Spatial Sorting and the Rise of Geographic Inequality

[Click here for the latest version]

Lukas Mann*

October 26, 2023

Abstract

I show that unobserved sorting patterns of firms and workers across space can account for the tight link between rising wage inequality and rising spatial inequality in West Germany. Two-sided sorting patterns of workers and firms interact with a change in technology to produce a spatially concentrated increase in inequality, driving up regional disparities. These sorting patterns are determined jointly in equilibrium and depend on theoretical objects that are difficult to measure in the data. This paper develops a novel bi-clustering method to recover these objects empirically and uses these results to structurally estimate a dynamic spatial search model with two-sided sorting. I find that regional sorting of firms is more pronounced than regional sorting of workers and the former is an important determinant of workers' job ladders and lifetime values. Compensating differentials between regions are large, driven in part by better labor market outcomes in rich places. The model allows me to consider the redistributive effects of spatial policy, which I find to be strong.

Keywords: Sorting, spatial inequality, spatial policy, two-sided heterogeneity

1 Introduction

Advanced economies have grown more unequal over the past half century. In many major economies, wage inequality has risen in the aggregate but also across space. Recent

*Department of Economics, Princeton University. I thank my advisor, Gianluca Violante, as well as John Grigsby and Richard Rogerson for their invaluable comments, guidance, and support. I also thank Steve Redding, Karthik Sastry, Ayşegül Şahin, Ernest Liu, Isaac Sorkin, Mark Aguiar, Katarína Borovičková, Claudia Macaluso, Nicolas Morales, Jeffrey Sun, Jacob Dorn, Lukas Freund, and various seminar participants at Princeton and other institutions for their comments and feedback. I gratefully acknowledge financial support from the Program for Research on Inequality.

work by Bauluz et al. (2023) shows that the upward trend of spatial inequality in developed countries mirrors the trend for aggregate inequality. Concretely, the standard deviation of regional mean log wages, a common measure of spatial inequality, has strongly increased over the last forty years in countries like the US, the UK, Canada, and Germany (Bauluz et al., 2023). Understanding why this has happened is important: Spatial disparities create social tensions and many policies are explicitly designed to reduce spatial inequality.[1] However, what forces shape the level and the change of spatial inequality is still an open research question. In large part, this question has remained open because it is difficult to answer. Unlike aggregate inequality, spatial inequality is shaped by regional sorting patterns of workers and firms that are hard to measure and might change over time. In principle, multiple channels can lead to an increase in spatial inequality, such as increased regional worker sorting (Diamond and Gaubert, 2022), increased regional sorting of firms, or spatially biased technological change. It is essential to separately identify these channels because each carries different implications for how spatial gaps affect individual employment outcomes. Spatial inequality that is driven by the spatial concentration of skilled workers is not necessarily economically meaningful, because changing spatial inequality by reducing such concentration does not affect worker-level outcomes. On the other hand, spatial inequality that is driven by firm sorting is economically important for workers, as heterogeneous job ladders make their labor market outcomes dependent on their location.

This paper sets out to answer the question why spatial inequality has increased. It does so by contributing a structural search model of two-sided spatial sorting and a novel way of quantifying this model. The new quantification method comes in the form of a new bi-clustering algorithm, which I call LANCE, short for "Likelihood-based Algorithm for Non-parametric Classification and Estimation". Unlike existing methods, LANCE allows me to measure spatial sorting patterns directly in the data, which can be used to estimate the model and to distinguish between changes in sorting and changes in technology. The estimated model then delivers insights into why spatial inequality has risen,

[1] A large set of policies in multiple countries is explicitly designed with the aim of closing spatial gaps. Examples of this are the European Union's regional policy, the "levelling up" policy in the United Kingdom or the "Solidarpakt" in Germany. From 1995 until 2019, this policy was used to support the economic convergence of East German regions but has since been phased out in favor of recent policies that target trailing regions in both East and West Germany.

the drivers of sorting, the role of spatial inequality in shaping workers' labor market outcomes, and the role of spatial policy.

I find that spatial sorting of workers and firms is key to reconciling three facts that I show in West German data - aggregate inequality has risen, spatial inequality has risen at the same time, and within-region inequality has risen faster in rich locations. The estimation reveals that there is strong positive assortative spatial sorting in the cross-section: Highly productive firms and highly productive workers concentrate disproportionately in rich locations. Quantitatively, I find that the former margin is more important: Sorting is significantly stronger on the firm than on the worker side. Perhaps surprisingly, sorting patterns do not change much over time. Nonetheless, they form a tight link between overall and spatial inequality. To see why, consider a highly productive class of firms, say financial firms, that are concentrated in a rich place, say Frankfurt. If the financial sector experiences technological improvements that lead to productivity gains, this leads to increases in aggregate inequality that are driven mostly by an increase in the wage in Frankfurt. As a consequence, aggregate and spatial inequality move in lockstep. Interactions of technological changes and spatial sorting can also account for the fact that the rise of inequality has been spatially concentrated. In the same example, wage increases in the financial sector drive apart the local wage distribution in Frankfurt, but not in other, poorer places. This type of effect explains why rich places have become more unequal relative to poor places.

Using the model, I quantify the importance of the spatial dimension of inequality by calculating the differences in worker's lifetime values that arise from being in a rich location versus a poor location. I find these differences to be quantitatively meaningful. Workers would on average be willing to pay 21% of perpetual average income in order to keep their job and costlessly switch locations (or to prevent such a switch). In the model, workers value rich locations for three reasons: First, job ladders can vary sharply across locations because cross-sectional sorting is strong on the firm side. Job ladders in rich locations enable more upward mobility and thus workers who live in these locations have superior and repeated access to good jobs. Second, workers may enjoy higher amenities in these locations. Lastly, workers who live in such locations expect to switch locations less often. I show that switching costs account for much of the value differential. However, workers switch only because it is optimal for them to do so - the benefits are large. I show that job ladders account for more than half of these benefits, taking up a larger role than locational amenities.

The model also has implications for spatial policy. In a world with two-sided sorting, there are many channels that can in principle govern the effects of policy. The quantified model shows that two of them are quantitatively salient. First, size effects are a central driver of firm-side spatial sorting. It is easier to fill a vacancy when there are many workers located near it. Firms with more valuable jobs have the strongest incentive to fill their jobs faster and thus locate disproportionately in rich locations. In the context of spatial policy, this means that setting incentives to increase the population of small locations, such as location-based transfers, will reduce the degree of spatial firm sorting in the economy. Second, firm sorting is the dominant driver of worker sorting. This finding implies that policymakers can affect access to rich locations by changing the degree of firm sorting directly or indirectly. Reducing firm sorting is generally good for low skill workers because it increases their access to rich locations. Conversely, high skill workers benefit from strong firm sorting. Thus, policies can redistribute progressively by expanding smaller locations or by incentivizing highly productive firms to locate in such places. Policies that increase firm sorting, such as subsidies to high-skill plants in rich locations, redistribute welfare in a regressive manner.

The paper contributes to at least three strands of literature. First, a number of authors have considered the forces that drive workers with different skills and characteristics to sort into different geographical regions. Examples of this are Eeckhout et al. (2014), Behrens et al. (2014), Fajgelbaum and Gaubert (2020), Diamond and Gaubert (2022), or Heise and Porzio (2021) in the context of a spatial search model. Another strand of literature has focused on the location decisions of firms, such as for example Gaubert (2018), Kleinman (2023) and Lindenlaub et al. (2022) who study a spatial search model in which firms make location choices. Unlike these papers, I consider two-sided geographic sorting, allowing both workers and firms to make location decisions. This allows me to separately quantify the role of firms and that of workers in generating spatial wage differentials. As outlined above, distinguishing between geographic sorting of workers and firms is useful because both have vastly different implications for the importance of spatial inequality on the worker level. My paper shows that spatial inequality is primarily a reflection of productive firms locating in rich places. This implies that locations are an important determinant of workers' employment outcomes.

I also contribute to the literature on the measurement of sorting. Measuring the geographic sorting patterns of workers and firms is challenging. Worker and firm characteristics that are relevant to the expected wage of a match are often unobserved, which substantially complicates any attempt to measure sorting based on these characteristics.[2] One solution has been measure the distribution of worker and firm fixed effects that arise from wage regressions following Abowd et al. (1999) (AKM). Some papers, most notably Dauth et al. (2022) and Card et al. (2023), have followed such approaches. However, employing an AKM approach requires putting major restrictions on complementarity patterns of workers and firms. This is a problem for any application in which one seeks to estimate a structural model of sorting, since complementarities between workers and firms are potentially strong forces of sorting (Shimer and Smith, 2000).

Since AKM's seminal paper, there have been advances in the measurement of sorting under the presence of complementarities, most notably Bonhomme et al. (2019) (BLM) and Lentz et al. (2023) (LPR). BLM show that in matched employer-employee data, worker types are identified from data on employment transitions and wages, conditional on a classification of firms and LPR build on this result. As I argue in this paper, the assumptions of BLM are not satisfied in a spatial setting and neither BLM's nor LPR's methodology is appropriate for quantifying the model. The reason is that firm types are not well identified when employment distributions are in part shaped by locations. I thus develop LANCE, a new bi-clustering algorithm that classifies workers and firms based on their joint wage distribution. LANCE allows for arbitrary patterns of worker-firm complementarities and can be applied in a spatial setting. Unlike BLM or LPR, LANCE imposes type stability over long stretches of time. However, it identifies worker and firm types solely on wages and not on transitions, making updating steps over worker and firm types fast.

[2] Figure 24 in appendix A.8 shows that in my data college workers sort positively into rich regions both before and after the increase in spatial inequality. However, a general increase in the college share makes it somewhat difficult to judge whether the degree of worker sorting has increased and whether the skills of college-educated workers is comparable over time. In wage regressions, observables such as education and occupation typically explain less than a third of the variance in wages. This is also the case for the data set used in this paper. For an empirical study on the distribution of observables across space, see e.g. Mion and Naticchioni (2009).

The algorithm yields direct estimates for regional sorting patterns of workers and allows me to use these estimates to inform the model.

The remainder of the paper proceeds as follows. Section 2 introduces the data and shows patterns of aggregate and spatial inequality in West Germany that the model can speak to. Section 3 introduces the model. Section 4 discusses how the the model is estimated and introduces LANCE. Section 5 discusses further restrictions on the data made in the context of estimation. Section 6 presents the results from the model estimation, including spatial sorting patterns in the data, the model fit, and various decompositions that shed light on the questions above. It also discusses implications for policy. Section 7 concludes.

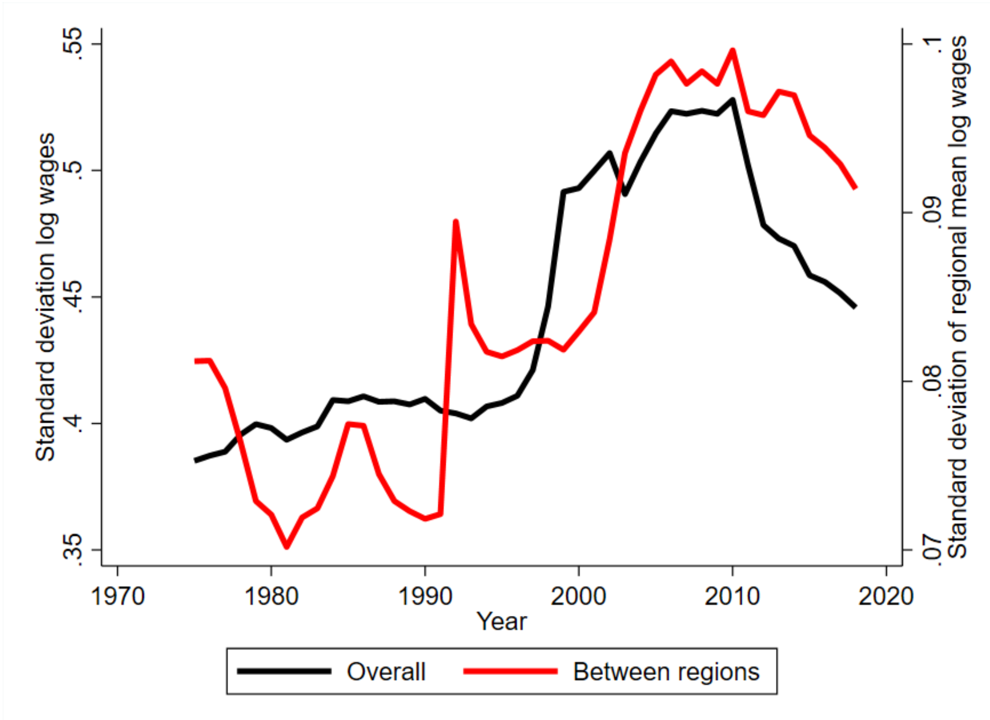
2 Inequality patterns in the data

Data

I begin by showing statistics on inequality within and across space, using matched employer-employee data from West Germany. Concretely, I use the Sample of Integrated Employer-Employee Data (SIEED), provided by the German Institute for Employment Research (IAB). This data set is constructed as follows: A random sample of 1.5% of all German establishments is selected by the IAB. Of those establishments, the SIEED the contains the full employment biography of every worker who has ever worked at one of these establishments, including the identifiers of any establishments not included in the 1.5% sample. For every employment spell, the SIEED contains information on the daily wage, updated at a frequency of at least once every year. The data also contains information on the location of an employment spell by recording the commuting zone of each establishment associated with a job. A commuting zone is a county or small collection of counties defined in Kosfeld and Werner (2012). Commuting zones are selected to minimize cross-border commuter flows. The full data contains 171 such areas, 108 of which are in West Germany.

In total, the data contains 176 million employment spells between 1975 and 2018. From this large data set, I keep only observations that are located in West Germany (excluding Berlin) and have achieved their highest education level recorded in the data. I also restrict the sample to full time jobs, since I observe daily wages in the data but not working hours.

Inequality within and across space, 1975-2018



Note: The left hand side scale corresponds to the black line which depicts aggregate wage inequality as measured by the standard deviation of aggregate wages in the SIEED. The right hand side scale corresponds to the red line which depicts wage inequality between regions as measured by the standard deviation of regional mean log wages.

Figure 1: West German wage inequality over time

I use the SIEED to document three simultaneous developments:

1. Aggregate wage inequality has risen
2. Spatial (i.e, between region) wage inequality has risen at the same time
3. Wage inequality has increased most in rich regions

Figure 1 documents the first two of the above facts in the SIEED. The figure depicts two dimensions of inequality over time, aggregate and spatial. The black line depicts aggregate inequality as measured by the standard deviation of log wages between 1975 and 2018, the full scope of the data set. Inequality in West Germany starts to rise in the mid-1990s and plateaus in the mid-2000s. The red line displays spatial inequality, defined as the standard deviation of mean log wages across commuting zones by year. Spatial inequality broadly follows the trajectory of aggregate inequality. There is a spike in the early 1990s which can be explained by a handful of outlier commuting zones located at the former inner German border.[3] Apart from this spike, aggregate and spatial inequality move in tandem and the two time series are highly correlated.

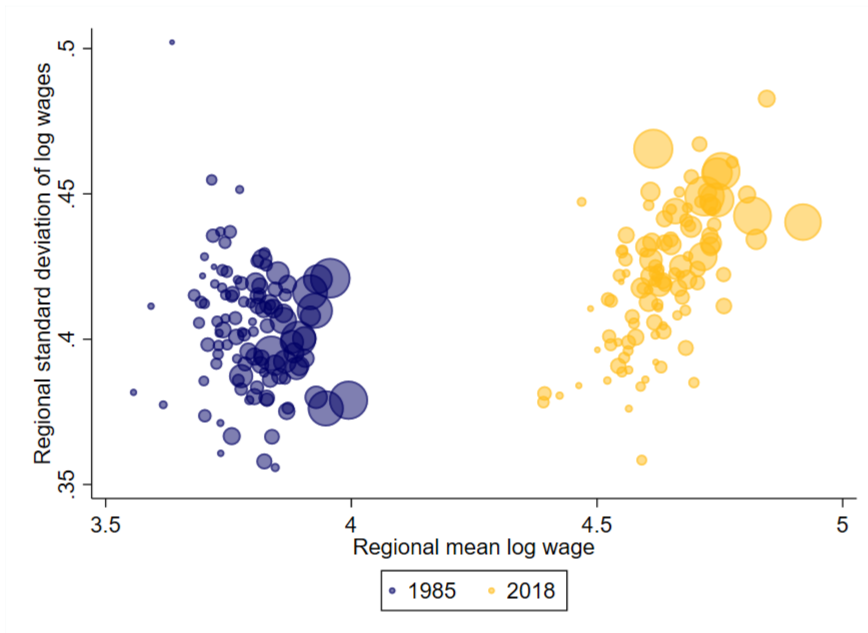


Figure 2: Income-inequality relationship

Figure 2 displays the third fact. The figure plots the level of within-region inequality for West German commuting zones, as measured by the standard deviation of log wages within the commuting zone, against their mean log wage. Thus, the x-axis captures a measure of how rich a region is, the y-axis the level of wage inequality within that region. In 1985, there is no clear relationship between income and inequality. In 2018, the relationship is positive: Rich regions are a lot more unequal. The size of each marker, capturing the size of each commuting zone, hints at the type of region that characterizes both extremes of this correlation - rich places tend to be larger. In recent times, they are also more unequal.

Figure 3 shows the same relationship in terms of changes in within-region inequality. Again, it is clear from the figure that rich places have seen a larger increase in inequality

[3]Removing these outliers retains the trend of the figure but removes the early-1990s spike.

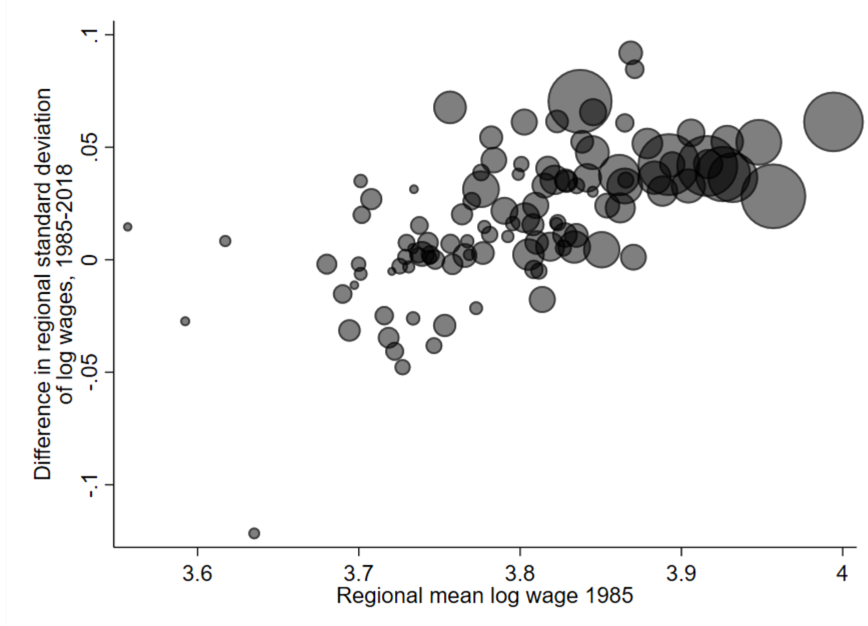


Figure 3: Inequality changes by regional income

than poor places.[4] Taking stock, three facts emerge from the data - aggregate inequality has risen, spatial inequality has risen, and inequality has risen mostly in rich places. To explain these trends jointly, the next section introduces a model of two-sided sorting.

3 Model

The model has two main objectives. The first is to provide a tractable framework to quantify the drivers of sorting for both workers and firms. This will allow us to make sense of the observed sorting patterns, judge their implications for the role of spatial inequality in generating life-time value differences, and to predict the effects of spatial policies on sorting, welfare and inequality. The second objective is to deliver a foundation for the structural estimation of key model parameters that drive sorting. In the model, both workers and firms make forward-looking location decisions, which complicates identification. The model helps with this problem by delivering tractable equations for worker mobility that can be estimated in the data.

[4]This pattern does not change is 2018 is taken as a base year for the x-axis, as shown in Figure 22 in appendix A.8.

Workers

Time is infinite and continuous. There is a unit measure of workers. Each worker belongs forever to one of K types, denoted $k \in \{1, \ldots, K\}$. At any time, the worker lives in one of L locations $l \in \{1, \ldots, L\}$ and either works in a job, or is non-employed. Workers of type k are born into non-employment in location l at exogenous rate $\chi_{k,l}$ and die stochastically at Poisson rate ρ. This implies that there there is a fixed distribution $f_k = \frac{\sum_l \chi_{k,l}}{\rho}$ of worker types in the population. Workers consume one unit of housing each, for which they pay p_l units of the numeraire to absentee landlords. Workers' utility is linear in consumption and equals $a_l - p_l$ when in non-employment and $w_{k,m,l} + a_l - p_l$ in employment, where $w_{k,m,l}$ is their wage, and a_l is the local amenity, i.e. how much a worker values living in location l. This value could consist, for example, of natural amenities, such as a beautiful landscape or a temperate climate. Alternatively, one can think of this value containing utility derived from local government-sponsored infrastructure, such as museums or roads.

Jobs

There is an endogenous mass $\mathcal{J}$ of jobs. A job is associated with a firm (or "plant") which belongs to one of M types, denoted by $m \in \{1, \ldots, M\}$.[5] In what follows, I denote non-employment as an additional firm type $m = 0$. Each job can only be filled by a particular worker type k and has zero productivity with any other worker. At any moment, a job can be filled or unfilled. If filled, the job counts towards employment. If unfilled, the job is labeled a "vacancy". The total number of jobs is thus given by employment and vacancies:

$$\mathcal{J}_{k,m,l} = \underbrace{e_{k,m,l}}_{\substack{\text{Filled jobs/} \\ \text{Employment}}} + \underbrace{\mathcal{V}_{k,m,l}}_{\substack{\text{Unfilled jobs/} \\ \text{Vacancies}}}$$

A filled job gets broken up at some rate $\delta_{k,m}$ that may depend on the worker and firm types of the match. An unfilled job disappears at rate δ^v which is common across all types.[6] Jobs are attached to a particular location forever. A job is therefore characterized

[5] A plant has no economic meaning in the model but does become meaningful when thinking about its identification. We can think of a plant as a collection of jobs that are observed to be of the same type.

[6] In this model, vacancies remain intact after a match ends. This is both for realism, as jobs are long-term investments that can survive one employment spell, and for tractability, see equation (10).

by its three immutable characteristics: its own type m, its worker type requirement k and its location l.

Production and wages

When a worker is matched to a job, they produce a flow value of the numeraire that depends on the product of two productivity components. The first component, $Z_{k,m}$, depends flexibly on the worker and firm type of a match. The second, H_l, varies by location. Thus,

$$Y_{k,m,l} = Z_{k,m} H_l$$

Wages are set by a simple flow sharing rule by which the worker gets a share β and the firm a share $1 - \beta$ of the flow value. Thus, log wages satisfy

$$\log w_{k,m,l} = \log \beta + z_{k,m} + h_l \tag{1}$$

where $z_{k,m} = \log Z_{k,m}$ and $h_l = \log H_l$. This wage setting protocol abstracts from the traditional pass-through of outside options onto wages. It is chosen for simplicity and to focus the model onto wage changes that arise from the type of matches formed rather than outside options. The assumption is also helpful in mapping the estimated wage function into a model object: $z_{k,m}$ and h_l are central objects that my methodology allows me to estimate directly in the data. Their empirical estimates are discussed in detail in section 6.1.

Matching and mobility

Workers and firms meet in a random search market. Encounters between them are generated in a global matching market according to a Cobb-Douglas meeting function $\mathcal{M} = m(\mathcal{V}, \mathcal{S}) = A\mathcal{V}^{\alpha}\mathcal{S}^{1-\alpha}$ where $\mathcal{V} = \sum_{k,m,l} \mathcal{V}_{k,m,l}$, $\mathcal{S} = \sum_{k,m,l} s^{\mathbb{1}(m \neq 0)} e_{k,m,l}$ and s is the (exogenous) relative search intensity of the employed relative to the non-employed.[7] The encounter rate for firms is thus $q = \mathcal{M}/\mathcal{V}$ and the encounter rate for workers is $\lambda = \mathcal{M}/\mathcal{S}$. Simple algebra implies that the encounter rates for firms (q) and workers (λ) are related

[7] In appendix A.6, I discuss in more detail why random search and global matching are convenient assumptions in the context of this paper.

by the following equation:

$$q = A^{\frac{1}{\alpha}} \lambda^{\frac{\alpha-1}{\alpha}} \tag{2}$$

For an encounter to be viable, the worker type has to be the correct type required by the vacancy. I further assume that there is an additional chance that a worker coming from job type m is unable to perform a job m'. That is, I assume that with probability $(1-\zeta_{m,m'})$, the prospective match is unproductive and thus not viable. Upon a viable new encounter (m', l'), workers draw two preference shocks, ε_1 and ε_2 from a Gumbel$(-\gamma\sigma^w, \sigma^w)$ distribution and then solve

$$\max\{W_{k,m,l} + \varepsilon_1, W_{k,m',l'} - \mathbb{1}(l' \neq l)c_k + \varepsilon_2\}$$

where $W_{k,m,l}$ is the value of a worker in state (k, m, l) and c_k denotes the cost that a worker of type k incurs when moving locations.[8] Like the local amenity, these costs can be monetary or non-monetary and might include the physical cost of planning and executing a move as well as the intangible costs associated with moving, such as the loss of established social networks and a familiar environment. For simplicity, and to enhance interpretability, I assume that these costs do not depend on which locations are its origin or destination.[9] A worker in state (k, m, l) therefore has a value that satisfies the following Bellman equation:

$$\begin{aligned}
&(r + \delta_{k,m} + \rho)W_{k,m,l} \\
=&w_{k,m,l} + a_l - p_l + \delta_{k,m}W_{k,0,l} \\
&+ \sum_{m' \neq 0, l'} s^{m \neq 0} \lambda \bar{v}_{k,m',l'} \zeta_{m,m'} \mathbb{E}\left[\max\{W_{k,m,l} + \varepsilon_1, W_{k,m',l'} - \mathbb{1}(l' \neq l)c_k + \varepsilon_2\}\right]
\end{aligned}$$

where $\bar{v}_{k,m,l} = \frac{\mathcal{V}_{k,m,l}}{\mathcal{V}}$ is the share of vacancies that are of type m in location l and require worker type k. Given the well-known properties of the Gumbel distribution, we can write

[8] Here, γ denotes the Euler-Mascheroni constant. The parameters are chosen such that $\mathbb{E}[\varepsilon_i] = 0$, which ensures that a worker is indifferent between an encounter they will never accept and no encounter at all.

[9] This assumption can be relaxed, but it ensures that the model abstracts from the geographic properties of the country on which the data has been estimated.

this as

$$
\begin{aligned}
&(r+\delta_{k,m}+\rho)W_{k,m,l} = w_{k,m,l} + a_l - p_l + \delta_{k,m}W_{k,0,l} \\
&+ \sum_{m'\neq 0,l'} s^{m\neq 0}\lambda \bar{v}_{k,m',l'}\zeta_{m,m'}\left[\sigma^w \log\left(\exp\left(\frac{W_{k,m,l}}{\sigma^w}\right)+\exp\left(\frac{W_{k,m',l'}-\mathbb{1}(l\neq l')c_k}{\sigma^w}\right)\right) - W_{k,m,l}\right]
\end{aligned} \tag{3}
$$

The transition probability conditional on an encounter takes the following logit form:

$$
P_{k,m,m',l,l'} = \frac{\exp\left(\frac{W_{k,m',l'}-\mathbb{1}(l\neq l')c_k}{\sigma^w}\right)}{\exp\left(\frac{W_{k,m,l}}{\sigma^w}\right)+\exp\left(\frac{W_{k,m',l'}-\mathbb{1}(l\neq l')c_k}{\sigma^w}\right)} = \frac{\tilde{c}_k^{\mathbb{1}(l\neq l')}\tilde{W}_{k,m',l'}}{\tilde{W}_{k,m,l}+\tilde{c}_k^{\mathbb{1}(l\neq l')}\tilde{W}_{k,m',l'}}
$$

where $\tilde{c}_k = \exp(-\frac{c_k}{\sigma^w})$ and $\tilde{W}_{k,m,l} = \exp(\frac{W_{k,m,l}}{\sigma^w})$. The resulting flow rate for a worker of type k in a job of type m in location l to a job of type $m' > 0$ in location l' is therefore

$$
\mu_{k,m,m',l,l'} = s^{m\neq 0}\lambda \bar{v}_{k,m',l'}\zeta_{m,m'}P_{k,m,m',l,l'} \tag{4}
$$

That is, the flow rate between any two jobs and locations is equal to the product of a worker's search intensity, encounter rate, vacancy distribution, probability of having the skills to perform the job and the probability of transitioning conditional on a viable encounter.

I vectorize the birth rates $\chi_{k,m,l}$ and the employment-unemployment distribution.[10] This allows us to write the KFEs in general matrix form:

$$
\dot{e} = (M' - \rho I)e + \chi \tag{5}
$$

where

$$
M = \begin{pmatrix} M_1 & 0 & \dots & 0 \\ 0 & M_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_L \end{pmatrix} \text{ and } M_l = \begin{pmatrix} -\sum_{m'\neq 0,k'} \mu_{l,0,m',1,k'} & \mu_{l,0,1,1,1} & \cdots \\ \mu_{l,1,0,1,1} & -\sum_{m'\neq 1,k'} \mu_{l,1,m',1,k'} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}
$$

Here, M is the matrix of worker flows across states. It is block-diagonal because workers never transition between different types k. When solving the model, I will focus on steady

[10] I add a subscript m for notational purposes, since the vectorization requires it. However, I retain the assumption that $\chi_{k,m,l} = 0$ for all $m > 0$.

states of equation (5). Thus, we can write down a system of steady state flow equations that takes the following form:

$$e = -(M' - \rho I)^{-1}\chi \tag{6}$$

Entry

There is an infinite mass of potential entrants who can decide to enter at will. Entry happens in four stages. First, upon entering, firms pay an entry cost, $\varrho \geq 0$. Second, they draw a firm type m at random, where type m is drawn with probability ψ_m. Third, they draw a cost of locating in any location l, given by $-u_l$ where $-u_l$ is Gumbel$(-\gamma\sigma^f, \sigma^f)$ distributed. Given this draw, they choose the location in which they create the job. Lastly, they draw the worker type requirement k with probability $\eta_{k,m}$.[11]

The ex-ante value of a vacancy can be found by working backwards through the four steps of entry. Denoting the value conditional on firm type m, location l, and worker requirement k as $\Omega_{k,m,l}$, we can aggregate up to find the value of an ex-ante entrant, which is equal to

$$\sum_m \psi_m \sigma^f \log\left(\sum_l \exp\left(\frac{\sum_k \eta_{k,m}\Omega_{k,m,l}}{\sigma^f}\right)\right) \tag{7}$$

Free entry to vacancy posting guarantees that the ex-ante value of entry equals the cost, i.e.

$$\sum_m \psi_m \sigma^f \log\left(\sum_l \exp\left(\frac{\sum_k \eta_{k,m}\Omega_{k,m,l}}{\sigma^f}\right)\right) = \varrho \tag{8}$$

Invoking again the properties of a Gumbel distribution and defining $\bar{\Omega}_{m,l} = \sum_k \eta_{k,m}\Omega_{k,m,l}$, the probability of creating a job in location k for a firm of type m is given by

$$\phi_{m,l} = \frac{\exp\left(\frac{\bar{\Omega}_{m,l}}{\sigma^f}\right)}{\sum_{l'} \exp\left(\frac{\bar{\Omega}_{m,l'}}{\sigma^f}\right)} \tag{9}$$

[11] The fact that firms make location decisions without having their worker type draw revealed to them is a conscious design choice of the model. While firms are not an economically meaningful unit in the model, one can think of plants as a large number of jobs of the same firm type that face the same location-specific job creation costs. Under this interpretation of the model, a plant will require an exogenous (and firm-type-specific) proportion of all worker types. This reflects the fact that plants typically employ workers of many types and skill-levels. The implications of this are economically meaningful as discussed in section 6.3.

Now, it is straightforward to find $\bar{v}_{k,m,l} = \frac{\mathcal{V}_{k,m,l}}{\mathcal{V}}$: In steady state, the number of filled vacancies equal the number of separations. Thus, denoting as ι the flow of entering vacancies, it is possible to show that the number of hires and separations for a particular vacancy type is irrelevant to the equilibrium stock, which is fully determined by the location choice probabilities determined at entry:

$$\begin{aligned} 0 =& d\mathcal{V}_{k,m,l} = \psi_m \phi_{m,l} \eta_{k,m} \iota - \delta^v \mathcal{V}_{k,m,l} + \underbrace{\text{hires}_{k,m,l} - \text{separations}_{k,m,l}}_{=0 \text{ in SS}} \\ =& \psi_m \phi_{m,l} \eta_{k,m} \delta^v \mathcal{V} - \delta^v \mathcal{V}_{k,m,l} \end{aligned}$$

which implies

$$\bar{v}_{k,m,l} = \frac{\mathcal{V}_{k,m,l}}{\mathcal{V}} = \psi_m \phi_{m,l} \eta_{k,m} \tag{10}$$

Once entered, vacancies encounter workers at rate q. They can only form matches with the required worker type, so they search until they have found the correct type that is willing to transition and then match with that worker. Let

$$\mu^{\text{ee}}_{k,m,l} = \sum_{m' \neq 0, l'} \mu_{k,m,m',l,l'}$$

denote the Poisson rate at which the worker leaves the firm voluntarily. Let $e_{k,m,l}$ denote the employment of workers of type k in firms of type m and location l and let

$$\xi_{k,m,l} = \sum_{m',l'} \frac{s^{\mathbb{1}(m' \neq 0)} \zeta_{m',m} e_{k,m',l'}}{\sum_{\hat{k},\hat{m},\hat{l}} s^{\mathbb{1}(\hat{m} \neq 0)} e_{\hat{k},\hat{m},\hat{l}}} P_{k,m',m,l',l}$$

denote the probability that an encounter of firm type m in location l requiring a worker of type k yields a successful match. Such a firm then has a post-match value of

$$(r + \delta_{k,m} + \mu^{\text{ee}}_{k,m,l}) J_{k,m,l} = (1-\beta) Y_{k,m,l} + (\delta_{k,m} + \mu^{\text{ee}}_{k,m,l}) \Omega_{k,m,l}$$

and thus the value of a vacancy of this type has value

$$(r + \delta^v) \Omega_{k,m,l} = q \xi_{k,m,l} (J_{k,m,l} - \Omega_{k,m,l})$$

We can combine these to write $\Omega_{k,m,l}$ as a function of the searcher distribution and parameters only:

$$\left(r + \delta^v + q\xi_{k,m,l} \frac{r}{r + \delta_{k,m} + \mu^{\text{ee}}_{k,m,l}} \right) \Omega_{k,m,l} = q\xi_{k,m,l} \left(\frac{(1-\beta) Y_{k,m,l}}{r + \delta_{k,m} + \mu^{\text{ee}}_{k,m,l}} \right) \tag{11}$$

We are now ready to define an equilibrium.

Equilibrium

A stationary equilibrium is defined as a distribution $e_{k,m,l}$, worker flow rates $\mu_{k,m,m',l,l'}$, firm value functions $\Omega_{k,m,l}$, worker and firm encounter rates λ and q, worker value functions $W_{k,m,l}$, and location choice probabilities $\phi_{m,l}$ such that

1. Workers optimize: The value function of the worker satisfies the Bellman equation (3) and worker flow rates are given by equation (4)

2. Firms optimize: The value functions of the firm satisfies the Bellman equation (11) and location choice probabilities of firms are given by equation (9)

3. There is free entry: The ex-ante firm value equals the entry cost as in equation (8)

4. The steady state distribution of workers across firm types and locations is consistent with flows, i.e. $e_{k,m,l}$ is given by equation (6)

5. Job finding and job filling rates are moderated by the matching function: q and λ are consistent with equation (2)

Note that the definition of an equilibrium makes no mention of housing supply, because I define the equilibrium conditional on a vector of house prices $(p_1, \ldots, p_l)$. This does not matter until section 6.4 where I introduce assumptions on the supply of housing, which provide an additional equilibrium condition.

Uniqueness of equilibrium is not guaranteed in this model. When taking the model to the data, I therefore numerically search for equilibria that imply employment distributions close to the employment distributions in the data.

4 Identification

I show that the parameters of the model can be identified using matched employer-employee data in which I can track workers across firms and locations and observe their wages. I proceed in three steps. First, I develop a novel bi-clustering algorithm to jointly classify workers and firms into their respective types. I then classify regions based on the local distribution of workers and firms.[12] Second, I use the resulting classification in

[12]This is the correct approach because in the model even ex-ante identical locations can produce different ex-post worker and firm distributions in equilibrium.

the data to estimate the mobility parameters of the model via maximum likelihood conditional on their type. Doing so requires a maximization-minorization (MM) approach which I detail in appendix A.3. Lastly, I choose the remaining parameters to minimize the distance between moments in the model and the data, matching some directly and minimizing the distance between others. Section 4.1 introduces the bi-clustering algorithm while sections 4.2 and 4.3 describe the identification of all remaining parameters.

4.1 Step 1: LANCE - a new bi-clustering algorithm

The following section introduces a new bi-clustering algorithm that is straightforward to implement, comparatively fast, and generalizable to other contexts that require two-sided classification based on an unknown distribution of repeated joint signals. I apply this algorithm to identify worker and firm types in the context of the model and simultaneously identify the model-implied joint wage function. Given that the algorithm maximizes a likelihood function, non-parametrically identifies the wage function and delivers a classification of workers and firms, I refer to it as "LANCE", which is short for "Likelihood-based Algorithm for Non-parametric Classification and Estimation". The algorithm can be viewed as a generalization of the famed wage regression of Abowd et al. (1999) (AKM) but, unlike AKM, makes no assumption on the additive separability of the wage function. By avoiding this assumption, it is possible to use LANCE to directly analyze changes in the joint wage function. By imposing additional assumptions, I then relate such changes to changes in the underlying production function.

The idea of LANCE is also closely related to similar attempts to estimate a joint wage function based on matched worker-firm data, most notably Bonhomme et al. (2019) (BLM) and Lentz et al. (2023) (LPR). BLM show that, conditional on a pre-classification of firms, data on wages and on worker mobility across two periods are enough to identify both a type-conditional wage function and worker mobility patterns. LPR extend the methodology in BLM by including a re-assignment step for firms that eventually allows firms to be identified based on the same likelihood function as workers. However, neither the BLM methodology nor the LPR variant translate well to the case where both the distribution and the joint wages of workers and firms critically depend on workers' and firms' location. For one, running the k-means pre-classification step from BLM on the universe of firms is no longer model-consistent in a world where the type conditional worker distribution of firms varies by location. The reason for this is that two firms that belong to the same type will nonetheless face different employment flows if they choose to locate in distinct places. For example, a firm that locates in a place with many workers of type k will on average employ more type-k workers, since these workers are more likely to transition within their home region. Second, the presence of location fixed effects in the wage equation implies that even conditional on having the same cross-sectional distribution of worker types, two firms in distinct locations may have distinct wage distributions.

Using the strategy from LPR alleviates this problem somewhat, since in their algorithm firms are successively re-assigned based on the likelihood function. However, firm re-assignments using LPR's methodology are slow, because the likelihood must include information on worker mobility. The inclusion of mobility data is necessary in LPR since like BLM, it is the only way to ensure identification. Including mobility information in the likelihood implies that changing one firm's assignment affects not only the likelihood terms corresponding to that firm's matches but also the likelihood of transitions away from and towards the firm. Thus, the order of re-assignments matters and re-assignments are computationally costly.

LANCE takes a different approach and omits information on mobility, only utilizing joint wages, i.e. static information (the likelihood of which is not affected by the assignments of other firms or workers). This comes at a cost: Unlike BLM, which requires only two periods for identification, LANCE requires long data on worker careers throughout their lifetime.[13] However, there is also a substantial benefit: The reliance on static information makes the algorithm much faster compared to LPR and thus suitable for a problem in which it is hard to obtain a good initial guess of firm or worker classification and frequent updating is necessary to obtain a good solution.

LANCE: Setup and general formulation

Formally, LANCE is an algorithm that can be used in situations where

1. One wishes to estimate a distribution of some variable $x \sim G(\cdot|\theta, \tau_1, \ldots, \tau_n)$ in a

[13] For noiseless identification, the number of jobs per worker and the number of matches per firm both have to be sufficiently large.

data set with a large number of observations x_s indexed by $s \in \{1, \ldots, S\}$

2. The distribution of x depends on the membership of the associated observation s to a finite number of types $(\tau_1, \ldots, \tau_n)$ where $\tau_i \in \{1, \ldots, T_i\} = \mathfrak{T}_i$ and T_i is assumed to be known. That is, there is an unobserved set of maps $a_i : s \mapsto \mathfrak{T}_i$ that maps observations s into their associated types. We call a_i an "assignment function".

3. Large subsets of observations s are known to be associated with the same type, i.e. for each type i there is a known partition $\{P^i_1, \ldots, P^i_{J_i}\}$ such that $s, s' \in P^i_j \implies a_i(s) = a_i(s')$

To interpret this setup, think of the application in this paper. In our application, $i \in \{w, f\}$, i.e. there are two kinds of types (worker types and firm types), with $T_w = K$ and $T_f = M$ known. The variable x will be the log wage, indexed by each employment spell s observed in the data.[14] The distribution of log wages depends on the parameters $z_{k,m}$ and h_l which are indexed by worker and firm types. The mapping from spells to types, which we call the assignment function, is unobserved. However, we can partition all observations into sets associated with the same worker or the same firm: The partitions $\{P^i_1, \ldots, P^i_{J_i}\}$ will simply be employment spells associated with individual workers ($i = w$) and individual firms ($i = f$) respectively. Intuitively, identification is possible because we impose type stability across repeated spells. Two employment spells that are associated with the same worker have to belong to the same worker type. Likewise, two employment spells that are associated with the same firm have to belong to the same firm type. Because of this, it is possible to rewrite the domain of each assignment function as individual workers or firms instead of spells, as we will do in the implementation below.

The goal of LANCE is to deliver estimates of θ as well as an estimate of every assignment function a_i in this situation. To do this, the idea is to solve the following likelihood maximization problem:

$$\max_{\theta, a_1, \ldots, a_n} \mathcal{L}(\theta, a_1, \ldots, a_n) \quad \text{where} \quad \mathcal{L}(\theta, a_1, \ldots, a_n) = \sum_s g(x_s | \theta, a_1(s), \ldots, a_n(s)) \tag{12}$$

This problem is impossible to solve by brute force because the discrete space over which assignments are optimized is too large for this to be computationally feasible. However,

[14] In principle, any static information on the match may be included, such as occupation. Since I am focusing here on earnings dynamics, I opt to solely classify based on wages.

LANCE solves this problem by applying coordinate descent (Bezdek et al., 1987) to the likelihood function, re-optimizing assignments and parameters in sequential order. The general algorithm is shown in appendic A.1.

Intuitively, as the algorithm progresses, both the parameters of the type-conditional distribution and the estimates for type assignments improve over time. In appendix A.2, I show that the algorithm converges to a local maximum of the likelihood in a finite number of iterations. While I do not formally prove in this paper that local maxima of the empirical likelihood in equation (12) necessarily converge to the parameters of the data generating process, Monte Carlo simulations show that this is typically the case when $|P_j^i|$ gets large for all i, j, i.e. when workers are observed over multiple spells and firms are observed with many workers. This is intuitive: Taking workers as an example, we obtain precise estimates of the type-conditional parameters of any worker that is observed across multiple spells with different firms. Since two workers of the same type produce draws from a distribution with identical parameters, the likelihood losses of classifying these two workers as the same type are small. On the other hand, giving two workers of different types the same assignments leads to large likelihood losses when these workers produce multiple draws in the data set.

LANCE: Implementation

I now apply LANCE to identify $z_{k,m}$ and h_l in my model and classify workers and firms. As part of the estimation procedure, I split the sample into two periods, which I refer as the "before and "after" periods. The period a spell belongs to is determined by whether the majority of the match duration was recorded prior to 1995, as this is the approximate inflection point for inequality dynamics in West Germany (see Figure 1). I use $t \in \{0, 1\}$ to denote what period an observation belongs to. $t = 1$ reflects matches that are mostly recorded after 1995. To apply LANCE, recall that the model predicts that log wages satisfy

$$\log w_{k,m,l} = \log \beta + z_{k,m} + h_l$$

Since I split the sample in the data, I allow $z_{k,m}$ to vary by period, i.e. I write $z_{k,m,t}$ in place of $z_{k,m}$. Moreover, to get a likelihood in a model with deterministic wages I assume that wages in the data are observed with a measurement error and that the match-specific component of each type-pair might have changed due to technology. That is, I assume

$$\log w_{k,m,l,t} = \log \beta + z_{k,m,t} + h_l + \varepsilon_s, \quad \log w_{k,m,l,t} \sim F\left(\cdot \mid \theta\right) \tag{13}$$

where $\theta = (z_{k,m,t}, h_l, \sigma)$. Concretely, I assume that $\varepsilon_s \sim \mathcal{N}\left(0, \sigma^2\right)$, and thus F is the cdf of a normal distribution with mean $\log \beta + z_{k,m,t} + h_l$ and standard deviation σ^2.

In the data, I observe a collection of employment spells indexed by s. Each spell consists of a worker identifier $i(s)$, a firm identifier $j(s)$, a location identifier $l(s)$, a period identifier $t(s)$ (which denotes whether the match belongs to the "before" or "after" sample), and a log wage $\log w(s)$. I define assignment functions $a^w(\cdot) : \{1, \ldots, I\} \mapsto \{1, \ldots, K\}$ and $a^f(\cdot) : \{1, \ldots, J\} \mapsto \{1, \ldots, M\}$ that map each worker and each firm into their type. Then, I follow section 4.1 in solving the likelihood maximization problem:

$$\max_{\theta, a^w, a^f} \mathcal{L}(\theta, a^w, a^f)$$
$$= \max_{\theta, a^w, a^f} \sum_{t=0}^{1} \sum_{k=1}^{K} \sum_{m=1}^{M} \sum_{l=1}^{L} \sum_{s} d_{k,m,l,t}(s, a^w, a^f) \cdot \log f_{\theta,k,m,l,t}(w_s)$$

where $f_{\theta,k,m,l,t}$ is the pdf of a normal distribution with mean $\log \beta + zk, m, t + h_l$ and standard deviation σ^2

$$d_{k,m,l,t}(s, a^w, a^f) = \mathbb{1}(a^w(i(s)) = k, a^f(j(s)) = m, l(s) = l, t(s) = t)$$

is an indicator function that equals one if the assigned worker type, firm type, location and period are equal to (k, m, l, t). As described above, LANCE starts with a guess for θ, a^f, a^w and proceeds to sequentially re-optimize a^w, a^f, and θ. This yields the updating rules

1. **Parameter updating:** $z_{k,m,t}, h_l, \sigma$ are determined by a regression of log wages on (k, m, t) and h_l fixed effects given the assignments, following equation (13). Note that at this stage, β is just a constant shifter, so we can set $\beta = 0.5$, which is a normalization.[15]

[15] In order to maximize the speed of this computation, it is a helpful property of the regression that $D_{k,m,l,t} = \sum_s d_{k,m,l,t}(s, a^w, a^f)$ is a sufficient statistic. Instead of including the complete data set in this regression, it is enough to perform an appropriately weighted regression with one observation per state (k, m, l, t).

2. **Worker type updating:** The updated assignment for each worker is given by

$$a^w(i) = \max_{k\in\{1,\dots,K\}} \sum_{s:i(s)=i} \sum_{t=0}^{1} \sum_{m=1}^{M} \sum_{l=1}^{L} d_{k,m,l,t}(s, a^w, a^f) \cdot \log f_{\theta,k,m,l,t}(w_s)$$

   This means that the new guess of each worker's type is the type that maximizes their likelihood given the current estimate of θ and a^f.

3. **Firm type updating:** The updated assignment for each firm is given by

$$a^f(j) = \max_{m\in\{1,\dots,M\}} \sum_{s:j(s)=j} \sum_{t=0}^{1} \sum_{k=1}^{K} \sum_{l=1}^{L} d_{k,m,l,t}(s, a^w, a^f) \cdot \log f_{\theta,k,m,l,t}(w_s)$$

   Again, the new guess of each firm's type is the type that maximizes their likelihood given the current estimate of θ and a^w.

LANCE can be viewed as a bi-clustering algorithm. Bi-clustering algorithms are algorithms that simultaneously cluster the rows and columns of a matrix. In this paper's application, workers and firms form the rows and columns of the matrix. Match-level wages form its entries.[16] Unlike most existing bi-clustering algorithms, which are NP-complex, the computational complexity of LANCE is comparatively low. The reason is that bi-clustering algorithms typically compare units of observations against each other, which yields a large number of potential comparisons. LANCE, on the other hand, can be viewed as comparing each unit against a "representative" agent as captured by the matrix $z_{k,m,t}$, which means that the number of potential comparisons is lower by several orders of magnitude.

Thinking of LANCE as a bi-clustering algorithm helps to understand the intuition surrounding the identification of worker and firm types. Loosely speaking, LANCE clusters workers into the same type if they have similar wages conditional on their firm types. So for example, two workers who share two employers and earn similar wages at both are likely to become part of the same cluster, i.e. are assigned the same worker type. Likewise, on the firm side, two firms that share some employees and pay them similar wages are likely to be classified as the same type. Firm types also share same match-conditional

[16]In this application, the overwhelming majority of entries is missing as most worker-firm pairs never match in the data.

productivity across periods. Therefore, two periods who see the same wage growth conditional on worker types across periods are likely to be classified as the same type. It is in part this property of the algorithm that allows me to separate firm and location components of the wage function, as plants do not change their location.

Identifying region types

In order to further reduce the dimensionality of the system and obtain transparent insights on sorting patterns, I restrict my attention to equilibria in which some of the locations l are ex-ante and ex-post identical. Concretely, I assume that each location belongs to one location type $n \in \{1, \ldots, N\}$ and that two locations l, l' belonging to the same location type n have the same productivity shifter $h_l = h_{l'}$ and identical distributions of workers and firms, $e_{k,m,l} = e_{k,m,l'}$. To estimate type membership of locations, I run a k-means clustering algorithm with N clusters on the employment distributions estimated by LANCE. That is, I select region clusters based on minimizing the within-group distance of $\frac{e_{k,m,l}}{\sum_{k,m} e_{k,m,l}}$. This is consistent with the model, since regions that are ex-post identical should have zero within group distance of their respective normalized employment distributions $\frac{e_{k,m,l}}{\sum_{k,m} e_{k,m,l}}$. I map h_l into a new parameter vector h_n by averaging over regions of the same type.

Given this simplification, in what follows, I switch the sub-indices corresponding to location l and instead index relevant parameters and variables using the index n, capturing the type of the region. I obtain transparent results for regional sorting by setting $N = 2$, so that I only consider two region types. This is a somewhat restrictive assumption but reducing the dimensionality of the problem in this way helps to shed light on sorting patterns that are otherwise hard to disentangle. As it turns out, the two region types do well in capturing two polar ends of the spectrum of locations. One is rich, the other poor. One is larger, the other smaller.

For the number of worker and firm types, I set $K = M = 7$. This is somewhat ad hoc but strikes a balance between generality, simplicity and computational feasibility. In terms of magnitude, it is comparable to the choice made in BLM to set $K = 6, M = 10$. I also order types in order of increasing average log wage, in order to be able to interpret higher type numbers as higher wage types.

4.2 Step 2: Identifying mobility parameters through maximum likelihood

Given the assignments for workers, firms and locations from step 1, I now proceed to estimate the technological sorting parameters $\eta_{k,m}$ and $\zeta_{m,m'}$, the mobility costs $\tilde{c}_{k,t}$, the firm type distribution $\psi_{m,t}$ the search intensity of the employed s, and the separation rate $\delta_{k,m}$ using transition data. To do so, I impose some of the structure that the model offers by imposing the parametric form of transitions. However, I do not impose all model assumptions, such as for example the Bellman equation that relates value functions and wages. Instead, I allow the endogenous objects of the model, $\phi_{m,n,t}$, λ_t and $\tilde{W}_{l,m,n,t}$ to vary freely in the estimation.[17] This yields estimates for those variables that are then discarded and replaced with values implied by the structure of the model.

A spell in the data contains two pieces of information - its length, and the direction of the worker's transition across firm types and locations. Given the Markov process implied by the model, the length of a spell is distributed according to an exponential distribution with rate $\mu_{k,m,n} = \sum_{m',n'} \mu_{k,m,m',n,n'}$ where $\mu_{k,m,m',n,n'}$ is the transition rate of a worker of type k at firm type m and region type n to firm type m' and region type n'.[18]. Defining $\bar{v}_{m,n} = \frac{\sum_k \mathcal{V}_{k,m,n}}{\mathcal{V}}$ and using $\mathfrak{N}_n$ to denote the number of regions of type n, we can re-write equation (4) and combine it with equation (10) to make it more amenable to the empirical estimation procedure:

$$\mu_{k,m,m',n,n',t} = \mathbb{1}(m' \neq 0) s^{\mathbb{1}(m \neq 0)} \lambda_t \psi_{m',t} \phi_{m',n',t} \eta_{k,m'} \zeta_{m,m'} \sum_{h=0}^{1} \mathfrak{k}_{n,n',h} P_{k,m,m',n,n',h,t} + \mathbb{1}(m' = 0) \delta_{k,m}$$

where

$$\mathfrak{k}_{n,n',h} = \begin{cases} (\mathfrak{N}_{n'} - \mathbb{I}(n = n')) & \text{if } h = 1 \\ \mathbb{I}(n = n') & \text{if } h = 0 \end{cases}$$

is the number of location-to-location transitions from n to n' that either require a location switch ($h = 1$) or do not require a location switch ($h = 0$). Note also that we have slightly

[17] In this way, we avoid the additional complication of having to solve the model for every parameter update. Trying to incorporate these restrictions would also lead to a much more complicated likelihood maximization problem.

[18] Recall equation (4) which features a version of this expression without the reference to region types

redefined the notation on

$$P_{k,m,m',n,n',0,t} = \frac{\tilde{W}_{k,m',n,t}}{\tilde{W}_{k,m,n,t} + \tilde{W}_{k,m',n',t}}, \quad P_{k,m,m',n,n',1,t} = \frac{\tilde{c}_{k,t}\tilde{W}_{k,m',n,t}}{\tilde{W}_{k,m,n,t} + \tilde{c}_{k,t}\tilde{W}_{k,m',n',t}}$$

to write the transition probability conditional on having to switch locations or not having to switch locations, captured in the new index $h \in \{0, 1\}$. Using the properties of Poisson processes, conditional on a transition, the probability of a transition in each direction (k, m, m', n, n') is given by $\frac{\mu_{k,m,m',n,n',t}}{\mu_{k,m,n,t}}$. Letting $r(s)$ denote the length of employment spell s, the log likelihood of the observed spells is thus

$$\mathcal{L}_{\text{Mobility}} = \sum_{k,m,m',n,n',t} \sum_{s} d_{k,m,m',n,n',t}(s) \Bigg[\underbrace{\log\left(\sum_{\tilde{m},\tilde{n}} \mu_{k,m,\tilde{m},n,\tilde{n},t}\right) - \sum_{\tilde{m},\tilde{n}} \mu_{k,m,\tilde{m},n,\tilde{n},t} r(s)}_{\text{Length}} + \underbrace{\log\left(\frac{\mu_{k,m,m',n,n',t}}{\sum_{\tilde{m},\tilde{n}} \mu_{k,m,\tilde{m},n,\tilde{n},t}}\right)}_{\text{Direction}} \Bigg]$$

Like with LANCE, I again employ a coordinate descent strategy to optimize this object over the parameters $(\tilde{c}_{k,t}, \psi_{m,t}, s, \delta_{k,m}, \phi_{m,n,t}, \lambda_t, \tilde{W}_{l,m,n,t})$: Starting with a guess for the parameters, one can successively update each parameter using its respective first order condition. Some of the first order conditions turn out to be intractable and thus, maximizing this likelihood requires some additional finesse. I follow the same strategy as Lentz et al. (2023) who estimate the same worker mobility model in discrete time. Like them, I use an MM algorithm, amending their estimation method to fit a continuous time context. Appendix A.3 derives all the resulting updating conditions for reach relevant parameter in my setting.

4.3 Step 3: Pinning down residual parameters

Given the model parameter β, the application of LANCE directly yields estimates for $z_{k,m,t}$ and h_n, a classification which enables maximum likelihood, as well as an employment distribution in the data, $e^{\text{data}}_{k,m,n,t}$. The maximum likelihood procedure further yields values for $\eta_{k,m}, \zeta_{m,m'}, \tilde{c}_{k,t}, s$ and $\delta_{k,m}$. The remaining parameters can be separated into groups that differ in terms of how I estimate or calibrate them. Table 1 summarizes the different parameter groups and their respective estimation method or calibration target.

Parameter	Symbol	Value	Estimation method	Target
Worker share of output	β	0.5	Normalization	
Elasticity of encounter function	α	0.5	Externally set	
Interest rate	r	5% p.a.	Externally set	
Worker death rate	ρ	$\frac{1}{30}$ p.a.	Externally set	
Match-specific productivity	$z_{k,m,t}$		LANCE	
Location-specific productivity	$h_{n,t}$		LANCE	
Required type distribution	$\eta_{k,m}$		Max. Likelihood	
Prob. that a worker is qualified	$\zeta_{m,m'}$		Max. Likelihood	
Mobility cost	$\tilde{c}_{k,t}/c_{k,t}$		Max. Likelihood	
Firm type probability at entry	$\psi_{m,t}$		Max. Likelihood	
Search intensity in employment	s	0.305	Max. Likelihood	
Separation rate	$\delta_{k,m}$		Max. Likelihood	
Birth rates	$\chi_{k,n,t}$		Min. distance	Employment dist.
Randomness in worker mob. decision	σ^w	$9.18 \cdot 10^5$	Min. distance	Employment dist.
Randomness in firm location choice	σ^f	$9.35 \cdot 10^5$	Min. distance	Employment dist.
Match efficiency	A	7.33	Matched moments	Job filling rate $= \frac{1}{30}$
Ex-ante entry value	ϱ	$1.34 \cdot 10^5$	Matched moments	Encounter rate $= 0.203$
Net location amenity	$a_{n,t} - p_{n,t}$	$(-5.91, 5.91)$	Matched moments	Location size
Vacancy death rate	δ^v	$\frac{1}{30}$	Matched moments	Vacated vacancy share $= \frac{1}{2}$

Table 1: Parameters and estimation strategy

First, under the assumptions made in the model setting β to any value is simply a normalization, since it simply rescales the firm or worker values - I set it to 0.5. Second, some parameters are set externally to values in reasonable ranges found elsewhere in the literature. The elasticity of the encounter function is set to $\alpha = 0.5$, the interest rate r to 5% per annum and the death rate of workers is set to $\frac{1}{30}$ per annum, to approximate a 30-year working life.

Next, I choose some parameters to hit specific moments in the "after" period. The meeting efficiency A is chosen in order to make the job filling rate equal to $\frac{1}{30}$ per day, or equal to one in monthly terms. The ex-ante entry value ϱ is set to match the estimated value of the worker encounter rate λ that arises from the maximum likelihood estimation, which is 0.203. $a_{n,t}$ and $p_{n,t}$ are not separately identified, since they jointly enter the worker's value as $a_{n,t} - p_{n,t}$. However, defining $\tilde{a}_{n,t} = a_{n,t} - p_{n,t}$, I chose this value to make the model-implied populations in the model consistent with the population implied by the data employment vector $e^{\text{data}}_{k,m,n,t}$. That is, I assume housing supply adjusts in a way that makes the model populations in the two region types consistent with the data.

Lastly, the vacancy death rate δ^v is chosen to equal $\frac{1}{30}$, like the job filling rate. Equating δ^v and the job filling rate ensures that the share of vacated vacancies (i.e. the share of vacancies that where vacated by a previous employee) equals exactly $\frac{1}{2}$, which is approximately consistent with the evidence presented in Qiu (2022).

The last remaining parameters to set are σ^w, σ^f, and $\chi_{k,n,t}$. I choose these parameters to minimize the distance between the model-implied distribution of workers and the one measured in the data. That is, I minimize

$$\sum_{k,m\geq 1,n,t} \left(\frac{e_{k,m,n,t} - e^{\text{data}}_{k,m,n,t}}{e^{\text{data}}_{k,m,n,t}} \right)^2$$

subject to $\chi_{k,n,t} \geq 0$.

5 Data

To estimate the model, I leverage the SIEED data set described in section 2. Since the model does not allow for commuting, commuting zones are the appropriate choice to map a location from the model to the data. However, mapping other aspects of the model into the data requires a few further modifications. First, the model abstracts from both aggregate wage growth over time and from seniority. I thus residualize log wages in the data by running the following regression

$$\log w_{it} = \delta_t + \beta_1 \text{age}_{it} + \beta_2 \text{age}^2_{it} + \varepsilon_{it}$$

which removes age and time effects. To keep wages at an interpretable scale, I set as new log wage equal to

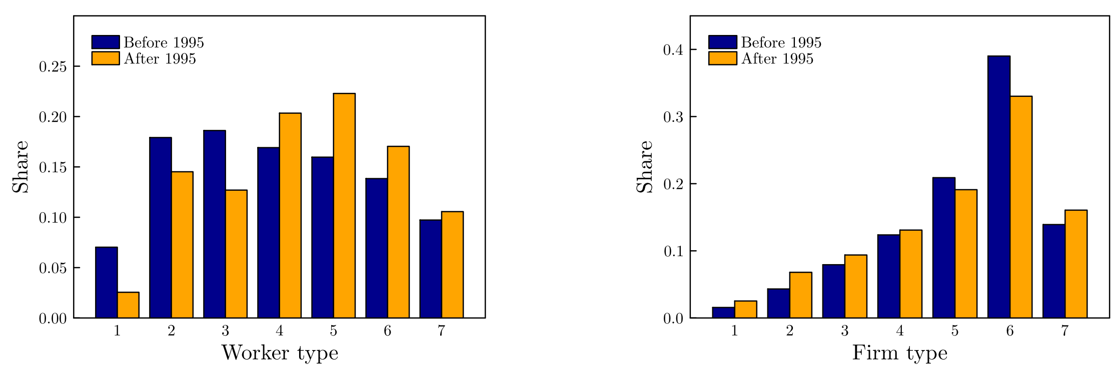
$$\log w'_{it} = \delta_{2018} + \beta_1 35 + \beta_2 35^2 + \varepsilon_{it}$$

which gives it the interpretation of the wage the worker would get if they were 35 years old in 2018. Hence, all monetary amounts in this paper can be interpreted as 2018-Euros. Since the model conceptualizes wages as being fixed for a match, I then assign a log wage to each match by averaging across its associated data entries. This procedure finally yields a data set with about 25 million continuous spells of 4.4 million workers and 2.6 million firms across 108 commuting zones. In order to estimate the model, I further impute non-employment spells for all workers by filling in periods during which these workers are not observed in employment.

6 Results

I now turn to the results from the estimation and simulation of the model. The model delivers answers to the central questions of this paper. Most importantly, it allows us to answer the motivating question: Why has spatial inequality increased? By estimating sorting patterns and match productivity, we can disentangle the drivers of the rise in spatial inequality. The model also offers a new view on how to understand the role of this dimension of inequality: As we will see, the model shows that spatial inequality reflects major differences in long-term worker values across locations. Finally, we can also use he model to better understand how the two-sided sorting patterns of firms and workers come about in equilibrium. The answer to this question is important to understand how sorting might drive aggregate outcomes under different kinds of spatial policy interventions.

6.1 Empirical results from LANCE



(a) Workers

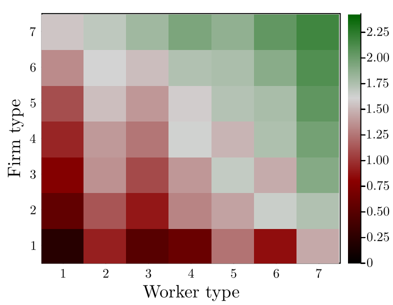
(b) Firms

Note: The two panels depict the marginal distributions $\sum_{m,n} e^{\text{data}}_{k,m,n,t}$ (left panel) and $\sum_{k,n} e^{\text{data}}_{k,m,n,t}$ (right panel) respectively.

Figure 4: Marginal distribution of types

I begin by describing the estimated marginal distributions of workers and firm types, displayed in Figure 4. We can see from the figure that LANCE assigns roughly similar numbers of workers to all types, with a slight shift towards higher worker types in the post-1995 period. Figure 25 in appendix A.8 shows that college educated workers are mostly concentrated in types 6 and 7. The proportion of each firm type, displayed in the second panel, does not significantly change over time. Firm type employment varies considerably across types, with type 6 being the most common.

Next, consider the clustering of regions. I assign 55 commuting zones as "type 1" regions and 53 commuting zones as "type 2" regions. Type 2 regions are richer and contain the majority of employment, about 75% before 1995 and 74% after 1995. h_n is estimated to equal 3.60 in the poor region and 3.59 in the rich region, implying that the composition of workers and firms fully accounts for the wage premium of rich regions. In general, the quantitatively small difference between the two places indicates a muted role for fundamentals. The mean log wages for the poor region type and the rich region type are 4.56 and 4.69 respectively, implying substantial ex-post differences between the two places.



(a) $z_{k,m,0}$ (before 1995)

(b) $z_{k,m,1}$ (after 1995)

(c) Difference

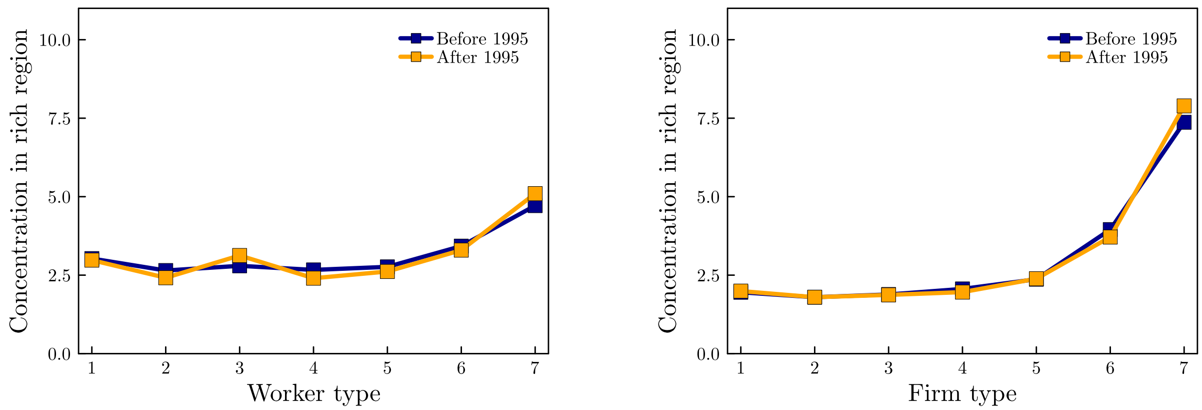
Note: The first two panels show the estimates of $z_{k,m,0}$ and $z_{k,m,1}$ delivered by LANCE, with the color scale shown on the right hand side. The third panel shows the difference $z_{k,m,1} - z_{k,m,0}$.

Figure 5: Match-specific component of the wage function

Next, I turn to the estimation of the technology shock. The first two panels of Figure 5 display the estimates for the match component of the log wage function (net of a constant), $z_{k,m,t}$, for both the "before" and the "after" period. Given the assumptions made on wages, this is also the match component of the production function. The plot is highly informative about the nature of worker-firm matching even outside a spatial context. Wages are mostly monotone in worker and firm types. These results support models in which workers and firms are mainly vertically differentiated.

The wage function is not quite log-additive, as imposed by AKM. This is visible by the differences in slope for different worker types when moving from lower to higher firm types. For example, according to Figure 5, in the "before" period, type 4 workers tend to benefit more from moving to higher firm types than type 5 workers do, signified by the steeper color gradient across firm types for the former. Generally, low-wage firms tend to pay low wages to all worker types and, in the "after" period, high-wage firms pay high wages to most worker types. The third panel of Figure 5 shows the difference $z_{k,m,1} - z_{k,m,0}$. The figure shows that wages rise especially at top firms for all workers. This is consistent with the hypothesis of "superstar firms" put forth by Autor et al. (2020): A small number of firms become very productive, offering higher wages to their workers. The figure also reveals that the highest worker type has seen increases in their expected wages for matches with all firms.

Next, I turn to the distribution $e^{\text{data}}_{k,m,n,t}$. This distribution is computed by summing up the spell lengths of all states as observed in the data, normalizing $\sum_{k,m,n} e^{\text{data}}_{k,m,n,t} = 1$. The marginal distributions by worker and firm type are shown in Figure 4. Figure 6 shows a measure of regional sorting for workers and firms respectively. The figure displays the respective worker and firm types on its x-axis and on the y-axis displays a measure of concentration in region type 2, which is chosen to be the richer of the two region types. Concretely, the y-axis displays the ratio of employment in region 2 and employment in region 1. Positive sorting between regions and workers (firms) is therefore indicated by an increasing line, since it means that higher type workers (firms) are more heavily concentrated in the rich regions. It is obvious from the figure that there is indeed positive regional sorting for both workers and firms, as the line is upward-sloping for workers and firms in both periods. However, firm sorting is much more pronounced than worker sorting, implying that the distribution of firms, and thereby job ladders, vary quite starkly between the two regions. However, the figure also shows that sorting has not changed very much over time. We can see at most a very moderate increase in firm and worker sorting over time. This means that changes in spatial inequality that arise from a change in technology must arise almost exclusively as a result of different



(a) Worker-region sorting

(b) Firm-region sorting

Note: The two panels show the regional concentration ratios $\sum_m e^{\text{data}}_{k,m,2,t} / \sum_m e^{\text{data}}_{k,m,1,t}$ (left panel) and $\sum_k e^{\text{data}}_{k,m,2,t} / \sum_k e^{\text{data}}_{k,m,1,t}$ (right panel) respectively.

Figure 6: Regional sorting in the data

exposures to the technology shock, not from changing sorting patterns.

In order to quantify the intuition in Figure 6 that firm sorting is indeed more pronounced than worker sorting, we can use tools from information theory. Concretely, I compare the reduction in entropy of the employment distribution across region types that is caused by revealing worker or firm types respectively. That is, I quantify the amount of information about the location type of a randomly sampled worker that is generated by revealing the worker or firm type of said worker. Intuitively, if all sorting was on the worker side, revealing information about the worker type would reveal information about the location type but revealing information about the firm would not. Conversely, in a world in which firms sort across locations but workers do not, one could only reveal information about a worker's location by revealing the firm type of their employer, not their own worker type. Thus, denoting the employment distribution in a given period as $e_{k,m,n}$, measures of the relative contribution of worker and firm sorting to regional sorting is given by

$$\omega^{\text{firm}} = \frac{I_e(n|m)}{I_e(n|m) + I_e(n|k)}, \quad \omega^{\text{worker}} = 1 - \omega^{\text{firm}}$$

where

$$I_e(n|m) = H(e_m) + H(e_n) - H(e_{m,n})$$
$$I_e(n|k) = H(e_k) + H(e_n) - H(e_{k,n})$$

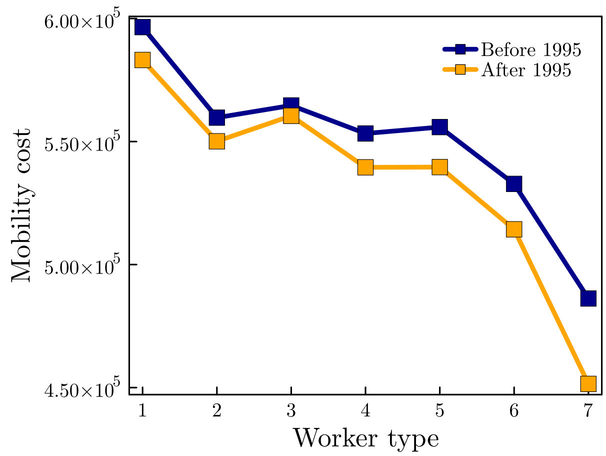
Here, dropped subscripts indicate the summing out of a dimension, e.g. $e_{m,n} = \sum_k e_{k,m,n}$ and H is the entropy operator $H(x) = -\sum_i x_i \log(x_i)$.

As Williams and Beer (2011) point out, $I(e_{k,m,n}|m)$ and $I(e_{k,m,n}|k)$ contain both the unique information content of worker and firm types respectively, but also the redundant content delivered by both variables. By the definition above, this implies that in a world with perfect worker-firm sorting (e.g. all workers of type k work for firm type k), $\omega^{\text{firm}} = \omega^{\text{worker}} = 0.5$ because the same information is revealed by either component.

Applying this decomposition to the measured distribution $e^{\text{data}}_{k,m,n,t}$, I find that $\omega^{\text{firm}} = 0.79$ in the more recent period ($\omega^{\text{firm}} = 0.83$ in the earlier period), confirming that most of the regional sorting comes from the firm side. This is an important result: Since most of the sorting occurs on the firm side, the ability to climb up the job ladder varies strongly depending on what location a worker lives in. Recall the question brought up in the introduction: Do spatial disparities occur from regional sorting of workers or regional sorting of firms? In the model presented here, spatial inequality does not matter if spatial disparities come mostly from worker sorting. The decomposition shows that this is not the case: Spatial disparities are mostly an expression of firm sorting. Spatial inequality thus becomes very important in generating life-time differences between workers. Consequently, there is a large potential for spatial policy in reshaping worker-level outcomes, as discussed in section 6.4.

6.2 Empirical results from Maximum Likelihood

Next, I consider the results from the maximum likelihood estimation step. The relative search intensity of employed workers s is estimated to be equal to 0.305. Figure 7 displays the estimated mobility cost c_k for each worker type for both periods. The scale is in 2018-Euros. We can see that the mobility cost ranges between around 450'000 Euros and 600'000 Euros for all workers. This is roughly consistent with the estimates in Kennan and Walker (2011), who estimate a moving cost of 312'000 2010-USD on average in US

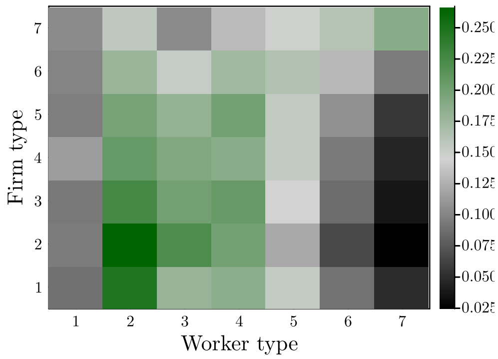


Note: The figure shows the estimates for the mobility cost $c_{k,t}$ of a worker type k in both periods.

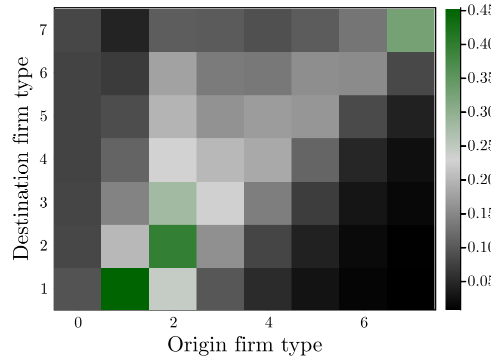
Figure 7: Mobility costs

data. As the figure shows, high worker types tend to be more mobile than poor worker types. In the model, realized mobility is driven by a trade-off between mobility costs and the benefits of switching locations. While these estimates seem large, their exact interpretation hinges in part on how one views variation in the preference shocks that govern a moving decision. If such shocks can be interpreted as including shocks to moving costs, workers are more likely to move if the moving costs are lower than their unconditional expectation. Furthermore, these estimates may capture non-pecuniary costs to moving, such as the loss of established social circles, a familiar environment, and the labor associated with administrative costs.

Figure 8 shows the estimates for the probability that a firm of type m requires a worker of type k ($\eta_{k,m}$) and the probability that a match is viable conditional on the origin and destination firm types ($\zeta_{m,m'}$). There is clear evidence for non-spatial segmentation in the form of $\eta_{k,m}$: Low firm types typically require low worker types and vice versa, signified by larger values close to the diagonal. In particular, the highest firm types is accessible almost exclusively to the highest worker type. The second panel, depicting $\zeta_{m,m'}$ reveals that job ladders follow a stepping-stone process: Firm types typically accept workers who previously worked at the same firm type. Again, the diagonal values are larger than off-diagonal ones, showing that firms are more likely to hire workers that previously worked



(a) $\eta_{k,m}$



(b) $\zeta_{m,m'}$

Note: The two panels show the estimates of $\eta_{k,m}$ and $\zeta_{m,m'}$ delivered by LANCE, with the color scale shown on the right hand side.

Figure 8: Maximum likelihood estimates for $\eta_{k,m}$ and $\zeta_{m,m'}$

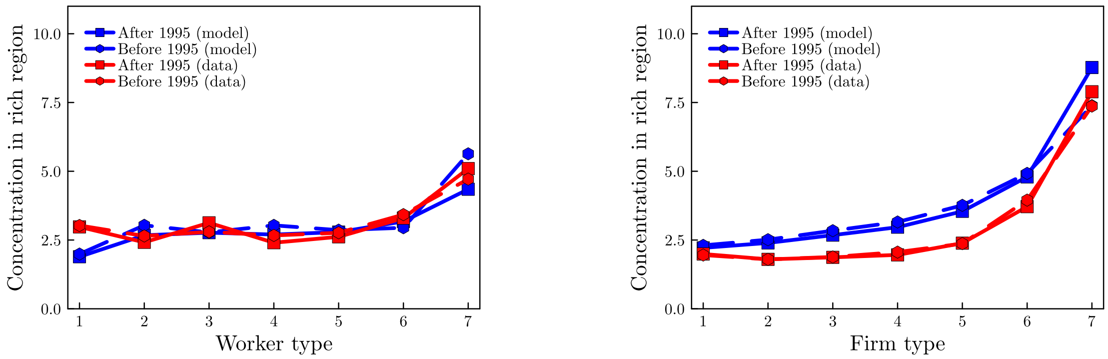
at the same firm type.

6.3 Model results

Finally, we turn to the results of the model. First, I verify that the model captures the cross-sectional patterns of spatial sorting well. Figure 9 overlays the model-predicted spatial sorting patterns from Figure 6 with the sorting patterns implied by the model. The model does a good job at capturing the cross section, both in the "before" and "after" periods. The model replicates the magnitudes of spatial sorting well: The firm side sees strong sorting by region. The worker side sees more muted sorting.

Since sorting has not changed very much, it is arguably more interesting to decompose the level of sorting than it is to decompose its change. The model allows us to decompose the drivers of both worker and firm sorting. The decomposition is done separately for workers and firms by changing the distributions of workers and firms they face in partial equilibrium.

For the firm decomposition, I modify the employment distribution facing firms, holding all other variables constant. Denoting marginal distributions by omitting subscripts, I consider a version with no regional sorting of workers or firms ($e^{\text{cf}}_{l,m,k} = e_{l,m} \cdot e_n / e$), a version with equal sized regions ($e^{\text{cf}}_{l,m,k} = e_{l,m,k} / (e_n \cdot n)$) and a version with both ($e^{\text{cf}}_{l,m,k} =$



(a) Workers

(b) Firms

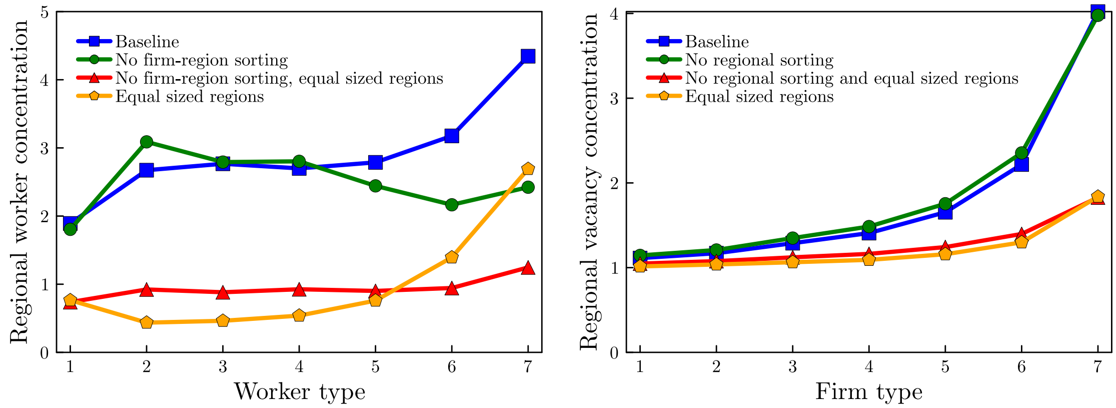
Note: The two panels show the regional concentration ratios $\sum_m e_{k,m,2,t} / \sum_m e_{k,m,1,t}$ (left panel) and $\sum_k e_{k,m,2,t} / \sum_m e_{k,m,1,t}$ (right panel) for both $e_{k,m,n} = e^{\text{data}}_{k,m,n}$ (red) and $e_{k,m,n} = e^{\text{model}}_{k,m,n}$ (blue) respectively.

Figure 9: Type sorting into rich regions

$e_{l,m} \cdot 1/n$). For the decomposition on the worker side, I vary the distribution of vacancies across regions, holding fixed all other variables. Again I consider three versions, one with no regional sorting of firms ($\lambda^{\text{cf}}_{m,n,t} = \lambda_m \cdot \lambda_n / \lambda$), a version with equal sized regions ($\lambda^{\text{cf}}_{m,n,t} = \lambda_{m,n,t} / (\lambda_n \cdot n)$) and a version with both ($\lambda^{\text{cf}}_{m,n,t} = \lambda_m \cdot (1/n)$). The results are displayed in Figure 10.

Consider firm sorting first. We can see from Figure 10b that firm sorting becomes significantly weaker as soon as the two regions are of equal size. The advantage of locating in a larger region consists mostly in being able to fill jobs more quickly, since the larger location size means that matches are more likely to be filled without a worker having to move locations, which is costly to the worker. This motive is more important to higher firm types, since their matches are worth comparatively more. The ease of filling a job is also the main reason for the residual sorting that remains after removing both regional sorting in the employment distribution and making the size of employment equal in both region types. Since all other variables are held constant, it still remains the case that workers are more willing to transition into the rich region type than into the poor region type, retaining a sorting incentive that arises from ease of filling.

Spatial sorting on the worker side contributes to firm sorting as well but only has a minor quantitative impact on firm sorting in my specification. There are at least two reasons



(a) Decomposition of worker-region sorting (b) Decomposition of firm-region sorting

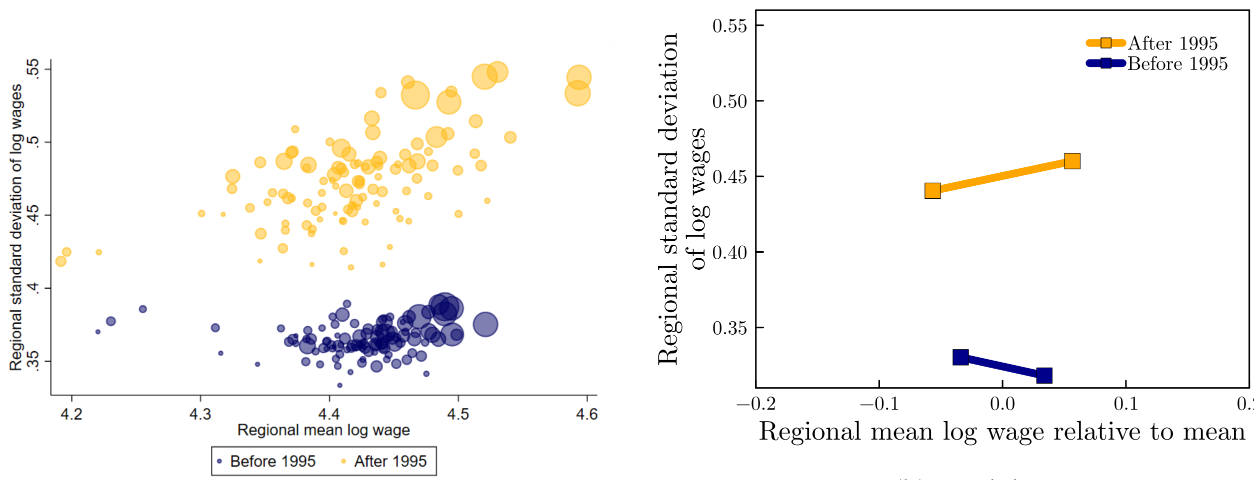
Note: The left panel shows the regional concentration ratio of workers $\sum_m e^{\text{model}}_{k,m,2} / \sum_m e^{\text{model}}_{k,m,1}$ in the "after" period under the different scenarios outlined in the main text. The right panel shows the concentration ratio of vacancies $(\lambda \bar{v}_{m,2})/(\lambda \bar{v}_{m,1})$ in the "after" period under the different scenarios outlined in the main text.

Figure 10: Sorting decompositions

for this. The first reason is that region-worker sorting is not very strong in the data and the model. The second reason is that when firms make location decisions, they have not yet realized their draw of the required worker type necessary to fill the position. While this is an imposed assumption of the model, it mirrors an important feature of reality which quantitatively limits the channel from worker sorting onto firm sorting: Plants typically require workers of all types. Even high-skill-intensive plants often employ a significant number of low-skill workers. Thus, even if worker sorting is significant, jobs need to consider regional sorting patterns of all worker types when making location decisions.

Sorting of workers is a different story and is entirely driven by the strong firm sorting. Since high type firms are more likely to locate in the rich region type, and are more likely to require high type workers, high type workers are more likely to find work in the rich region type. This accounts for essentially all sorting on the worker side, as revealed by Figure 10a. This is another important result and has direct implications for policy, as discussed in section 6.4: As firm sorting drives worker sorting, changing firm sorting incentives translates directly into changes in the distribution of workers across space. The redistributive effects of policy stem primarily from this mechanism. Workers like living in a rich region but access to those regions is primarily determined by the type composition of local jobs. As we will see, policies that reduce firm sorting thus tend to be progressive and benefit low-skill workers at the expense of high-skill workers.

Next, we test whether the model is capable of replicating the dynamics of inequality outlined in section 2: Spatial inequality has increased, and the cross-sectional relationship between the income level of a region and its inequality has become positive.



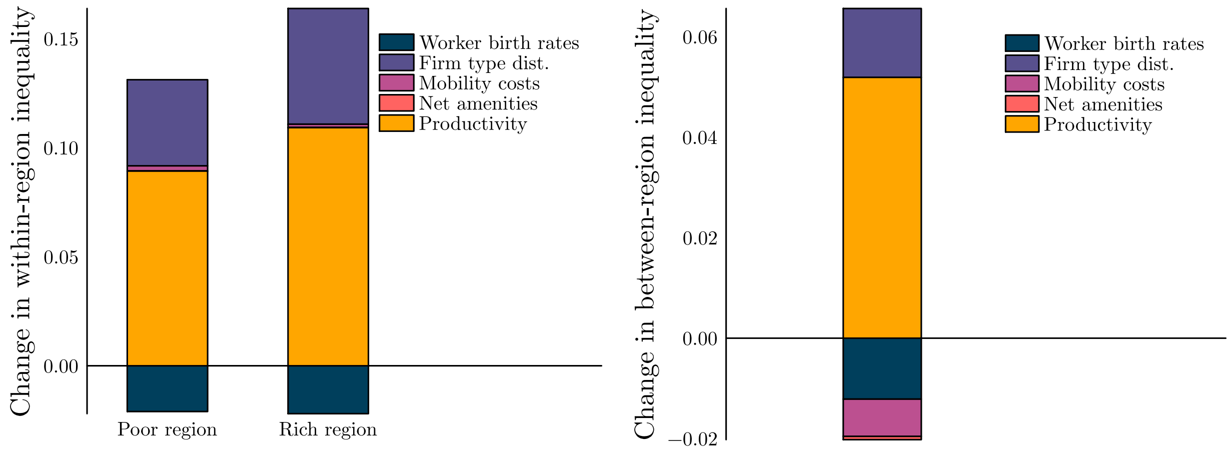
(a) Data

(b) Model

Note: The left panel shows the standard deviation of match log wages in each commuting zone plotted against the regional mean log wage of that commuting zone, with the size of the circle proportional to the size of the commuting zone. The right panel plots the standard deviation of match log wages in a region type in the model against the mean log wage of that region type in the model relative to the average across both region types.

Figure 11: Income-inequality relationship

Figure 11b tests these properties of the model. For both periods, Figure 11b plots the two region types' coordinates in income-inequality space and contrasts this outcome of the model outcome against its counterpart in the data (Figure 11a). The model replicates two important features of the data. First, inequality across space rises significantly between the "before" and "after" periods, visible by the increasing distance between the two dots in the x-dimension. Second, the model replicates the emerging income-inequality relationship we can see in the data. Whereas there is a weakly negative relationship between income and inequality in the "before" period, the dots for the "after" period exhibit a clear positive correlation, much like in the data.



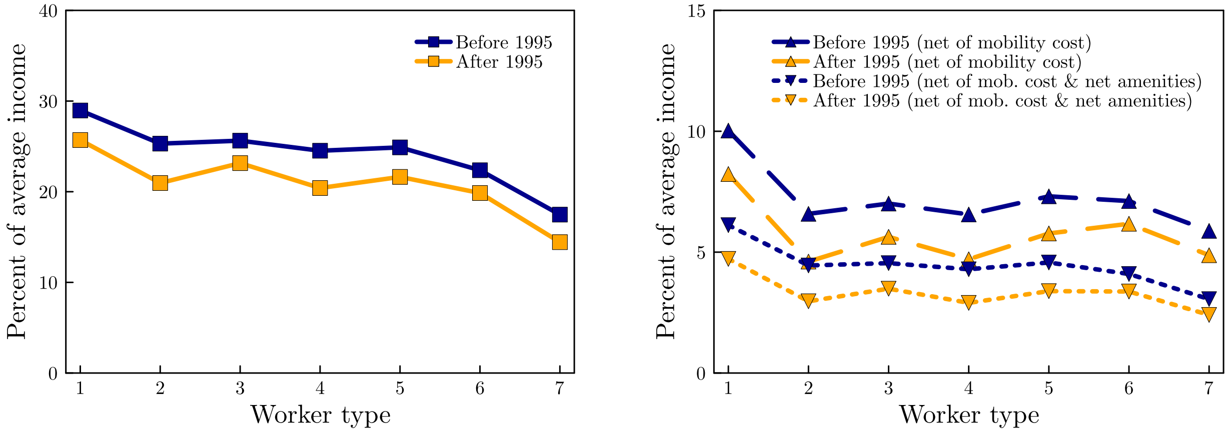
(a) Within regions

(b) Between regions

Note: The left panel shows the difference in the standard deviation of match log wages for each region type arising from the different equilibria described in the main text. The right panel shows the difference in the regional difference of regional mean log wages corresponding to these equilibria.

Figure 12: Model-based decomposition of the change in inequality

Figure 12 shows how the model rationalizes this development. The figure is based on different equilibria which are generated by successively adjusting the different exogenous variables that are allowed to change across periods: Worker birth rates $\chi_{k,m,n,t}$, The firm type distribution $\psi_{m,t}$, mobility costs $c_{k,t}$, net amenities $a_{n,t} - p_{n,t}$ and, lastly, match productivity $z_{k,m,t}$. The figure shows that, through the lens of the model, the change in match productivity accounts for the lion share of the inequality dynamics both within and between regions. As seen above, changes in sorting are extremely muted, implying that most of this contribution arises from interactions. This finally delivers an important insight into the root causes of changing between regions and the differential developments of increasing spatial inequality: To understand the changing dynamics of spatial inequality over time, we need two main ingredients - strong regional firm sorting driven by size effects on the one hand and an increase in the productivity of high-type firms on the other. As high-type firms experience productivity gains, these firms are primarily located in rich regions. This means that the gains from this productivity increase become visible primarily in the wage distribution of these regions - spatial disparities rise, and within-region wage inequality rises in those regions.



(a) Total

(b) Decomposition

Note: In the left panel, the y-axis displays the average regional compensating differential in percentage terms of mean income, i.e. $(\sum_{m,n} e^{\text{model}}_{k,m,n}(r+\rho)(W_{k,m,2}-W_{k,m,1}))/\bar{w}$ where $\bar{w}$ is the average wage level in the economy. In the right panel, the y-axis displays the same compensating differential based on $W^{\text{net mc}}_{k,m,n}$ and $W^{\text{net mc,am}}_{k,m,n}$ instead of $W_{k,m,n}$, which are the worker values net of mobility costs and net of mobility costs and amenities, as described in the main text.

Figure 13: Compensating differential, rich versus poor region type

Next, we turn to the question whether the increase in spatial inequality matters for worker values. To answer this question, I compute the spatial compensating differential for every worker in every state. The spatial compensating differential is defined as the amount of money a worker would be willing to pay to move costlessly from a poor to a rich region (while retaining their job), or to prevent such a move in the opposite direction. Formally, it is defined as

$$\text{CD}_k = \sum_m e_{k,m,n}(r+\rho)(W_{k,m,2}-W_{k,m,1}) \tag{14}$$

The first panel of Figure 13 shows this compensating differential as an annuity, expressed as a share of mean income for both periods and for each worker type. The figure clearly displays three important facts: First, compensating differentials are large and even larger for lower worker types. In the more recent period, workers would on average be willing to give up between 14% (type 7) and 26% (type 1) of average income if they could costlessly switch from the poor to the rich location (or prevent a switch in the opposite direction) while retaining their jobs, with an average compensating differential of 21%. Second, compensating differentials are decreasing in worker types, meaning that spatial disparities are more important for low-skill workers. Third, compensating differentials have not increased in the more recent period. On the contrary - despite the rise of cross-sectional spatial inequality, we see a mild decrease in the compensating differential. This mirrors mostly the decrease in mobility costs, which is easy to rationalize - in a model with costly worker mobility, the benefits of moving should in equilibrium equal the costs. Thus, the dynamics of mobility costs from Figure 7 are also reflected in Figure 13.

The large compensating differentials reflect three benefits of being in a rich location. First, being in a rich location entails less expected location switching in the future. Second, rich locations feature higher net amenities (see Table 1). Third, living in a rich region allows the worker to gain access to the preferable job ladder there. To evaluate the relative role of these three components, I first calculate a counterfactual value function $W^{\text{net mc}}_{k,m,n}$ that nets out the mobility cost from the value function and recompute the compensating differential from equation (14). Details are in appendix A.7. It turns out that large mobility costs explain much of the compensating differential. Netting out mobility costs, the average compensating differential shrinks from 21% to 5.4% of permanent average income. This is also reflected in Figure 13b, which displays the counterfactual compensating differential by worker type. Yet, the contribution of mobility costs to compensating differentials are only part of the story. Workers are not forced to move but do so when it is beneficial for them. Benefits of living in a rich location include access to better jobs and access to higher amenities. Thus, to understand the high compensating differentials it is important to understand whether workers move for the former or the latter reason. Figure 13b delivers the answer to this question. Netting out both mobility costs and net amenities from the value function ($W^{\text{net mc,am}}_{k,m,n}$, see appendix A.7 for details), we can see that the compensating differential declines further to about 3.2% of permanent average income. Thus, better job ladders account for more than half of the benefits that make rich locations desirable.

It is worth noting that compensating differentials are very large despite the fact that much of the variance of wages arises within, not between regions, a fact that has sometimes been used to argue against the importance of spatial inequality.[19]. The reason that both facts can be true at the same time is that compensating differentials are determined by forward-looking workers that internalize the cost of moving and the different job ladders that each location offers. Since the model presented here endogenously generates heterogeneous job ladders across space, spatial inequality in worker values can be far larger than spatial inequality in measured wages, as values take into account the large benefit of sampling from an improved pool of job offers and the cost of accessing this pool. These dynamic benefits of relocation underscore the importance of studying spatial inequality in a dynamic setting.

Summing up, the above exercises shed light on the causes and consequences of spatial sorting: Location size drives firm sorting and firm sorting drives worker sorting. Spatial inequality is due to both but firm sorting is more pronounced than worker sorting. Changes in spatial inequality come from productivity gains of the highest firm types and to some extent from the highest worker types. As an example, consider, say, Deutsche Bank with offices in Frankfurt.[20] The model predicts that a high-productivity company like Deutsche Bank creates jobs in these locations because they are easy to fill - lots of workers look for jobs in large cities like Frankfurt. Since most jobs at Deutsche Bank are high-skill, high-skill workers disproportionately live in these cities which explains part of the wage gap between a place like Frankfurt and a smaller, poorer commuting zone. Through the lens of the model, a technology shock that makes Deutsche Bank more productive drives up inequality across regions, as cities like Frankfurt experience disproportionate wage increases. It also makes rich regions more unequal, as wage dispersion in Frankfurt rises.

The fact that much of the spatial inequality comes from firm sorting means that good jobs are harder to come by in poor regions. Thus, workers in poor regions have to move to access the top of the job ladder, which is costly. Places like Frankfurt tend to be places of opportunity for workers precisely because companies like Deutsche Bank locate most of their jobs there. The model shows that it is mostly the poor who lose from these spatial disparities, as they are more mobility-constrained than their rich peers. As the model shows, their losses from being in a poor region are large.

[19]See Figure 23 in appendix A.8

[20]This is a purely fictional example and is in no way related to the SIEED.

6.4 Policy experiments

The previous section has shown new insights about the drivers of sorting - size effects drive firm sorting and firm sorting drives worker sorting. In this section, I explore the implications of this insight for place-based policy. The rich two-sided heterogeneity of the model enables us to focus particularly on the redistributive properties of policy. I consider two possible policies that are commonly used in practice to support spatial redistribution. First, I analyze the effect of a transfer to workers in poor regions, financed by a lump-sum tax borne by all workers. Such transfers can either be thought of as a direct monetary transfer, or can be interpreted as a more indirect subsidy of local amenities and infrastructure. I model the policy as a transfer τ_n that is paid out conditional on living in a region of type n. Consequently, under a policy τ_n^{reg}, a worker living in a location of type n now enjoys a flow value of $w_{k,m,n} + a_n - p_n + \tau_n$. The equilibrium value function of the worker is modified accordingly.

Second, I consider spatial policies that incentivize job creation by particular firm types in poor regions. These kinds of policies have been subject to previous research (see e.g. Suárez Serrato and Zidar (2016) for a static model with firm heterogeneity), and are prevalent throughout many developed economies, such as the United States and Germany. For example, in the US, tax incentives for firms are primarily regulated at the state level, making business taxes and tax breaks a powerful tool of place-based policy. To bypass the question of whether the level of job creation is efficient in my model, I only consider tax perturbations that are budget-balanced. That is, I consider policies that subsidize job creation in one region but tax job creation of the same firm type in another region. I model this policy as a transfer $\tau_{m,n}^{\text{job}}$ that is paid to firms upon job creation in region type n. This means that $\bar{\Omega}_{m,n} + \tau_{m,n}^{\text{job}}$ replaces $\bar{\Omega}_{m,n}$ in all equilibrium conditions.

I focus on perturbations of the equilibrium caused by marginal changes in policy. Any equilibrium outcome of the model can be written as a function of a vector of endogenous firm-side objects and house prices on the one hand, and the vector of policies on the other. Denoting the vector of endogenous objects as $x = (\{\bar{\Omega}_{m,n}\}_{m,n}, q, \{p_n\}_n)$ and a vector of outcomes as y, we can write:

$$y = f(x, \tau)$$

Furthermore, the vector of endogenous objects x must adjust to keep a vector of equilibrium conditions satisfied. That is, an equilibrium is given by some function g for which

$$0 = g(x, \tau)$$

Denoting the Jacobians of f and g as F and G respectively, we can therefore analyze a perturbation of the steady state as follows:

$$dy = F_x dx + F_\tau d\tau$$
$$0 = G_x dx + G_\tau d\tau$$

which implies

$$dy = \underbrace{-F_x G_x^{-1} G_\tau}_{\text{GE effects}} d\tau + \underbrace{F_\tau}_{\text{PE effects}} d\tau$$

Thus, any policy perturbation can be decomposed into a partial equilibrium (PE) and a general equilibrium (GE) component. The PE component captures changes that arise directly from the policy change, holding x constant. The GE component captures outcome changes arising from second round effects: The policy change might induce a reallocation of workers, for instance, that results in changing firm values, house prices, or job creation.

To analyze the two policies in question, I take a Rawlsian perspective and evaluate policies by how they change the expected value of a prospective unborn worker conditional on their type.[21] That is I define the "birth value" of a worker type k as

$$\text{BirthValue}_k = \sum_n \chi_{k,n} (r + \rho) W_{k,0,n}$$

and use changes in this birth value as the relevant welfare criterion.

To analyze the effects of policy, we also need to take a stance on the supply curve of housing. I assume that the housing stock is owned by absentee landlords who offer housing according to an exogenous supply curve. I further assume that the supply curve of housing can be approximated to a first order by

$$de_n = \omega_n dp_n$$

[21] An alternative criterion would be the average value of workers in the model. However, as discussed in appendix A.4, this welfare criterion has undesirable properties in my setting.

where ω_n denotes the sensitivity of the housing supply to price changes.[22] For every exercise, I will need to take a stance on the values for ω_n. The model features a form of observational equivalence of different vectors ω_n. As shown in appendix A.5, conditional on some policy change $d\tau$, for any vector ω_n, there exists a vector $\omega_{n'}$ that implies the same change of worker and firm allocations across states but differs in its implication for prices and thus worker values. As a simple example of this in the $N = 2$ case, consider the two cases where $\omega_n = (\infty, 0)$ on the one hand and $\omega_n = (0, \infty)$ on the other. Subsidizing workers in region 1 by $d\tau_1^{\text{reg}}$ leads to an instantaneous adjustment of prices to keep all allocations constant. In the first case, this means that prices in region 2 fall by $d\tau_1^{\text{reg}}$. In the second case, prices in region 1 rise by $d\tau_1^{\text{reg}}$. All allocations of workers and firms remain unaffected in both cases but the implications for prices and values differ strongly. By this logic, a policymaker who cares about the welfare of workers has a strong motive to subsidize the market with elastic housing supply to lower prices in the market with inelastic housing supply. This motive is extremely strong from a quantitative perspective, since it directly affects every worker in the economy.

While this mechanism is interesting in its own right, the welfare consequences of policy can be transmitted through other channels which can be obfuscated by price effects. To sidestep this, in what follows I net out price effects by subtracting from any value response the effects on the average birth value that arises from the change in prices.[23] For each policy I then consider two extreme cases: First, I consider a completely inelastic housing market where $\omega_n \to \infty$ for all n. Second, I consider a completely elastic housing market where $\omega_n = 0$ for all n.

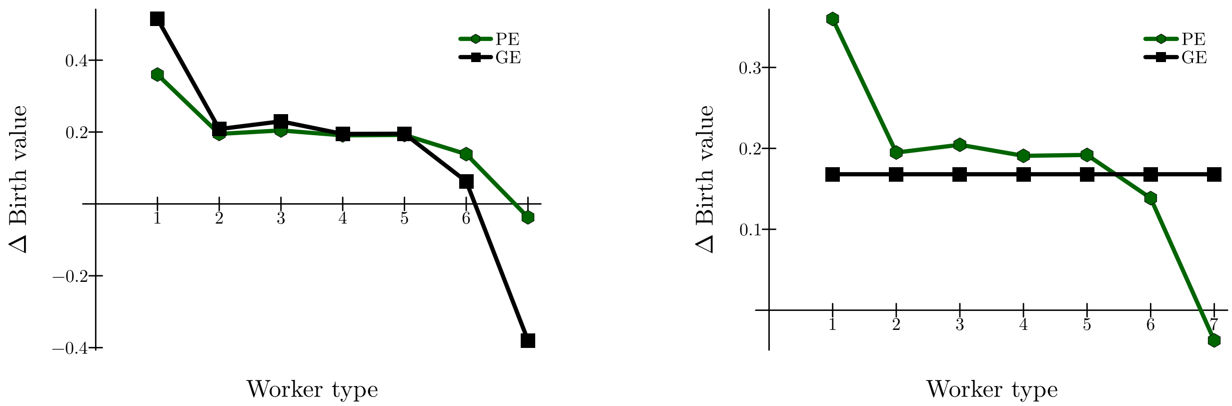
Regional subsidies

I now turn to analyzing regional subsidies. I consider a subsidy to workers living in poor regions, financed by a lump tax on all workers. I scale the responses to correspond to a daily subsidy to the poor region of 1 Euro, financed by a daily 35 cent tax on workers in

[22] ω_n is conceptually similar to the price elasticity of housing. However, since the level of house prices is undetermined in the model, the relevant object to govern the model's behavior is the slope of the supply curve, not its elasticity. In a slight abuse of terminology, I will sometimes refer to ω_n as the elasticity of housing in the remainder of the paper.

[23] From the logic outlined in appendix A.5, this is equivalent to choosing a vector of elasticities ω_n for which the change in expected birth value arising from price changes is zero.

rich regions. Figure 14 plots the changes in birth values for every worker type, separating out the two cases of elastic and inelastic housing markets. PE effects are identical across both and show the redistributive nature of such transfers: Low-type workers gain, while high-type workers lose. This is due to existing sorting. Low-type workers spend more time in poor regions and thus benefit more from the policy. Likewise, high-type workers are concentrated in rich markets and are less likely to reap the benefits of the policy.



(a) Elastic housing

(b) Inelastic housing

Figure 14: Changes in birth value, 1 Euro subsidy to poor regions

Next, consider the GE responses. In the inelastic case, prices adjust to keep allocations constant. Thus, inelastic housing markets *dampen* the GE effects of policy.[24] In the case of elastic housing however, things look quite different. Here, GE effects *amplify* the redistributive effects of policy. This is because these transfers have strong effects on the scale of each location. Under elastic housing, subsidizing poor regions increases the size of poor regions. By the mechanism illustrated in Figure 10b, this leads to decreases in firm sorting. Figure 15 shows this for the concrete policy in question: Vacancies are reallocated towards poorer regions, particularly those of high-skill firms. In other words, the strength of firm sorting across space decreases. From a welfare perspective, this reduced sorting is bad for high-skill workers, since it makes it harder for them to find work in rich regions. On the other hand, low-skill workers benefit since they have higher mobility costs and benefit from increased high-skill jobs in the poor region.

[24] GE effects do not fully net out to zero, since in the model workers are disproportionally born in poor regions and transition to rich regions later. Discounting thus leads them to value transfers to poorer regions since these kinds of transfers are temporally closer to the moment of birth.

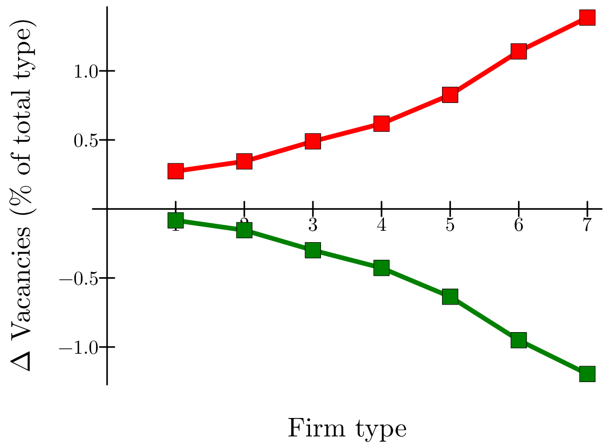


Figure 15: Changes in vacancies, 1 Euro subsidy to poor regions (inelastic housing)

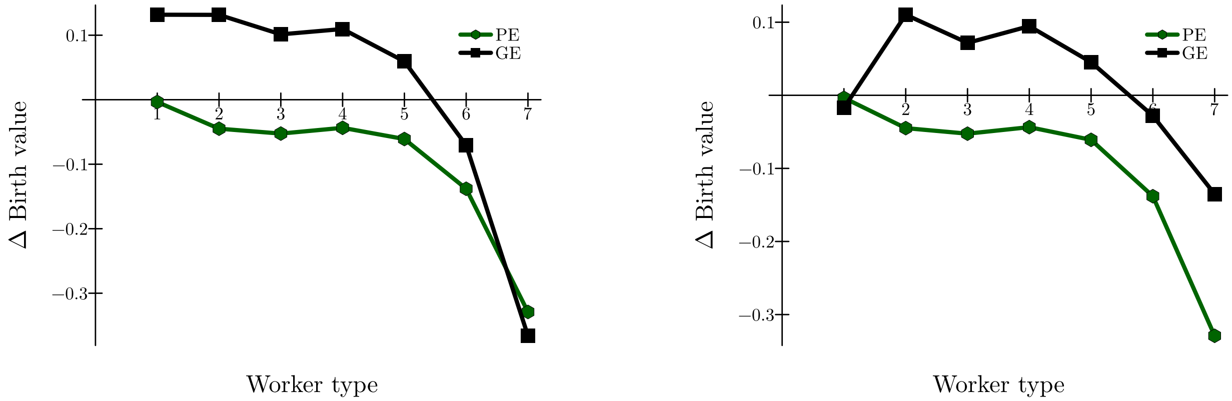
Job creation policy

Next, we turn to policies that affect the propensity of particular firm types to create jobs in poor versus rich regions. I consider pairs of job creation subsidies and taxes $(\tau_{m,1}, \tau_{m,2})$ that are budget-balanced, i.e. require no worker-side funding.[25]

Consider first subsidies for the most productive firm type, $m = 7$. Figure 16 depicts the birth value effects of a marginal change in $\tau_{7,1}$ (incentivizing job creation in poor regions) offset by a budget-balancing decrease in $\tau_{7,2}$ (disincentivizing job creation in rich regions). Responses are scaled to correspond to a one-time subsidy of 10000 Euros per job created in a poor region. This subsidy is financed by a tax of 2486.25 Euros per job created in rich regions. The figure illustrates that such policies are again highly redistributive: Low-type workers benefit somewhat while a few high-type workers experience substantial losses from the policy. This is true for both elastic and inelastic housing markets. Like in the case of regional subsidies, less elastic housing moderates the redistributive effects. However, the decrease in firm sorting caused by the policy change leads to progressive redistribution even under inelastic housing markets.

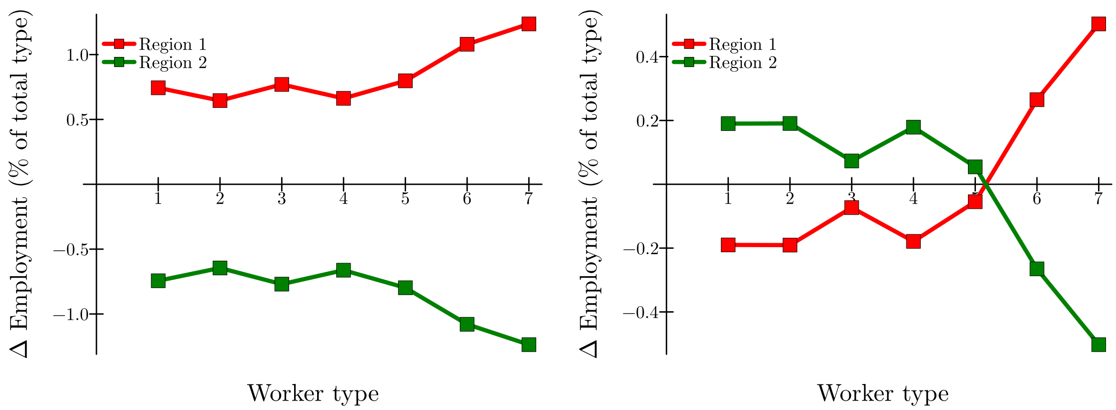
As Figure 17 demonstrates, when the housing supply is elastic, the policy again creates second round effects that reduce the size of the rich location, decrease firm sorting, and thereby generate losses for high-type workers at the benefit of their low-type peers.

[25]This approach has the advantage that job creation incentives cannot be employed to rectify inefficient *levels* of job creation, which are not the focus of this paper.



(a) Elastic housing (b) Inelastic housing

Figure 16: Changes in birth value, firm type 7 location incentive (10000 Euros)

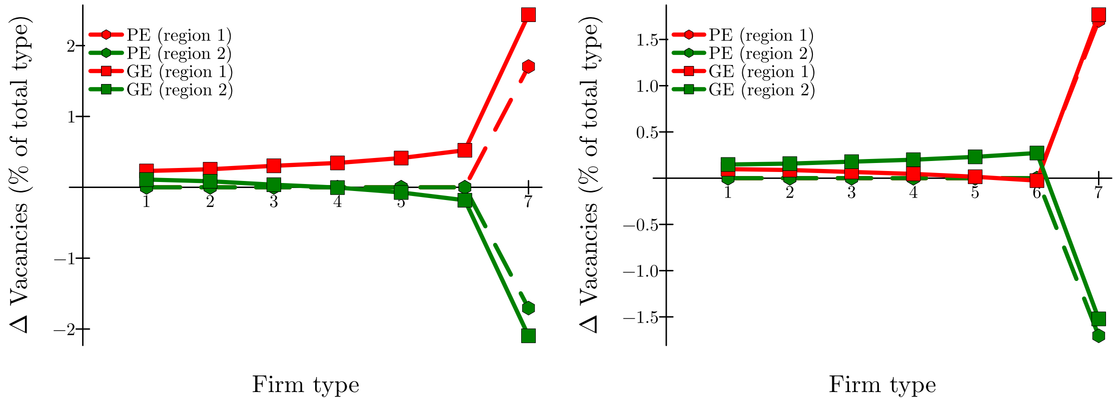


(a) Elastic housing (b) Inelastic housing

Figure 17: Changes in employment, firm type 7 location incentives (10000 Euros)

If housing is inelastic, the location size does not change and so there are no additional sorting effects which offsets some of these losses for the rich worker type. This can also be seen in Figure 18. The increased presence in high type vacancies below type 7 in the rich region moderates the losses incurred from the exodus of type 7 vacancies.

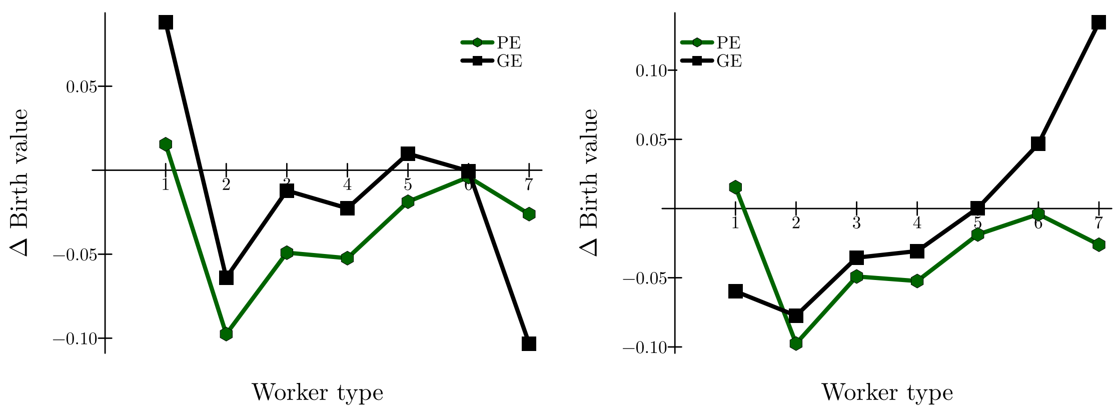
Lastly, I analyze subsidies for low-skill jobs, i.e. subsidies that incentivize firm type 1 to locate in poorer regions. The scale of the policy again corresponds to a 10000 Euro subsidy for jobs located in poor regions, financed by a 8987.01 Euro tax on type-1 jobs in rich regions. Figure 19 illustrates that the redistributive properties of this policy depend on the supply of housing. If housing is inelastic, such policy is progressive. However, if



(a) Elastic housing

(b) Inelastic housing

Figure 18: Changes in the vacancy distribution, firm type 7 location incentives (10000 Euros)

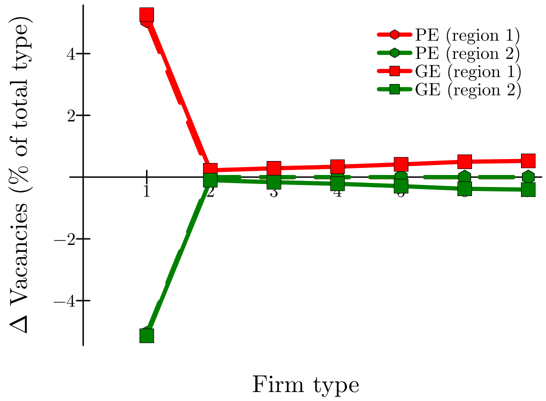


(a) Elastic housing

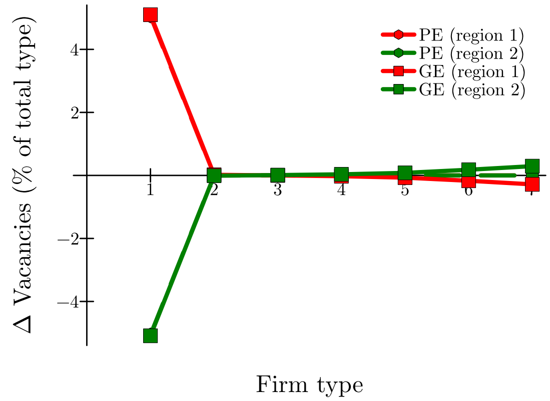
(b) Inelastic housing

Figure 19: Changes in birth value, firm type 1 location incentives (10000 Euros)

housing is elastic, the policy is highly regressive. This again is a result that stems directly from the effects on sorting. If housing supply constraints hold the location sizes constant, the policy causes an increase in firm sorting, as shown in Figure 20. This change happens to the benefit of high type workers but to the detriment of low type workers. Sorting changes are more ambiguous when housing is fully elastic and spatial sorting of firms actually decreases at the upper end of the firm distribution as the location size shrinks. Figure 21 also shows that the population of high worker types in the rich region increases when housing is inelastic but shrinks when it is elastic, which also explains part of the redistributive effects.

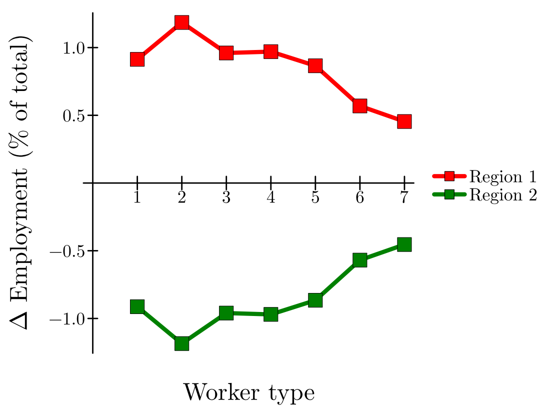


(a) Elastic housing

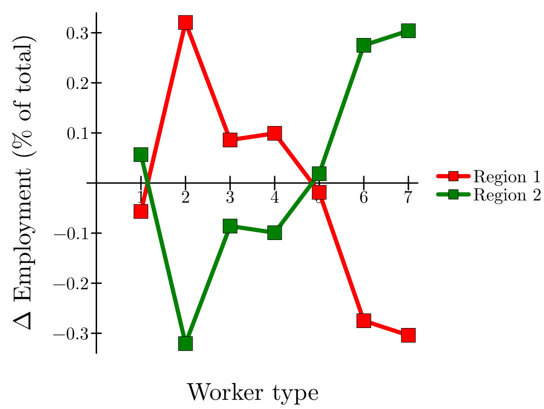


(b) Inelastic housing

Figure 20: Changes in the vacancy distribution, firm type 1 location incentives (10000 Euros)



(a) Elastic housing



(b) Inelastic housing

Figure 21: Changes in employment, firm type 1 location incentives (10000 Euros)

Overall, policies that make the poor location larger tend to be progressive and tend to reduce firm sorting. When the supply curve of housing constrains size effects, policies that increases firm sorting, such as policies incentivizing low-type firms to locate in the poor region type, are regressive. On the other hand, policies that decreases firm sorting, such as policies incentivizing high-skill jobs to move to poor regions, are progressive. Overall, such place-based policy has strong distributional effects that can only be measured once the heterogeneity of both workers and firms is taken into account.

7 Conclusion

This paper delivers an explanation for the rise of spatial inequality in developed countries and draws conclusions about the nature of spatial inequality, the drivers of spatial sorting, and the role of spatial policy. I find that spatial sorting of workers and firms plays an important role in linking aggregate inequality and spatial inequality. High-productivity firms and high-skill workers both sort disproportionally into locations that are rich and big. In the cross-section, firm sorting is the main reason why jobs in high-income locations pay higher wages. While cross-sectional sorting patterns have stayed approximately constant over time, technological change has produced spatially biased outcomes: Spatial inequality has risen and within-region inequality has risen particularly in rich regions.

The strong regional sorting patterns of firms imply that job ladders differ strongly for workers in different locations. This reveals that spatial inequality is quite important for workers when one considers forward-looking values, as opposed to a static view that focuses on the cross-sectional distribution of wages. Workers are willing to give up the equivalent of between 21% of average income if they could costlessly change locations while retaining their job.

The model can be used to study spatial policy and its redistributive effects. Spatial policy has large redistributive effects across worker types and often creates both winners and losers. The approach to spatial policy may thus depend on the normative preferences of the policymaker. Progressive policies generally set incentives that increase the location size of poor regions and reduce the degree of spatial firm sorting. A policymaker interested in redistributing from the rich to the poor can use spatial policy tools, such as subsidies to poor regions and subsidies to job creation in poor regions, particularly of highly productive firms. If housing is elastic, policies that increase the size of the poor location are generally progressive. However, subsidizing low-skill jobs in poor regions can be regressive when housing is inelastic, as it increases firm sorting.

The insights generated by this paper are possible because the model features a new combination of ingredients: A dynamic job ladder, two-sided heterogeneity and sorting of workers and firms across locations. We have seen that in such a model the effects of policies may depend on the ways in which firms and workers sort towards each other and across space. We have also seen that these sorting patterns cannot be understood in isolation; their interplay is important when analysing policy. Since firm sorting is a main quantitative driver of worker sorting, the analysis of spatial policies that target firms, such as location-based tax incentives, is incomplete without incorporating worker heterogeneity and sorting.

I leave a more complete analysis of these and other spatial policies to future research. This paper may serve as a first stepping stone towards such research and highlights some of the ways in which we might benefit from viewing spatial policies and spatial inequities through the lens of search models with two-sided heterogeneity and sorting. Future research may incorporate additional mechanisms that this paper abstracts from, such as learning on the job, trade, wage bargaining or productivity-enhancing agglomeration effects. Such extensions serve as exciting research directions that may further our understanding of the interplay between geographic sorting, dynamic labor markets, and policy.

References

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

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Bauluz, L., P. Bukowski, M. Fransham, A. Lee, M. López Forero, F. Novokmet, S. Breau, N. Lee, C. Malgouyres, M. Schularick, et al. (2023). Spatial wage inequality in north america and western europe: changes between and within local labour markets 1975-2019. Technical report, Kiel Working Paper.

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

Bezdek, J., R. Hathaway, R. Howard, C. Wilson, and M. Windham (1987). Local convergence analysis of a grouped variable version of coordinate descent. *Journal of Optimization Theory and Applications 54*, 471–477.

Bonhomme, S., T. Lamadon, and E. Manresa (2019). A distributional framework for matched employer employee data. *Econometrica 87*(3), 699–739.

Card, D., J. Rothstein, and M. Yi (2023). Location, location, location. Technical report, National Bureau of Economic Research.

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum (2022). Matching in cities. *Journal of the European Economic Association 20*(4), 1478–1521.

Diamond, R. and C. Gaubert (2022). Spatial sorting and inequality. *Annual Review of Economics 14*, 795–819.

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial sorting. *Journal of Political Economy 122*(3), 554–620.

Fajgelbaum, P. D. and C. Gaubert (2020). Optimal spatial policies, geography, and sorting. *The Quarterly Journal of Economics 135*(2), 959–1036.

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

Heise, S. and T. Porzio (2021). The aggregate and distributional effects of spatial frictions. Technical report, National Bureau of Economic Research.

Kennan, J. and J. R. Walker (2011). The effect of expected income on individual migration decisions. *Econometrica 79*(1), 211–251.

Kleinman, B. (2023). Wage inequality and the spatial expansion of firms. Technical report, Working Paper.

Kosfeld, R. and A. Werner (2012). Deutsche arbeitsmarktregionen–neuabgrenzung nach den kreisgebietsreformen 2007–2011. *Raumforschung und Raumordnung | Spatial Research and Planning 70*(1), 49–64.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2023). The anatomy of sorting: Evidence from danish data. *Working paper*.

Lindenlaub, I., R. Oh, and M. Peters (2022). Firm sorting and spatial inequality. Technical report, National Bureau of Economic Research.

Mion, G. and P. Naticchioni (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue canadienne d'économique 42*(1), 28–55.

Qiu, X. (2022). Vacant jobs. *Working Paper*.

Shimer, R. and L. Smith (2000). Assortative matching and search. *Econometrica 68*(2), 343–369.

Suárez Serrato, J. C. and O. Zidar (2016). Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review 106*(9), 2582–2624.

Williams, P. L. and R. D. Beer (2011). Generalized measures of information transfer. *arXiv preprint arXiv:1102.1507*.

A Appendix

A.1 LANCE: Algorithm

LANCE: Algorithm
Initialization Choose an initial guess of assignment functions $(a_1, \ldots, a_n)$.
Repeat steps (1)-(4) until termination.
(1) Record objective value $\mathcal{L}$.
(2) Solve $\max_\theta \mathcal{L}(\theta, a_1, \ldots, a_n)$ given the current guess for $a_1, \ldots, a_n$. Update θ.
(3) **For** $i \in (1, \ldots, n)$ Solve $\max_{a_i} \mathcal{L}(\theta, a_1, \ldots, a_n)$ subject to $a_i(s) = a_i(s') \quad \forall s, s' \in P^i_j$. This can be written as follows: **For** $j \in (1, \ldots, n)$ For all $s \in P^i_j$, set $a_i(s) = \text{argmax}_{\tau_i \in \mathfrak{T}_i} \sum_{s \in P^i_j} \sum_s g(x_s \vert \theta, \tau_i, a_{-i})$ **end** **end**
(4) Record objective value $\mathcal{L}'$.
(5) Terminate if $\mathcal{L}' = \mathcal{L}$. Otherwise, go to (1)

A.2 LANCE: Convergence theorem

Theorem (LANCE convergence): *Suppose that for any assignment functions $(a_1, \ldots, a_n)$, $\mathcal{L}(\theta, a_1, \ldots, a_n)$ has a unique maximum. Then the LANCE algorithm converges to a local maximum of $\mathcal{L}(\theta, a_1, \ldots, a_n)$ in a finite number of steps.*

Proof: By assumption, $\mathcal{L}$ assumes its maximum on Θ conditional on any assignment $(a_1, \ldots, a_n)$. For any assignment, let $\tilde{\Theta} = \{\text{argmax}_{\theta \in \Theta} \mathcal{L}(\theta, a_1, \ldots, a_n)\}_{a_1, \ldots, a_n}$ be the associated finite set of unique maximizers. Since both the set of possible assignment functions and the set $\tilde{\Theta}$ are finite, the set

$$\{|\mathcal{L}(\theta, a_1, \ldots, a_n) - \mathcal{L}(\theta', a'_1, \ldots, a'_n)|\}_{\theta \in \tilde{\Theta}, a_1, \ldots, a_n, \theta' \in \tilde{\Theta}, a'_1, \ldots, a'_n} \setminus \{0\}$$

is also finite and therefore assumes its minimum $\Delta > 0$, which is positive by construction. By construction of the algorithm, any time the algorithm reaches step (1), $\mathcal{L} \in \{\mathcal{L}(\theta, a_1, \ldots, a_n)\}_{\theta \in \tilde{\Theta}, a_1, \ldots, a_n}$. At step (4), the same holds true. Thus, either $|\mathcal{L}' - \mathcal{L}| = 0$

(covergence), or $|\mathcal{L}' - \mathcal{L}| \geq \Delta$. Since $\mathcal{L}$ is bounded above, the latter implies that convergence is reached after a finite number of steps. By construction, there cannot be a local improvement: θ maximizes $\mathcal{L}$ given $a_1, \ldots, a_n$ and vice versa. Thus, the algorithm converges to a local maximum in finite time. $\square$

A.3 Estimation of the mobility parameters

The mobility parameters are estimated with a strategy that essentially adapts the Bradley-Terry algorithm in Lentz et al. (2023) to a continuous time context and adds extra step in order to account for mobility costs. Using the notation

$$Q_{k,m,m',n,n',h,t} = \mathfrak{k}_{n,n',h} P_{k,m,m',n,n',h,t}$$
$$\mathfrak{k}_{n,n',h} = \begin{cases} (\mathfrak{N}_{n'} - \mathbb{I}(n = n')) & \text{if } h = 1 \\ \mathbb{I}(n = n') & \text{if } h = 0 \end{cases}$$
$$P_{k,m,m',n,n',0,t} = \frac{\tilde{W}_{k,m',n',t}}{\tilde{W}_{k,m,n,t} + \tilde{W}_{k,m',n',t}}$$
$$P_{k,m,m',n,n',1,t} = \frac{\tilde{c}_{k,t}\tilde{W}_{k,m',n',t}}{\tilde{W}_{k,m,n,t} + \tilde{c}_{k,t}\tilde{W}_{k,m',n',t}}$$
$$d_{k,m,m',n,n',h,t}(s) = \mathbb{1}\left(k(s) = k, m(s) = m, m'(s) = m', n(s) = n, n'(s) = n', t(s) = t\right)$$

and

$$\mu_{k,m,n} = \left(\sum_{m',n'} s^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t}\right) + \delta_{k,m}$$

we can write the mobility log likelihood as

$$\begin{aligned}
\mathcal{L}_{\text{Mobility}} &= \sum_{k,m,m',n,n',h,t} \sum_s d_{k,m,m',n,n',h,t}(s) \Bigg[\log\left(\mu_{k,m,n}\right) - \mu_{k,m,n} r_s + \\
&\quad \mathbb{1}(m' \neq 0) \log\left(\frac{s^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t}}{\mu_{k,m,n}}\right) + \mathbb{1}(m' = 0) \log\left(\frac{\delta_{k,m}}{\mu_{k,m,n}}\right) \Bigg] \\
&= \left(-\sum_{k,m,n} \sum_s r_s d_{k,m,n,t}(s) \mu_{k,m,n}\right) + \sum_{k,m,m',n,n',h,t} \sum_s d_{k,m,m',n,n',h,t}(s) \Bigg[\\
&\quad \mathbb{1}(m' \neq 0) \log\left(s^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t}\right) + \mathbb{1}(m' = 0) \log\left(\delta_{k,m}\right) \Bigg]
\end{aligned} \tag{15}$$

I maximize equation (15) as follows: Starting with a guess for

$$\theta = (\eta_{k,m'}, \zeta_{m,m'}, \lambda_{m',n',t}, \delta_{k,m}, \tilde{W}_{k,m,n,t}, \tilde{c}_{k,t}, s)$$

I optimize the likelihood by coordinate descent but use a minorize-maximization approach to optimize over $\tilde{W}_{k,m,n,t}, \tilde{c}_{k,t}$. Below are the FOCs that deliver the algorithm:

FOC for $\eta_{k,m'}$

First, establish

$$\begin{aligned}
&\sum_{k,m,n,t} \mu_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \\
&= \sum_{k,m,n,t} \left(\sum_{m',n',h} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} + \delta_{k,m} \right) \sum_s r_s d_{k,m,n,t}(s) \\
&= \sum_{k,m,n,t} \left(\sum_{m'} \eta_{k,m'} \sum_{n',h} s^{\mathbb{1}(m\neq 0)} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} + \delta_{k,m} \right) \sum_s r_s d_{k,m,n,t}(s) \\
\implies &\frac{\partial}{\partial \eta_{\hat{k},\hat{m}}} \left(\sum_{k,m,n,t} \mu_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \right) = \sum_{m,n,t} \left(\sum_{n',h} s^{\mathbb{1}(m\neq 0)} \zeta_{m,\hat{m}} \lambda_{\hat{m},n'} Q_{\hat{k},m,\hat{m},n,n',h,t} \right) \sum_s r_s d_{\hat{k},m,n,t}(s)
\end{aligned}$$

Then, the FOC is

$$\begin{aligned}
0 = &- \sum_{m,n,t} \left(\sum_{n',h} s^{\mathbb{1}(m\neq 0)} \zeta_{m,\hat{m}} \lambda_{\hat{m},n'} Q_{\hat{k},m,\hat{m},n,n',h,t} \right) \sum_s r_s d_{\hat{k},m,n,t}(s) \\
&+ \sum_{m,n,n',h,t} \sum_s d_{\hat{k},m,\hat{m},n,n',h,t}(s) \mathbb{1}(\hat{m} \neq 0) \eta_{\hat{k},\hat{m}}^{-1}
\end{aligned}$$

which implies

$$\eta_{\hat{k},\hat{m}} = \frac{\sum_{m,n,n',h,t} \sum_s d_{\hat{k},m,\hat{m},n,n',h,t}(s) \mathbb{1}(\hat{m} \neq 0)}{\sum_{m,n,t} \left(\sum_{n',h} s^{\mathbb{1}(m\neq 0)} \zeta_{m,\hat{m}} \lambda_{\hat{m},n'} Q_{\hat{k},m,\hat{m},n,n',h,t} \right) \sum_s r_s d_{\hat{k},m,n,t}(s)}$$

FOC for $\zeta_{m,m'}$

Similar to before, we have

$$\begin{aligned}
&\sum_{k,m,n,t} \mu_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \\
&= \sum_{k,m,n,t} \left(\sum_{m',n',h} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} + \delta_{k,m} \right) \sum_s r_s d_{k,m,n,t}(s) \\
&= \sum_{k,m,n,t} \left(\sum_{m'} \zeta_{m,m'} \sum_{n',h} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} + \delta_{k,m} \right) \sum_s r_s d_{k,m,n,t}(s) \\
\implies &\frac{\partial}{\partial \zeta_{\hat{m},\hat{m}'}} \left(\sum_{k,m,n,t} \mu_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \right) = \sum_{k,n,t} \left(\sum_{n',h} s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{k,\hat{m}'} \lambda_{\hat{m}',n',t} Q_{k,\hat{m},\hat{m}',n,n',h,t} \right) \sum_s r_s d_{\hat{k},m,n,t}(s)
\end{aligned}$$

The FOC is thus

$$\begin{aligned}
0 = &- \sum_{k,n,t} \left(\sum_{n',h} s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{k,\hat{m}'} \lambda_{\hat{m}',n',t} Q_{k,\hat{m},\hat{m}',n,n',h,t} \right) \sum_s r_s d_{\hat{k},m,n,t}(s) \\
&+ \sum_{k,n,n',h,t} \sum_s d_{k,\hat{m},\hat{m}',n,n',h,t}(s) \mathbb{1}(\hat{m}' \neq 0)(s) \zeta^{-1}_{\hat{m},\hat{m}'}
\end{aligned}$$

which implies

$$\zeta_{\hat{m},\hat{m}'} = \frac{\sum_{k,n,n',h,t} \sum_s d_{k,\hat{m},\hat{m}',n,n',h,t}(s) \mathbb{1}(\hat{m}' \neq 0)(s)}{\sum_{k,n,t} \left(\sum_{n',h} s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{k,\hat{m}'} \lambda_{\hat{m}',n',t} Q_{k,\hat{m},\hat{m}',n,n',h,t} \right) \sum_s r_s d_{\hat{k},m,n,t}(s)}$$

FOC for $\lambda_{m',n',t}$

Proceed as before:

$$\begin{aligned}
&\sum_{k,m,n,t} \mu_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \\
&= \sum_{k,m,n,h,t} \left(\sum_{m',n'} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} + \delta_{k,m} \right) \sum_s r_s d_{k,m,n,t}(s) \\
\implies &\frac{\partial}{\partial \lambda_{\hat{m},\hat{n},\hat{t}}} \left(\sum_{k,m,n} \mu_{k,m,n,\hat{t}} \sum_s r_s d_{k,m,n,\hat{t}}(s) \right) = \sum_{k,m,n,h} \left(s^{\mathbb{1}(m\neq 0)} \eta_{k,\hat{m}} \zeta_{m,\hat{m}} Q_{k,m,\hat{m},n,\hat{n},h,\hat{t}} \right) \sum_s r_s d_{k,m,n,\hat{t}}(s)
\end{aligned}$$

Then, the FOC is

$$
\begin{aligned}
0 = &- \sum_{k,m,n,h} \left(s^{\mathbb{1}(m \neq 0)} \eta_{k,\hat{m}} \zeta_{m,\hat{m}} Q_{k,m,\hat{m},n,\hat{n},h,t} \right) \sum_{s} r_s d_{k,m,n,\hat{t}}(s) \\
&+ \sum_{k,m,n,h} \sum_{s} d_{k,m,\hat{m},n,\hat{n},h,\hat{t}}(s) \mathbb{1}(m' \neq 0) \lambda^{-1}_{\hat{m},\hat{n},\hat{t}}
\end{aligned}
$$

which implies

$$
\lambda_{\hat{m},\hat{n},\hat{t}} = \frac{\sum_{k,m,n,h} \sum_s d_{k,m,\hat{m},n,\hat{n},h,\hat{t}}(s) \mathbb{1}(m' \neq 0)}{\sum_{k,m,n,h} \left(s^{\mathbb{1}(m \neq 0)} \eta_{k,\hat{m}} \zeta_{m,\hat{m}} Q_{k,m,\hat{m},n,\hat{n},h,\hat{t}} \right) \sum_s r_s d_{k,m,n,\hat{t}}(s)}
$$

FOC for $\delta_{k,m}$

Taning the FOC directly yields

$$
0 = - \sum_{n,t} \left(\sum_{s} r_s d_{\hat{k},\hat{m},n,t}(s) \right) + \sum_{m',n,n',h,t} \left(\sum_{s} d_{\hat{k},\hat{m},m',n,n',h,t}(s) \mathbb{1}(m' = 0) \right) \frac{1}{\delta_{\hat{k},\hat{m}}}
$$

and therefore the updating condition is given by

$$
\delta_{\hat{k},\hat{m}} = \frac{\sum_{n,n',h,t} \left(\sum_s d_{\hat{k},\hat{m},0,n,n',h,t}(s) \right)}{\sum_{n,t} \left(\sum_s r_s d_{\hat{k},\hat{m},n,t}(s) \right)}
$$

FOC for s

The FOC for s is given by

$$
\begin{aligned}
0 = &- \sum_{k,m \neq 0,n,t} \left(\sum_{s} r_s d_{k,m,n,t}(s) \right) \left(\sum_{m',n',h} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} \right) \\
&+ \sum_{k,m \neq 0,m',n,n',h,t} \left[\sum_{s} d_{k,m,m',n,n',h,t}(s) \mathbb{1}(m' \neq 0) \right] \frac{1}{s}
\end{aligned}
$$

and therefore the updating condition is

$$
s = \frac{\sum_{k,m \neq 0,m',n,n',h,t} \left[\sum_s d_{k,m,m',n,n',h,t}(s) \mathbb{1}(m' \neq 0) \right]}{\sum_{k,m \neq 0,n,t} \left(\sum_s r_s d_{k,m,n,t}(s) \right) \left(\sum_{m',n'} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} Q_{k,m,m',n,n',h,t} \right)}
$$

Minorizing the objective to get estimates for $\tilde{W}_{k,m,n,t}$ and $\tilde{c}_{k,t}$

Recall that the initial objective function is given by

$$
\begin{aligned}
&-\sum_{k,m,n,t}\sum_{s} r_s d_{k,m,n,t}(s)\left(\left(\sum_{m',n',h} s^{\mathbb{1}(m\neq 0)}\eta_{k,m'}\zeta_{m,m'}\lambda_{m',n',t}Q_{k,m,m',n,n',h,t}\right)+\delta_{k,m}\right)\\
&+\sum_{k,m,m',n,n',h,t}\sum_{s} d_{k,m,m',n,n',h,t}(s)\Bigg[\mathbb{1}(m'\neq 0)\log\left(s^{\mathbb{1}(m\neq 0)}\eta_{k,m'}\zeta_{m,m'}\lambda_{m',n',t}Q_{k,m,m',n,n',h,t}\right)\\
&+\mathbb{1}(m'=0)\log\left(\delta_{k,m}\right)\Bigg]
\end{aligned}
$$

Since here we are optimizing over $\tilde{W}_{k,m,n,t}$ and $\tilde{c}_{k,t}$ only, we can drop any terms that do not affect the derivative w.r.t. $Q_{k,m,m',n,n',h,t}$:

$$
\begin{aligned}
&\left(-\sum_{k,m,n,t}\sum_{s} r_s d_{k,m,n,t}(s)\left(\left(\sum_{m',n',h} s^{\mathbb{1}(m\neq 0)}\eta_{k,m'}\zeta_{m,m'}\lambda_{m',n',t}Q_{k,m,m',n,n',h,t}\right)\right)\right)\\
&+\sum_{k,m,m',n,n',h,t}\left[\sum_{s} d_{k,m,m',n,n',h,t}(s)\mathbb{1}(m'\neq 0)\right]\left[\log\left(Q_{k,m,m',n,n',h,t}\right)\right]
\end{aligned}
$$

Since the direct FOC is intractable, we proceed by using a minorization-maximization (MM) strategy.[26] MM algorithms are used to solve optimization problems over some function f for which the FOC is intractable. The idea of the MM algorithm is to construct, for any x_0, a function g_{x_0} with $g_{x_0}(x_0)=f(x_0), g_{x_0}(x)\leq f(x)\forall x\neq x_0$ where the FOC of g is tractable. If one can find such a function, it becomes possible to construct an iterative algorithm that maximizes g in place of f and then uses the resulting optimizer x_1 as a new starting guess, iterating until convergence. Implementing this approach in my case, I start by using the fact that for $0<x<1, a\in\mathbb{R}_+$:

$$
-xa \geq (1-\bar{x})a\log\left(\frac{1-x}{1-\bar{x}}\right)-\bar{x}a
$$

[26] This idea is adapted from Lentz et al. (2023) who use a corresponding strategy in the discrete time case, which requires a different minorizing function.

with equality when $x = \bar{x}$. Again, dropping constant terms, we can use this to maximize

$$\sum_{k,m,n,t} \left[\sum_s r_s d_{k,m,n,t}(s) \right] \Bigg(\sum_{m',n',h} s^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \cdot$$
$$\mathfrak{t}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n',h,t}\right) \log\left(1 - P_{k,m,m',n,n',h,t}\right) \Bigg)$$
$$+ \sum_{k,m,m',n,n',h,t} \left[\sum_s d_{k,m,m',n,n',h,t}(s) \mathbb{1}(m' \neq 0) \right] \left[\log\left(P_{k,m,m',n,n',h,t} \right) \right]$$

in place of the initial objective. Writing out the probabilities yields

$$\sum_{k,m,n,t} \left[\sum_s r_s d_{k,m,n,t}(s) \right] \Bigg(\sum_{m',n',h} s^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \cdot$$
$$\left(\mathfrak{t}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n',h,t}\right) \log\left(\frac{\tilde{W}_{k,m,n,t}}{\tilde{W}_{k,m,n,t} + \tilde{c}_l^h \tilde{W}_{k,m',n',t}} \right) \right) \Bigg)$$
$$+ \sum_{k,m,m',n,n',h,t} \left[\sum_s d_{k,m,m',n,n',h,t}(s) \mathbb{1}(m' \neq 0) \right] \cdot \left[\log\left(\frac{\tilde{c}_l^h \tilde{W}_{k,m',n',t}}{\tilde{W}_{k,m,n,t} + \tilde{c}_l^h \tilde{W}_{k,m',n',t}} \right) \right]$$

Then, we use the trick from Hunter (2004) to minorize this function once more, using the following fact:

$$-\log(x) \geq 1 - \log(\bar{x}) - \frac{x}{\bar{x}}$$

with equality when $x = \bar{x}$. Applying this to our problem, we get as the final objective

$$\sum_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \Bigg[\sum_{m',n',h} s^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \mathfrak{t}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n'}\right) \cdot$$
$$\left(\log\left(\tilde{W}_{k,m,n,t} \right) - \frac{\tilde{W}_{k,m,n,t} + \tilde{c}_{k,t}^h \tilde{W}_{k,m',n',t}}{\bar{V}_{k,m,n,t} + \bar{C}_{k,t}^h \bar{V}_{k,m',n',t}} \right) \Bigg]$$
$$+ \sum_{k,m,m',n,n',h,t} \sum_s d_{k,m,m',n,n'}(s) \left[\mathbb{1}(m' \neq 0) \cdot \left(\log\left(\tilde{c}_{k,t}^h \tilde{W}_{k,m',n',t} \right) - \frac{\tilde{W}_{k,m,n,t} + \tilde{c}_{k,t}^h \tilde{W}_{k,m',n',t}}{\bar{V}_{k,m,n,t} + \bar{C}_{k,t}^h \bar{V}_{k,m',n',t}} \right) \right]$$

Taking the FOC with respect to $\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}$ yields:

$$
\begin{aligned}
0 = \sum_{k,m,n,t} \left[\sum_s r_s d_{k,m,n,t}(s)\right] & \left[\sum_{m',n',h,t} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \mathfrak{t}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n',h,t}\right) \right. \\
& \left(\mathbb{1}(k=\hat{k}, m=\hat{m}, n=\hat{n}, t=\hat{t}) \left(\frac{1}{\tilde{W}_{k,m,n,t}} - \frac{1}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right. \\
& \left. \left. - \mathbb{1}(k=\hat{k}, m'=\hat{m}, n'=\hat{n}, t=\hat{t}) \left(\frac{\tilde{c}^h_{k,t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right) \right] \\
& + \sum_{k,m,m',n,n',h,t} \left[\sum_s d_{k,m,m',n,n',h,t}(s) \cdot \mathbb{1}(m' \neq 0) \right] \cdot \Bigg[\\
& \mathbb{1}(k=\hat{k}, m'=\hat{m}, n'=\hat{n}, t=\hat{t}) \left(\frac{1}{\tilde{W}_{k,m',n',t}} - \frac{\tilde{c}^h_{k,t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \\
& - \mathbb{1}(k=\hat{k}, m=\hat{m}, n=\hat{n}, t=\hat{t}) \left(\frac{1}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \Bigg]
\end{aligned}
$$

This can be simplified to

$$
\begin{aligned}
0 = \sum_{m',n',h} & \left[\sum_s r_s d_{\hat{k},\hat{m},\hat{n},\hat{t}}(s)\right] s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{\hat{k},m'} \zeta_{\hat{m},m'} \lambda_{m',n',\hat{t}} \mathfrak{t}_{\hat{n},n',h} \left(1 - \bar{P}_{\hat{k},\hat{m},m',\hat{n},n',h,\hat{t}}\right) \cdot \\
& \left(\frac{1}{\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}} - \frac{1}{\bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m',n',\hat{t}}} \right) \\
& - \sum_{m,n,h} \left[\sum_s r_s d_{\hat{k},m,n,\hat{t}}(s)\right] s^{\mathbb{1}(m\neq 0)} \eta_{\hat{k},\hat{m}} \zeta_{m,\hat{m}} \lambda_{\hat{m},\hat{n},\hat{t}} \mathfrak{t}_{n,\hat{n},h} \left(1 - \bar{P}_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}\right) \cdot \\
& \left(\frac{\tilde{c}^h_{\hat{k},\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}}} \right) \\
& + \sum_{m,n,h} \left[\sum_s d_{\hat{k},m,\hat{m},n,\hat{n},\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0) \right] \cdot \left(\frac{1}{\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}} - \frac{\tilde{c}^h_{\hat{k},\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}}} \right) \\
& - \sum_{m',n',h,\hat{t}} \left[\sum_s d_{\hat{k},\hat{m},m',\hat{n},n',\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0) \right] \cdot \left(\frac{1}{\bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m',n',\hat{t}}} \right)
\end{aligned}
$$

and then to

$$
\begin{aligned}
0 = \sum_{m,n,h} \Bigg\{ & \left[\sum_s r_s d_{\hat{k},\hat{m},\hat{n},\hat{t}}(s)\right] s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{\hat{k},m} \zeta_{\hat{m},m} \lambda_{m,n,\hat{t}} \mathfrak{t}_{\hat{n},n,h} \left(1 - \bar{P}_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}\right) \cdot \\
& \left(\frac{1}{\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}} - \frac{1}{\bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m,n,\hat{t}}}\right) \\
& - \left[\sum_s r_s d_{\hat{k},m,n,\hat{t}}(s)\right] s^{\mathbb{1}(m\neq 0)} \eta_{\hat{k},\hat{m}} \zeta_{m,\hat{m}} \lambda_{\hat{m},\hat{n},\hat{t}} \mathfrak{t}_{n,\hat{n},h} \left(1 - \bar{P}_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}\right) \cdot \\
& \left(\frac{\tilde{c}^h_{\hat{k},\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}}}\right) \\
& + \left[\sum_s d_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0)\right] \cdot \left(\frac{1}{\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}} - \frac{\tilde{c}^h_{\hat{k},\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}}}\right) \\
& - \left[\sum_s d_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0)\right] \cdot \left(\frac{1}{\bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m,n,\hat{t}}}\right) \Bigg\}
\end{aligned}
$$

Separating out $\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}$ yields

$$
\begin{aligned}
& \frac{1}{\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}}} \Bigg(\sum_{m,n,h} \left[\sum_s r_s d_{\hat{k},\hat{m},\hat{n},\hat{t}}(s)\right] s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{\hat{k},m} \zeta_{\hat{m},m} \lambda_{m,n,\hat{t}} \mathfrak{t}_{\hat{n},n,h} \left(1 - \bar{P}_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}\right) \\
& + \left[\sum_s d_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}(s) \cdot \mathbb{1}(\hat{m} \neq 0)(s)\right] \Bigg) \\
= \sum_{m,n,h} \Bigg\{ & \Bigg(\left[\sum_s r_s d_{\hat{k},\hat{m},\hat{n},\hat{t}}(s)\right] s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{\hat{k},m} \zeta_{\hat{m},m} \lambda_{m,n,\hat{t}} \mathfrak{t}_{\hat{n},n,h} \left(1 - \bar{P}_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}\right) \\
& + \left[\sum_s d_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0)\right] \Bigg) \cdot \left(\frac{1}{\bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m,n,\hat{t}}}\right) \\
& + \Bigg(\left[\sum_s r_s d_{\hat{k},m,n,\hat{t}}(s)\right] s^{\mathbb{1}(m\neq 0)} \eta_{\hat{k},\hat{m}} \zeta_{m,\hat{m}} \lambda_{\hat{m},\hat{n},\hat{t}} \mathfrak{t}_{n,\hat{n},h} \left(1 - \bar{P}_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}\right) \\
& + \left[\sum_s d_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}(s) \cdot \mathbb{1}(\hat{m} \neq 0)(s)\right] \Bigg) \cdot \left(\frac{\tilde{c}^h_{\hat{k},\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}}}\right) \Bigg\}
\end{aligned}
$$

which finally gives us the following updating condition:

$$
\begin{aligned}
\tilde{W}_{\hat{k},\hat{m},\hat{n},\hat{t}} = & \left(\sum_{m,n,h} \left[\sum_s r_s d_{\hat{k},\hat{m},\hat{n},\hat{t}}(s) \right] s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{\hat{k},m} \zeta_{\hat{m},m} \lambda_{m,n,\hat{t}} \mathfrak{k}_{\hat{n},n,h} \left(1 - \bar{P}_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}\right) \right. \\
& \left. + \left[\sum_s d_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}(s) \cdot \mathbb{1}(\hat{m} \neq 0)(s) \right] \right) \cdot \\
& \left[\sum_{m,n,h} \left\{ \left(\left[\sum_s r_s d_{\hat{k},\hat{m},\hat{n},\hat{t}}(s) \right] s^{\mathbb{1}(\hat{m}\neq 0)} \eta_{\hat{k},m} \zeta_{\hat{m},m} \lambda_{m,n,\hat{t}} \mathfrak{k}_{\hat{n},n,h} \left(1 - \bar{P}_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}\right) \right. \right. \right. \\
& \left. + \left[\sum_s d_{\hat{k},\hat{m},m,\hat{n},n,h,\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0) \right] \right) \cdot \left(\frac{1}{\bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m,n,\hat{t}}} \right) \\
& + \left(\left[\sum_s r_s d_{\hat{k},m,n,\hat{t}}(s) \right] s^{\mathbb{1}(m\neq 0)} \eta_{\hat{k},\hat{m}} \zeta_{m,\hat{m}} \lambda_{\hat{m},\hat{n},\hat{t}} \mathfrak{k}_{n,\hat{n},h} \left(1 - \bar{P}_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}\right) \right. \\
& \left. \left. \left. + \left[\sum_s d_{\hat{k},m,\hat{m},n,\hat{n},h,\hat{t}}(s) \cdot \mathbb{1}(\hat{m} \neq 0)(s) \right] \right) \cdot \left(\frac{\tilde{c}^h_{\hat{k},\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}^h_{\hat{k},\hat{t}} \bar{V}_{\hat{k},\hat{m},\hat{n},\hat{t}}} \right) \right\} \right]^{-1}
\end{aligned}
$$

Now, turn to the mobility cost $\tilde{c}_{k,\hat{t}}$. Recall the objective that we want to maximize:

$$
\begin{aligned}
& \sum_{k,m,n,t} \sum_s r_s d_{k,m,n,t}(s) \left[\sum_{m',n',h} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \mathfrak{k}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n'}\right) \cdot \right. \\
& \left. \left(\log\left(\tilde{W}_{k,m,n,t}\right) - \frac{\tilde{W}_{k,m,n,t} + \tilde{c}^h_{k,t} \tilde{W}_{k,m',n',t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right] \\
& + \sum_{k,m,m',n,n',h,t} \sum_s d_{k,m,m',n,n'}(s) \left[\mathbb{1}(m' \neq 0) \cdot \left(\log\left(\tilde{c}^h_{k,t} \tilde{W}_{k,m',n',t}\right) - \frac{\tilde{W}_{k,m,n,t} + \tilde{c}^h_{k,t} \tilde{W}_{k,m',n',t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right]
\end{aligned}
$$

Taking the FOC with respect to $\tilde{c}_{\hat{k},\hat{t}}$ yields:

$$
\begin{aligned}
0 = \sum_{k,m,n,t} & \left[\sum_s r_s d_{k,m,n,t}(s)\right] \left[\sum_{m',n',h} s^{\mathbb{1}(m\neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \mathfrak{t}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n',h,t}\right) \cdot \right. \\
& \left. \mathbb{1}(k=\hat{k}, h=1, t=\hat{t}) \left(-\frac{\tilde{W}_{k,m',n',t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}}\right)\right] \\
& + \sum_{k,m,m',n,n',h,t} \left[\sum_s d_{k,m,m',n,n',h,t}(s) \cdot \mathbb{1}(m' \neq 0)\right] \cdot \\
& \left[\mathbb{1}(k=\hat{k}, h=1, t=\hat{t}) \left(\frac{1}{\tilde{c}_{k,t}} - \frac{\tilde{W}_{k,m',n',t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}}\right)\right]
\end{aligned}
$$

which can be rewritten as

$$
\begin{aligned}
0 = \sum_{m,m',n,n'} & \left[\sum_s r_s d_{\hat{k},m,n,\hat{t}}(s)\right] \left[s^{\mathbb{1}(m\neq 0)} \eta_{\hat{k},m'} \zeta_{m,m'} \lambda_{m',n',\hat{t}} \mathfrak{t}_{n,n',1} \left(1 - \bar{P}_{\hat{k},m,m',n,n',1,\hat{t}}\right) \cdot \right. \\
& \left. \left(-\frac{\tilde{W}_{\hat{k},m',n',\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m',n',\hat{t}}}\right)\right] \\
& + \sum_{m,m',n,n'} \left[\sum_s d_{\hat{k},m,m',n,n',1,\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0)\right] \cdot \left[\left(\frac{1}{\tilde{c}_{\hat{k},\hat{t}}} - \frac{\tilde{W}_{\hat{k},m',n',\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m',n',\hat{t}}}\right)\right]
\end{aligned}
$$

and thus yields

$$
\begin{aligned}
\tilde{c}_{\hat{k},\hat{t}} = & \left(\sum_{m,m',n,n'} \left[\sum_s d_{\hat{k},m,m',n,n',1,\hat{t}}(s) \cdot \mathbb{1}(m' \neq 0)\right]\right) \cdot \\
& \left[\sum_{m,m',n,n'} \left[\sum_s r_s d_{\hat{k},m,n,\hat{t}}(s)\right] \left[s^{\mathbb{1}(m\neq 0)} \eta_{\hat{k},m'} \zeta_{m,m'} \lambda_{m',n',\hat{t}} \mathfrak{t}_{n,n',1} \left(1 - \bar{P}_{\hat{k},m,m',n,n',1,\hat{t}}\right) \cdot \right.\right. \\
& \left.\left. \left(\frac{\tilde{W}_{\hat{k},m',n',\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m',n',\hat{t}}}\right)\right] + \left[\sum_s d_{\hat{k},m,m',n,n',1}(s) \cdot \mathbb{1}(m' \neq 0)\right] \cdot \left[\frac{\tilde{W}_{\hat{k},m',n',\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}} + \bar{C}_{\hat{k},\hat{t}} \bar{V}_{\hat{k},m',n',\hat{t}}}\right]\right]^{-1}
\end{aligned}
$$

A.4 Discussion of the welfare criterion

In the model, workers are born in some locations and potentially transition throughout different locations during their lifetime. Using the mean worker value as a criterion incentivizes the planner to redistribute from states workers reach early to states that they are likely to reach later in life. As an example, consider an OLG model where every worker lives for two periods and receives an exogenous income stream of 0 in both periods. Obviously, the value function of the worker is 0 in the beginning of both periods. However, now consider a planner who redistributes an income of 1 from the young to the old in every period. Assume no discounting. A worker in the beginning of period 1 will receive a net income stream of (-1,1), so their value is 0. A worker who enters period 2 however receives an expected income stream of 1. Thus, with this simple transfer, the planner has increased the mean value in the economy by 1/2.

While this example is highly stylized, it illustrates the problem with taking the mean forward looking value of workers in the economy as a welfare measure. In the previous example, it is hard to argue that the planner has truly improved outcomes in the economy. This paper thus takes a more Rawlsian view of welfare and asks how policy measures affect the value of newborn workers. This resolves the paradox from the previous example: The value of newborn workers is 0 both with and without the intervention. However, with discounting, it now incentivizes the planner to distribute towards states the worker is likely to reach earlier in life. This leads to the positive GE effect of the policy in the inelastic housing case displayed in Figure 14, as most workers are likely to reach the poor region earlier in life and transition to the rich region later. This is important to keep in mind when interpreting the results.

A.5 Observational equivalence of different house price elasiticies

Let N=2 and let $d\tau$ be an arbitrary policy change. Let $\overrightarrow{\omega} = (\omega_1, \omega_2)$ be the slope coefficients of the house price supply curve. Let dv be the change in values arising from the policy change.

Claim: For any $c \in \mathbb{R}$, there exists a vector (ω_1', ω_2') that generates the same allocational changes (i.e. de and $d\mu$ are identical) but changes values by $dv + c$.

Proof: It is easy to verify from the model equations that a uniform price increase by $\frac{c}{\rho+r}$ in both markets raises worker values in all states by 1 and leaves all allocations constant. Let dp be the price change implied by $d\tau$. Now, let $\omega_n' = \omega_n \frac{dp_n}{dp_n - \frac{c}{\rho+r}}$. Then $de_n = \omega_n'(dp_n - \frac{c}{\rho+r})$ which leaves all model equations satisfied for a price change of $dp_n - \frac{c}{\rho+r}$ which raises all values by c.

A.6 Discussion of matching assumptions

In the model, I make two important assumptions on matching: First, matches are produced by a global matching function which also implies that search is random, not directed. Second, cross-regional transitions are moderated by mobility costs only, not by search frictions. Here, I provide a short discussion of both assumptions.

Let me begin by discussing the latter assumption, which also yields insight into the former. Relaxing the assumption that cross-regional transitions are moderated by mobility costs only, an alternative assumption would be that cross-regional transitions can be moderated by search frictions, e.g. because some of the matches that are created across regions are destroyed. This would not be a technical problem in the context of the model - we would simply need to include an additional term $x^{\mathbb{1}(l \neq l)}$ in equation (4) and in the corresponding equations on the firm side. However, separately quantifying x and the mobility cost term c_k is difficult in the data, as both terms have very similar observational implications, especially given the fact that values are also unobserved in the estimation. Consequently, allowing for both terms in the likelihood maximization step means that the algorithm does not converge. While one could thus easily impose the assumption that cross-regional mobility is moderated *only* by search frictions, doing so would eliminate any mechanisms by which changing incentives for cross-regional transitions (e.g. because of higher value differentials) would translate into changing rates of cross-regional transitions. I thus opt to attribute these frictions to costs, which allows workers to make economic trade-offs when deciding whether to transition from one region to another.

Second, the assumption that search is random is useful for similar reasons in the context of measurement. If one allows for freely directed search, the search intensity of a worker of type k in some local market l' depends on their current state (m, l). This means that equation (4) contains an additional term with indices k, m, l, l'. When taking this equation to the data, we would thus need to include a free term with these indices. This is problematic for two reasons. First, as mentioned before, separately quantifying this term and the mobility cost is a difficult problem. Second, this approach would likely worsen the empirical fit of the model, since we would attribute all worker-firm sorting to this endogenous object of the model ($\eta_{l,m'}$ and $\zeta_{m,m'}$ would not be separately identified from this term and would need to be dropped). Considering that the model can only serve its intended objectives if it can replicate sorting patterns along all 3 dimensions, I therefore leave the problem of directed search to future work.

A.7 Construction of counterfactual value functions

To net out the mobility cost from equation 3, one can solve the following fixed point equation:

$$
\begin{aligned}
&(r+\delta_{k,m}+\rho)W_{k,m,l}^{\text{net mc}} = w_{k,m,l} + a_l - p_l + \delta_{k,m}W_{k,0,l} \\
&+ \sum_{m'\neq 0,l'} s^{m\neq 0}\lambda \bar{v}_{k,m',l'}\zeta_{m,m'}\left[\sigma^w \log\left(\exp\left(\frac{W_{k,m,l}}{\sigma^w}\right) + \exp\left(\frac{W_{k,m',l'} - \mathbb{1}(l\neq l')c_k}{\sigma^w}\right)\right)\right. \\
&\left. + P_{k,m,m',l,l'}\mathbb{1}(l\neq l')c_k - W_{k,m,l}\right]
\end{aligned}
$$

which corresponds to the value a worker would get if they did not adjust their transition decisions, but would receive a compensation equal to the mobility cost upon every location switch.

To further net out net amenities, we simply need to remove them from the same equation:

$$
\begin{aligned}
&(r+\delta_{k,m}+\rho)W_{k,m,l}^{\text{net mc,am}} = w_{k,m,l} + \delta_{k,m}W_{k,0,l} \\
&+ \sum_{m'\neq 0,l'} s^{m\neq 0}\lambda \bar{v}_{k,m',l'}\zeta_{m,m'}\left[\sigma^w \log\left(\exp\left(\frac{W_{k,m,l}}{\sigma^w}\right) + \exp\left(\frac{W_{k,m',l'} - \mathbb{1}(l\neq l')c_k}{\sigma^w}\right)\right)\right. \\
&\left. + P_{k,m,m',l,l'}\mathbb{1}(l\neq l')c_k - W_{k,m,l}\right]
\end{aligned}
$$

A.8 Additional Figures

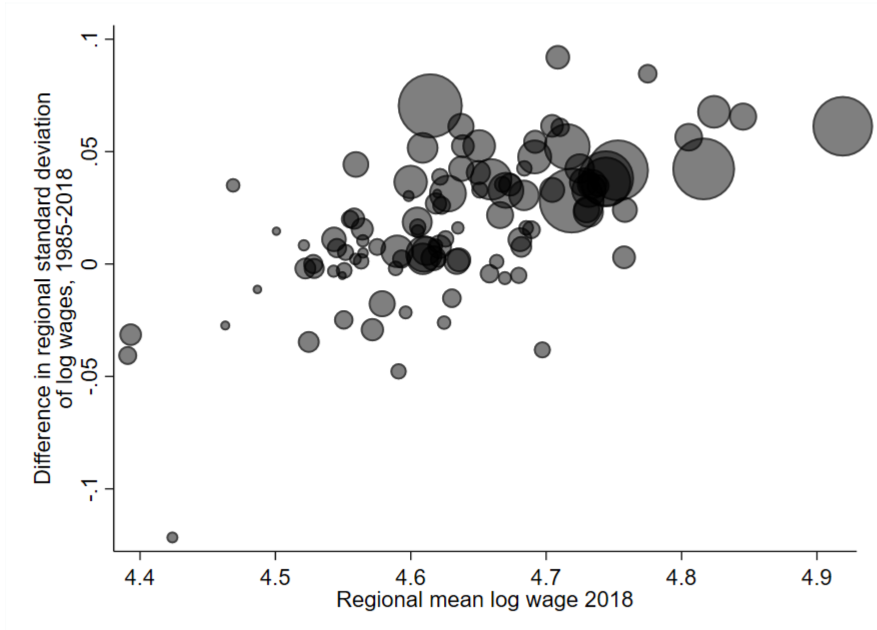


Figure 22: Inequality changes by 2018 regional income

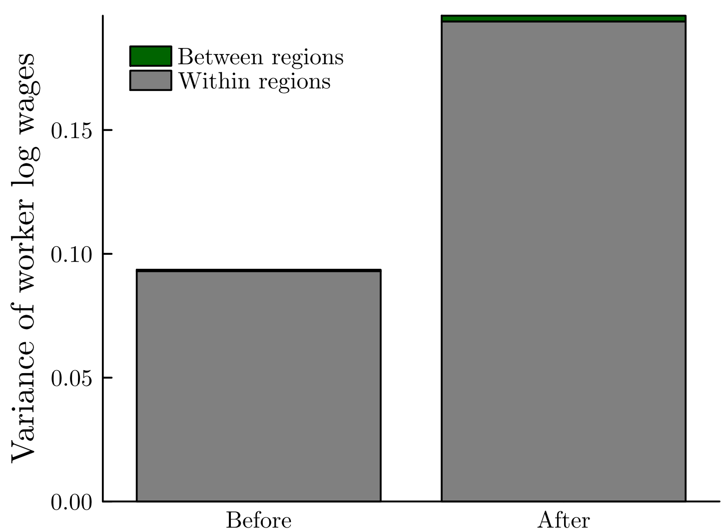


Figure 23: Variance decomposition of log wages

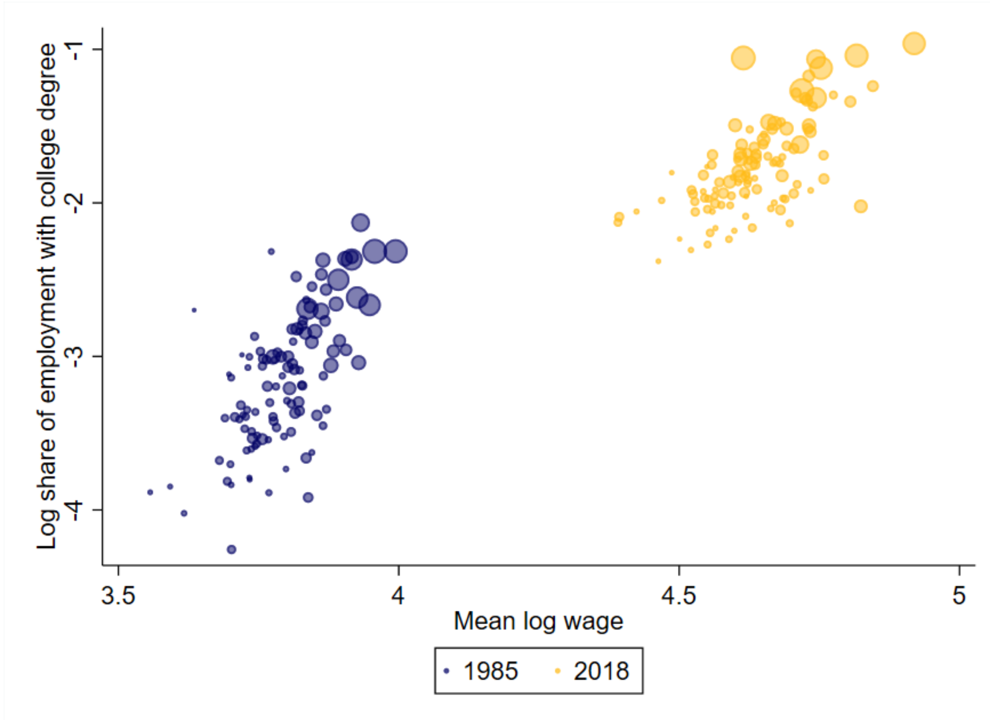


Figure 24: Worker sorting on observables

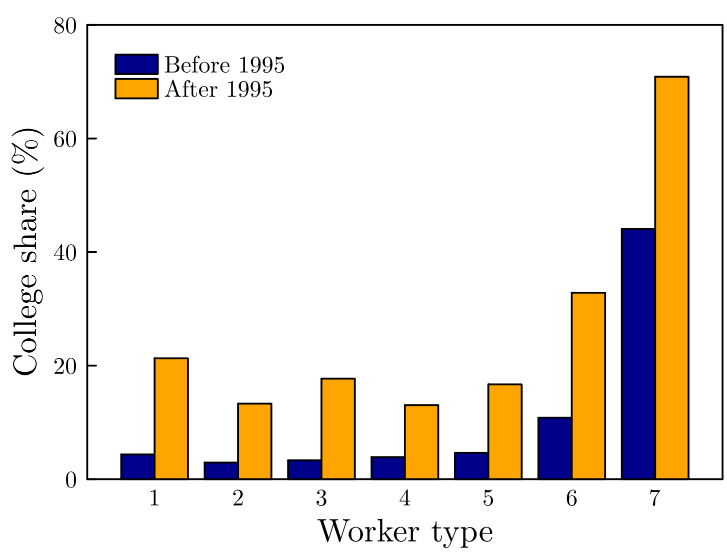


Figure 25: Mean education by worker type