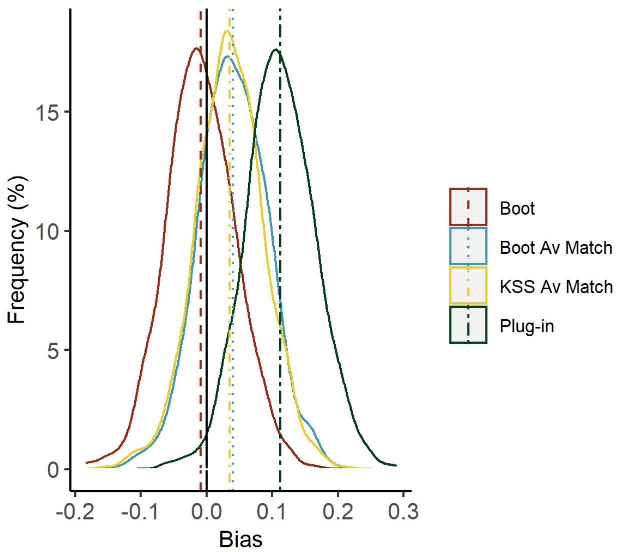
(b) Bias of $\widehat{\sigma}^2_\psi$



(c) Bias of $\widehat{\sigma}_{\theta,\psi}$

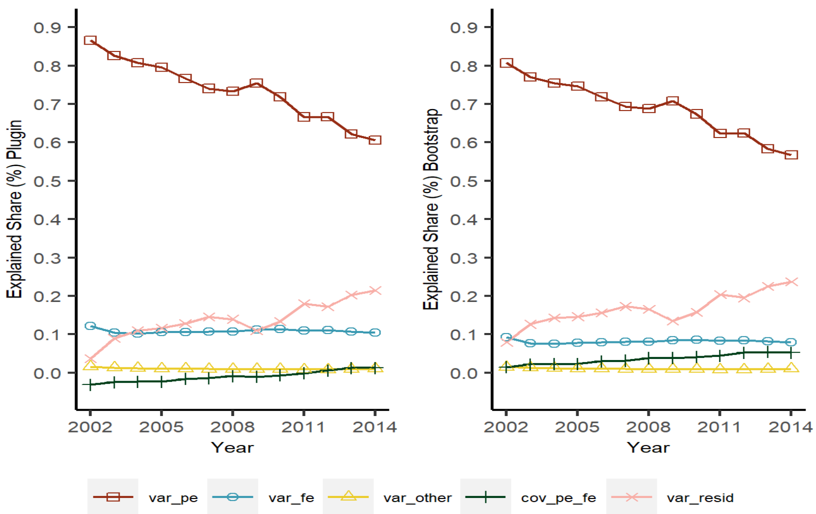
Notes: These figures present the distributions of the bias for the naive plug-in estimate and the bias of corrected moments for *KSS* and our method. Simulated errors are heteroscedastic and labor mobility is low.

Figure 35 – Model Comparison: Serial Correlation with heteroscedasticity.



Notes: This figure presents the distributions of the bias of $\widehat{\sigma}^2_{\psi}$ for the naive plug-in estimate and the corrected moments for the different methods. Simulated errors have serial correlation, true innovations are heteroscedastic and labor mobility is high.

Figure 36 – Application. Evolution of the explained shares.



Notes: This figure presents the year-to-year evolution of the explained shares of the total log wage variance of the plugin and corrected estimates of the person and firm fixed effects, their covariance and the variances of other covariates and the residual.

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.

Adão, R. (2015): "Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil," *Unpublished paper, MIT*, 98.

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*.

——— (2020a): "On the Equilibrium Properties of Network Models with Heterogeneous Agents," Tech. rep., National Bureau of Economic Research.

Allen, T., C. Arkolakis, and Y. Takahashi (2020b): "Universal gravity," *Journal of Political Economy*, 128, 393–433.

Almagro, M. and T. Dominguez-Iino (2020): "Location Sorting and Endogenous Amenities: Evidence from Amsterdam," .

Anderson, J. E. and E. Van Wincoop (2003): "Gravity with gravitas: A solution to the border puzzle," *American economic review*, 93, 170–192.

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.

Artuç, E., S. Chaudhuri, and J. McLaren (2010): "Trade shocks and labor adjustment: A structural empirical approach," *American economic review*, 100, 1008–45.

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.

Bell, R. M. and D. F. McCaffrey (2002): "Bias reduction in standard errors for linear regression with multi-stage samples," *Survey Methodology*, 28, 169–182.

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.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A distributional framework for matched employer employee data," *Econometrica*, 87, 699–739.

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.

Bryan, G. and M. Morten (2019): "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 127, 2229–2268.

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 O. Danieli (2018): "Outside options in the labor market," *Unpublished manuscript*.

Caldwell, S. and N. Harmon (2019): "Outside Options, Bargaining, and Wages: Evidence from Coworker Networks," .

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

Caliendo, L., F. Monte, and E. Rossi-Hansberg (2015): "The anatomy of French production hierarchies," *Journal of Political Economy*, 123, 809–852.

Caliendo, L., L. D. Opromolla, F. Parro, and A. Sforza (2020): "Goods and factor market integration: a quantitative assessment of the eu enlargement," Tech. rep.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008): "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 90, 414–427.

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.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018): "Inference in linear regression models with many covariates and heteroscedasticity," *Journal of the American Statistical Association*, 113, 1350–1361.

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.

Combes, P.-P., M. Lafourcade, and T. Mayer (2005): "The trade-creating effects of business and social networks: evidence from France," *Journal of international Economics*, 66, 1–29.

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 (2018): "Global market power," Tech. rep., National Bureau of Economic Research.

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

Dhaene, G. and K. Jochmans (2015): "Split-panel jackknife estimation of fixed-effect models," *The Review of Economic Studies*, 82, 991–1030.

Diamond, R. (2016): "The determinants and welfare implications of US workers' diverging location choices by skill: 1980-2000," *American Economic Review*, 106, 479–524.

Dingel, J. I. and F. Tintelnot (2020): "Spatial Economics for Granular Settings," .

Dodini, S., M. Lovenheim, K. G. Salvanes, and A. Willén (2020): "Monopsony, Skills, and Labor Market Concentration," .

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

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2018): "Monopsony in online labor markets," Tech. rep., National Bureau of Economic Research.

Eaton, J. and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

Eckert, F. (2019): "Growing Apart: Tradable Services and the Fragmentation of the U.S. Economy," *Unpublished manuscript*.

Edmond, C., V. Midrigan, and D. Y. Xu (2018): "How costly are markups?" Tech. rep., National Bureau of Economic Research.

Fack, G. (2006): "Are housing benefit an effective way to redistribute income? Evidence from a natural experiment in France," *Labour Economics*, 13, 747–771.

Falch, T. (2010): "The elasticity of labor supply at the establishment level," *Journal of Labor Economics*, 28, 237–266.

Fally, T. (2015): "Structural gravity and fixed effects," *Journal of International Economics*, 97, 76–85.

Friggit, J. (2013): "Loyers et revenus depuis les années 1970," *CGEDD, juin*, 15.

Galle, S., A. Rodríguez-Clare, and M. Yi (2017): "Slicing the pie: Quantifying the aggregate and distributional effects of trade," Tech. rep., National Bureau of Economic Research.

Gathmann, C., I. Helm, and U. Schönberg (2017): "Spillover effects of mass layoffs," *Journal of the European Economic Association*.

Gaubert, C., P. Kline, and D. Yagan (2020): "Place-Based Redistribution," .

Gaure, S. (2014): "Correlation bias correction in two-way fixed-effects linear regression," *Stat*, 3, 379–390.

Guimaraes, P., O. Figueirdo, and D. Woodward (2003): "A tractable approach to the firm location decision problem," *Review of Economics and Statistics*, 85, 201–204.

Gutierrez, G. (2017): "Investigating global labor and profit shares," *SSRN Scholarly Paper ID*, 3040853.

Gutiérrez, G. and T. Philippon (2016): "Investment-less growth: An empirical investigation," Tech. rep., National Bureau of Economic Research.

Hafner, F. (2020): "Labor Market Competition, Wages and Worker Mobility," .

Hayashi, F. (2000): "Econometrics. 2000," *Princeton University Press. Section*, 1, 60–69.

Head, K. and T. Mayer (2014): "Gravity equations: Workhorse, toolkit, and cookbook," in *Handbook of international economics*, Elsevier, vol. 4, 131–195.

Heise, S. and T. Porzio (2019): "Spatial Wage Gaps and Frictional Labor Markets," *FRB of New York Staff Report*.

Hellerstein, J. K., M. J. Kutzbach, and D. Neumark (2019): "Labor market networks and recovery from mass layoffs: Evidence from the Great Recession period," *Journal of Urban Economics*, 113, 103192.

Hershbein, B., C. Macaluso, and C. Yeh (2020): "Monopsony in the US Labor Market," Tech. rep., Working Paper.

Hornbeck, R. and E. Moretti (2020): "Estimating Who Benefits From Productivity Growth: Local and Aggregate Effects of City TFP Shocks on Wages, Rents, and Inequality," .

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2019): "The allocation of talent and us economic growth," *Econometrica*, 87, 1439–1474.

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.

Imbens, G. W. and M. Kolesar (2016): "Robust standard errors in small samples: Some practical advice," *Review of Economics and Statistics*, 98, 701–712.

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.

Jäger, S., B. Schoefer, S. Young, and J. Zweimüller (2020): "Wages and the Value of Nonemployment," *The Quarterly Journal of Economics*, 135, 1905–1963.

Jarosch, G., I. Sorkin, J. S. Nimczik, et al. (2019): "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*.

Jochmans, K. and M. Weidner (2019): "Fixed-Effect Regressions on Network Data," *Econometrica*, 87, 1543–1560.

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.

Kennan, J. and J. R. Walker (2011): "The effect of expected income on individual migration decisions," *Econometrica*, 79, 211–251.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88, 1859–1898.

Kucheryavyy, K. (2012): "Continuity of a model with a nested CES utility function and Bertrand competition," *Economics Letters*, 117, 473–476.

Lagakos, D. and M. E. Waugh (2013): "Selection, agriculture, and cross-country productivity differences," *American Economic Review*, 103, 948–80.

Lamadon, T., M. Mogstad, and B. Setzler (2018): "Imperfect competition and rent sharing in the us labor market," *University of Chicago mimeo*.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2018): "On worker and firm heterogeneity in wages and employment mobility: Evidence from danish register data," *PIER Discussion Papers*, 91.

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.

Macaluso, C. (2017): "Skill remoteness and post-layoff labor market outcomes," in *2017 Meeting Papers*, Society for Economic Dynamics, 569.

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.

Manning, A. and B. Petrongolo (2017): "How local are labor markets? Evidence from a spatial job search model," *American Economic Review*, 107, 2877–2907.

Marinescu, I. E., I. Ouss, and L.-D. Pape (2020): "Wages, hires, and labor market concentration," *NBER Working Paper*.

Monras, J. (2020): "Economic shocks and internal migration," Tech. rep.

Monte, F., S. J. Redding, and E. Rossi-Hansberg (2018): "Commuting, migration, and local employment elasticities," *American Economic Review*, 108, 3855–90.

Moretti, E. (2011): "Local labor markets," in *Handbook of labor economics*, Elsevier, vol. 4, 1237–1313.

Morlacco, M. (2018): "Market Power in Input Markets: Theory and Evidence from French Manufacturing," .

Naouas, A. and F. Romans (2014): "La négociation salariale d'entreprise de 2004 à 2010. Entre renforcement de l'obligation de négocier et baisse de l'implantation des délégués syndicaux," .

Nimczik, J. S. (2017): "Job mobility networks and endogenous labor markets," .

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

Osborne, M. J. and A. Rubinstein (1990): *Bargaining and markets*, Academic press.

Pérez-Cervantes, F. (2014): "Railroads and economic growth: A trade policy approach," Tech. rep., Working Papers.

Redding, S. J. (2016): "Goods trade, factor mobility and welfare," *Journal of International Economics*, 101, 148–167.

Redding, S. J. and E. Rossi-Hansberg (2017): "Quantitative spatial economics," *Annual Review of Economics*, 9, 21–58.

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.

Roback, J. (1982): "Wages, rents, and the quality of life," *Journal of political Economy*, 90, 1257–1278.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb (2019): "Fast and wild: Bootstrap inference in Stata using boottest," *The Stata Journal*, 19, 4–60.

Rosen, S. (1979): "Wage-based indexes of urban quality of life," *Current issues in urban economics*, 74–104.

Schmutz, B. and M. Sidibé (2019): "Frictional labour mobility," *The Review of Economic Studies*, 86, 1779–1826.

Schubert, G., A. Stansbury, and B. Taska (2020): "Employer Concentration and Outside Options," .

Silva, J. S. and S. Tenreyro (2006): "The log of gravity," *The Review of Economics and statistics*, 88, 641–658.

Simonovska, I. and M. E. Waugh (2014): "The elasticity of trade: Estimates and evidence," *Journal of international Economics*, 92, 34–50.

Sokolova, A. and T. Sorensen (2018): "Monopsony in labor markets: A meta-analysis," .

Sotelo, S. (2019): "Practical aspects of implementing the multinomial pml estimator," Tech. rep., Mimeo.

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.

Train, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

Webber, D. A. (2015): "Firm market power and the earnings distribution," *Labour Economics*, 35, 123–134.

Weidner, M. and T. Zylkin (2020): "Bias and Consistency in Three-way Gravity Models," .

White, H. (1980): "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica: journal of the Econometric Society*, 817–838.

Wong, H. C. (2019): "Wage Dispersion Across Firms," .

Young, A. (2013): "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 128, 1727–1785.

——— (2014): "Structural transformation, the mismeasurement of productivity growth, and the cost disease of services," *American Economic Review*, 104, 3635–67.

Zabek, M. (2020): "Local Ties in Spatial Equilibrium," .