The Birthplace Premium

Miguel Zerecero*

Toulouse School of Economics

JOB MARKET PAPER

This version: April 21, 2021

Click here for the latest version

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.

*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 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.[1] 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.[2]

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

[1] Bryan and Morten (2019) and Caliendo, Dvorkin, and Parro (2019) have models with costly adjustment of labor across regions; Ahlfeldt, Redding, Sturm, and Wolf (2015) and Monte, Redding, and Rossi-Hansberg (2018) propose a model where commuting is costly.

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

standard migration costs from the home bias. Third, I use the estimated model to quantify the 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 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 95 *départements* in continental France with great variation in size and connectivity.[3] 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 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 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 migra-

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

tion costs along with identifying location-specific characteristics, like productivities and amenities, 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.[4]

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.[5] 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 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.[6] 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.[7] 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.[8] Using the identified migration costs and count data on labor flows I esti-

[4]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.

[5]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.

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

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

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

mate the underlying migration probabilities via maximum likelihood. I show that the solution to this maximization problem is equivalent to solving for the 'source-country effects' of a balanced trade condition from a gravity-trade model.[9] 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.[10] 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, Chaudhuri, and McLaren (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 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

[9]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'.

[10]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.

of locations—means that birthplace is a big determinant of expected lifetime utility.

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.[11] 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.[12] 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

[11] 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.

[12] 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.

mobility as a force for welfare equalization.

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.[13] 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.[14] 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.[15] 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).

[13]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.

[14]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, Rodríguez-Clare, and Yi (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, Hurst, Jones, and Klenow (2019), using a model of occupational choice due to heterogeneous skills, study how discrimination of minorities affected aggregate productivity in the U.S.

[15]For example, see Redding (2016), Diamond (2016), Monte et al. (2018), Caliendo et al. (2019), Caliendo, Opromolla, Parro, and Sforza (2020), Schmutz and Sidibé (2019) and Monras (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 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, Figueirdo, and Woodward (2003), who show that there is an equivalence relation between the likelihood function of the conditional logit and the Poisson regression.[16] 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.[17] 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.[18] 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%.[19] 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

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

[17] 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.

[18] 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.

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

location. The distribution of these home preferences are the same conditional on birthplace, but in equilibrium, depressed places are going to endogenously retain workers that value their home 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.[20] 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.

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.[21] Appendix 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 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 *Départe-*

[20]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.

[21]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.

ments et Régions d'Outre-Mer (DROM), and the island of Corsica. In continental France there are 95 departements, which vary very much in size and connectivity among each other. For example, 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.[22]

To make the geographical unit of analysis comparable, instead of using directly the departments, 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.[23] 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.[24] 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.

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

[23] 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.

[24] 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.

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.

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.

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

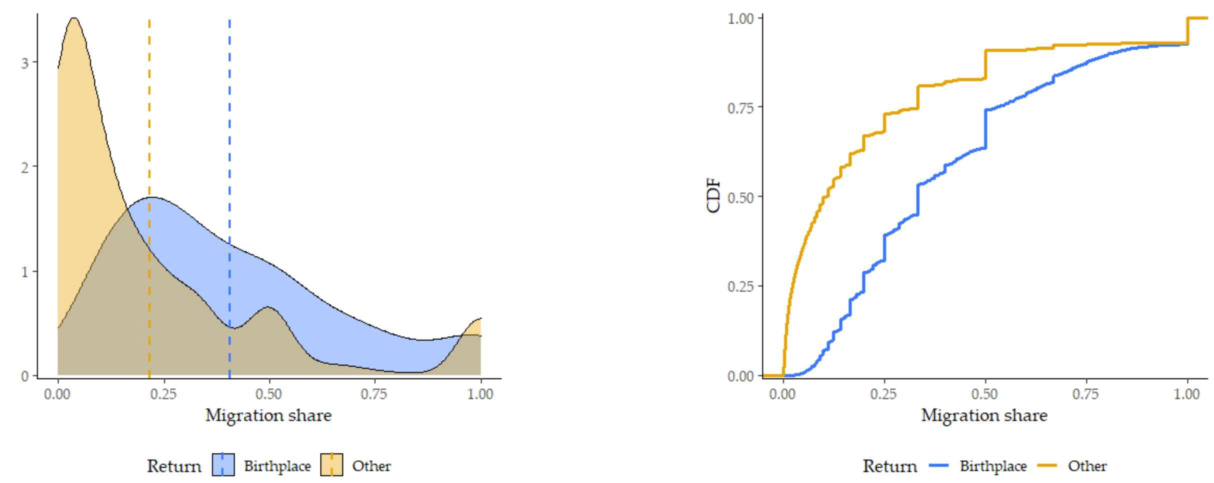


Figure 2: Distribution of conditional migration shares. These are defined as $\tilde{s}^{i,j}_{b,t} = \frac{L^{i,j}_{t,b}}{\sum_{k\neq i} L^{i,k}_{t,b}}$, where $L^{i,j}_{t,b}$ is the number of workers who were born in location b and that moved from location i to j at year t. Both plots distinguish between 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.[25]

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

[25]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.[26] 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.[27]

The first three columns of Table 2 show the results using different variables for distance.[28] 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.[29]

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.[30]

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 H. I estimate both models using *département* as locations instead of the aggregated regions I used here.

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

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

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

[29] 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. $\max_{b,j}(\alpha_1 + \beta_1 \log(\mathrm{d}^{b,j})) < 0$.

[30] 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, $L_{t,b}^{i,j}$ PPML			Migration shares, $L_{t,b}^{i,j} / \sum_k L_{t,b}^{i,k}$ 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, $L_{b,t}^{i,j}$ 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

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

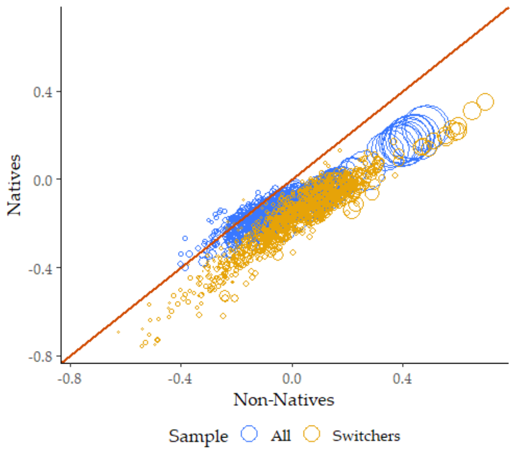
$$\mathbb{E}\left[\text{wage}|\text{No Native}\right] > \mathbb{E}\left[\text{wage}|\text{Native}\right],$$

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.[31]

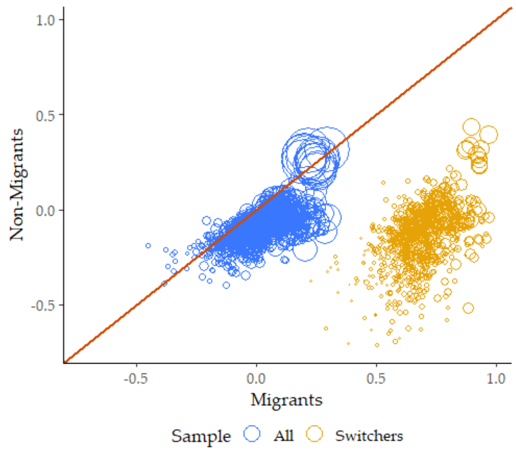
Figure 3a shows a plot where the y-axis corresponds to the mean of (log) wages in a particular location/period of natives after a normalization, whereas the x-axis does the same but for non-natives.[32] I include the 45 degree line to compare the relative magnitudes. Each of the blue circles in the Figure 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

[31] 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.

[32] 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.[33]

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

[33] In Appendix 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.

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.[34] 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.[35] 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 A contains the detailed derivations of the expressions in this section.

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

[34] 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.

[35] 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:

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

$$\beta(1-\rho)\max_k\left[\mathbb{E}_t\left(\mathrm{v}_{t+1,b}^k(\theta_{t,\iota}^k, \mathbf{\Theta}_{t+1,\iota})\right) - \tau^{i,k}\right]. \tag{2}$$

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

$$V^i_{t,b} = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho\overline{V}^i_{t+1,b} + \beta(1-\rho)\mathbb{E}_{\Theta_t}\left(\max_k\left[\overline{V}^k_{t+1,b} - \tau^{i,k} + \frac{\theta^k_{t,\iota}}{1-\beta\rho}\right]\right). \quad (3)$$

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 (3) as

$$V^i_{t,b} = B^i + \log\left(\frac{w^i_t}{P^i_t}\right) - \kappa^i_b + \beta\rho\overline{V}^i_{t+1,b} + \beta(1-\rho)\lambda\log\left(\sum_k \exp\left(\overline{V}^k_{t+1,b} - \tau^{i,k}\right)^{1/\lambda}\right). \quad (4)$$

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

$$p^{i,j}_{t,b} = \frac{\exp(\overline{V}^j_{t+1,b} - \tau^{i,j})^{\frac{1}{\lambda}}}{\sum_{k\in\mathcal{N}}\exp(\overline{V}^k_{t+1,b} - \tau^{i,k})^{\frac{1}{\lambda}}}. \quad (5)$$

The parameter $\lambda \equiv \delta/(1-\beta\rho)$, which appears in expressions (4) and (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 (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.[36] 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

[36] 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

$$\mathbb{E}_{\iota}(\exp(\theta_{t,\iota})|i \to j) = \frac{\Gamma(1-\delta)}{\exp(\gamma\delta)}(p^{i,j}_{t,b})^{-\delta}, \tag{6}$$

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 (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.

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.[37]

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 $\tilde{H}$.[38] 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 non-housing 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 non-housing 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 non-housing 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 non-housing good at location j is

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

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

[38] 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 $\left(\Gamma\left(1+(1-\sigma)/\varphi\right)\right)^{1/(1-\sigma)}$ 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.

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 E_t^j be the total expenditure in location j on non-housing goods. Also, let Y_t^j be the total income of intermediate firms in location j. Then, non-housing goods market clearing implies

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

The labor market clearing condition implies

$$w_t^jN_t^j=(1-\eta)Y_t^j.$$

while the market clearing condition for housing is

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

Finally, I assume trade is balanced, meaning

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

Substituting into the non-housing goods market clearing condition

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

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 A.3 contains the detailed derivations to get the expression above.

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 $\{W^i\}$

$$\left(W^i\right)^{\tilde{\varphi}\varphi(1+\varphi)}\left(N^i\right)^{(1+\eta\varphi)(1-\tilde{\varphi}(1+\varphi))} = \sum_k \tilde{\psi}^{k,i}\left(\tilde{A}^i\right)^{\varphi}\left(\frac{\tilde{A}^k}{\tilde{A}^i}\right)^{\varphi\tilde{\varphi}(1+\varphi)}\left(W^k\right)^{\varphi(\tilde{\varphi}(1+\varphi)-1)}\left(N^k\right)^{1-\tilde{\varphi}(1+\varphi)}. \tag{7}$$

The total efficiency units in a location

$$N^i = \sum_b N^i_b. \tag{8}$$

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

$$\left(U^i_b\right)^{1/\lambda} = \mathcal{B}^i\left(W^i\right)^{\frac{1-\alpha}{\delta}}\left(N^i\right)^{-\alpha/\delta}K^i_b\left(\Omega^i_b\right)^{\frac{\beta(1-\rho)}{\delta}}. \tag{9}$$

The option value of living in location i is equal to

$$\left(\Omega^i_b\right)^{1/\lambda} = \sum_k T^{i,k}\left(U^k_b\right)^{1/\lambda}. \tag{10}$$

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

$$L^i_b\left(U^i_b\right)^{-1/\lambda} = \sum_k T^{i,k}\left(\Omega^k_b\right)^{-1/\lambda}L^k_b. \tag{11}$$

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,

$$L_b = \sum_k L^k_b. \tag{12}$$

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

$$N^i_b\left(U^i_b\right)^{\frac{\delta-1}{\lambda}} = \sum_k\left(T^{i,k}\right)^{1-\delta}\left(\Omega^k_b\right)^{\frac{\delta-1}{\lambda}}L^k_b. \tag{13}$$

Appendix 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 $\{U^i_b\}_{b,i\in\mathcal{I}}$, option values $\{\Omega^i_b\}_{b,i\in\mathcal{I}}$, labor flows $\{L^i_b\}_{b,i\in\mathcal{I}}$ and efficiency units per birthplace cohort/location $\{N^i_b\}_{b,i\in\mathcal{I}}$, such that equations (7)-(13) are satisfied for all $i, b\in\mathcal{I}$.*

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).[39] The internal trade costs, $(\psi^{i,j})^{-\varphi}$ are taken from Combes, Lafourcade, and Mayer (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 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.

[39]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.[40] 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.[41]

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

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

$$\mathbb{E}_t(\ell_{t,b}^{i,j}) = p_{t,b}^{i,j} L_{t,b}^i = \exp\left(\mathcal{O}_{t,b}^i + \mathcal{D}_{t,b}^j - \tau^{i,j}/\lambda\right), \tag{14}$$

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 (14). This moment condition is equivalent to the first order condition of a Poisson regression (or Poisson Pseudo Maximum Likelihood).

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

[41] 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 $\tau^{i,j}$.[42] 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

$$\frac{p^{i,j}_{t,b}}{p^{i,i}_{t,b}} \frac{p^{j,i}_{t,b}}{p^{j,j}_{t,b}} = \exp\left(\mathcal{D}^j_{t,b} - \mathcal{D}^i_{t,b} - \tau^{i,j}/\lambda\right) \exp\left(\mathcal{D}^i_{t,b} - \mathcal{D}^j_{t,b} - \tau^{i,j}/\lambda\right) = \exp(-2\tau^{i,j}/\lambda). \tag{15}$$

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 (14) and (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-

[42] 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.

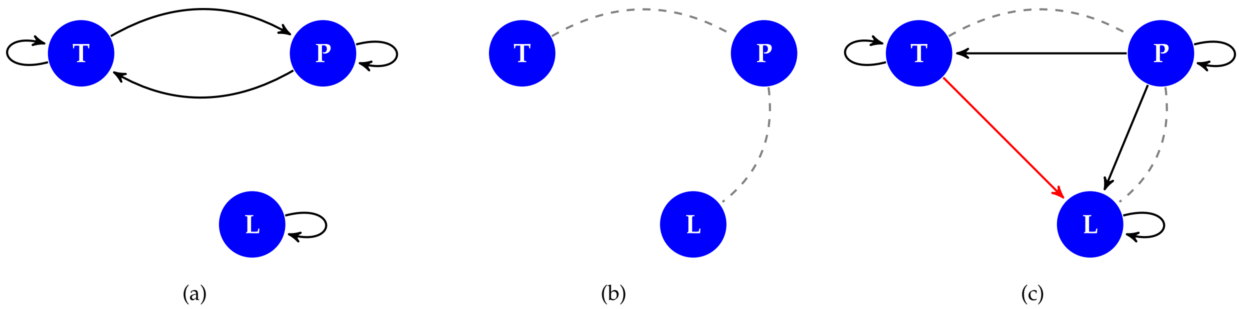


Figure 4: Identification of Migration Costs in a Three Locations Example. The three locations are Toulouse (T), Paris (P) and Lyon (L). The left and right panel plot the graph representation of data where each (solid) edge represents some positive worker flow. The middle panel is an undirected graph 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 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 $\{L_{t,b}^{n,m}\}_{n,m\in\{T,P,L\}}$ 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 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

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

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

$$\frac{p^{T,L}_{t',b'}}{p^{T,T}_{t',b'}} \frac{p^{P,T}_{t',b'}}{p^{P,L}_{t',b'}} = \exp(-\tau^{T,L}/\lambda - \tau^{P,T}/\lambda + \tau^{P,L}/\lambda). \tag{16}$$

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 (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.[43] Therefore, for my context, I need identification conditions that are simple and easy to verify. Proposition 7 in Appendix 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 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

[43]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

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

is a bias-corrected estimate of the migration cost, where $\frac{\widehat{\tau}^{i,j}}{\lambda}$ correspond to the estimate with the whole sample, and $\frac{\widehat{\tau}^{i,j}_{1}}{\lambda}$ and $\frac{\widehat{\tau}^{i,j}_{2}}{\lambda}$ correspond to the estimates for each half of the panel.[44]

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 6.2.

5.2 Wage dispersion parameter δ

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

$$\mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,b}\right)\right) = \log(w^{j}_{t}) - \delta \log(p^{i,j}_{t-1,b}). \tag{17}$$

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

[44] 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.

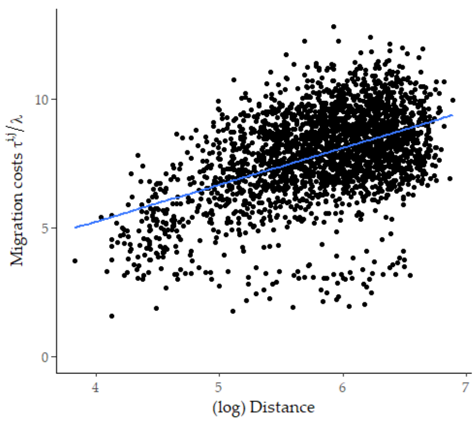


Figure 5: Migration Costs vs Distance. The graph plots the migration costs vs (log) geodesic distance. Each point corresponds to a mobility cost and (log) geodesic distance of a pair of locations. The lines correspond to fitted a linear model. The slope corresponding is 1.43, s.e. 0.05 and the R^2 is 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 $-\delta$.[45] Variation in migration costs imply variation in migration probabilities, which ultimately identifies δ. Substituting 5 into 17, I can write the previous expression as

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

where $\tilde{\mathcal{O}}_{t,b}^{i}$ and $\tilde{\mathcal{D}}_{t,b}^{j}$ 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.[46]

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

[45]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.

[46]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. [47] Details of the first stage regression are in Appendix 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}).[48] 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.[49] Taking their estimate as a benchmark, this means that, in absence of a persistence force, i.e. $\rho = 0$, the migration elasticity in France $1/0.145 \approx 7$ is four 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.

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:

$$\log \mathcal{L} = \sum_t \sum_b \sum_{i,j} \ell_{t,b}^{i,j} \log \left(\frac{\exp\left(\mathcal{D}_{t+1,b}^{j} - \tau^{i,j}/\lambda\right)}{\sum_k \exp\left(\mathcal{D}_{t+1,b}^{k} - \tau^{i,k}/\lambda\right)} \right), \tag{18}$$

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 (18) are the same that solve the following system of equations*

$$\sum_i \ell_{t,b}^{i,j} = \sum_i \frac{\exp\left(\mathcal{D}_{t+1,b}^{j} - \tau^{i,j}/\lambda\right)}{\sum_k \exp\left(\mathcal{D}_{t+1,b}^{k} - \tau^{i,k}/\lambda\right)} \sum_h \ell_{t,b}^{i,h}, \quad \forall i,j \in \mathcal{I}. \tag{19}$$

[47] 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.

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

[49] Bryan and Morten (2019) estimate the dispersion parameter by running a regression of average log wages against the (log) migration probabilities corresponding to equation (17).

Proof. See Appendix 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 $L^{j,\text{dest}}_{t,b} = \sum_i p^{i,j}_{t,b} L^{i,\text{orig}}_{t,b}$, where the inward labor flow is $L^{j,\text{dest}}_{t,b} = \sum_i \ell^{i,j}_{t,b}$ and the outward labor flow is $L^{i,\text{orig}}_{t,b} = \sum_h \ell^{i,h}_{t,b}$ 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, $p^{i,j}_{t,b} L^{i,\text{orig}}_{t,b}$, have to be equal to the total labor observed in that destination, $L^{j,\text{dest}}_{t,b}$. 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.[50]

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 $\mathcal{D}^j_{t+1,b}$ and migration costs $\tau^{i,j}/\lambda$ to maximize the conditional likelihood (18) is equivalent to doing a Poisson-Pseudo-Maximum Likelihood (PPML) estimation adding origin fixed effects, whose moment condition is equal to equation (14).[51] This moment condition is derived from the equilibrium expression of labor flows, which corresponds to a 'general gravity' framework.[52] 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.[53] 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

[50] In Appendix 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.

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

[52] 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.

[53] 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.[54] 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.[55] 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 (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.

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 (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.

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

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

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

[55]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 (17), conditional migration probabilities (5), and lifetime utility (4), after some algebra I obtain that

$$\zeta_{t,b}^{i,j} = \mathcal{C}_t^j - \mathcal{C}_t^b + \left(\frac{(1-\beta)\delta}{\lambda}\right)(\tau^{i,j} - \tau^{i,b}) + \frac{\delta}{\lambda}\kappa_b^j + \text{exp. error,} \tag{20}$$

where $\mathcal{C}_t^j$ captures all the terms related to location j that are independent of birthplace.[56]

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 (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

[56]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.[57]

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,*

$$\frac{\zeta_{t,b}^{i,j} + \zeta_{t,j}^{i,b}}{2\delta} = \frac{1}{\lambda}\kappa_b^j + exp.\ error. \tag{21}$$

Proof. It follows from equation (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 (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.[58]

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.[59] 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.

[57] 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.

[58] 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.

[59] 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.

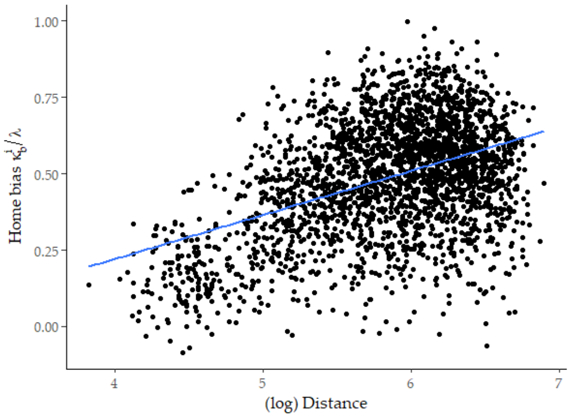


Figure 6: Home Bias vs Distance. The graph plots the home bias vs (log) geodesic distance. Each point corresponds to a mobility cost and (log) geodesic distance of a pair of locations. The lines correspond to fitted a linear model. The slope corresponding is 0.14 s.e. 0.006 and the R^2 is 0.18.

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 (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.[60] 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.[61]

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.

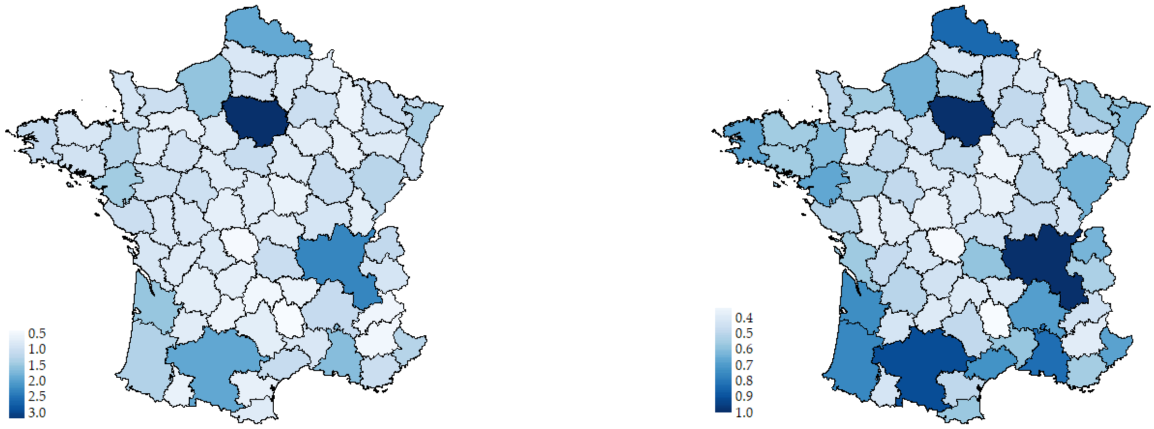
Combining the goods and labor market clearing conditions we get:

$$w_t^j N_t^j = \sum_{i \in \mathcal{I}} \frac{S^j \tilde{\psi}^{i,j}}{\sum_{k \in \mathcal{N}} S^k \tilde{\psi}^{i,k}} w_t^i N_t^i, \tag{22}$$

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

[60] 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.

[61] 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 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 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.

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 (3) we get

$$V^j_{t,b} = \tilde{B}^j - \alpha \log\left(N^j_t\right) + (1-\alpha)\log\left(\frac{w^j_t}{P^j_{T,t}}\right) + \beta\rho V^j_{t+1,b}$$

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

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:

$$\log\left(\left(\frac{p^{i,j}_{t,b}}{p^{i,x}_{t,b}}\right)\left(\frac{p^{i,x}_{t+1,b}}{p^{i,j}_{t+1,b}}\right)^{\beta\rho}\left(\frac{p^{x,i}_{t+1,b}}{p^{j,i}_{t+1,b}}\right)^{\beta(1-\rho)}\right) = \frac{\tilde{B}^j}{\lambda} - \frac{\alpha}{\lambda}\log\left(\frac{N^j_{t+1}}{N^x_{t+1}}\right) + \frac{(1-\alpha)}{\lambda}\log\left(\frac{w^j_{t+1}}{P^j_{T,t+1}}\frac{P^x_{T,t+1}}{w^x_{t+1}}\right)$$

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

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.

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 B.

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 $v_b^{j,i}(\theta_\iota^i) = V_b^i + \theta_\iota^i / (1 - \beta\rho)$.[62] Then, the expected utility of workers living in i is

$$\tilde{V}_b^i = V_b^i - \frac{\lambda}{L_b^i} \sum_j \log\left(p_b^{j,i}\right) p_b^{j,i} L_b^j. \tag{23}$$

The second term to the right corresponds to an average selection term for workers living in i born in b. Appendix 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.*[63]

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

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

[62] This expression comes from combining equations (1) and (3).

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

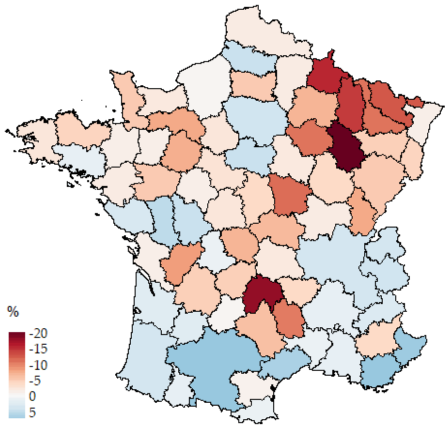


Figure 8: Birthplace Premium. The map shows the different birthplace premia ζ_b for the different birth cohorts. The birthplace premium is the excess welfare, in consumption terms, that each birth cohort has on excess to the national average.

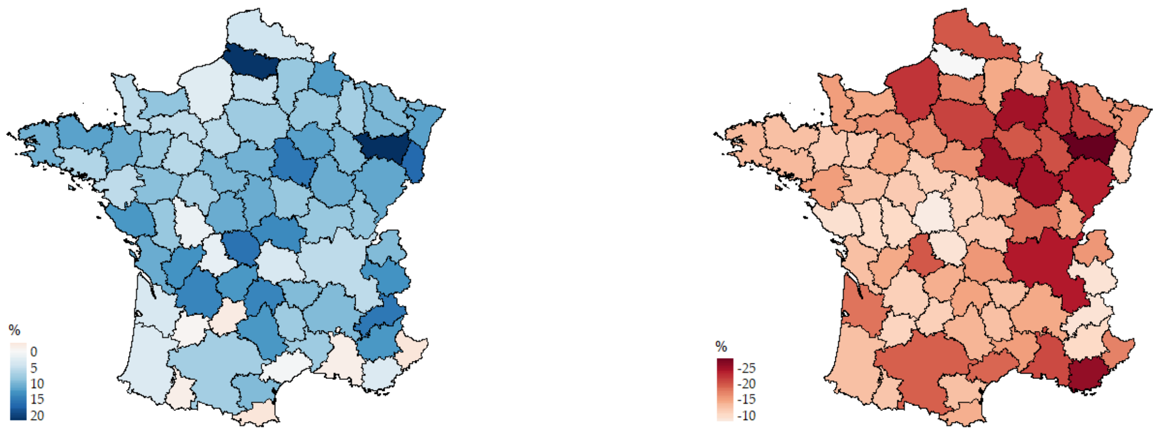
for the different cohorts. A location in the map represents a *birth* location and the color within a location represents the birthplace premium of the cohort born in such location. In absence of the home bias, these premia should all be equal to zero.

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.'[64] 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

[64] 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.[65] 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

[65]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 $\bar{V}_b^j$, 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.[66] 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.[67] 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.[68] When increasing the home bias, the mobility patterns differ more, causing inequality to increase as well as the importance of the between-birthplace component.[69] 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—along with the home bias—that determines the outside option of workers and influence their loca-

[66] 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 D

[67] 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.

[68] 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.

[69] 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.

tion 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.

6.2 Comparison of migration costs and home bias

In the model sketched in Section 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 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 (4) and the migration probability equation (5) into the expected wage equation (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

$$\xi_{t,b}^{j}(\kappa) = \exp\left(\kappa_b^j - \beta(1-\rho)(1-\beta\rho)\mathbb{E}_t\sum_{s=0}^{\infty}(\beta\rho)^{s-1}\left(\log\left(\Omega_{t+s,b}^{j}\right) - \log\left(\Omega_{t+s,j}^{j}\right)\right)\right) - 1. \quad (24)$$

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*

$$\frac{1}{2}\left(\log(1+\xi^{j}_{t,b}(\kappa)) + \log(1+\xi^{b}_{t,j}(\kappa))\right) = \frac{1}{2}\left(\begin{array}{c} \mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,b}\right)\right) - \mathbb{E}\left(\log\left(\text{wage}^{i,j}_{t,j}\right)\right) + \\ \mathbb{E}\left(\log\left(\text{wage}^{i,b}_{t,j}\right)\right) - \mathbb{E}\left(\log\left(\text{wage}^{i,b}_{t,b}\right)\right) \end{array}\right) \tag{25}$$

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 (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.[70] 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

[70]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 (18).[71]

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.[72] Such migration costs might look implausible a priori, but when interpreted as average wage differences between

[71]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.

[72]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 (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

$$\frac{\exp(\kappa_b^j) - 1}{\exp\left(\tau^{b,j}\right)^{(1-\beta\rho)} - 1} \tag{26}$$

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 (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.

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 $Y^i = A_L^i L^i$, where A_L^i is the labor productivity in location i. Labor productivity A_L^i 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

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

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

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

The covariance term $\widetilde{\text{cov}}(\cdot)$ gives a measure of how concentrated is the population in the most productive locations.[73]

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

$$\begin{aligned}
\mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \mathcal{Y}(\mathbf{A}_L, \mathbf{L}) &= \mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \left(\overline{A}_L + \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L})\right) \\
&= \mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \left(\overline{A}_L + \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}')\right) + \left(\widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}') - \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L})\right) \\
&= \underbrace{\mathcal{Y}(\mathbf{A}'_L, \mathbf{L}') - \mathcal{Y}(\mathbf{A}_L, \mathbf{L}')}_{\text{Change Selection}} + \underbrace{\left(\widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L}') - \widetilde{\text{cov}}(\mathbf{A}_L, \mathbf{L})\right)}_{\text{Change Composition}}. \qquad (27)
\end{aligned}$$

The first term in equation (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 term captures the composition effect: how much of the output change be explained by changes

[73] 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.

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.[74]

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, removing home bias can lead to large reallocation of labor towards more productive regions, but

[74] 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.

the removal of migration costs allows people to sort better but mostly close to their home location.

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

$$-\kappa_b^i + \beta(1-\rho)\left(\left(\log(\Omega_b^i) - \overline{\log(\Omega^i)}\right) - \left(\log(\Omega_b^b) - \overline{\log(\Omega^b)}\right)\right), \tag{28}$$

where $\overline{\log(\Omega^i)}$ is the across-birthplace average of option values for location i.[75] 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.

[75] 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 (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.

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.[76] Appendix 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

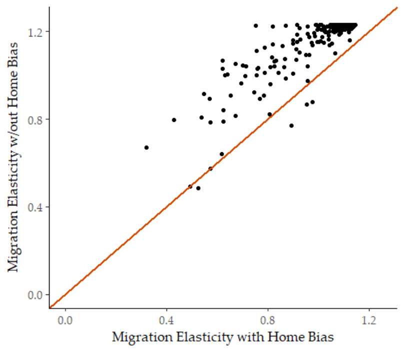
$$\epsilon^{i,j} = \frac{1}{\lambda}\left(1 - \sum_b \frac{L_b^{i,j}}{L^{i,j}} p_b^{i,j}\right),$$

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 $\epsilon_b^{i,j} = \frac{1}{\lambda}\left(1 - p_b^{i,j}\right)$.

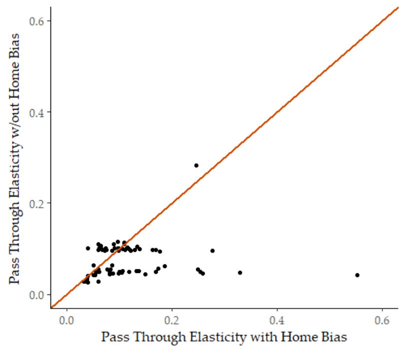
Figure 10a compares the different elasticities for the models with and without home bias in the steady state. Each dot corresponds to the migration elasticity for a pair of locations i, j. In almost all of the cases the elasticities are larger in the model without home bias. Thus, the predicted migration 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

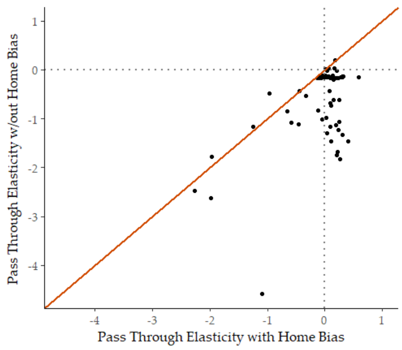
[76] I estimate the migration costs doing the same Poisson regression as in section 5.1 but the dependent variable is the total flow from i to j, i.e. $L^{i,j}$ and the origin and destination fixed effects do not depend on birthplace.



(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.[77] 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.[78] 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.[79] Neglecting the home bias might lead to wrong conclusions when analyzing counterfactual scenarios.

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.

[77] 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.

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

[79] 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.

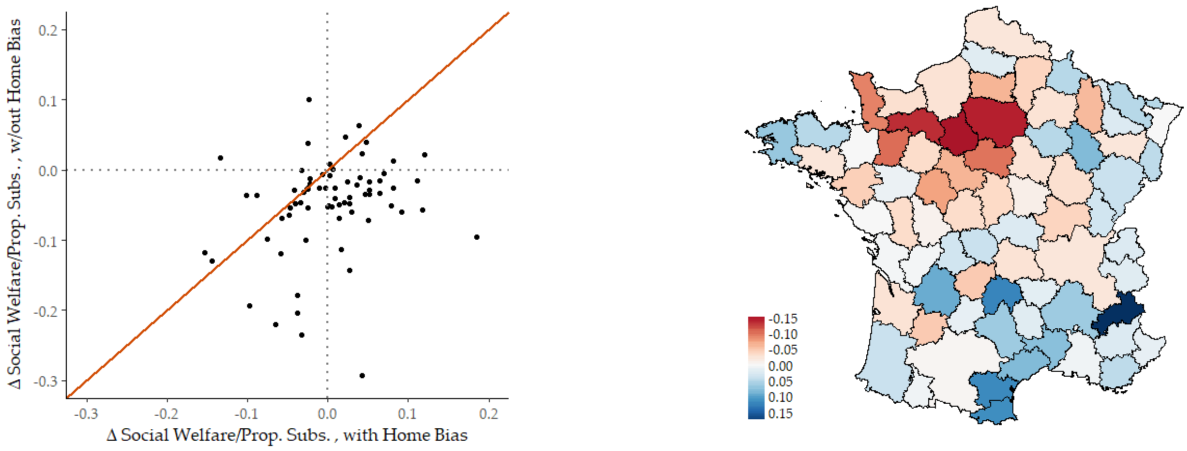
I compare the effects on social welfare of a 10% place-based labor subsidy between the model estimated with and without home bias.[80] In the exercise, I subsidize each location, one-by-one, in both models and compare the changes in social welfare. I find that the model without home bias—where the efficiency 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.[81]

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.

[80] 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.

[81] 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 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, Kline, and Yagan (2020), can be important to not overestimate, either the effects, or the costs, of implementing such policies.

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.

References

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 (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.

Artuç, E., S. Chaudhuri, and J. McLaren (2010): "Trade shocks and labor adjustment: A structural empirical approach," *American economic review*, 100, 1008–45.

Bryan, G. and M. Morten (2019): "The aggregate productivity effects of internal migration: Evidence from Indonesia," *Journal of Political Economy*, 127, 2229–2268.

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., L. D. Opromolla, F. Parro, and A. Sforza (2020): "Goods and factor market integration: a quantitative assessment of the eu enlargement," Tech. rep.

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.

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," .

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

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*.

Fack, G. (2006): "Are housing benefit an effective way to redistribute income? Evidence from a natural experiment in France," *Labour Economics*, 13, 747–771.

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.

Gaubert, C., P. Kline, and D. Yagan (2020): "Place-Based Redistribution," .

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.

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*.

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.

Kennan, J. and J. R. Walker (2011): "The effect of expected income on individual migration decisions," *Econometrica*, 79, 211–251.

Lagakos, D. and M. E. Waugh (2013): "Selection, agriculture, and cross-country productivity differences," *American Economic Review*, 103, 948–80.

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.

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

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.

Roback, J. (1982): "Wages, rents, and the quality of life," *Journal of political Economy*, 90, 1257–1278.

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.

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.

SOTELO, S. (2019): "Practical aspects of implementing the multinomial pml estimator," Tech. rep., Mimeo.

TRAIN, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

WEIDNER, M. AND T. ZYLKIN (2020): "Bias and Consistency in Three-way Gravity Models," .

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," .

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.

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

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

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

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

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 $\exp(-\delta\overline{\gamma})\left(p^{i,j}_{t,b}\right)^{-\delta}$. [82] 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 $-\delta(\log(p^{i,j}_{t,b} + \overline{\gamma}))$.[83] 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

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

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, Arkolakis,

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

[83] 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.

and Takahashi (2020b), 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 Y^i in location i is equal to

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

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

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

Using the assumption that trade is balanced, i.e. $Y_i = E_i$ 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

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

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

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

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.[84] This is the expression in the main text.

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

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

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

where $\tilde{C}$ is a constant. Recall that $W_t^i \equiv \frac{w_t^i}{P_{T,t}^i}$. Defining $\tilde{B}^i \equiv B^i + \tilde{C} + \alpha \log\left(H^i\right)$ we get the expression in the main text up to a constant

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

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

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

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

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

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

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

$$N^i = \sum_b N_b^i, \quad (40)$$

$$L_b = \sum_k L_b^k. \quad (41)$$

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.

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.[85] Formally, we need to find $\theta^j_{1,t-1} - \theta^j_{2,t-1}$ such that

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

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

[85] 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

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

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.[86] 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.

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.

[86] 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

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

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

A.8 Model with Place-Based Policies

In section 6.6 I introduce place based policies in the form of direct labor subsidies. These subsidies are entirely financed by a labor tax that is levied among all workers in all locations. This would generate trade imbalances as the expenditure on non-housing goods—which equals the demand for tradable intermediate inputs—would differ from the income of tradable intermediate firms, whenever the labor tax and subsidies differ.

In this section I show how the model changes with the introduction of such policies. There are two main changes. First, the presence of trade imbalances prevents the characterization of the static equilibrium into one single equation. Second, the overall price level in a location is affected as it affects the demand for housing, thus changing the expression for the lifetime utility derived in the previous sections.

In what follows I consider an exogenous local labor subsidy S^i and a labor tax tax_w, where I omitted the time subscript as I only consider the steady-state. Although in the main text they are not considered, I include also some exogenous location-specific taxes on housing (levied to house owners) tax^i_H. This more general case might be interested to some readers. To solve for the case in the main text one just needs to adjust all the expressions by putting these housing taxes equal to zero.

The total expenditures of non-housing goods in location i are

$$E^i = \underbrace{(1-\alpha)w^iN^i(1+S^i-\text{tax}_w)}_{\text{Demand workers}} + \underbrace{(1-\text{tax}^i_H)\left(\frac{\eta}{1-\eta}w^iN^i + \alpha w^iN^i(1+S^i-\text{tax}_w)\right)}_{\text{Demand House Owners}},$$

where the first term inside the brackets of the demand from house owners is the expenditure of housing from intermediate firms, where I use the fact that $Y^i = w^iN^i/(1-\eta)$. The second term is the demand from workers. After some algebra, total expenditures is simplified to

$$E^i = Y^i(1+\overline{S^i}),$$

where

$$\overline{S^i} \equiv (1-\eta)(S^i - \text{tax}_w)(1-\alpha\text{tax}^i_H) - \text{tax}^i_H(\eta + \alpha(1-\eta)).$$

Thus, $\overline{S^i} = (E^i - Y^i)Y^i$ is equal to the trade deficit as proportion of local tradable output.

Define

$$\widetilde{S^i} \equiv S^i - \text{tax}_w \tag{42}$$

as the net labor subsidy. Thus, given some taxes on housing, there is a direct correspondence between the net labor subsidy and the trade deficit $\overline{S^i}$

$$\widetilde{S^i} = \frac{\overline{S^i} + \text{tax}^i_H(\eta + \alpha(1-\eta))}{(1-\eta)(1-\alpha\text{tax}^i_H)}. \tag{43}$$

In the case where there are no housing taxes ($\text{tax}^i_H = 0$) and no demand for housing from firms ($\eta = 0$) then $\widetilde{S^i} = \overline{S^i}$.

Total income for tradable firms in location i is

$$Y^i = \sum_j \pi^{j,i}E^j,$$
$$\leftrightarrow w^iN^i = \sum_j \pi^{j,i}w^jN^j(1+\overline{S^j}).$$

By similar arguments as the ones developed above we have that, given a vector of efficiency units per location N^i and trade deficits $\overline{S^i}$, the static equilibrium can be characterized by

$$w^iN^i = \sum_j \left(\tilde{A}^i\right)^{\varphi}\left(w^i\left(N^i\right)^{\eta}\right)^{-\varphi}\tilde{\psi}^{j,i}\left(P^j_T\right)^{\varphi}\boldsymbol{C}^{-\varphi}w^jN^j(1+\overline{S^j}), \tag{44}$$

$$\left(P^i_T\right)^{-\varphi} = \boldsymbol{C}^{-\varphi}\sum_j \left(\tilde{A}^j\right)^{\varphi}\left(\psi^{i,j}\right)^{-\varphi}. \tag{45}$$

As done above, to rewrite the lifetime utility as a function of the wage deflated by the price of tradables, I need to rewrite the overall price level of a location as a function of both the price of tradables, the wage bill and the total housing supply. First, note that with the inclusion of subsidies and taxes , the total demand for housing in location i is equal to

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

Then, the price for the final good in location i is

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

The budget constraint for an individual with skill θ_ι is also changed to include the net labor subsidy. Thus the total disposable labor income in location i is $w^i N^i (1 + \widetilde{S}^i)$.

The lifetime utility is then

$$\begin{aligned} V_b^i &= B^i + \tilde{\boldsymbol{C}} + \alpha \log\left(H^i\right) - \alpha \log\left(N^i\right) - \alpha \log(\eta + \alpha(1-\eta)(1+\widetilde{S}^i)) + \log(1+\widetilde{S}^i) \\ &+ (1-\alpha) \log\left(\frac{w^i}{P_T^i}\right) - \kappa_b^i + \beta\rho V_b^i + \beta(1-\rho)\lambda \log\left(\sum_k \exp\left(V_b^k - \tau^{i,k}\right)^{1/\lambda}\right). \end{aligned}$$

From the above expression is clear that the flow utility from a positive net subsidy is increased directly, as expressed by the term $\log(1+\widetilde{S}^i)$. However, there is also a negative effect as the price of housing increases, reflected in the term $\alpha \log(\eta + \alpha(1-\eta)(1+\widetilde{S}^i))$ which enters negatively. Defining $\tilde{B}^i \equiv B^i + \tilde{\boldsymbol{C}} + \alpha \log\left(H^i\right)$ we get

$$\begin{aligned} V_b^i &= \tilde{B}^i - \alpha \log\left(N^i\right) - \alpha \log(\eta + \alpha(1-\eta)(1+\widetilde{S}^i)) + \log(1+\widetilde{S}^i) \\ &+ (1-\alpha) \log\left(\frac{w^i}{P_T^i}\right) - \kappa_b^i + \beta\rho V_b^i + \beta(1-\rho)\lambda \log\left(\sum_k \exp\left(V_b^k - \tau^{i,k}\right)^{1/\lambda}\right). \end{aligned}$$

Now, define $\widetilde{S^i}_{aux} \equiv \eta + \alpha(1-\eta)(1+\widetilde{S}^i)$. Using the same definitions for the characterization of the steady-state above, we can rewrite above's expression as

$$\left(U_b^i\right)^{1/\lambda} = \mathcal{B}^i \left(\widetilde{S^i}_{aux}\right)^{\frac{-\alpha}{\delta}} \left(1+\widetilde{S}^i\right)^{\frac{1}{\delta}} \left(w^i/P_T^i\right)^{\frac{1-\alpha}{\delta}} \left(N^i\right)^{-\alpha/\delta} K_b^i \left(\Omega_b^i\right)^{\frac{\beta(1-\rho)}{\delta}}. \tag{46}$$

Finally, the labor tax is set such that the government budget is balanced. The government budget can be written as

$$\sum_i w^i N^i (S^i - \text{tax}_w) = \sum_i \text{tax}_H^i w^i N^i \left(\frac{\eta}{1-\eta} + \alpha(1 + S^i - \text{tax}_w) \right).$$

Solving for the labor tax we get

$$\text{tax}_w = \frac{\sum_i w^i N^i (S^i - \text{tax}_H^i(\alpha(1+S^i) + \frac{\eta}{1-\eta}))}{\sum_i w^i N^i (1 - \alpha \text{tax}_H^i)}. \tag{47}$$

Given some labor subsidies and housing taxes $\{S^i, \text{tax}_H^i\}$, the steady-state continuous population equilibrium with place based policies and housing taxes are the vectors of location-specific wages, prices, supply of efficiency units, trade deficits, and net subsidies $\{w^i, P_T^i, N^i, \overline{S^i}, \widetilde{S}^i\}$, as well as location-birthplace-specific labor supplies (both in number of workers and efficiency units), the lifetime utilities, and the option values $\{L_b^i, N_b^i, U_b^i, \omega_b^i\}$ and labor tax tax_w such that they solve the system form by 37-47 for all i, b.

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 C.2. Note than in the case of $K = 1$ the result is just an application of a standard contraction mapping.

B.1 Solution for baseline model

I will describe how the sequence of the algorithm goes in the baseline model without trade deficits. 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[87]

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

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

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

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

$$\mathbf{L}_b = \mathbf{P}_b'\mathbf{L}_b. \tag{48}$$

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

Note that system of equations (48) 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 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.[88] 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

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

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 $\{N^{i,(0)}\}_{i\in\mathcal{I}}$

[88] I borrowed this idea from Eckert (2019).

B.2 Solution of model with place-based policies

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 (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 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.

C.1 Migration Costs

Using labor flows $\ell_{t,b}^{i,j}$ or migration shares $s_{t,b}^{i,j}$ 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 $\mathcal{E}_{t,b} = \{(i,j) | (i,j) \in \mathcal{I}^2 \text{ and } s_{t,b}^{i,j} > 0\}$.*

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.

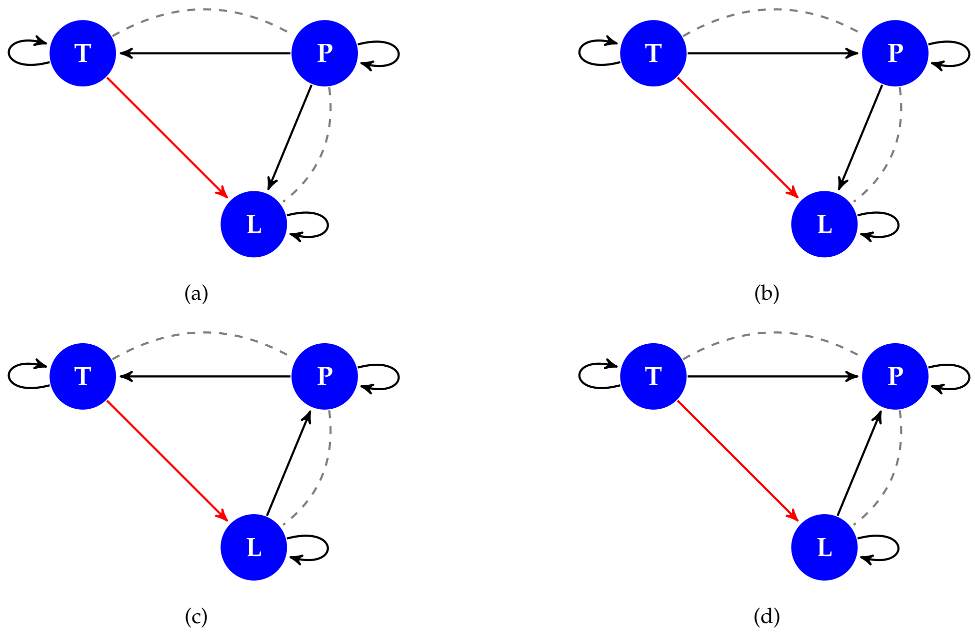


Figure 12: Different cases for weakly connectivity between T and L through P.

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 $\mathcal{G}_{t,b}$.*

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 $\mathcal{P}_{t,b}(i, j)$. 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.

C.1.1 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.

The algorithm requires some explanation. It can be separated in two main parts. First it begins by finding the directly identified pairs. Second it initializes the recursive algorithm to find identified pairs by using previously identified pairs. The directly identified pairs just initialize this algorithm.

From lines 2 to 13, the algorithm creates the symmetric matrix $\tilde{M}_0$ where coordinate (i, j) equal to 1 means that the migration costs between locations i, j are directly identified for some t, b. The for loop started in line 4 just checks whether there are flows that stay within the same location, meaning that the diagonal has a non-zero elements. If there are no flows it eliminates all the connections in the original adjacency matrix. The matrix $\mathcal{B}^d_{t,b}$ indicates whether two locations are directly identified for t, b, meaning there are flows from both directions and there are flows staying within each of the locations. The matrix $\tilde{M}_0$ would just collect all the information regarding the direct identified matrices $\mathcal{B}^d_{t,b}$ for all t, b.

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**
  for all $i \in \{1, ..., N\}$ **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
end for
$\mathcal{B}^{d}_{t,b} \leftarrow \mathcal{A}^{d}_{t,b} * \left(\mathcal{A}^{d}_{t,b}\right)'$, for all t, b.
$\tilde{M}_0 \leftarrow \sum_{t,b} \mathcal{B}^{d}_{t,b}$.
$\tilde{M}_0 \leftarrow \tilde{M}_0 > 0$.
$diff = 1$
while $diff \neq 0$ **do** $M_0 \leftarrow \tilde{M}_0$.
  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.
    $M_1 \leftarrow \mathcal{P}^{aux}_{t,b} * \mathcal{A}^{wc}_{t,b} + M_0$.
    $M_1 \leftarrow M_1 > 0$.
    $M_0 \leftarrow M_1$.
  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

The recursive algorithm starts in line 15. Using a matrix of previously identified pairs $\tilde{M}_0$, the algorithm first creates the matrices for all t, b of weakly connected nodes that satisfy the requirement that both locations have flows that stay (lines 17 and 18). Then it does an element-wise multiplication with the matrix that stores the information of previously identified pairs. Thus $\mathcal{A}^{aux}_{t,b}$ would indicate whether two nodes are weakly connected if those nodes are previously identified. Then using such matrix we can find whether there is an undirected path between any two nodes. If such a path exists, then the second condition of Proposition 7 would be satisfied. As it could be any path, in practice what I do is to use Dijkstra's algorithm to find the minimum length separating any two nodes. If the distance is infinite, then no path exists between the two nodes. I re-actualize the information in matrix M_1 by checking whether two nodes have are weakly connected and there is path between the two that consist of previously identified pairs. The algorithm then loops for all t, b where it re-actualizes the information of which pairs are actually identified. After looping for all t, b it starts again until there is no change in the information of which pairs are identified.

C.2 Conditional migration probabilities

In this section I give the proof for Proposition 3. From the maximization of the log-likelihood 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. □

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.

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.

D Estimation details

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.

D.2 Conditional migration probabilities

Estimating via maximum likelihood the destination fixed effects on (18) boils down to solving different systems of equations for each birth cohort/period

$$L^{j,\text{dest}}_{t,b} = \sum_i \frac{\exp(\mathcal{D}^j_{t+1,b} - \tau^{i,j}/\lambda)}{\sum_k \exp(\mathcal{D}^k_{t+1,b} - \tau^{i,k}/\lambda)} L^{i,\text{orig}}_{t,b}, \quad \text{for all } i, j, h \in \mathcal{I}. \tag{49}$$

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.[89]

Before, jumping to the description of the algorithm, note that (49) 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.

D.3 Trade costs

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

$$X^{i,j} = \left(x^i_t \psi^{j,i} / A^i\right)^{-\varphi} \left(P^j_t\right)^{\varphi} \overline{\Gamma}^{\varphi} E^j_t. \tag{50}$$

[89] In his application, Pérez-Cervantes (2014) computes the source effects, as in (22), for a single system with more than 3,000 equations.

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**

In logarithms

$$\log X^{i,j} = \mathcal{O}_t^i + \mathcal{D}_t^j - \varphi \log(\psi^{j,i}), \tag{51}$$

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.

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.

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.[90] 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

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

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.

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.

F Is Birthplace a good 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 I 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%.

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.

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 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.

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.[91]

[91] 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.

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 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.[92] 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.

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.

[92]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.

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

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

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 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.[93]

To see the equivalence between the likelihood and the Poisson regression, consider $\ell_{t,b}^{i,j}$ to be independently distributed with

$$\mathbb{E}\left(\ell_{t,b}^{i,j}\right) = \tilde{\mu}_{t,b}^{i,j} = \exp\left(\mathcal{O}_{t+1,b}^{i} + \mathcal{D}_{t+1,b}^{j} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right).$$

The log-likelihood would then be written as

$$\begin{aligned}\log \mathcal{L}_P &= \sum_t \sum_b \sum_{i,j} \left(-\tilde{\mu}_{t,b}^{i,j} + \ell_{t,b}^{i,j} \log \tilde{\mu}_{t,b}^{i,j} - \log \ell_{t,b}^{i,j}!\right) \\ &= \sum_t \sum_b \sum_{i,j} \left(-\exp\left(\mathcal{O}_{t+1,b}^{i} + \mathcal{D}_{t+1,b}^{j} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right) + \ell_{t,b}^{i,j}\left(\mathcal{O}_{t+1,b}^{i} + \mathcal{D}_{t+1,b}^{j} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right) - \log \ell_{t,b}^{i,j}!\right).\end{aligned}$$

Taking the first order conditions with respect to $\mathcal{O}_{t+1,b}^{i}$ we obtain

$$\frac{\partial \log \mathcal{L}_P}{\partial \mathcal{O}_{t+1,b}^{i}} = \sum_j \left(\ell_{t,b}^{i,j} - \exp\left(\mathcal{O}_{t+1,b}^{i} + \mathcal{D}_{t+1,b}^{j} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right)\right) = 0$$

and therefore

$$\exp\left(\mathcal{O}_{t+1,b}^{i}\right) = \frac{\ell_{t,b}^{i}}{\sum_j \exp\left(\mathcal{D}_{t+1,b}^{j} - (1-\beta\rho)\frac{\tau^{i,j}}{\delta}\right)},$$

where $\ell_{t,b}^{i} = \sum_j \ell_{t,b}^{i,j}$. Concentrating out the origin fixed effects $\mathcal{O}_{t+1,b}^{i}$ we get

$$\begin{aligned}\log \mathcal{L}_{Pc} &= \sum_t \sum_b \sum_{i,j} \ell_{t,b}^{i,j} \log p_{t,b}^{i,j} + \sum_t \sum_b \sum_{i,j} \ell_{t,b}^{i,j} \log \ell_{t,b}^{i} - \sum_t \sum_b \sum_i \ell_{t,b}^{i} \\ &= \sum_t \sum_b \sum_{i,j} \ell_{t,b}^{i,j} \log p_{t,b}^{i,j} + C_{\mathcal{L}_P},\end{aligned}$$

[93] 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. $\ell_{t,b}^{i,j} = 0$ 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.

where $C_{\mathcal{L}_P}$ 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.[94]

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 $\times$ 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 $\{p^{i,j}_{t,b}\}$, which I can then use to identify the home bias.

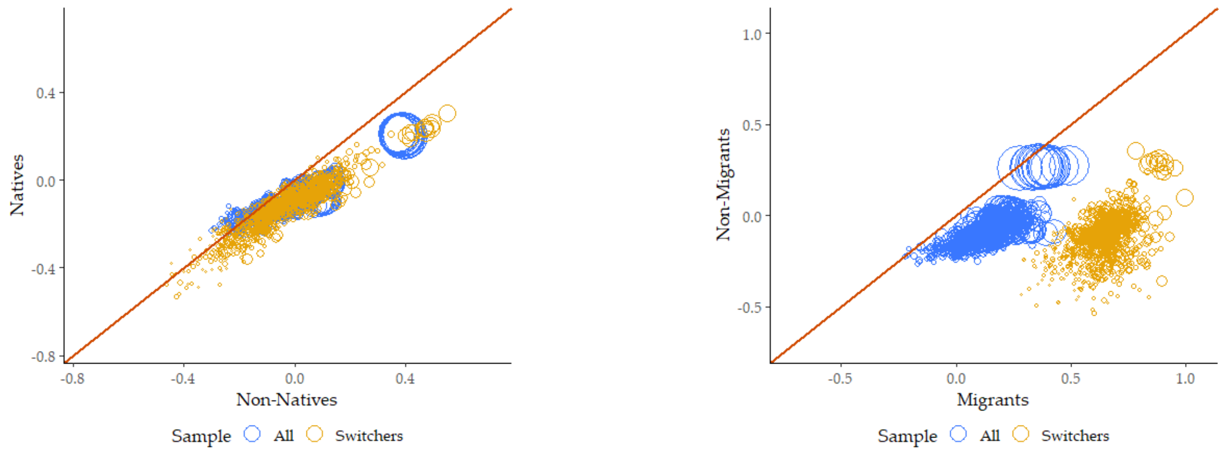
H Additional Figures and Tables

Table 10: Gravity regression using *départements*

	(log) Labor flows, $\log L^{i,j}_{t,b}$ *PPML*			(log) Migration shares, $\log\left(L^{i,j}_{t,b}/\sum_k L^{i,k}_{t,b}\right)$ *PPML*		
	(1) Geodesic (km)	(2) Driving (km)	(3) Driving (hours)	(4) Geodesic (km)	(5) Driving (km)	(6) Driving (hours)
$\mathbf{1}(j \neq b)$	2.318*** (0.154)	2.921*** (0.179)	−2.625*** (0.031)	-0.127*** (0.003)	-0.124*** (0.218)	−0.145*** (0.059)
$\mathbf{1}(j \neq b)\log(\mathrm{d}^{b,j})$	−1.289*** (0.030)	−1.323*** (0.033)	−1.687*** (0.029)	−0.004*** (0.000)	−0.005*** (0.000)	−0.005*** (0.000)
$\mathbf{1}(j \neq n)$	−2.554*** (0.124)	−2.052*** (0.137)	−7.452*** (0.029)	-1.871*** (0.128)	-1.353*** (0.139)	−7.050*** (0.03)
$\mathbf{1}(j \neq i)\log(\mathrm{d}^{i,j})$	−1.299*** (0.025)	−1.320*** (0.026)	−1.837*** (0.029)	−1.349*** (0.024)	−1.368*** (0.025)	−1.773*** (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, $L^{i,j}_{b,t}$ 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

[94] I use the package *fixest* for R.



(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*** (0.0002)	0.145*** (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

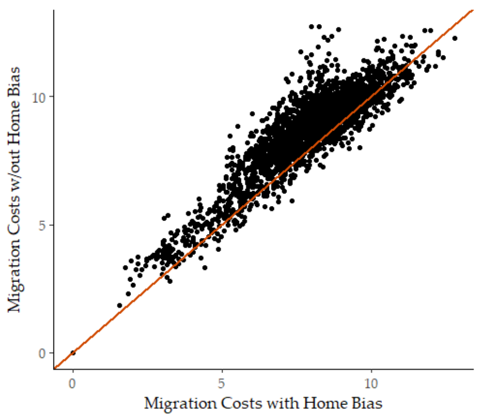


Figure 14: Comparison Migration Costs. The plot compares the migration costs estimated in a model with home bias and a model without.

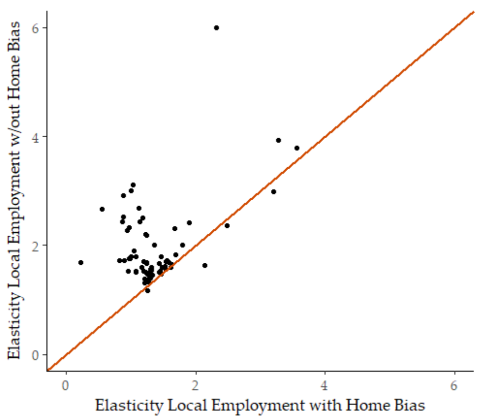


Figure 15: Local Employment Response to a Productivity Shock. The plot compares the local employment elasticities to a productivity shock for a model estimated with home bias and a model without.

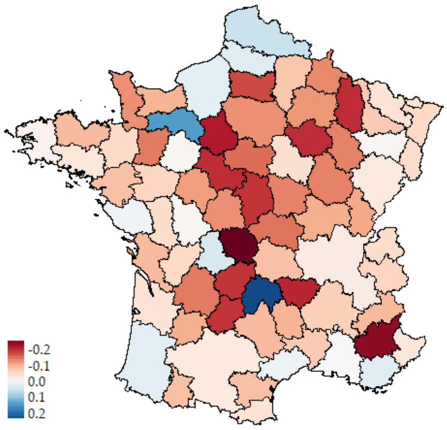


Figure 16: Response to Place Based Policies, no Home Bias. The map shows the change in overall social welfare by subsidizing each location, normalized by the subsidies as a proportion of output, in an economy without home bias. Locations in red mean that when subsidizing such locations, overall social welfare decreased.