Spatial Sorting and the Rise of Geographic Inequality

[Click here for the latest version]

Lukas Mann*

October 16, 2023

Abstract

I show that unobserved sorting patterns of firms and workers across space can account for the tight link of 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 way to recover these objects empirically by developing a novel bi-clustering method and uses these results to structurally estimate a dynamic spatial search model with two-sided sorting in which policy simulations can be conducted. 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 - on average 21 percent of mean income. The model also 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. Inequality has risen along at least two dimensions: First, as a long line of research has emphasized,

*Department of Economics, Princeton University. I thank my advisor, Gianluca Violante, as well as John Grigsby and Richard Rogerson for their invaluable comments 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, 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.

aggregate wage inequality has seen a marked rise in many countries. Second, inequality has also grown along the geographic dimension. Recent 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 the US, the UK, in Canada, and in Germany (Bauluz et al., 2023).

This development throws up both a positive and a normative question. In a positive sense, understanding the origins and implications of increasing spatial disparities is important because policymakers and society care about spatial inequality: Geographic income gaps create social tensions across regions and, in absence of factor mobility, can be a symptom of misallocation.[1] Understanding the origins of a rise in such disparities is therefore an important research question on its own.

In addition, the origins of spatial disparities also have strong normative implications: The question how much, if at all, policymakers and society *should* care about spatial inequality depends fundamentally on the drivers of spatial inequality. To see this, consider two worlds. In world one, workers are homogeneous but firms are heterogeneous and pay constant wages according to their productivity irrespective of location. In this world, spatial inequality is a symptom of productive plants concentrating in rich regions. Here, the labor market outcomes of geographically immobile workers depend crucially on the place of their birth. Spatial policy that reallocates workers has strong consequences for equity and might affect efficiency. Now consider world two: Firms are homogeneous, but workers are heterogeneous and get paid constant wages irrespective of location. In this world, spatial inequality arises because productive workers concentrate in rich regions. However, moving workers around space is of no fundamental consequence to worker outcomes. Spatial policy is futile.

This paper thus answers three related questions. First: Why have we seen spatial inequality rise in developed economies? Second: Why, if at all, should we care about spatial inequality? Third: What do the drivers of spatial inequality imply for policy? To

[1] A large set of policies in multiple countries is explicitly designed with the aim of closing such 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.

answer these questions, I frame the issue in terms of the productive units of the economy - workers and firms[2]. I build a structural model of spatial search in which there is two-sided sorting: Both workers and firms make location decisions. This model enables us to think about the positive and normative aspects of spatial inequality as well as policy implications.

In contrast to much previous work on spatial sorting, the model is structural and can be estimated directly on the data. Conditional on a type-classification of workers and firms in the data, data on worker mobility and wages identifies key parameters of the model. However, existing methods for classification of workers and firms are not well-suited for this task, as key assumptions of those methods are violated in a spatial setting. I therefore contribute a new classification method in the form of a bi-clustering algorithm that I call LANCE. LANCE generalizes the two-way fixed effect regression of Abowd et al. (1999) in a way that does not impose log-additivity of the wage function. The method is similar in spirit to recent work by Bonhomme et al. (2019) and Lentz et al. (2023) but, unlike those papers, identifies workers and firms based on static signals instead of mobility patterns. This requires imposing permanent types but makes the algorithm fast for large data sets and suitable for a spatial setting, in which the assumptions of Bonhomme et al. (2019) are not satisfied. LANCE is capable of jointly classifying workers and firms and of estimating the type-conditional wage distribution of workers and firms flexibly. This yields direct estimates for sorting patterns of workers and firms and allows me to test the sorting implications of the model against the data. I use the model to explore the drivers of spatial sorting, to assess the importance of spatial inequality for workers at different points of the wage distribution and across geographic locations, and to consider the redistributive effects of spatial policy.

Under two-sided sorting, three different mechanisms can link aggregate and spatial inequality. First, as shocks changes the incentives for sorting, worker sorting patterns might change. This channel is in line with the claims of some recent papers which have suggested that the geographic concentration of high-skill workers has increased in US data (Diamond and Gaubert, 2022)[3]. Second, firm sorting patterns can change. For ex-

[2]I use the term firm loosely in this paper. Sometimes, the word firm will refer to an establishment in the data. In the model, the concept of a firm will be an arbitrary bundle of jobs of the same type.

[3]Figure 24 in appendix A.2 shows that in my data college workers sort positively into rich regions both

ample, productive plants can become more concentrated in rich markets because of size effects or changing worker-firm complementarities. Third, a technology change that increases inequality by changing the productivity of specific worker or firm types can increase spatial inequality even when their sorting patterns stay approximately constant. For example, if some industry (say, Finance) is concentrated in a particular rich place (say, New York City) then increases in the productivity of plants producing financial services translate into rising spatial disparities.

The estimated model allows me to distinguish between these three explanations. I find that sorting patterns are strong but approximately constant over time. Once these constantly strong spatial sorting patterns are taken into account, they form a tight link between overall and spatial inequality. Thus, a change in technology that pushes up aggregate inequality does not affect all places equally. As a consequence, aggregate and spatial inequality move in lockstep. I further show that interactions of technological changes and sorting can also account for the fact that the rise of inequality has been spatially concentrated. I then study the drivers of firm and worker sorting across space, and show that firm sorting is driven in large part by thick market effects and it is, in turn, an important determinant of worker sorting.

The finding that firm sorting is quantitatively important implies that job ladders can vary strongly across locations. Motivated by this finding, I ask whether spatial disparities generate differences in worker values across locations. More concretely, conditional on the worker's firm type, I quantify differences in value that arise from being in a rich location type versus a poor location type. I find these differences to be quantitatively meaningful. Workers would on average be willing to pay a penalty equivalent to 21% of average income in order to keep their job and costlessly switch locations (or to prevent such a switch). This means that locations are important determinants of worker-level outcomes. This result puts in context the finding that cross-sectional wage variance decompositions imply a small role of spatial inequality in generating aggregate inequality. The key conceptual insight is that rich locations enable upward mobility for workers and thus matter when longer horizons are considered.

In the final part of the paper, I use the model to analyze a number of policies that the

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.

literature has previously considered. The rich heterogeneity allowed for my the model makes it a potent tool to consider the redistributive effects of policy that targets worker and firm location incentives. I find that spatial policies generally produce winners and losers and can be either regressive or progressive. Progressive policy expands poor regions and reduces the amount of firm sorting in the data. I find that one such policy is to incentivize high-skill plants to locate in poor locations. However, the redistributive properties of spatial policies generally depend on the elasticity of housing in both rich and poor regions.

The paper contributes to various strands of literature. 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) and more recently Lindenlaub et al. (2022) who like me study a spatial search model in which firms make location choices. This paper overcomes two limitations of this strand of literature.

First, 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 the urban wage premium and identify their respective role in generating wage inequality. Distinguishing between these two types of geographic sorting is particularly important in my setting, since both have vastly different implications from each other for the importance of spatial inequality on the worker level. Geographic firm sorting shapes job ladders and makes spatial policy a potent redistributive tool. Geographic worker sorting on the other hand opens less of a role for policy. Studying a model of search helps in estimating the impact of spatial inequality on individual workers' outcomes - workers visit multiple parts of the wage distribution during their career but their location may determine which job ladders they climb.

Second, the careful measurement of geographic sorting patterns in the data has so far remained beyond reach for the literature. The reason for this is twofold. First, under 3-dimensional sorting (workers, firms and locations), it is a challenge to transparently describe sorting patterns of workers, firms and locations. This problem is compounded by the second reason: The geographic sorting patterns of workers and firms are in fact quite hard to measure. Worker and firm characteristics that are relevant to the expected wage of a match are often unobserved[4], which substantially complicates any attempt to measure sorting based on these characteristics. Other papers, most notably Dauth et al. (2022) and Card et al. (2023), have previously run different versions of AKM regressions to obtain patterns of worker-firm sorting across space. However, employing AKM requires putting major restrictions on complementarity patterns of workers and firms. This is a major 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). In contrast, LANCE allows for arbitrary patterns of worker-firm complementarities and simultaneously allows me to measure changing sorting patterns directly in the data. In developing the new algorithm, this paper also contributes to the literature on the measurement of sorting workers-firm sorting in the data, in particular the pioneering paper by Bonhomme et al. (2019) and more recent work by Lentz et al. (2023).

The remainder of the paper proceeds as follows. Section 2 illustrates data patterns of aggregate and spatial inequality in West Germany. Section 3 introduces the model. Section 4 discusses the way model parameters are estimated. Section 5 introduces the data set and sample selection. Section 6 presents the results from the measurement exercise, the model fit and the counterfactuals. Section 7 concludes.

2 Inequality within and across space, 1975-2018

I begin by using data from West Germany to document three simultaneous developments:

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

[4] 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)

To show these facts, I use the Sample of Integrated Employer-Employee Data (SIEED) from the German Institute for Employment Research (IAB). I select only West German full time employment spells. Further details on this data are discussed in section 5.

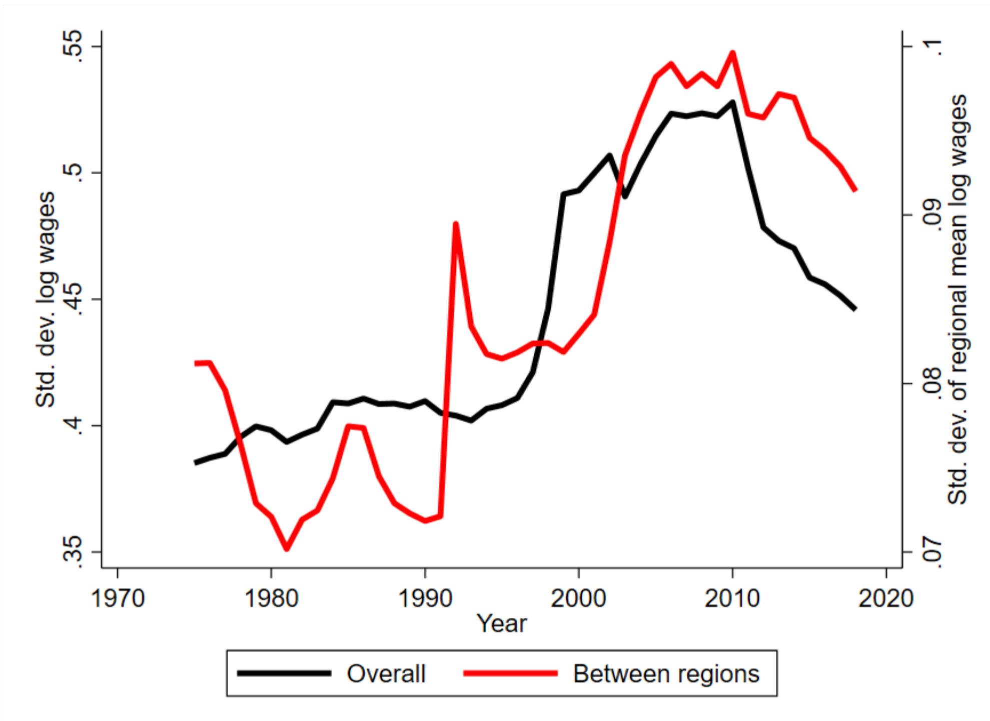


Figure 1: West German wage inequality over time

Figure 1 documents the first two of these facts. 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.[5] Apart from this spike, aggregate and spatial inequality move in tandem and the two time series are highly correlated.

Figure 2 displays the third fact. The figure plots the level of within-region inequality

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

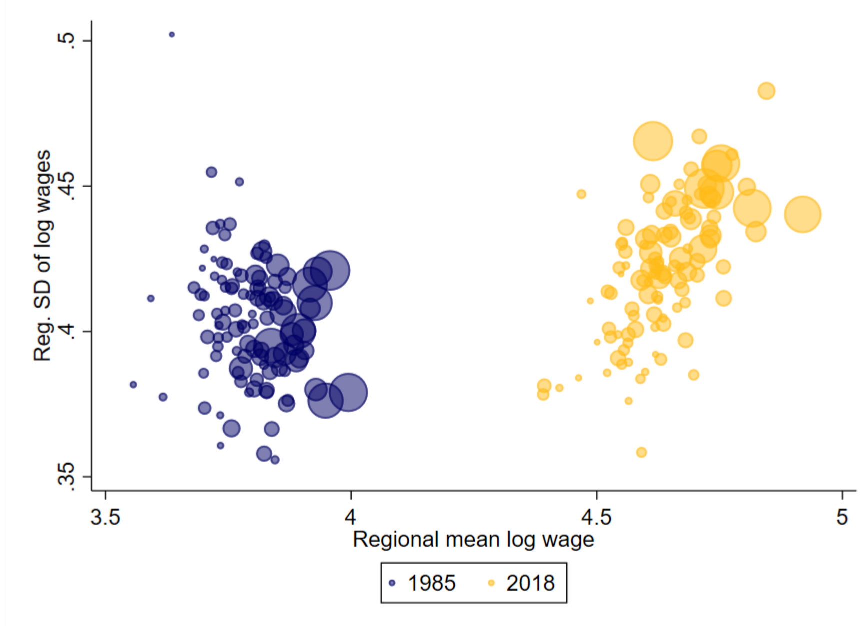


Figure 2: Income-inequality relationship

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 wealthy a region is, the y-axis the change in wage inequality within that region. In 2018, there is a clear relationship. Rich regions tend to become a lot more more unequal. In 1985, the data does not exhibit this relationship. 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 also tend to be more unequal.

Figure 22 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 than poor places.[6] To explain these trends, the next section introduces a model of two-sided sorting.

[6]This pattern does not change is 2018 is taken as a base year for the x-axis, as shown in appendix A.2.

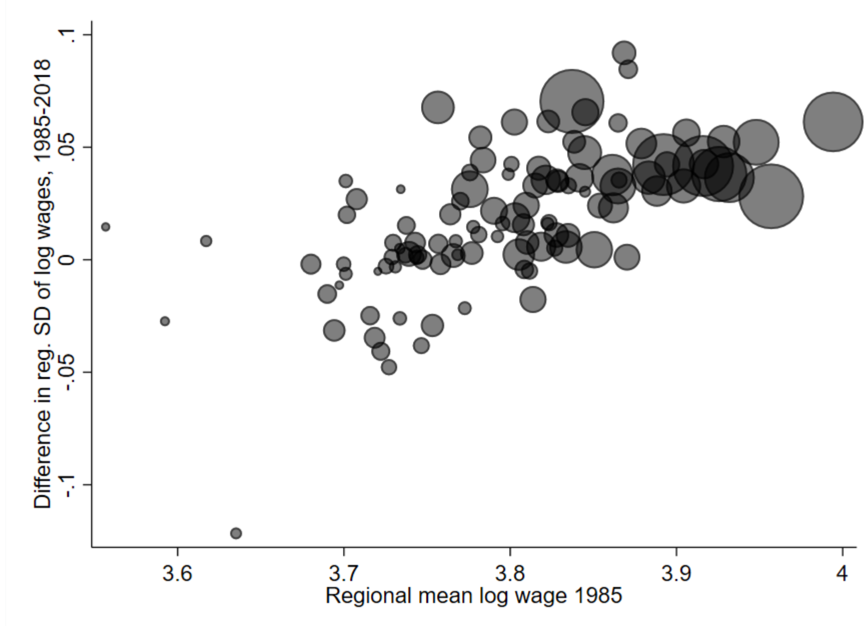


Figure 3: Inequality changes by regional income

3 Model

The model has two main objectives. The first is to provide a tractable framework to think about the sources of sorting for both workers and firms, both of whom make forward-looking location decisions. The second is to deliver a foundation for a structural estimation of key model parameters that drive sorting. Given the relatively high number of parameters, the latter part is key for the quantitative credibility of the model. Section 4 discusses the identification process in detail. The paper's methodological contribution, discussed in detail in section 4.1, enables not only this estimation but also allows me to measure sorting patterns in the data. This provides a natural test for the model, since model-implied spatial sorting patterns can be directly contrasted with data estimates.

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 are born in non-employment with exogenous rate $\chi_{k,l}$ and die stochastically at rate ρ, implying 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 $b + a_l - p_l$ when in non-employment and $w_{k,m,l} + a_l - p_l$ in employment, where b is the worker's flow value of being idle, $w_{k,m,l}$ 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 number $\mathcal{J}$ of jobs. A job is associated with a firm (or "plant") which belongs to one of M types[7], denoted by $m \in \{1, \ldots, M\}$. 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 at any moment is either filled by a worker of this type or unfilled. If filled, the job count towards employment. If unfilled, the job is labeled a "vacancy". The total number of jobs is thus given by employment and vacancies:

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

A filled job gets broken up at some rate $\delta_{k,m}$. An unfilled job disappears at rate δ^v. Jobs are attached to a particular location forever. A job is therefore characterized 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 both the worker and firm types k, m in the match and the match location l:

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

[7] 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.

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.

Matching and mobility

Workers and firms search randomly for jobs. 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 $\mathcal{V} = \sum_{k,m,l} \mathcal{V}_{k,m,l}$, $\mathcal{S} = \sum_{k,m,l} \nu^{\mathbb{1}(m \neq 0)} e_{k,m,l}$ and ν is the (exogenous) relative search intensity of the non-employed. 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 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 $(1 - \zeta_{m,m'})$ that a worker does not have the required skills to perform a job, rendering the match unviable. 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[8] and then solve

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

where $v_{k,m,l}$ is the value of a worker in state (k, m, l) and c_k denotes the cost that a worker incurs when moving locations. 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

[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.

networks and a familiar environment. For simplicity, and to enhance interpretability, I assume that these costs are independent of the distance of a move and do not depend in any way on which locations are its origin or destination.

Given the well-known properties of the Gumbel distribution, a worker in state (k, m, l) has a value that satisfies the following Bellman equation:

$$\begin{aligned}
&(r + \delta_{k,m} + \rho) v_{k,m,n} \\
=& w_{k,m,n} + a_n - p_n + \delta_{k,m} v_{k,0,n} \\
&+ \sum_{m' \neq 0, n'} \nu^{m \neq 0} \eta_{k,m'} \lambda s_{m',n'} \mathbb{E}\left[\max\{v_{k,m,l} + \varepsilon_1, v_{k,m',l'} - \mathbb{1}(l' \neq l) c_k + \varepsilon_2\}\right] \\
=& w_{k,m,n} + a_n - p_n + \delta_{k,m} v_{k,0,n} \\
&+ \sum_{m' \neq 0, n'} \nu^{m \neq 0} \eta_{k,m'} \lambda s_{m',n'} \left[\sigma^w \log\left(\exp\left(\frac{v_{k,m,n}}{\sigma^w}\right) + \exp\left(\frac{v_{k,m',n'} - \mathbb{1}(l \neq l') c_k}{\sigma^w}\right)\right) - v_{k,m,n}\right]
\end{aligned} \tag{3}$$

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

$$P_{k,m,m',l,l'} = \frac{\exp\left(\frac{v_{k,m',l'} - \mathbb{1}(l \neq l') c_k}{\sigma^w}\right)}{\exp\left(\frac{v_{k,m,l}}{\sigma^w}\right) + \exp\left(\frac{v_{k,m',l'} - \mathbb{1}(l \neq l') c_k}{\sigma^w}\right)} = \frac{C_k^{\mathbb{1}(l \neq l')} V_{k,m',l'}}{V_{k,m,l} + C_k^{\mathbb{1}(l \neq l')} V_{k,m',l'}}$$

where $C_k = \exp(-\frac{c_k}{\sigma^w})$ and $V_{k,m,l} = \exp(\frac{v_{k,m,l}}{\sigma^w})$. 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' in location l' is therefore:

$$\mu_{k,m,m',l,l'} = \nu^{m \neq 0} \lambda s_{k,m',l'} \zeta_{m,m'} P_{k,m,m',l,l'} \tag{4}$$

Here, $s_{k,m,l} = \frac{\mathcal{V}_{k,m,l}}{\mathcal{V}}$ is the share of vacancies that are of type m in location l and require worker type k.

Vectorizing the birth rates $\chi_{k,m,l}$[9] and the employment-unemployment distribution, we can write the system of steady state flow equations in matrix form:

$$e = -(M' - \rho I)^{-1} \chi \tag{5}$$

[9] 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 $m > 0$.

where

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

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 $\eta_{k,m}$.[10]

The ex-ante value of a vacancy (which has to equal the entry cost ϱ in equilibrium) 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 entrant:

$$\bar{\Omega}_{m,l} = \sum_l \eta_{k,m} \Omega_{k,m,l} \tag{6}$$

$$\bar{\bar{\Omega}}_m = \mathbb{E}\left[\max_l \{\bar{\Omega}_{m,l} - u_l\}_{l=1}^L\right] = \sigma^f \log\left(\sum_l \exp\left(\frac{\bar{\Omega}_{m,l}}{\sigma^f}\right)\right) \tag{7}$$

$$\bar{\bar{\bar{\Omega}}} = \sum_m \psi_m \bar{\bar{\Omega}}_m \tag{8}$$

In equilibrium, the ex-ante value of entry has to equal the cost, i.e.

$$\bar{\bar{\bar{\Omega}}} = \varrho \tag{9}$$

Invoking again the properties of a Gumbel distribution, given $\bar{\Omega}_{m,l}$, the probability of

[10] 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.

creating a job in location k for a firm of type m is given by

$$\phi_{m,l} = \frac{\exp\left(\frac{\bar{\Omega}_{m,l}}{\sigma^f}\right)}{\sum_{l'} \exp\left(\frac{\bar{\Omega}_{m,l'}}{\sigma^f}\right)} \tag{10}$$

Now, it is straightforward to find $s_{k,m,l} = \frac{\mathcal{V}_{k,m,l}}{\mathcal{V}}$: The fact that jobs become vacant upon separation implies that, 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

$$s_{k,m,l} = \frac{\mathcal{V}_{k,m,l}}{\mathcal{V}} = \psi_m \phi_{m,l} \eta_{k,m} \tag{11}$$

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

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

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:

$$\left(r+\delta^v + q\xi_{k,m,l}\frac{r}{r+\delta_{k,m}+\mu^{\text{ee}}_{k,m,l}}\right)\Omega_{k,m,l} = q\xi_{k,m,l}\left(\frac{(1-\beta)Y_{k,m,l}}{r+\delta_{k,m}+\mu^{\text{ee}}_{k,m,l}}\right) \tag{12}$$

We are now ready to define an equilibrium.

Equilibrium

An equilibrium is defined as a distribution $e_{k,m,l}$, worker flow rates $\mu_{k,m,m',l,l'}$, firm value functions $(\Omega_{l,m,n}, \bar{\Omega}_{l,m}, \bar{\bar{\Omega}}_m, \bar{\bar{\bar{\Omega}}})$, worker and firm encounter rates λ and q, worker value functions $v_{k,m,l}$, and location choice probabilities $\phi_{m,l}$ such that

1. The employment-unemployment distribution is given by the flow equation (5)
2. Flow rates are given by equation (4)
3. The value function of the worker satisfies the Bellman equation (3)
4. The value functions of the firm satisfies the Bellman equation (12) and the aggregation equations (6), (7), and (8)
5. Location choice probabilities of firms are given by equation (10)
6. The ex-ante firm value equals the entry cost as in equation (9)
7. q and λ are consistent with the matching function as specified by 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.

4 Identification

I identify the model using matched employer-employee data on matches between workers and firms in West Germany between 1975 and 2018. Section 5 provides further details on the data.

I proceed in three steps. First, I develop a novel bi-clustering algorithm for the classification of workers and firms. I use this algorithm to jointly classify workers and firms into their respective types, and then classify regions based on the local distribution of workers and firms. Second, I use the resulting classification in 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.1. Lastly, I choose the remaining parameters to minimize the distance between the model-implied employment distribution and the employment distribution identified by the bi-clustering algorithm in the data. Section 4.1 introduces the bi-clustering algorithm while sections 4.2 and 4.3 describe the remaining steps of the estimation.

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 vanilla 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[11]. 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.

Setup and general idea

LANCE solves a likelihood maximization problem in which the distribution of some set of signals depends on type membership of associated signal-emitting entities. The likelihood function takes the following form:

$$\max_{\theta,a_1,a_2} \mathcal{L}(\theta, a_1, a_2)$$

where θ is a set of parameters and $\tilde{a}_n(\cdot) : \mathcal{I}_n \mapsto \mathcal{T}$ is an "assignment function" that maps a signal-emitting unit into its type. More concretely in my case, the signal-emitting units are workers and firms, and the signals are wages[12]. The distribution of the signals depends on the respective type membership of workers and firms, whose type membership is unknown. The estimated type membership, which I call assignment, in turn depends on the type-conditional distribution of signals. Thus, the maximization problem has to be solved jointly. Although the margins of optimization interact, they cannot be solved jointly at once, because the discrete space over which assignments are optimized is too large for this to be computationally feasible.

The idea of LANCE is to solve this problem by applying coordinate descent[13] to the likelihood function, re-optimizing assignments and parameters in sequential order. Thus, as the algorithm progresses, the both the parameters of the type-conditional distribution of signals and the estimates for type assignments become better over time. It is easy to show that this version of the algorithm converges to a local maximum of the likelihood in a finite number of iterations.

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

[12] 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.

[13] See Bezdek et al. (1987).

Applying LANCE to the identification problem

To apply this idea to the present setting, recall that the model predicts that log wages satisfy

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

where $\gamma_{k,m} = \log \beta + z_{k,m}$. To map from the data into the model, 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

$$\log w_{k,m,l,t} = \gamma_{k,m,t} + h_l + \varepsilon_i, \quad \log w_{k,m,l,t} \sim F\left(\cdot \mid \theta\right)$$

where $\theta = (\gamma_{k,m,t}, h_l, \sigma)$. 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). Thus, in the data I estimate $\gamma_{k,m,t}$ where $t \in \{0,1\}$ and $t = 1$ reflects matches conceived after 1995.

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 solve the likelihood maximization problem:

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

where

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

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:** $\gamma_{k,m}, h_l, \sigma$ are determined by a regression of log wages on (k, m, t) and h_l fixed effects given the assignments.[14]

2. **Worker type updating:**

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

3. **Firm type updating:**

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

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[15]. 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

[14]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).

[15]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 propery 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-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. 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 $\eta_{k,m}$, $\zeta_{m,m'}$, $C_{k,t}$, $\psi_{m,t}$, ν and $\delta_{k,m}$ 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, $\phi_{m,n}$, λ and $V_{l,m,n}$ to vary freely in the estimation. 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 defined in equation (4). Defining $s_{m,n} = \frac{\sum_k \mathcal{V}_{k,m,n}}{\mathcal{V}}$ and using $\mathfrak{n}_{n'}$ to denote the number of regions of type n, we can re-write equation (4) to make it more amenable to the empirical estimation procedure:

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

and

$$Q_{k,m,m',n,n',h,t} = \mathfrak{k}_{n,n',h} P_{k,m,m',n,n',h,t}$$

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

$$P_{k,m,m',n,n',0,t} = \frac{V_{k,m',n,t}}{V_{k,m,n,t} + V_{k,m',n',t}}, \quad P_{k,m,m',n,n',1,t} = \frac{C_{k,t}V_{k,m',n,t}}{V_{k,m,n,t} + C_{k,t}V_{k,m',n',t}}$$

Using the properties of Poisson processes, conditional on a transition, the probability of each directions 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

$$\mathcal{L}_{\text{Mobility}} = \sum_{k,m,m',n,n',h,t} \sum_{s} d_{k,m,m',n,n',h,t}(s) \Bigg[\overbrace{\log\left(\mu_{k,m,n,t}\right) - \mu_{k,m,n,t} r(s)}^{\text{Length}} + \underbrace{\mathbb{1}(m' \neq 0) \log\left(\frac{\nu^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \lambda_t \psi_{m',t} \phi_{m',n',t} \zeta_{m,m'} Q_{k,m,m',n,n',h,t}}{\mu_{k,m,n,t}} \right) + \mathbb{1}(m' = 0) \log\left(\frac{\delta_{k,m}}{\mu_{k,m,n,t}} \right)}_{\text{Direction}} \Bigg]$$

Like with LANCE, I again employ a coordinate descent strategy to optimize this object: 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.1 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 $z_{k,m}$ and h_n, a classification which enables maximum likelihood, as well as an employment distribution in the data, $e^{\text{data}}_{k,m,n}$. The maximum likelihood procedure further yields values for $\eta_{k,m}$, C_k, ν 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.

First, 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 flow value of non-employment to 0, and the interest rate r to 5% per annum. The death rate of workers is set to $\frac{1}{30}$ per annum, to approximate a 30-year working life. Finally, β is set to 0.5[16].

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

[16]This is not important and just a normalization. A different value for β would yield an observationally equivalent model.

Parameter	Symbol	Value	Estimation method	Target
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	$C_{k,t}/c_{k,t}$		Max. Likelihood	
Firm type probability at entry	$\psi_{m,t}$		Max. Likelihood	
Search intensity in employment	ν		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		Min. distance	Employment dist.
Randomness in firm location choice	σ^f		Min. distance	Employment dist.
Elasticity of encounter function	α	0.5	Externally set	
Flow value of non-employment	b	0	Externally set	
Interest rate	r	5% p.a.	Externally set	
Worker death rate	ρ	$\frac{1}{30}$ p.a.	Externally set	
Worker share of output	β	0.5	Externally set	
Match efficiency	A		Calibrated	Job filling rate $= \frac{1}{30}$
Ex-ante entry value	ϱ		Calibrated	Encounter rate $= 0.203$
Net location amenity	$a_{n,t} - p_{n,t}$		Calibrated	Location size
Vacancy death rate	δ^v		Calibrated	Vacated vacancy share $= \frac{1}{2}$

Table 1: Parameters and estimation strategy

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 $e^{\text{data}}_{k,m,n,t}$, which. 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 high-quality German administrative micro-data provided by the Institute for Employment Research (IAB), the so-called Sample of Integrated Employer-Employee Data (SIEED). 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, as well as information on the wage for each match, updated at a frequency of at least once every year.

The data also contains information on the location of a match 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. Since the model does not allow for commuting, they are a good choice to map a location from the model to the data. 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 satisfy the following conditions

- Spells are located in West Germany (excluding Berlin)
- Workers have achieved their highest education level recorded in the data
- The job is full time

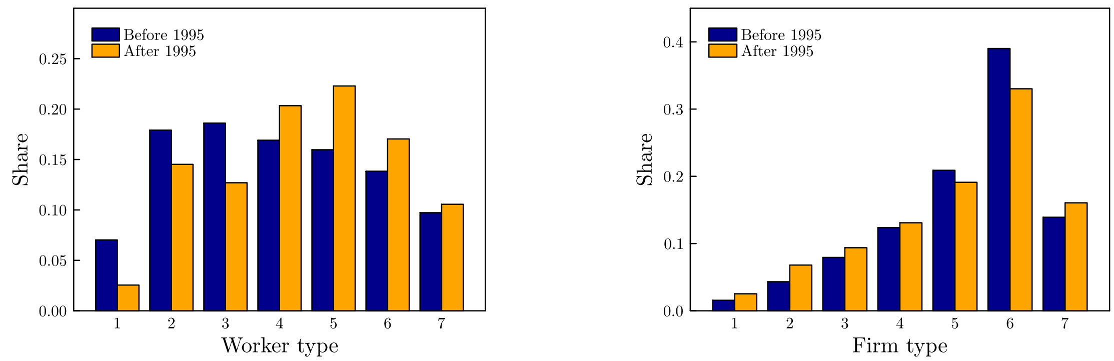
After filtering out these observations, I residualize the log wages by running the following regression

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

removing age and time effects. I use as the new wage the predicted wage for a representative 35 year old worker in 2018, so that all monetary amounts in this paper can be interpreted as 2018-Euros. 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 both the estimation procedure and the model. LANCE and the maximum likelihood estimation both return results that are informative from an empirical perspective even without the context of the model. Thus, section 6.1 discusses the results that arise directly from LANCE and section 6.2 summarizes the results from the maximum likelihood step. Section 6.3 discusses the results that are obtained from the model, such as the fit of the model, counterfactual sorting patterns, and compensating differentials. Finally, section 6.4 discusses implications for policy.



(a) Workers

(b) Firms

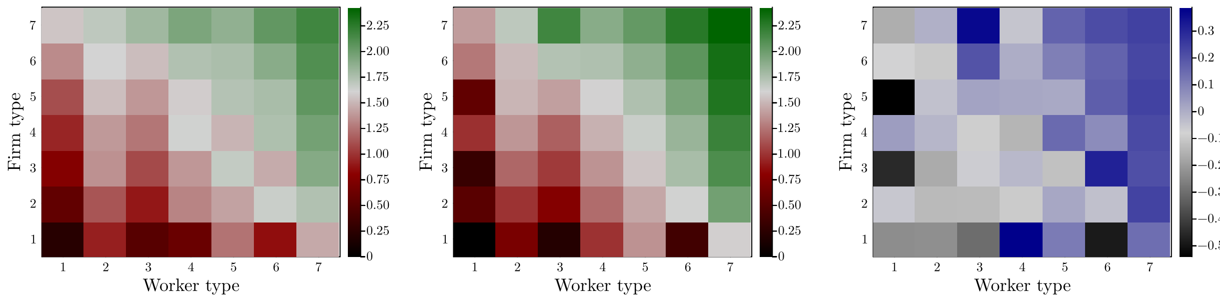
Figure 4: Marginal distribution of types

6.1 Empirical results from LANCE

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 worker types roughly uniformly to all types, with a slight shift towards higher worker types in the post-1995 period. Figure 25 in appendix A.2 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 regions as "type 1" regions and 53 regions 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 more than accounts for the urban wage premium. In general, the quantiatively 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 4.558 and 4.688 respectively, implying substantial ex-post differences between the two places.

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



(a) $z_{k,m,0}$ (before 1995) (b) $z_{k,m,1}$ (after 1995) (c) Difference

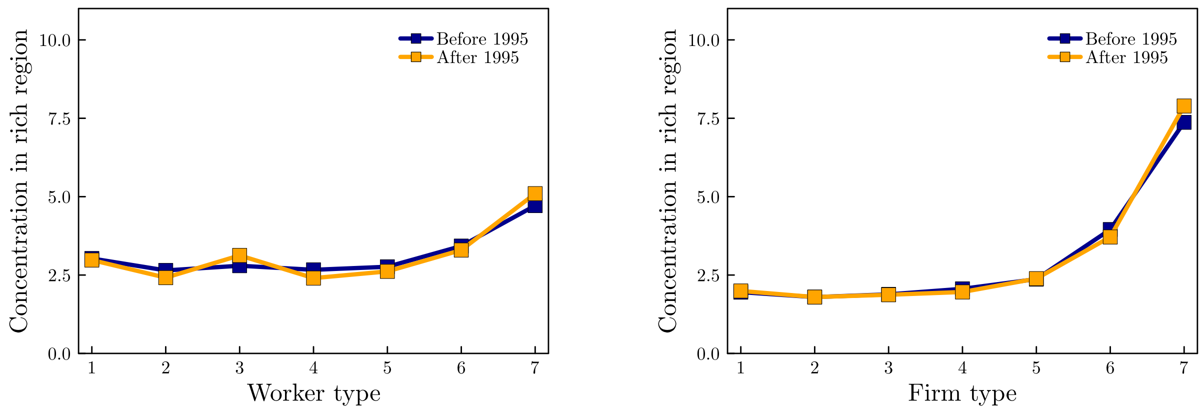
Figure 5: Match-specific component of the wage function

context. Given the large degree of horizontal differentiation in worker skills one might well expect a production (or wage) function to be non-monotone in the type of worker or firm. Such a production function would have wildly different implications for sorting. In a world characterized primarily by horizontal differentiation, ranking workers would be a lot harder. In terms of location choice, one would have to keep track of a skill dimension that not necessarily corresponds to an average wage. For example, one could expect actors and film studios to coordinate in Los Angeles, investment managers and hedge-funds to coordinate in New York, and farmers and farms to coordinate in Nebraska, even without substantial differences in average pay.

Figure 5 shows that for the aggregate spatial wage distribution, this type of spatial coordination is of secondary importance. Wages are mostly monotone in worker and firm types. Instead, the results support models in which workers and firms are mainly horizontally differentiated. The wage function is not quite log-additive, as imposed by AKM. 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



(a) Worker-region sorting

(b) Firm-region sorting

Figure 6: Regional sorting in the data

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 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. The converse is also true. 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.

6.2 Empirical results from Maximum Likelihood

Next, I consider the results from the maximum likelihood estimation step. ν 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 data. As the figure shows, high worker types tend to be more

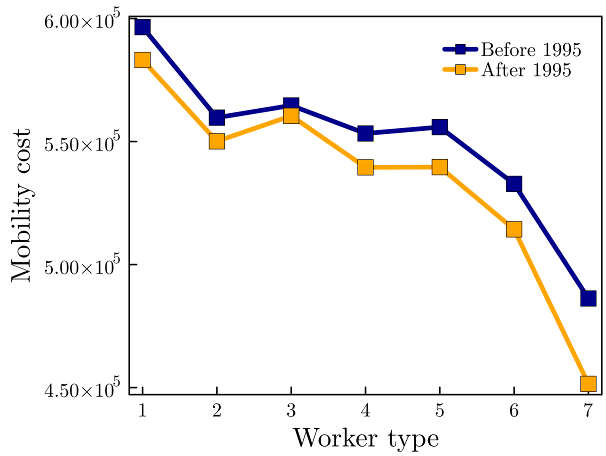
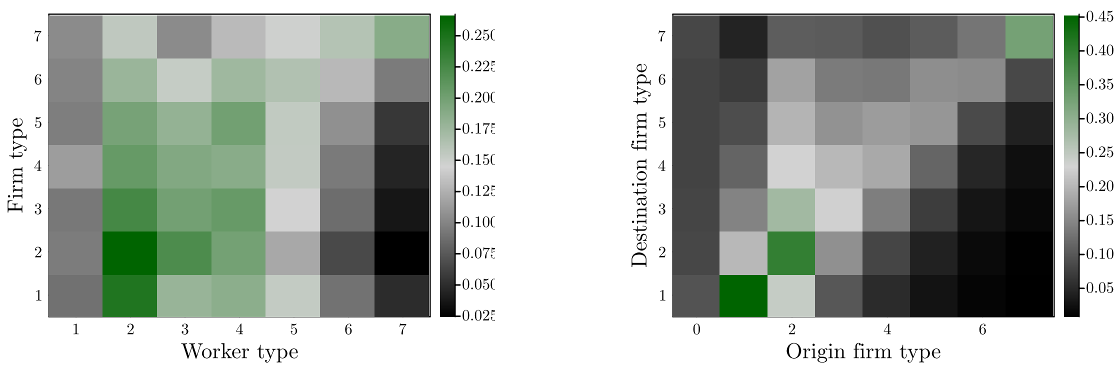


Figure 7: Mobility costs

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.



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

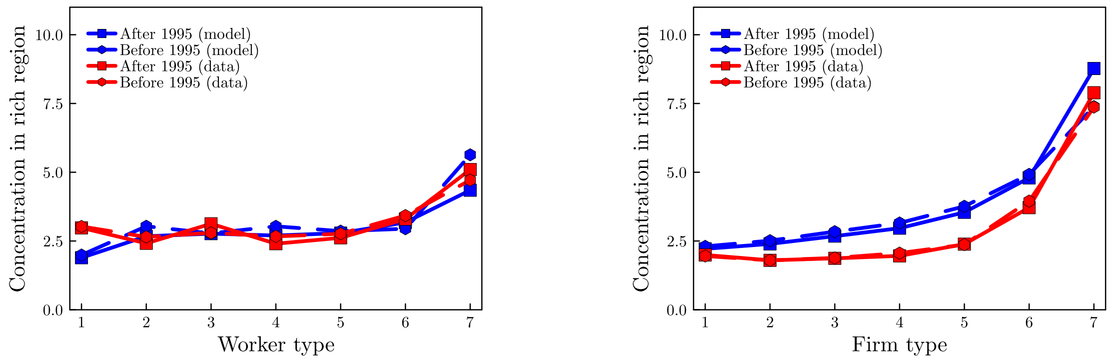
(b) $\zeta_{m,m'}$

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

Figure 8 shows the estimates for $\eta_{k,m}$ and $\zeta_{m,m'}$. There are clear signs for non-spatial segmentation in the form of $\eta_{k,m}$: Low firm types typically require low worker types and vice versa. 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.

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



(a) Workers

(b) Firms

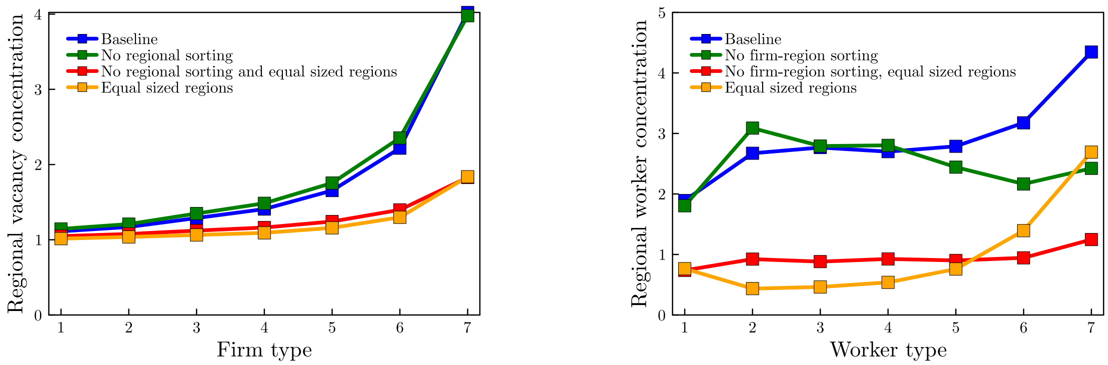
Figure 9: Type sorting into rich regions

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 slightly overestimates the concentration of middle firm types in the rich region and does not capture the mild increase in worker sorting. However, the broad patterns fit well - in particular the magnitude of sorting is well replicated. 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 ($e^{\text{cf}}_{l,m,k} = e_{l,m} \cdot 1/n$). 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

$(\lambda^{cf}_{m,n,t} = \lambda_{m,n,t}/(\lambda_n \cdot n))$ and a version with both $(\lambda^{cf}_{m,n,t} = \lambda_m \cdot (1/n))$. The results are displayed in Figure 10.



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

Figure 10: Sorting decompositions

Consider firm sorting first. We can see from Figure 10a 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 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 10b.

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.

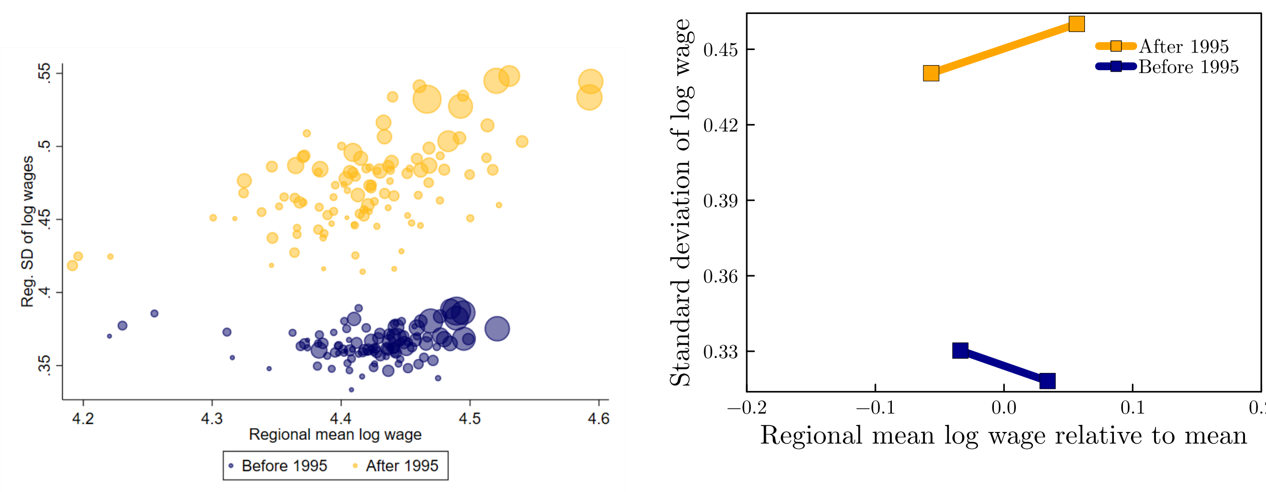
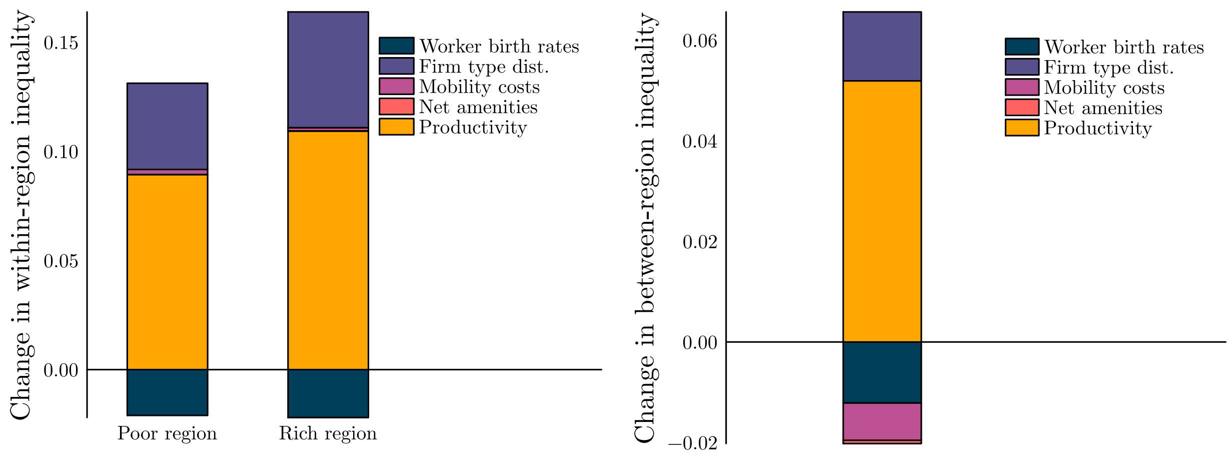


Figure 11: income-inequality relationship

Figure 11 tests these properties of the model. For both periods, Figure 11 plots the two region types' coordinates in income-inequality space. 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.

Figure 12 shows how the model rationalizes this development. The figure is based on different equilibria which are generated by successively adjusting the different exoge-



(a) Within regions

(b) Between regions

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

nous 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 these firms on the other. As a particular set of firms experiences 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.

Next, we turn to the question whether the increase in spatial inequality matters for worker values. To answer this question, I compute the compensating differential for every worker in every state. 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

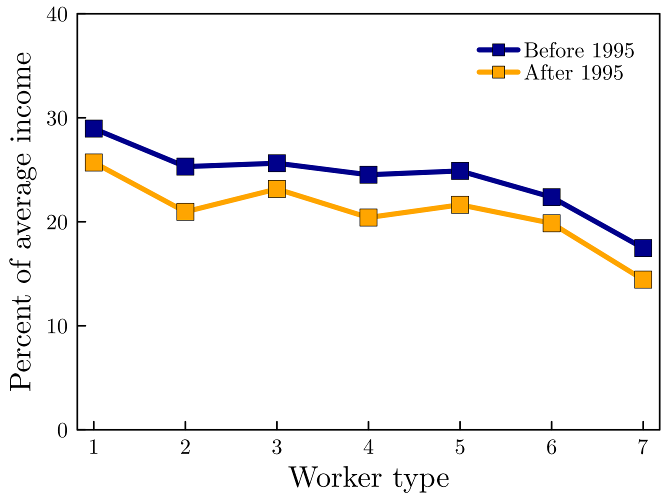


Figure 13: Compensating differential, rich versus poor region type

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 locations 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.

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[17], a fact that has sometimes been used to argue against the importance of spatial inequality. 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. This underscores the importance of studying the role of locations in a dynamic setting - a static view based on the data underestimates the impor-

[17] See Figure 23 in appendix A.2

tance of locations.

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. Changes in spatial inequality come from productivity gains of the highest firm and worker types. Much of the urban premium comes from firm sorting, meaning that good jobs are harder to come by in poor regions. It is mostly the poor who lose from these spatial disparities, as they are more mobility-constrained than their rich peers. Their losses from being in a poor region are large.

Finally, we turn to the question of policy. The implications of spatial sorting vary for different worker types and so might the implications of policy. The rich two-sided heterogeneity of the model enables us to consider redistributive properties of policy. The next section does so by analyzing a number of place-based policies that the literature has discussed.

6.4 Policy experiments

In this section I consider the effects of place-based policy in the context of the model as estimated in the more recent period. 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 in poor regions by particular firm types. These kinds of policies 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 on 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 $\tau^{\text{job}}_{m,n}$ that is paid to firms upon job creation in region type n. This means that $\bar{\Omega}_{m,n} + \tau^{\text{job}}_{m,n}$ replaces $\bar{\Omega}_{m,n}$ in all equilibrium conditions.

While versions of such policies have in the past been analyzed in the context of other spatial models, the rich heterogeneity embedded in the model allows me to focus on their redistributive effects. To do this, 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:

$$\begin{aligned} dy &= F_x dx + F_\tau d\tau \\ 0 &= G_x dx + G_\tau d\tau \end{aligned}$$

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 $x = (\{\bar{\Omega}_{m,n}\}_{m,n}, q, \{p_n\}_n)$ 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 effects of policy, we 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

$$de_n = \omega_n dp_n$$

where ω_n denotes the sensitivity of the housing supply to price changes.[18]

To analyze the two policies in question, we now need to decide by which criterion to evaluate them. This question is is less straightforward to answer than it first seems: For example, one potential welfare criterion is the average value of workers in the model. However, this turns out to create somewhat counter-intuitive incentives for a policymaker trying to maximize this object as discussed in more detail in appendix A.3. To sidestep the issues outlines in the appendix, I adopt a Rawlsian view of welfare: I evaluate policies by how they change the expected value of a prospective unborn worker conditional on its type. That is I define the "birth value" of a worker type k as

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

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

Lastly, I need to take a stance on the values for ω_n. As it turns out, the model features a peculiar form of observational equivalence of different vectors ω_n. As shown in appendix A.4, 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

[18] 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. However, in a slight abuse of terminology, I will sometimes refer to ω_n as the elasticity of housing in the remainder of the paper.

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 motif 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.[19]. 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.

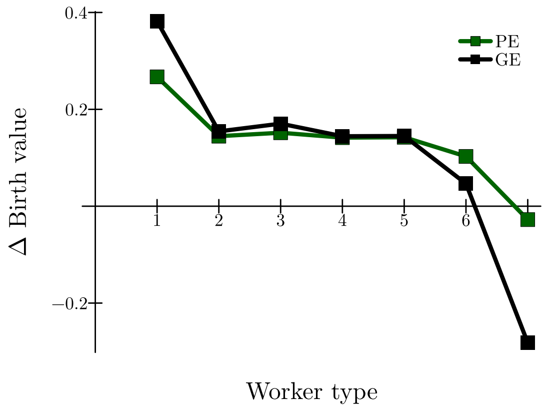
6.5 Regional subsidies

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 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: Poor workers gain, while rich workers lose. This is due to existing sorting. Poor workers spend more time in poor regions and thus benefit more from the policy. Likewise, high skill workers are concentrated in rich markets and are less likely to reap the benefits of the policy.

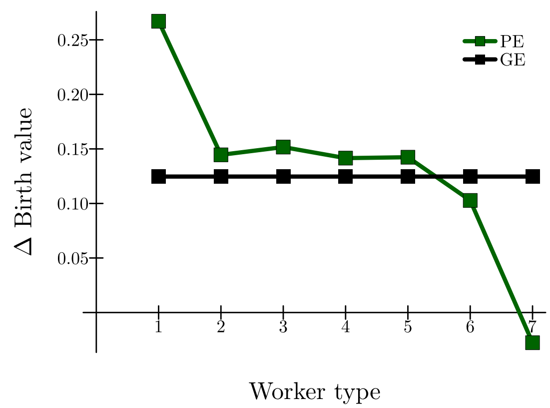
Next, consider the GE responses. In the inelastic case, prices adjust to keep allocations constant. Thus, inelastic housing markets *dampen* the GE effects of policy.[20] In the case of elastic housing however, things look quite different. Here, GE effects *amplify* the redistributive effects of policy. 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 10a, this leads to decreases in firm

[19] From the logic outlines in appendix A.4, 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.

[20] 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.



(a) Elastic housing



(b) Inelastic housing

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

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.

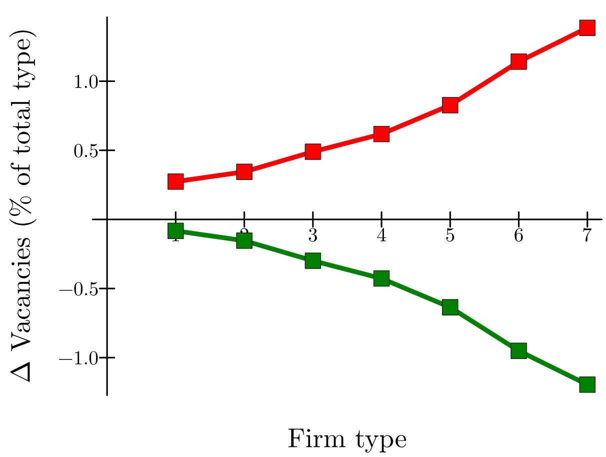
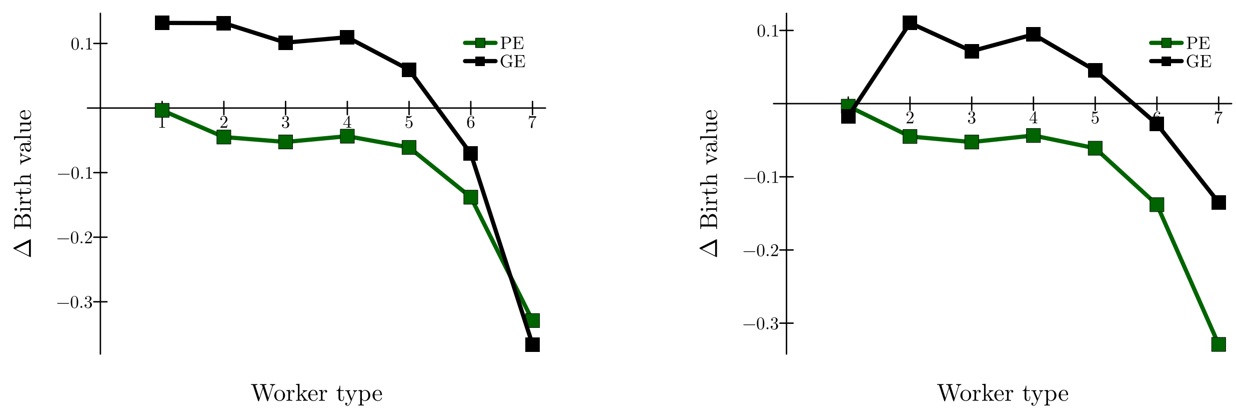


Figure 15: Changes in vacancies, 1 Euro subsidy to poor regions (inelastic housing)

6.6 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.[21]

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: Poor



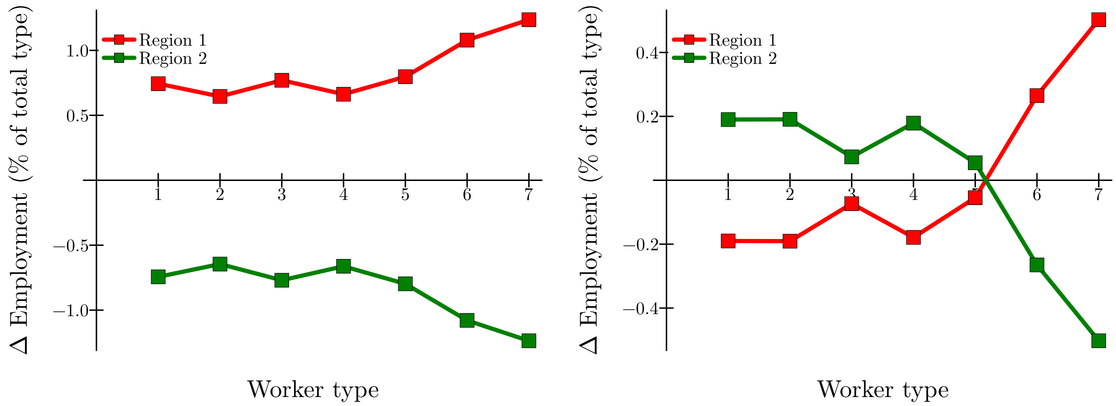
(a) Elastic housing (b) Inelastic housing

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

workers benefit somewhat while a few rich 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 rich workers at the benefit of their poorer peers. If housing

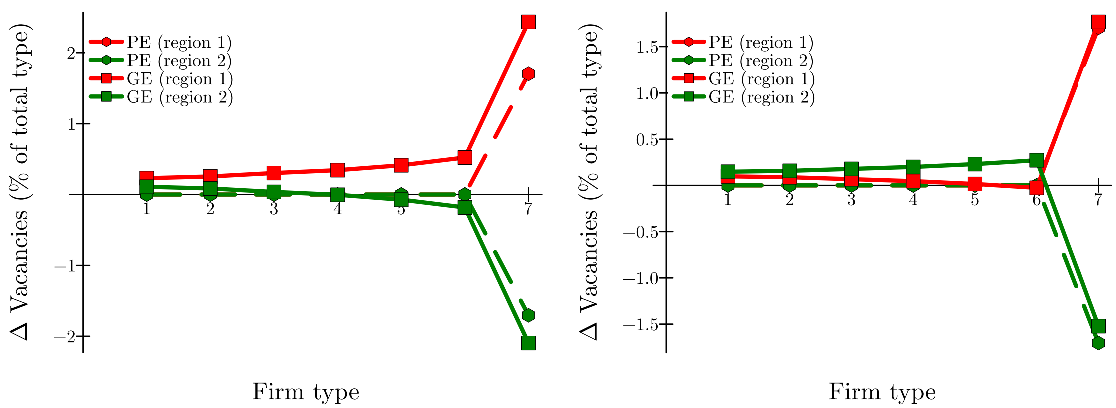
[21] This approach has the advantage that job creation incentives cannot be employed to rectify inefficient *levels* of job creation, which are not the focus of this paper.



(a) Elastic housing (b) Inelastic housing

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

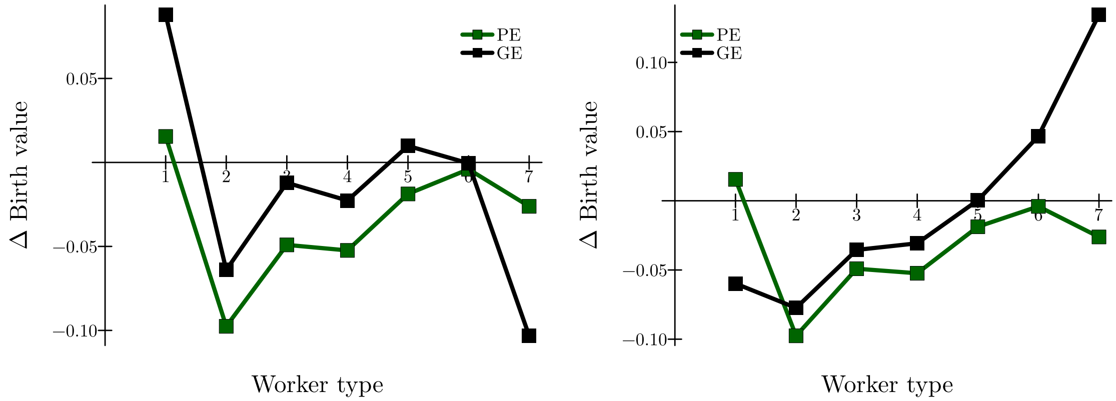
is inelastic, the location size does not change and so there is no additional sorting effects which offsets some of these losses for the rich worker type. This can also be seen in Figure 18. The increased presence in high type vacancies below type 7 in the rich region moderates the losses incurred from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



(a) Elastic housing (b) Inelastic housing

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

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

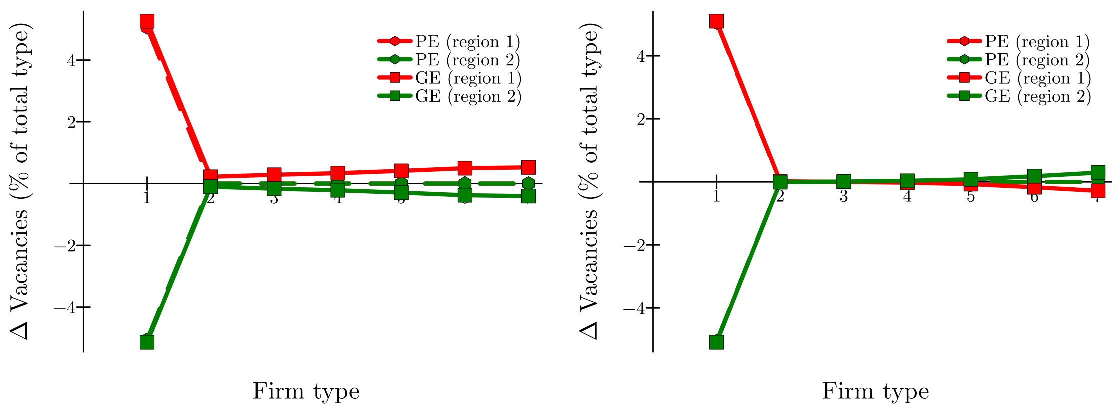


(a) Elastic housing

(b) Inelastic housing

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

on the supply of housing. If housing is inelastic, such policy is progressive. However, if housing is elastic, the policy is highly regressive. This again is a result that stems directly

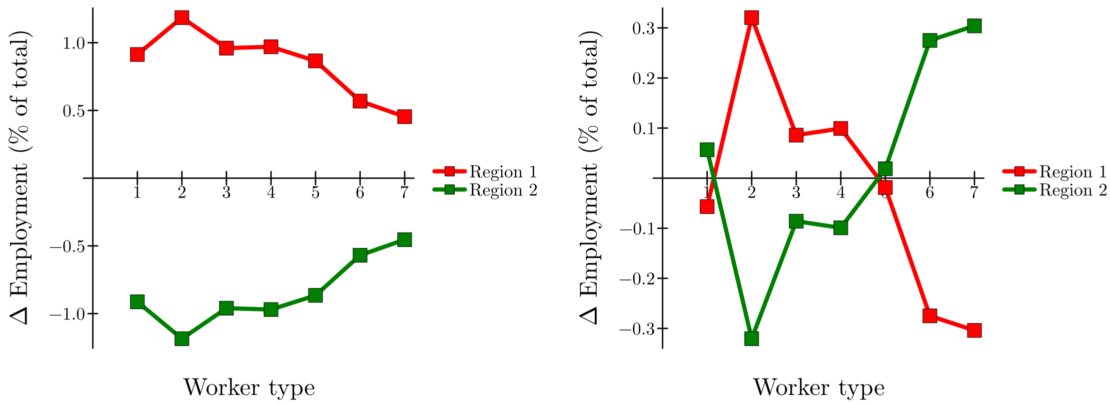


(a) Elastic housing

(b) Inelastic housing

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

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 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 rich underlying heterogeneity of workers and firms is taken into account.

7 Conclusion

This paper measures and models the spatial sorting patterns of workers and firms and draws important conclusions about the nature of spatial inequality. A structural model of spatial search delivers insights into the driving forces of geographic worker-firm sorting and its consequences for individual worker outcomes. A novel quantification method that can double as model-independent measurement allows us to fit a model that delivers empirically realistic sorting patterns of workers and firms across space.

I find that spatial sorting of workers and firms plays an important role in linking aggregate inequality and spatial inequality, explaining the joint rise of both in West German data. Firms in particular sort strongly into locations that are wealthy and big. Workers also see some spatial sorting which is quantitatively more muted and driven almost entirely by firm sorting. Despite the fact that the joint wage function of workers and firms has seen large changes, sorting patterns have stayed approximately constant. Nonetheless, technological change can produce spatially biased outcomes. Productive firms are disproportionally located in rich regions. As their productivity rises, the resulting wage increases are thus concentrated in rich regions. This leads to two developments that we observe in the data: Spatial inequality rises and within-region inequality rises particularly in rich regions.

The strong sorting patterns of firms also imply that job ladders differ strongly for workers in different locations. This reveals that spatial inequality is quite important for workers, especially when considering forward-looking values. Compared to a static view that focuses on wages, the spatial dimension of the job market becomes much more important when considering values: Workers are willing to give up the equivalent of between 14% and 26% of average income if they could costlessly change locations while retaining their job.

Finally, the model can be used to study spatial policy and its redistributive effects. Spatial policy has large redistributive effects across worker types. Progressive policies generally 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.

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.

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.

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

A Appendix

A.1 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

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

we can write the mobility log likelihood as

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

where

$$d_{k,m,m',n,n',h,t}(s) = \mathbb{1}\left(k(s) = k, m(s) = m, m'(s) = m', n(s) = n, n'(s) = n', t(s) = t\right)$$

and

$$d_{k,m,n,t}(s) = \mathbb{1}\left(k(s) = k, m(s) = m, n(s) = n, t(s) = t\right)$$

are indicator functions that assume one if the type and transition indicators of the spell are (k, m, m', n, n', h, t) and (k, m, n, t) respectively.

The assumptions on the mobility model made in the main text imply the following form for J2J transitions conditional on an encounter:

$$
\begin{aligned}
Q_{k,m,m',n,n',h,t} &= \mathfrak{k}_{n,n',h} P_{k,m,m',n,n',h,t} \\
\text{where } \mathfrak{k}_{n,n',h} &= \begin{cases} (\mathfrak{n}_{n'} - \mathbb{I}(n=n')) & \text{if } h = 1 \\ \mathbb{I}(n=n') & \text{if } h = 0 \end{cases} \quad \text{and} \\
P_{k,m,m',n,n',0,t} &= \frac{V_{k,m',n',t}}{V_{k,m,n,t} + V_{k,m',n',t}} \\
P_{k,m,m',n,n',1,t} &= \frac{C_{k,t} V_{k,m',n',t}}{V_{k,m,n,t} + C_{k,t} V_{k,m',n',t}}
\end{aligned}
$$

I maximize this quantity as follows: Starting with a guess for

$$
\theta = (\eta_{k,m'}, \zeta_{m,m'}, \lambda_{m',n',t}, \delta_{k,m}, V_{k,m,n,t}, C_{k,t}, \nu)
$$

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

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

First, establish

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

Then, the FOC is

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

which implies

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

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

Similar to before, we have

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

The FOC is thus

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

which implies

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

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

Proceed as before:

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

Then, the FOC is

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

which implies

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

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 ν

The FOC for ν is given by

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

and therefore the updating condition is

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

Minorizing the objective to get estimates for $V_{k,m,n,t}$ and $C_{k,t}$

Recall that the initial objective function is given by

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

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

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

Since the direct FOC is intractable, we proceed by using a minorization-maximization (MM) strategy[22]. MM algorithms are used to solve optimization problems over some

[22] 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.

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 $0 < x < 1, a \in \mathbb{R}_+$:

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

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

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

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

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

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

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

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

$$0 = \sum_{k,m,n,t} \left[\sum_{s} r_s d_{k,m,n,t}(s) \right] \left[\sum_{m',n',h,t} \nu^{\mathbb{1}(m \neq 0)} \eta_{k,m'} \zeta_{m,m'} \lambda_{m',n',t} \mathfrak{k}_{n,n',h} \left(1 - \bar{P}_{k,m,m',n,n',h,t}\right) \right.$$
$$\left(\mathbb{1}(k = \hat{k}, m = \hat{m}, n = \hat{n}, t = \hat{t}) \left(\frac{1}{V_{k,m,n,t}} - \frac{1}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right.$$
$$\left. \left. - \mathbb{1}(k = \hat{k}, m' = \hat{m}, n' = \hat{n}, t = \hat{t}) \left(\frac{C^h_{k,t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right) \right]$$
$$+ \sum_{k,m,m',n,n',h,t} \left[\sum_{s} d_{k,m,m',n,n',h,t}(s) \cdot \mathbb{1}(m' \neq 0) \right] \cdot \left[\vphantom{\frac{1}{1}} \right.$$
$$\mathbb{1}(k = \hat{k}, m' = \hat{m}, n' = \hat{n}, t = \hat{t}) \left(\frac{1}{V_{k,m',n',t}} - \frac{C^h_{k,t}}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right)$$
$$\left. - \mathbb{1}(k = \hat{k}, m = \hat{m}, n = \hat{n}, t = \hat{t}) \left(\frac{1}{\bar{V}_{k,m,n,t} + \bar{C}^h_{k,t} \bar{V}_{k,m',n',t}} \right) \right]$$

This can be simplified to

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

and then to

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

Separating out $V_{\hat{k},\hat{m},\hat{n},\hat{t}}$ yields

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

which finally gives us the following updating condition:

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

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

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

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

$$0=\sum_{k,m,n,t}\left[\sum_{s} r_s d_{k,m,n,t}(s)\right]\left[\sum_{m',n',h} \nu^{\mathbb{1}(m\neq 0)}\eta_{k,m'}\zeta_{m,m'}\lambda_{m',n',t}\mathfrak{k}_{n,n',h}\left(1-\bar{P}_{k,m,m',n,n',h,t}\right)\cdot\right.$$
$$\left.\mathbb{1}(k=\hat{k},h=1,t=\hat{t})\left(-\frac{V_{k,m',n',t}}{\bar{V}_{k,m,n,t}+\bar{C}^h_{k,t}\bar{V}_{k,m',n',t}}\right)\right]$$
$$+\sum_{k,m,m',n,n',h,t}\left[\sum_{s} d_{k,m,m',n,n',h,t}(s)\cdot\mathbb{1}(m'\neq 0)\right]\cdot$$
$$\left[\mathbb{1}(k=\hat{k},h=1,t=\hat{t})\left(\frac{1}{C_{k,t}}-\frac{V_{k,m',n',t}}{\bar{V}_{k,m,n,t}+\bar{C}^h_{k,t}\bar{V}_{k,m',n',t}}\right)\right]$$

which can be rewritten as

$$0=\sum_{m,m',n,n'}\left[\sum_{s} r_s d_{\hat{k},m,n,\hat{t}}(s)\right]\left[\nu^{\mathbb{1}(m\neq 0)}\eta_{\hat{k},m'}\zeta_{m,m'}\lambda_{m',n',\hat{t}}\mathfrak{k}_{n,n',1}\left(1-\bar{P}_{\hat{k},m,m',n,n',1,\hat{t}}\right)\cdot\right.$$
$$\left.\left(-\frac{V_{\hat{k},m',n',\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}}+\bar{C}_{\hat{k},\hat{t}}\bar{V}_{\hat{k},m',n',\hat{t}}}\right)\right]$$
$$+\sum_{m,m',n,n'}\left[\sum_{s} d_{\hat{k},m,m',n,n',1,\hat{t}}(s)\cdot\mathbb{1}(m'\neq 0)\right]\cdot\left[\left(\frac{1}{C_{\hat{k},\hat{t}}}-\frac{V_{\hat{k},m',n',\hat{t}}}{\bar{V}_{\hat{k},m,n,\hat{t}}+\bar{C}_{\hat{k},\hat{t}}\bar{V}_{\hat{k},m',n',\hat{t}}}\right)\right]$$

and thus yields

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

A.2 Additional Figures

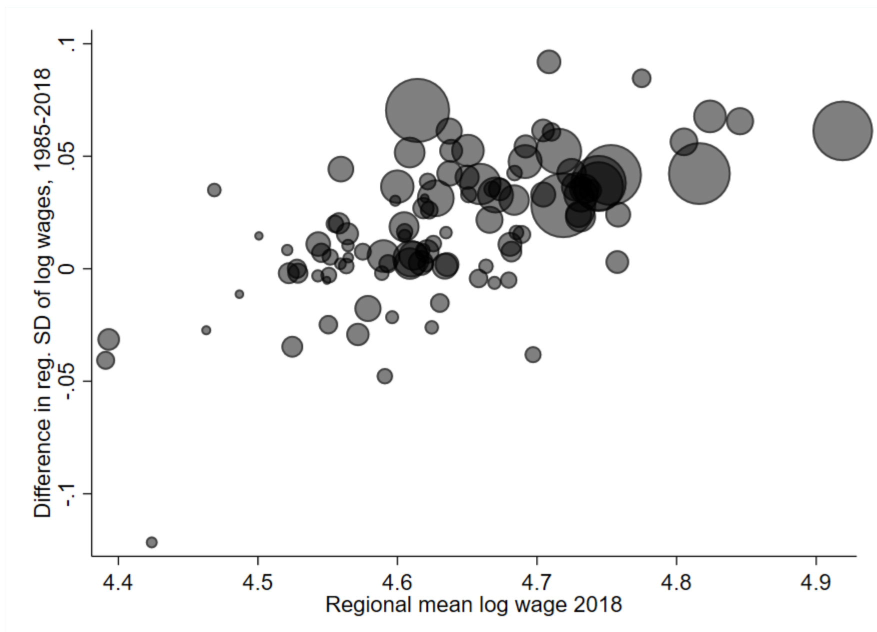


Figure 22: Inequality changes by 2018 regional income

A.3 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

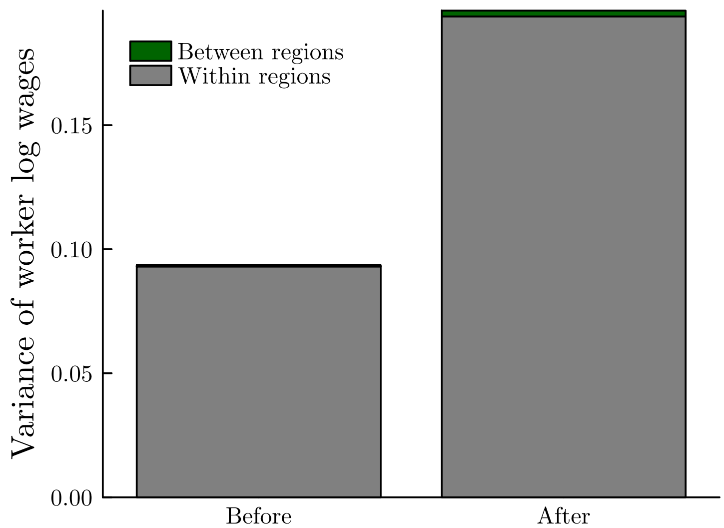


Figure 23: Variance decomposition of log wages

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.

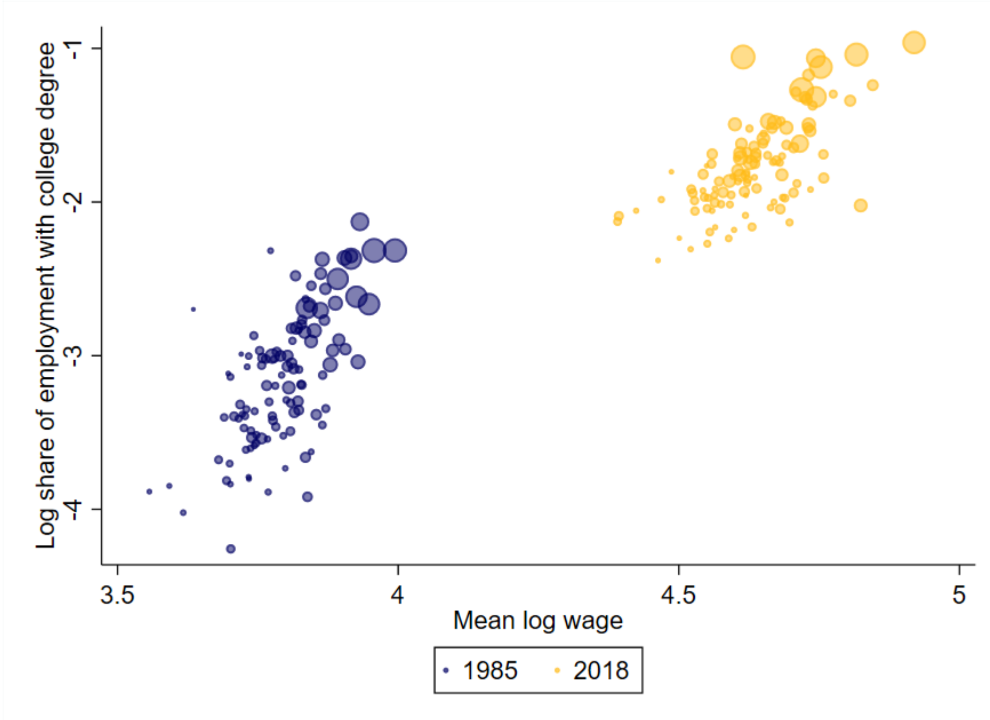


Figure 24: Worker sorting on observables

A.4 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.

<u>Claim:</u> 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 $dv + c$.

<u>Proof:</u> 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.

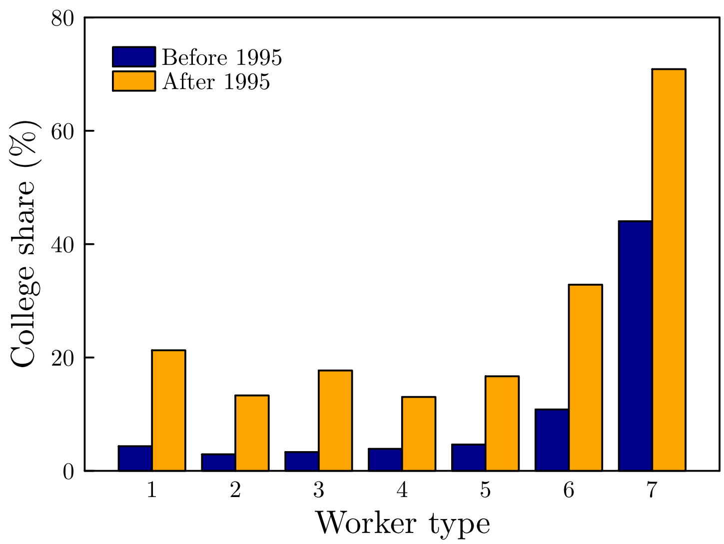


Figure 25: Mean education by worker type