Firm Sorting and Spatial Inequality*

Ilse Lindenlaub — *Yale University and NBER*

Ryungha Oh — *Yale University*

Michael Peters — *Yale University and NBER*

2 November 2022

Abstract

We study the importance of firm sorting for spatial inequality. If productive locations are able to attract the most productive firms, then firm sorting acts as an amplifier of spatial inequality. We develop a novel model of spatial firm sorting, in which heterogeneous firms first choose a location and then hire workers in a frictional local labor market. Firms' location choices are guided by a fundamental trade-off: Operating in productive locations increases output per worker, but sharing a labor market with other productive firms makes it hard to poach and retain workers, and hence limits firm size. We show that sorting between firms and locations is positive—i.e., more productive firms settle in more productive locations—if firm and location productivity are complements and labor market frictions are sufficiently large. We estimate our model using administrative data from Germany and find that highly productive firms indeed sort into the most productive locations. In our main application, we quantify the role of firm sorting for wage differences between East and West Germany, which reveals that firm sorting accounts for 17%-27% of the West-East wage gap.

*Lindenlaub: ilse.lindenlaub@yale.edu. Oh: ryungha.oh@yale.edu. Peters: m.peters@yale.edu.

We thank audiences at UCLA, Berkeley, Sciences Po, University of Minnesota, Dartmouth, Columbia, Johns Hopkins, the NBER Summer Institute 2022, SED 2022 and the EEA 2022 for useful comments. We are grateful to Gabriel Ahlfeldt for sharing his data on German property prices. We thank Juan Gambetta and Meili Wang for outstanding research assistance. The paper uses Linked-Employer-Employee-Data of the Institute for Employment Research (IAB) in Germany. Data access was provided via on-site use at the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the IAB and subsequently also remotely.

1 Introduction

Economic outcomes in developed economies have been starkly unequal since the early 1980s. Whereas most research focuses on inequality across people, a recent literature highlights disparate economic fortunes across space. Like inequality at the individual level, the spatial nature of growth has changed. Traditionally, many economies have been characterized by spatial convergence, but in past decades, spatial inequality has been increasing, with poor and rural locations falling further behind. Take Germany as an example: Urban areas are characterized by around 18% higher wages and value added, and West Germany still has 28% higher wages and 30% higher value added than East Germany. The phenomenon of a spatial divide has been attributed to the spatial sorting of more productive workers to cities or to the productivity advantages of areas with an increasingly larger population. By contrast, much less is known about the role of firm sorting for inequality across space.

In this project, we aim to fill this gap. We develop a theory of how heterogeneous firms sort across space and analyze the implications for spatial inequality. In our model, firms first decide where to produce and then compete for workers in frictional local labor markets that are characterized by job search of both unemployed and employed workers. If more productive firms sort into more productive locations, the local job ladder in these locations steepens, which amplifies spatial wage inequality. We use our framework to shed new light on spatial wage disparities in Germany, with special focus on the East-West divide. Our findings indicate that East Germany has significantly lower wages because it lacks highly productive firms.

Our choice to place local employment-to-employment transitions and local job ladders at the center of our theory is motivated by two observations from German administrative data. First, firms predominantly hire from their local labor market (around two-thirds of hires come from the commuting zone a firm is located in) and a substantial share of firms' hires (around 50%) are workers who were already employed in other firms.[1] This highlights the importance of firms' ability to poach workers from other local firms, and therefore the importance of firms' competitiveness in the local labor market. Second, on the worker side, we document that the wage growth from an employment-to-employment (EE) transition in the richest local labor market is more than twice as large as the wage gains in the poorest location. The fact that there is significant wage growth from switching jobs in all locations—but that, at the same time, these job ladders are steeper in rich labor markets—suggests that search frictions and on-the-job search are potentially important drivers of spatial inequality.

[1] For the detailed statistics, see Table A.2 in Appendix E.1.

We capture these observed features in a new model, in which the sorting of heterogeneous firms to heterogeneous locations is determined in equilibrium. Firms, which differ in productivity, first make their location choice. Then, they hire workers in a frictional local labor market, which is characterized by wage posting and an endogenous firm (or job) ladder. In each local labor market, both unemployed and employed workers search for jobs, all aiming to climb the local job ladder toward increasingly higher wages. Once matched, the worker-firm pair produces output that depends on the productivity of both the firm and location.

Firms face a fundamental trade-off when making their location choice. On the one hand, firms "like" productive locations with high TFP, because they boost output. These are locations with good fundamentals, which we interpret broadly; for instance, they can stem from modern infrastructure, productive spillovers, existing input-output networks, and workers' human capital. On the other hand, firms are hesitant to sort into such locations if many highly productive firms also choose to locate there. The presence of other productive firms pushes the firm into a low position on the local job ladder, which reduces its competitiveness in the local labor market and makes it difficult to poach and retain workers. Hence, firms' location decisions balance two considerations: local productivity and local competitiveness. While high location productivity increases firm output and per-worker profit, fierce competition with other firms reduces a firm's own capability to hire and keep workers, which curbs firm size. To our knowledge, this is the first model that integrates on-the-job search with firms' location choices, and highlights this novel trade-off.

We derive sufficient conditions for monotone firm sorting across space. Sorting is positive—i.e., better firms locate in more productive locations—if firm and location productivity are sufficiently complementary in production or if local labor market frictions are sufficiently large. Productive complementarities ensure that highly productive firms have greater willingness to pay for land in more productive places. In turn, sufficiently large labor market frictions (i.e., small job-to-job flows) ensure that the competition motive is of limited importance and does not outweigh this productivity consideration.

We show that under the conditions for monotone sorting, an equilibrium exists and is unique. Moreover, the main sorting trade-off, as well as the sufficient conditions for sorting, is robust to several extensions: endogenous labor mobility, endogenous location productivity spillovers that are determined by the local firm composition, and endogenous vacancy posting.

Our theory makes precise predictions on how firm sorting affects spatial inequality. First, positive firm sorting steepens the wage (job) ladder in more productive locations. Second, it leads to a stochastically better employment composition in more productive locations, so that more workers are employed in highly productive firms that pay higher wages. Both factors amplify the spatial wage premium of places that—due to higher local TFP—are more productive to start with.

To quantitatively assess the importance of spatial firm sorting for spatial inequality, we estimate our model on administrative data from Germany. A key aspect of our empirical strategy is to separately identify firm sorting from the fundamental productivity of a location. We prove that we can achieve identification from data on local labor shares and local value added. First, the variation in local labor shares identifies the extent of firm sorting. Because productive firms have more monopsony power within their local labor market, locations with more productive firms have lower labor shares. Empirically, we document the novel fact that the local labor share is *decreasing* in local GDP per capita, which is the variable we use to rank locations. Through the lens of our model, this implies that highly productive firms are in top-ranked locations—i.e., there is *positive sorting* between firms and locations. Second, the spatial variation in local value added per worker that is *not* due to differences in firm composition then pins down the fundamental productivity of each location.

We use our estimated model to better understand the sources of inequality between East and West Germany. To gauge the importance of spatial firm sorting, we conduct a counterfactual in which we match firms *randomly* to locations. We find that firm sorting can account for between 17% and 27% of the wage gap between East/West. The East is therefore not only disadvantaged because of poor economic fundamentals, but this weakness is amplified by the fact that low-productivity firms tend to cluster there. As for policy, integrating our segmented local labor markets into one market with a single economy-wide job ladder is an effective tool for reducing spatial inequality.

Related Literature. Our project merges two strands of the literature that have largely existed in isolation: the literature on frictional labor markets and cross-sectional wage dispersion and the urban literature on spatial inequality.

With respect to the literature on labor search and frictional wage dispersion (e.g., Burdett and Mortensen, 1998, Postel-Vinay and Robin, 2002), two important findings are that, conditional on worker heterogeneity, firm heterogeneity accounts for a sizable share of the cross-sectional wage dispersion (15-30%); and that search frictions and on-the-job search do as well (10-40%).[2] Despite this evidence on the importance of firms and search for inequality,

[2] Using a structural model, Postel-Vinay and Robin (2002) find a contribution of firm heterogeneity to wage dispersion of

there has been no attempt to link *spatial* inequality to the sorting of firms into local labor markets that are characterized by search frictions and job ladders. Our paper aims to fill this gap.

Second, there is a large literature on the sources of spatial wage inequality, with special focus on the urban wage premium in the U.S. (Glaeser and Maré, 2001, Duranton and Puga, 2004, Gould, 2007, Baum-Snow and Pavan, 2011, Moretti, 2011); in Spain (De La Roca and Puga, 2017); in France (Combes et al., 2008); and in Germany (Dauth et al., 2022). In turn, Heise and Porzio (2022) analyze the East-West German wage gap using a job ladder model and focus on *worker* mobility and preference frictions as the main source of the spatial divide. In contrast to our work, what these papers do not account for is the (endogenous) *spatial allocation of firms* as a driver of spatial inequality.[3]

A smaller but growing literature analyzes firms' location choices but differs from ours in focus and modeling choices. Combes et al. (2012) disentangle firm selection from agglomeration economies in the productivity advantage of cities by studying shifts and truncations of local productivity distributions. Behrens et al. (2014) develop a model of worker sorting, firm selection, and agglomeration economies to rationalize several stylized facts on the urban wage premium. Gaubert (2018) builds a model of spatial firm sorting in which firms trade off agglomeration economies and wage costs to analyze the efficiency impact of place-based policies. All of these papers feature friction*less* labor markets.[4] In turn, Bilal (2022) is the first paper that analyzes the effect of firms' location choices on spatial unemployment differences in a model that features labor market frictions. In contrast to our approach, he abstracts from on-the-job search, an ingredient that (as we will show) interacts with firm sorting in important ways to shape spatial inequality.

2 The Model

2.1 Environment

Time $t \in \mathbb{R}_+$ is continuous and the economy is in steady state. There is a continuum of locations (i.e., local labor markets) and a continuum of firms and workers.

Locations are indexed by ℓ and differ in exogenous productivity $A(\ell)$. We assume that $A(\ell)$

around 30% and a contribution of search frictions of around 40% in France. Bagger and Lentz (2019) find a firm contribution of 18%, while the contribution of the search channel is 10% in Denmark. Using a non-structural approach (two-way fixed effect regressions), firm effects typically explain around 20% of the variance of log-earnings or slightly less when correcting for limited mobility bias (Bonhomme et al., 2022). For instance, Card et al. (2013) find that 21% of the wage variance can be accounted for by workplace heterogeneity in Germany, and its importance for wage inequality has increased over time.

[3] Dauth et al. (2022) point out that empirically, more productive firms are in German cities as opposed to rural areas, but this is based on the distribution of AKM firm fixed effects, which conflate firm productivity with location productivity.

[4] Oberfield et al. (2022) analyze how firms with multiple plants sort across space in a frictionless context.

is strictly positive for all ℓ and continuously differentiable, and that locations are ordered by productivity, i.e., $\partial A(\ell)/\partial \ell > 0$. Each location has an exogenous amount of land, distributed with the continuously differentiable cdf R on $[\underline{\ell}, \overline{\ell}]$; $r > 0$ is the corresponding density.

In each location ℓ, there is a unit mass of risk-neutral homogeneous workers who are spatially immobile, something we relax below. Unemployed workers in ℓ receive flow benefit $b(\ell)$ and search for jobs, while employed workers receive a wage and do on-the-job search (OJS).

Firms are risk-neutral and differ in productivity p. We assume $p \sim Q(p)$, where $p \in [\underline{p}, \overline{p}]$ and Q is a continuously differentiable cdf with corresponding density $q > 0$. We call p the *ex ante productivity* of firms, based on which location choices are made. After settling in location ℓ, each firm with attribute p draws *ex post productivity* $y \in [\underline{y}, \overline{y}]$ from cdf $\Gamma(y|p)$ where Γ is continuously differentiable in both y and p. We assume that $\partial\Gamma(y|p)/\partial p < 0$ for all $y \in (\underline{y}, \overline{y})$, so that more productive firms ex ante draw their ex post productivity from better distributions in the first-order stochastic dominance (FOSD) sense. We distinguish between ex ante and ex post productivity so that, even with pure sorting between ex ante firm types and locations, we obtain a non-degenerate distribution of firm productivity in each location.

In order to produce in location ℓ, firms need to buy one unit of land at price $k(\ell)$ and post a wage to hire local workers. The returns to land accrue to a set of local landowners who operate in the background. Firms have no capacity constraint when employing workers, so they hire any worker that yields a positive profit. Firm y in location ℓ produces output $z(y, A(\ell))$ per worker hired. We assume that z is C^2 and strictly increasing in each argument. Note that while the ex ante productivity of firms p determines the distribution of ex post productivity y, p is irrelevant for production conditional on y. Hence, after entry, firms are fully characterized by their ex post productivity realization y. We assume that z is the output of the same homogeneous good in all locations, whose price is normalized to one. All agents discount the future at rate ρ.

In each location there is a frictional labor market, in which workers and firms face search frictions and search is random. In the baseline model, we assume that meeting rates are exogenous and constant across locations. Firms meet workers at Poisson rate λ^F. Employed workers' meeting rate is given by λ^E and unemployed workers' meeting rate by λ^U. Matches are destroyed at rate δ. We also denote the meeting rates of employed workers, unemployed workers, and firms *relative* to the job destruction rate by $\varphi^E \equiv \lambda^E/\delta$, $\varphi^U \equiv \lambda^U/\delta$ and $\varphi^F \equiv \lambda^F/\delta$. In our quantitative analysis we endogenize these meeting rates through endogenous labor mobility and a local matching function.

In terms of wage setting, we assume that firms post wages with commitment as in Burdett and Mortensen (1998). We denote the wage paid by firm y in location ℓ by $w(y,\ell)$. Hence, firm y in location ℓ receives flow profit $\pi(y,\ell) = z(y, A(\ell)) - w(y,\ell)$ when employing a worker.

We impose the following assumptions.

Assumption 1.

1. *(Common support) The distributions of ex post productivity $\Gamma(y|p)$ have common support: $\forall p, y \in [\underline{y}, \overline{y}]$.*
2. *(Zero profits for marginal firm type) In each location ℓ, firms with the lowest ex post productivity, $\underline{y}$, make zero profits.*

Together, these assumptions imply that—despite the endogenous sorting of firms across locations—the resulting distributions of firm productivity have common support across locations, with the marginal firm $\underline{y}$ making zero profits. While not strictly necessary, part 1. not only simplifies our analytical arguments but is also supported by the evidence.[5] Part 2. will be guaranteed by making sure that the output of the least productive firm equals the reservation wage, i.e., $w^R(\ell) = z(\underline{y}, A(\ell))$ for all ℓ. One way to ensure this property is by appropriately choosing non-employment utility $b(\ell)$ (a primitive) across locations.

2.2 Equilibrium

We now discuss the agents' decisions; namely, the job acceptance decisions of workers as well as firms' location choice and wage-posting decision. Finally, we specify the steady-state flow balance and market-clearing conditions.

Workers. Workers face two decisions: First, whether to accept a job offer when unemployed, and second, whether to accept a job offer when employed. We discuss each briefly, since both are standard.

Consider first a worker who is employed at wage w. The value of being employed at wage w in location ℓ, $V^E(w,\ell)$, solves the recursive equation

$$\rho V^E(w,\ell) = w + \delta(V^U(\ell) - V^E(w,\ell)) + \lambda^E \left[\int_{\underline{w}}^{\overline{w}} \max\{V^E(t,\ell), V^E(w,\ell)\} dF_\ell(t) - V^E(w,\ell)\right],$$

where F_ℓ is the endogenous wage-offer distribution in location ℓ and $V^U(\ell)$ denotes the value

[5] Using French data, Combes et al. (2012) find that firm productivity distributions across space do *not* vary in their left truncation, indicating that productivity of the least productive firms is similar across locations.

of unemployment, which is given by

$$\rho V^U(\ell) = b(\ell) + \lambda^U \left[\int_{\underline{w}}^{\overline{w}} \max\{V^E(t,\ell), V^U(\ell)\} dF_\ell(t) - V^U(\ell) \right]. \tag{1}$$

Note that, as is well known (and straightforward to show), V^E is increasing in w, so the optimal strategy of employed workers is to accept any wage higher than the current one.

In turn, the optimal strategy of unemployed workers is given by a reservation wage strategy. We obtain the reservation wage from a worker who is indifferent between accepting and rejecting a job, $V^E(w^R(\ell), \ell) = V^U(\ell)$, which—after simplifying the value functions—gives

$$w^R(\ell) = b(\ell) + (\lambda^U - \lambda^E) \left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta + \lambda^E(1 - F_\ell(t))} dt \right]. \tag{2}$$

Note that we can always sustain our Assumption 1.2—whereby $w^R(\ell) = z(\underline{y}, A(\ell))$—based on the appropriate choice of primitive function $b(\cdot)$, once F_ℓ is pinned down. For instance, in the special case of $\lambda^U = \lambda^E$, $b(\ell) = z(\underline{y}, A(\ell))$ satisfies our assumption.

Firms. Firms face two decisions. First, they choose location ℓ to maximize expected discounted profits, taking competition from other firms and land prices as given. Second, conditional on the location choice, firms post a wage to maximize profits. We solve backwards.

Wage Posting. When posting a wage, a firm in location ℓ trades off profit per worker against firm size, which is given by[6]

$$l(w, \ell) := \frac{\varphi^F}{\left(1 + \varphi^E \left(1 - F_\ell(w)\right)\right)^2}. \tag{3}$$

Firms that are higher ranked in local wage distribution F_ℓ are larger—since they poach more and are being poached less—compared with lower-ranked firms. Conversely, holding the firm's wage w fixed, its size is smaller if the local wage distribution, F_ℓ, is stochastically better. Importantly, the (relative) EE rate, φ^E, governs the extent to which firm size depends on local competition. If labor market frictions are severe, $\varphi^E \to 0$, then the competition channel is mitigated and firm size is independent of the local wage distribution.

The firm's wage-posting problem is then to maximize per-worker profit times its size,[7]

[6] Firm size can be derived as the hiring rate times the expected match duration; here we set $\rho \to 0$. Taking into account a consistency condition on λ^F (relating it to λ^E and λ^U), this is equivalent to the firm size definition in Burdett and Mortensen (1998), which is the measure of workers employed at y divided by the measure of firms with y. See Appendix A.2.

[7] To simplify exposition, we set $\rho \to 0$ for the remainder of the analysis.

$$\tilde{J}(y,\ell) = \max_{w \geq w^R(\ell)} l(w,\ell)(z(y,A(\ell)) - w), \tag{4}$$

whereby a higher wage increases firm size but reduces flow profits.[8]

Equation (4) highlights the fact that location matters to firms in two distinct ways, which already hints at their trade-off between local productivity and competition. On the one hand, choosing a high ℓ increases location TFP $A(\ell)$ and thus output and flow profits. On the other hand, if many productive firms sort into high-ℓ locations, competition is fierce (the wage offer distribution F_ℓ is stochastically better), and the size of any given firm y becomes compressed.

The firm's objective function (4) is supermodular in (w, y), which—in combination with a continuum of productivity levels—implies that w is strictly increasing in y. Therefore, the local distribution of wage offers coincides with the local distribution of firm productivity, $F_\ell(w(y,\ell)) = \Gamma_\ell(y)$, where Γ_ℓ is the *endogenous* productivity cdf of firms in location ℓ. Cdf Γ_ℓ encapsulates the spatial sorting of firms and is thus the crucial object in our model. In what follows, we will use Γ_ℓ instead of F_ℓ.

Making this substitution, we solve the firm's problem to obtain the well-known wage function under wage posting (Burdett and Mortensen, 1998),

$$w(y,\ell) = z(y,A(\ell)) - \left(1+\varphi^E(1-\Gamma_\ell(y))\right)^2 \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{\left(1+\varphi^E(1-\Gamma_\ell(t))\right)^2} dt, \tag{5}$$

with the exception that in our model, there is one such wage function in each location ℓ; it depends on both location productivity $A(\ell)$ and the endogenous distribution of firms in that location Γ_ℓ—the object we will turn to next.

Location Choice. Given the wage function for each location ℓ, we can now specify the firm's location choice problem. The ex ante expected value of firm p to settle in location ℓ is given by

$$\overline{J}(p,\ell) = \int \tilde{J}(y,\ell) d\Gamma(y|p) - k(\ell),$$

while taking into account that, when this choice is made, the firm still has to draw ex post productivity y and needs to pay land price $k(\ell)$ when settling in ℓ. Using $\tilde{J}(y,\ell)$ from (4) and wage function (5), $\overline{J}(p,\ell)$ can be expressed as

$$\overline{J}(p,\ell) = \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y} l(t,\ell) dt \, d\Gamma(y|p) - k(\ell). \tag{6}$$

[8] We show in Appendix A.1 that an equivalent way of formulating the wage-posting problem is to have firms maximize their hiring rate times their discounted flow profit.

The expected value for firm p of settling in location ℓ is given by the expected value of employment net of the price of the land. Hence, when choosing their location, firms balance local productivity $A(\ell)$ (which determines output $z(y, A(\ell))$), local competition Γ_ℓ (which determines their size $l(y,\ell)$), and land prices $k(\ell)$. Formally, the firm's location choice problem is

$$\max_{\ell} \overline{J}(p,\ell). \tag{7}$$

The solution to problem (7) describes firms' location decisions and is at the center of our analysis. The FOC of this problem highlights the fundamental location choice trade-off faced by firms and can be expressed as follows (see Appendix A.4):

$$\int_{\underline{y}}^{\overline{y}} \left(\frac{\partial \ln \left(\frac{\partial z(y,A(\ell))}{\partial y} \right)}{\partial \ell} + \frac{\partial \ln l(y,\ell)}{\partial \ell} \right) \frac{\partial z(y,A(\ell))}{\partial y} l(y,\ell)\left(1-\Gamma(y|p)\right) dy = \frac{\partial k(\ell)}{\partial \ell}, \tag{8}$$

where $\frac{\partial \ln l(y,\ell)}{\partial \ell}$ is the (semi-)elasticity of firm size wrt location ℓ and $\frac{\partial \ln \left(\partial z(y,A(\ell))/\partial y\right)}{\partial \ell}$ is the (semi-)elasticity of the firm's marginal product wrt ℓ.

FOC (8) reflects firms' trade-off between profitability and firm size when choosing the optimal ℓ. Locations with higher ℓ, by virtue of having higher productivity $A(\ell)$, push up output and thus firm profits per employee. But if these locations attract many productive firms, competition in high-ℓ locations is fierce; poaching and retaining workers is then difficult, which reduces firm size. At the optimal location choice, this marginal (net) benefit of choosing a higher ℓ equals its marginal cost, which is the increase in the price of land. If high-ℓ locations are overall more desirable, then $\partial k(\ell)/\partial \ell > 0$, which reflects the fact that these locations command higher land prices.

This FOC—along with land market clearing—pins down the equilibrium allocation of firms to locations, captured by Γ_ℓ. That is, for all ℓ,

$$\Gamma_\ell(y) = \int_{\underline{p}}^{\overline{p}} \Gamma(y|p) m_p(p|\ell) dp \qquad \forall y \in [\underline{y}, \overline{y}], \tag{9}$$

where we define by $m(\ell,p)$ the endogenous joint matching density between (ℓ,p) with conditional densities $m_\ell(\ell|p)$ and $m_p(p|\ell)$.[9] In addition, the FOC pins down the land price schedule, $k(\cdot)$, that sustains this allocation. It is obtained by solving differential equation (8), evaluated at the equilibrium assignment (see Appendix A.5). The land price in location ℓ is given by

[9]Under a measure-preserving matching between firms p and locations ℓ, the marginal densities of m are given by r and q.

the cumulative marginal contributions of land to the match surplus (between firms and land) in locations that are weakly less productive than ℓ.[10]

Land Market Clearing. The land market clearing condition is given by

$$R(\ell) = \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\overline{p}} m(\tilde{\ell}, \tilde{p}) d\tilde{p} d\tilde{\ell} = \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\overline{p}} m_\ell(\tilde{\ell}|\tilde{p}) q(\tilde{p}) d\tilde{p} d\tilde{\ell}, \tag{10}$$

which ensures that the mapping between firms' productivity distribution Q and land distribution R is measure-preserving.

Good Market Clearing. In each location ℓ, workers, firms and land owners consume their entire income. Total income thus equals total consumption, which in turn equals total output, so that the good market clears in each ℓ,

$$\int_{\underline{y}}^{\overline{y}} z(y, A(\ell)) l(y, \ell) d\Gamma_\ell(y) = \int_{\underline{y}}^{\overline{y}} w(y, \ell) l(y, \ell) d\Gamma_\ell(y) + \bar{J}(\mu(\ell), \ell) + k(\ell), \tag{11}$$

where we use firm size, $l(y, \ell)$, defined in (3).

Flow-Balance Conditions. We have two flow-balance conditions in steady state, which pin down the equilibrium unemployment rate and the distribution of employment in each location.

First, the inflow into and outflow out of unemployment must balance, which pins down the unemployment rate, $u(\ell)$ (which in our baseline model does not vary across ℓ):

$$\delta(1 - u(\ell)) = u(\ell)\lambda^U \quad \Rightarrow \quad u(\ell) = \frac{1}{1 + \varphi^U}. \tag{12}$$

Second, the inflow into and outflow out of employment in firms with productivity below y must balance (for all y), where we take into account the optimal decision of employed workers to accept the job offer of any firm that is more productive than their current one. This determines the cdf of employment in location ℓ, G_ℓ:

$$u(\ell)\lambda^U \Gamma_\ell(y) = (\delta + \lambda^E(1 - \Gamma_\ell(y))) G_\ell(y)(1 - u(\ell)) \quad \Rightarrow \quad G_\ell(y) = \frac{\Gamma_\ell(y)}{1 + \varphi^E(1 - \Gamma_\ell(y))}. \tag{13}$$

Note that the outflow of workers from firms with productivity below y, $G_\ell(y)(1-u(\ell))$, has two sources: exogenous job destruction (driven by δ) and endogenous on-the-job search, which induces workers to quit for better jobs when they find them (which happens at rate $\lambda^E(1 - \Gamma_\ell(y))$).

[10] In this competitive land market, firms that maximize expected profits and landowners who maximize land prices will result in the same allocation of firms to locations, which is why we detail only one side's decision: the one by firms.

Steady-State Equilibrium. We can now define a steady-state equilibrium.

Definition 1. *A steady-state equilibrium is a tuple* $(w(\cdot,\ell), k(\ell), m(\ell,p), \Gamma_\ell(\cdot), l(\cdot,\ell), G_\ell(\cdot), u(\ell), w^R(\ell))$, *such that for all* $\ell \in [\underline{\ell}, \overline{\ell}]$ *and* $p \in [\underline{p}, \overline{p}]$:

1. *Walrasian equilibrium in the land market: The pair* $(k(\ell), m(\ell,p))$ *is a competitive equilibrium of the land market, pinning down* Γ_ℓ *and also* $l(\cdot,\ell)$;
2. *Optimal wage posting:* $w(\cdot,\ell)$ *is consistent with* (4) *for all firm types* $y \in [\underline{y}, \overline{y}]$;
3. *Optimal worker behavior: Employed workers accept job offers from more productive firms; unemployed workers accept any job* y *with* $w(y,\ell) \geq w^R(\ell)$, *where* $w^R(\ell)$ *is pinned down by* (2);
4. *Flow balance conditions* (12) *and* (13) *hold, pinning down* $u(\ell)$ *and* G_ℓ;
5. *Good market clearing* (11) *holds.*

3 Equilibrium Analysis

3.1 Spatial Firm Sorting

We now analyze the patterns of firm sorting that occur in equilibrium. In particular, we provide conditions under which more productive firm types p sort into more productive locations ℓ. This is an allocation with positive assortative matching (PAM), which, as we show below, is the empirically relevant case.[11]

Sufficient Conditions for Positive Sorting. We focus on pure assignments between (p, ℓ), in which any two firms of the same type are matched to the same location type (and vice versa). Assignment $m_p(p|\ell)$ can then be captured by a matching *function* $\mu : [\underline{\ell}, \overline{\ell}] \to [\underline{p}, \overline{p}]$. We define positive sorting in a standard way.

Definition 2 (Positive Sorting of Firms to Locations)**.** *There is positive sorting in* (p, ℓ) *if matching function* μ *is strictly increasing.*

Under positive sorting, more productive firms sort into more productive locations. Moreover, $m_p(p|\ell)$ has positive mass only at a single point $p = \mu(\ell)$, and we can simplify the endogenous distribution of firms in location ℓ in (9) to[12]

$$\Gamma_\ell(y) = \Gamma(y|\mu(\ell)),$$

[11]For completeness, we analyze the allocation with negative assortative matching (NAM) in Online Appendix OA.1.

[12]Cdf $M_p(\cdot|\ell)$ (corresponding to density $m_p(\cdot|\ell)$) is a Dirac measure that concentrates its mass at $p = \mu(\ell)$ and (9) becomes

$$\Gamma_\ell(y) = \int_{\underline{p}}^{\overline{p}} \Gamma(y|p) m_p(p|\ell) dp = \int_{\underline{p}}^{\overline{p}} \Gamma(y|p) dM_p(p|\ell) = \Gamma(y|\mu(\ell)).$$

so that high-ℓ locations have ex post productivity distributions that are stochastically better.

To obtain *sufficient* conditions for positive sorting, recall that firm p chooses location ℓ to maximize $\overline{J}(p, \ell)$ given in (7). Based on results from the literature on monotone comparative statics (Milgrom and Shannon, 1994), the optimal location choice is (weakly) increasing in p if $\overline{J}(p, \ell)$ satisfies a strict single-crossing property in (p, ℓ). Then, due to the assumption of strictly positive densities r and q, μ is indeed *strictly* increasing. Note that strict supermodularity of $\overline{J}(p, \ell)$ in (p, ℓ) is sufficient for the strict single-crossing property. This discussion leads to our first result on the sufficient conditions for positive sorting.

Proposition 1 (Spatial Sorting of Firms I)**.** *Sorting is positive in equilibrium if* $\overline{J}(p, \ell)$ *is strictly supermodular in* (p, ℓ).

The spirit of Proposition 1 is familiar: Complementarities lead to positive sorting. We now derive conditions that guarantee this property of $\overline{J}(p, \ell)$. We postulate that firms anticipate positive sorting when making their location choices, and check that their optimal behavior indeed induces PAM.

Recalling how $\overline{J}(p, \ell)$ varies with ℓ (see FOC (8)) and using the assumption that p shifts $\Gamma(y|p)$ in the FOSD sense, we note that the supermodularity of $\overline{J}(p, \ell)$—and thus firm sorting—is controlled by the location choice trade-off between productivity gains and competition:[13]

$$\frac{\partial^2 \overline{J}(p, \ell)}{\partial p \partial \ell} > 0 \qquad \text{if} \qquad \underbrace{\frac{\partial \ln \left(\frac{\partial z(y, A(\ell))}{\partial y} \right)}{\partial \ell}}_{\text{Productivity Gains}} + \underbrace{\frac{\partial \ln l(y, \ell)}{\partial \ell}}_{\text{Competition}} > 0. \tag{14}$$

Whereas the local productivity gains from settling into high-ℓ locations are positive if production technology z is supermodular (first term in (14)), the local competition effect is negative under positive sorting since productive firms cluster in the best locations (second term in (14)). Positive sorting thus requires that the productivity benefits that boost profits per worker outweigh the costs from competition that translate into lower expected firm size.

To guarantee the supermodularity of $\overline{J}(p, \ell)$ in (p, ℓ) in terms of primitives, we use (14) and the expression for firm size $l(y, \ell)$ in (3) to obtain the following sufficient condition for PAM (note that Γ_ℓ can be expressed in terms of primitives only):[14]

[13] This sufficient condition is equivalent to the supermodularity of $\tilde{J}$, i.e. $\partial^2 \tilde{J}(y, \ell)/\partial y \partial \ell > 0$.

[14] When agents conjecture positive sorting, the endogenous firm distribution in location ℓ is given by $\Gamma_\ell(y) = \Gamma(y|\mu(\ell)) = \Gamma(y|Q^{-1}(R(\ell)))$, where we made use of candidate equilibrium matching function $\mu(\ell) = Q^{-1}(R(\ell))$, derived from land market clearing (10); see Appendix A.6. Therefore, we obtain $\frac{\partial \Gamma_\ell(y)}{\partial \ell} = \frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p} \frac{r(\ell)}{q(Q^{-1}(R(\ell)))} < 0$ (i.e., better locations have better firms in a FOSD sense) and (15) is a condition for PAM that only depends on primitives.

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\varphi^E}{1+\varphi^E\left(1-\Gamma_\ell\left(y\right)\right)} \left(-\frac{\partial \Gamma_\ell\left(y\right)}{\partial \ell}\right). \tag{15}$$

In order for PAM to materialize, the marginal productivity gains of settling in a better location (LHS) have to outweigh the costs of tougher competition by highly productive firms (RHS). Productivity gains are large if productivity differences across locations are large (the A-schedule is steep), and when complementarities of z in $(y, A(\ell))$ are strong. In turn, the cost of local competition that depresses firm size depends on two forces: first, on how endogenous firm distribution Γ_ℓ varies across space and, second, on the degree of labor market frictions φ^E that determines the impact of changes in Γ_ℓ on firm size. The cost of sorting into a high-ℓ region is low when λ^E—the rate at which employed workers meet firms—is small, since then poaching and competition do not matter much. The cost is also low if δ is large, so that match duration is mainly determined by workers who separate into unemployment as opposed to quitting. In this case, hiring results predominantly from unemployment and, again, poaching considerations carry less weight. The ratio φ^E captures both of these forces. A small φ^E weakens the competition channel so that it does not interfere with the productivity motive for positive spatial sorting. In the absence of on-the-job search, $\varphi^E = 0$, complementarities in production are enough to sustain positive sorting.

Formally, we need to ensure that the minimum of the elasticity of firms' marginal product with respect to location (productivity channel), denoted by

$$\varepsilon^P \equiv \min_{\ell,y} \frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}},$$

is positive and large enough; or that the competition channel (which gains importance in a fluid labor market when $\varphi^E = \lambda^E/\delta$ is large) is sufficiently weak.

Proposition 2 (Spatial Sorting of Firms II)**.** *If z is strictly supermodular, and either the productivity gains from sorting into higher ℓ, ε^P, are sufficiently large, or the competition forces are sufficiently small (i.e., φ^E is sufficiently small), then there is positive sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(R(\ell))$.*

The proof is in Appendix B.1, where we make the statements regarding "sufficiently large ε^P" and "sufficiently small φ^E" precise. Note that under the conditions of Proposition 2, the productivity gain from settling into high-ℓ locations outweighs the cost from competition for firms of *all* y-types. But—due to productive complementarities between (A, y)—the net benefit is especially high for those firms with high y, which high ex ante productivity p yields stochastically. Highly productive firms are thus willing to pay higher land prices, outbidding the less productive firms in the competition for land in high-ℓ locations. As a consequence, positive sorting arises, whereby high-ℓ locations have more productive firms in a FOSD sense, $\partial\Gamma_\ell/\partial\ell = \big(\partial\Gamma(y|\mu(\ell))/\partial p\big)\,\mu'(\ell) \leq 0$.

3.2 Existence and Uniqueness

We also show that when sorting is positive, a unique equilibrium exists.

Proposition 3. *Assume that the conditions from Proposition 2 hold. Then, a unique equilibrium (up to a constant of integration) exists.*

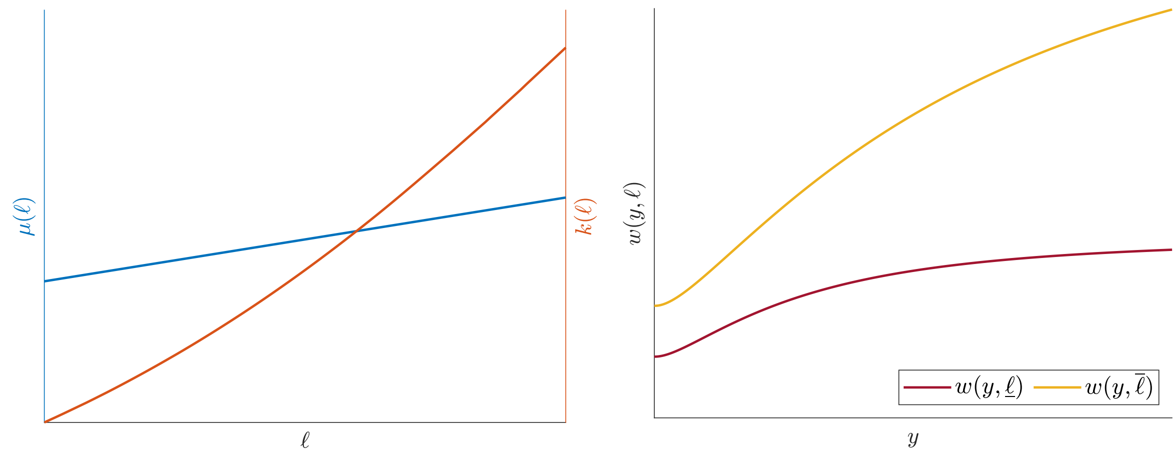
The proof is in Appendix B.2. We show the existence of a fixed point in Γ_ℓ by construction. In turn, uniqueness arises because, under the conditions on primitives stated in Proposition 2, the impact of endogenous firm composition on firms' value function leaves the complementarity properties of $\overline{J}$ unchanged.

3.3 Illustrating the Properties of the Equilibrium

In our model, spatial firm sorting, local land prices, and location-specific wages are jointly determined in equilibrium. Figure 1 illustrates the main properties of the equilibrium with positive firm sorting. In the left panel, which depicts the land market equilibrium, we plot both the matching function, μ, and the land price schedule, k, as a function of the location index ℓ. Positive sorting of firms is captured by the fact that μ is upward sloping: Firms with higher ex ante productivity p are matched to locations with higher ℓ (and thereby higher productivity, $A(\ell)$). Equilibrium land price schedule k sustains this allocation: Its positive slope ensures that high-ℓ locations are sufficiently expensive so that only firm types with the highest willingness to pay—i.e., the most productive ones—settle there.

The right panel of Figure 1 displays the equilibrium wage schedule, $w(\cdot, \ell)$, which we also call the *local job ladder*, as a function of firms' ex post productivity y for the top and bottom location, $\overline{\ell} > \underline{\ell}$. Two properties stand out. First, the wage intercept is increasing in ℓ. Second, while more productive firms pay higher wages everywhere, the local job ladder is steeper in high-ℓ locations. Below, we show that the intercept reflects spatial differences in location TFP A, whereas the differential steepness is also impacted by firm sorting.

Figure 1: Land Market (left) and Local Labor Market Equilibrium (right): An Illustration



Notes: The left panel displays matching function μ (blue) and land price schedule k (orange). The right panel displays wage function $w(\cdot,\ell)$ for a high-productivity ($\ell = \overline{\ell}$, yellow) and a low-productivity ($\ell = \underline{\ell}$, red) location.

We will show that these qualitative features of our equilibrium strongly resemble what we observe in the data. In our empirical and quantitative analyses, we investigate to what extent firms sort positively across space (Figure 1, left) and the implications for cross-location wage disparities (Figure 1, right). But before turning to the data, we clarify the relationship between firm sorting and spatial wage inequality in theory.

3.4 Spatial Firm Sorting and Spatial Wage Inequality

We now turn to the economic implications of firm sorting. First, we highlight the link between spatial sorting and spatial wage inequality. Second, we show that firm sorting yields distinct testable predictions on the difference between firms' global and local productivity ranks. We examine both of these features in our empirical analysis below.

The Spatial Wage Premium. The differences in wage schedules across locations (right panel of Figure 1) suggest that more productive locations pay higher wages. To systematically investigate this issue, as well as the role firm sorting plays in it, we measure spatial inequality as follows. We consider the average wage in location ℓ relative to the least productive location $\underline{\ell}$, which we call the *spatial wage premium*:

$$\frac{\mathbb{E}[w(y,\ell)|\ell]}{\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]} = \frac{w(\underline{y},\ell) + \int_{\underline{y}}^{\overline{y}} \frac{\partial w(y,\ell))}{\partial y}(1-G_\ell(y))dy}{w(\underline{y},\underline{\ell}) + \int_{\underline{y}}^{\overline{y}} \frac{\partial w(y,\underline{\ell})}{\partial y}(1-G_{\underline{\ell}}(y))dy}.$$

To illustrate the drivers of spatial inequality, we consider how this measure varies as we increase ℓ, comparing the average wage of more productive locations with the wage of the least productive one. The derivative of the spatial wage premium wrt ℓ is equal in sign ($\overset{\text{S}}{=}$) to

$$\frac{\partial \frac{\mathbb{E}[w(y,\ell)|\ell]}{\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]}}{\partial \ell} \overset{\text{S}}{=} \underbrace{\frac{\partial z(\underline{y}, A(\ell))}{\partial \ell}}_{\text{intercept of job ladder}} + \int_{\underline{y}}^{\overline{y}} \Bigg(\underbrace{\frac{\partial^2 w(y,\ell)}{\partial y \partial \ell}}_{\text{steepness of job ladder}} (1 - G_\ell(y)) + \underbrace{\frac{\partial w}{\partial y}\Bigg(- \frac{\partial G_\ell(y)}{\partial \ell}\Bigg)}_{\text{employment composition}} \Bigg) dy. \quad (16)$$

There are two fundamental differences between locations ℓ and $\underline{\ell}$: Location ℓ has higher TFP, $A(\ell)$, and—in our equilibrium with positive sorting—a better distribution of firms, Γ_ℓ. These differences drive the three factors that underlie cross-location wage inequality. The first two are displayed in Figure 1: First, higher-ℓ locations have a higher intercept of the wage function and thus job ladder, due to higher location TFP. Second, there is larger spatial inequality if higher-ℓ locations have a steeper wage function (i.e., a *steeper job ladder*). This steepness is driven by their higher A (TFP) along with complementarities between A and y in production, as well as by positive firm sorting, since tougher competition for workers among highly productive firms bids up wages. Finally, under positive spatial sorting, higher-ℓ locations have better firms and hence a stochastically better employment distribution: More employment is clustered at the upper part of the wage schedule, where wages are higher. We summarize:

Proposition 4. *Assume the conditions from Proposition 2 hold. There exists a spatial wage premium, $\mathbb{E}[w(y,\ell)|\ell]/\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}] > 1$, since compared with location $\underline{\ell}$, location ℓ has*
(i) a higher intercept of the wage function, $z(\underline{y}, A(\ell))$;
(ii) a steeper wage function, provided that the complementarities of z in (y, ℓ) are large enough;
(iii) a stochastically better firm, and thereby employment distribution G_ℓ.

In Appendix B.3, we make the sufficient condition on z in (ii) precise. This decomposition of spatial wage inequality will guide our quantitative analysis below.

Detecting Firm Sorting: Global vs. Local Rank. Given the importance of spatial firm sorting for spatial inequality in our model, a natural question is how to detect firm sorting in the data. We show that firm sorting has distinct implications for the relationship between the local and the global (economy-wide) productivity rank of firms.

We define the difference between firm y's global rank and its (average) local rank as

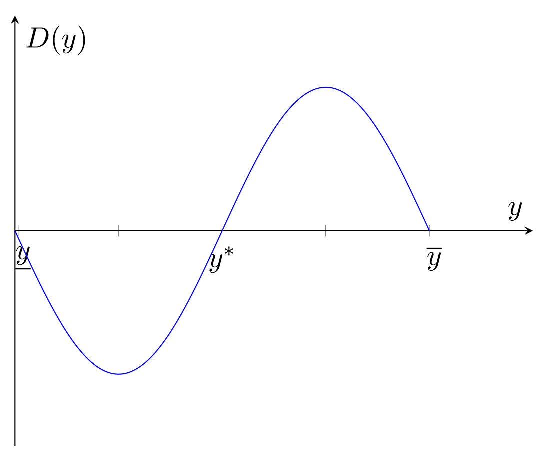
$$D(y) := \underbrace{\int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) r(\ell) d\ell}_{\text{Global Rank}} - \underbrace{\int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) \frac{\gamma(y|\mu(\ell)) r(\ell)}{\int_{\underline{\ell}}^{\overline{\ell}} \gamma(y|\mu(\hat{\ell})) r(\hat{\ell}) d\hat{\ell}} d\ell}_{\text{Average Local Rank}}.$$

The global rank reflects the firm's position in the economy-wide productivity ranking. By contrast, the local rank reflects the firm's position in the productivity ranking of its location. It takes into account that firms of a given type y can be found in all locations but, because of sorting, they are more prevalent in some locations than others. We therefore average the local rank of firm type y, $\Gamma_\ell(y)$, across locations using the density that describes the distribution of y across space (see Appendix B.4 for the detailed derivation of the local rank).

Spatial sorting by firms has distinct implications for the shape of D. In particular, if sorting is monotone, there is a concentration of highly productive firms in some locations and of much less productive firms in others. As a consequence, the local rank of highly productive firms is *low* relative to their global rank, which yields $D > 0$. The opposite is true for the least productive firms who are surrounded by other low-productivity peers in their locations. As a result, their local rank tends to be *high* compared with their global rank, with $D < 0$. Finally, $D(\underline{y}) = D(\overline{y}) = 0$ because the worst (best) firm economy-wide is also the worst (best) firm in any local labor market.

Figure 2 depicts D for a parametric example with spatial sorting that we detail in Appendix B.4 (see also Proposition 6 in that appendix for a general statement on the shape of D under sorting, which shows that the features of Figure 2 are common). Note that the difference between global and local ranks is absent (i.e., $D(y) = 0$ for all y) if there is no firm sorting.

Figure 2: Spatial Firm Sorting and the Difference between Global and Local Productivity Ranks



4 Empirical Analysis

First, we empirically assess our model's *qualitative* predictions on firm sorting and spatial wage inequality. Below, we also highlight the *quantitative* implications for spatial inequality.

4.1 Data and Measurement

For our individual-level analysis, we draw from linked employer-employee data (LIAB) in Germany provided by the Research Data Centre (FDZ) of the German Federal Employment Agency; see Appendix D.2 for details. These data have a panel dimension and are based on workers' social security records. We make particular use of information on wages and worker flows. Our main sample covers the period 2010-2017.

We base our analysis on 257 commuting zones (CZ)—our local labor markets—which are similar to US commuting zones (see Appendix D.4). They are defined for the year 2017 by the German Federal Office for Building and Regional Planning. In Table A.1 (Appendix E.1), we report some basic statistics, which show that average wages, firm size, and value added differ substantially across commuting zones.

4.2 Spatial Firm Sorting and Spatial Wage Inequality: Evidence

A key question for us is to what extent firm sorting amplifies spatial wage inequality. Spatial heterogeneity in wages is substantial. In particular, a set of local labor market fixed effects explains more than 20% of the total dispersion of log wages (see Table A.5 in Appendix E.3). Note that throughout, we deflate wages using a nationwide CPI.[15]

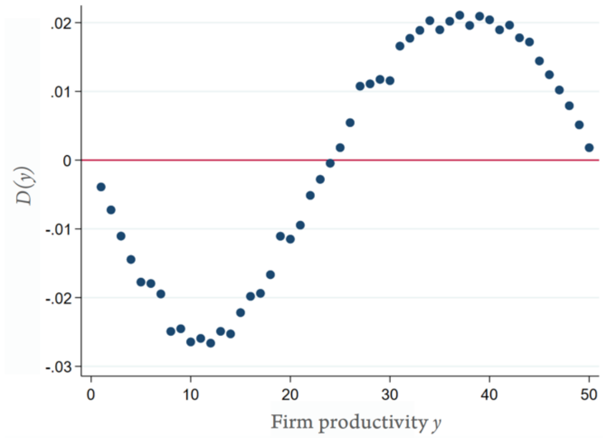
Our theory, and especially decomposition (16), highlights the fact that this spatial variation in wages is impacted by differences in the local job ladder across space. Positive spatial sorting by firms implies that productive locations attract better firms, which leads to both a steeper job ladder and an employment distribution that is concentrated at more productive firms who pay higher wages. Both of these forces boost spatial wage inequality.

To detect spatial sorting in the data, our model (Proposition 6 in Appendix B.4 and Figure 2) suggests a simple test: If there is monotone spatial sorting, there is an S-shaped relationship between the difference in firms' global and local ranks, $D(y)$, and productivity y.[16] In Figure 3, we plot this relationship in the data. On the horizontal axis, we order firms by their global productivity rank and categorize them into 50 equally sized bins (based on percentiles of the global productivity distribution). On the vertical axis, we display the average of the difference between global and local ranks for each productivity bin. The resemblance to our theoretical result, displayed in Figure 2, is noticeable.

[15]This is consistent with our model, which features nominal wages (the good's price is normalized to 1). Importantly, we show in Table A.4 (Appendix E.3) that spatial inequality remains substantial even when adjusting for regional price deflators.

[16]We proxy firm productivity y by a residualized AKM firm fixed effect. The residualization aims to purge location productivity $A(\ell)$ from the fixed effect, so that it only captures firm characteristics; see Appendix D.2. Note that our structural estimation below will allow us to directly identify firm productivity, which circumvents the need for this procedure.

Figure 3: Difference between Global and Local Productivity Rank



Notes: Data source: LIAB-BHP. We rank firms by their residualized AKM fixed effects and group them in 50 bins of equal size. For each bin, we measure local rank and global rank and plot the average difference, denoted by $D(y)$.

Globally unproductive firms sort into locations with a high concentration of unproductive competitors. Hence, their global rank is below their average local rank, i.e., $D(y) < 0$. In turn, for globally productive firms, the opposite pattern arises: They co-locate with other productive firms—i.e., within their local labor market they are relatively unproductive compared to their economy-wide productivity—and therefore $D(y) > 0$. Recall that if there is no spatial firm sorting, we would observe that D is a horizontal line and zero everywhere.

We now provide (indirect) evidence on how the documented firm sorting affects spatial inequality through its impact on local job ladders.[17] To assess how local job ladders vary across locations, we need to measure them. First, we focus on the ratio between a location's average wage and the wage of those who just made an unemployment-to-employment (UE) move.[18] This ratio is increasing in local GDP per capita—our measure of location prosperity that orders ℓ—which suggests that workers face a steeper job ladder in high-ℓ places (Figure 4, left).[19]

As a second measure of the local job ladder, we focus on the wage growth associated with an EE transition. To assess how these gains vary across locations, we run the following regression:

$$\frac{w_{i\ell,t} - w_{i\ell,t-1}}{w_{i\ell,t-1}} = \sum_{\ell=1}^{257} \beta_\ell + \sum_{\ell=1}^{257} \beta_\ell^{EE}\, \mathrm{EE}_{i\ell,t} + \sum_{\ell=1}^{257} \beta_\ell^{EXT}\, \mathrm{EXT}_{i\ell,t} + \varepsilon_{i\ell,t}, \tag{17}$$

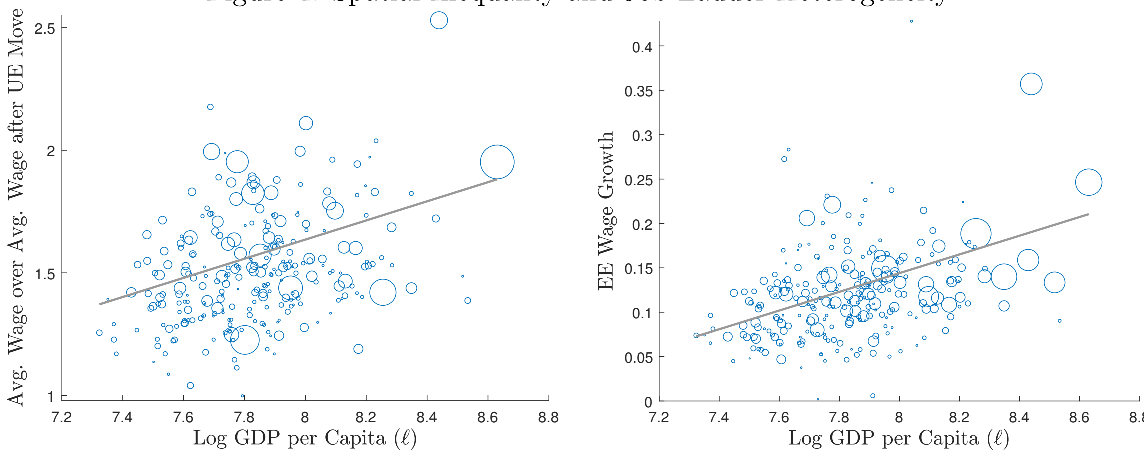
where $w_{i\ell,t}$ is the wage of individual i in CZ ℓ and month t, $EE_{i\ell,t}$ indicates that individual i made an EE move to a job in CZ ℓ between months t and $t-1$, and $EXT_{i\ell,t}$ indicates an EE

[17] In Table A.3 (Appendix E.2), we provide direct evidence for the importance of firms' local competitiveness for their decisions. In line with our theory (see Lemma O1 in Online Appendix OB), firms that are higher up in the *local job ladder*—as opposed to the global one—have higher poaching shares; pay higher wages; and are larger.

[18] One advantage of this ratio is that it controls for spatial differences in location TFP $A(\ell)$. Specifically, if production function z is multiplicative, as assumed below, then $A(\ell)$ cancels from this ratio for each ℓ.

[19] We obtain data on local GDP per capita from the German Federal Statistical Office; see Appendix D.1.

Figure 4: Spatial Inequality and Job Ladder Heterogeneity



Notes: Data source: LIAB-BHP. Both panels plot a linear fit. CZs are weighted by their number of wage observations (left) or EE moves (right), indicated by different sizes of the CZ-dots. The right panel plots β_ℓ^{EE} across ℓ based on (17).

transition to a job in CZ ℓ from a job outside of ℓ. The coefficients $(\beta_\ell, \beta_\ell^{EE}, \beta_\ell^{EXT})$ are CZ fixed effects and CZ-specific returns to EE moves from within/outside the CZ, respectively.

We are interested in the effect of an EE transition *within* local labor market ℓ on wage growth, i.e., β_ℓ^{EE}, which we plot against ℓ in the right panel of Figure 4. Importantly, wage growth is higher in high-ℓ locations. Quantitatively, these differences are meaningful: A single job-to-job move in the richest German local labor market increases wages by around 20%, which is more than twice as much compared with the poorest location.[20] These effects remain virtually unchanged when we also control for individuals' age, gender, and education; see Figure A.1 in Appendix E.3. We conclude that more productive locations indeed have steeper job ladders, as is the case under positive firm sorting in our theory.

To get a first sense of the quantitative impact of this job ladder heterogeneity on spatial wage inequality, we perform a statistical decomposition of the spatial wage gap in lifetime earnings into the parts driven by (i) starting wages, (ii) wage growth due to EE moves, (iii) wage growth during continuing job spells, and (iv) wage growth of the frequently unemployed. We follow a single cohort from 2002 to 2017 in two regions: the poorest 25% of locations in terms of GDP per capita and the richest 25%. We determine how much of the spatial earnings gap that emerges 15 years into workers' careers is due to differential EE wage growth and thus different local job ladders. We find that 24% of spatial inequality in lifetime earnings is due to this channel, which underscores the important role of local job ladders in the analysis of spatial wage inequality. See Table A.6 in Appendix E.3 for details on this exercise.

[20]In contrast to the evidence presented in Figure 4 (left), our finding in the right panel is not subject to the concern that there may be more human capital accumulation in high-ℓ locations (a mechanism that is not in our model), which would generate a positive slope in Figure 4 (left) and thus spatial wage inequality even in the absence of spatial firm sorting.

5 Estimation

To rigorously evaluate the quantitative importance of firm sorting for spatial wage inequality, we estimate our model. To this end, we enrich our model along four dimensions and show that it is identified. We then discuss our estimation strategy, results, and model fit.

5.1 Bringing our Model to the Data

Setting. We make four main changes to render our model suitable for estimation while preserving its key mechanism. First, we relax the assumption of fully immobile labor and allow unemployed workers to settle in any location. This feature is important, since even though we observe a high degree of local hiring, local labor markets are not perfectly segmented in the data.[21] Second, we introduce a residential housing market in each location, so that workers now use their flow income to consume not only the final good but also housing. Third, we introduce local amenities that can vary across space. These amenities scale the real consumption utility of workers, which allows us to fit observed residential house prices while ensuring that unemployed workers' value of search is the same across space. Last, we allow job separation rates δ to vary (exogenously) across locations.

This setting endogenizes local population size, $L(\ell)$, and thereby also the local meeting rates of workers and firms. We assume that in each location there is a labor market matching function with constant returns to scale, so that local meeting rates are determined by local market tightness, $\theta(\ell) = \mathcal{V}(\ell)/\mathcal{U}(\ell)$. The measure of vacancies per unit of land in ℓ still satisfies $\mathcal{V}(\ell) = 1$, since the measure of vacancies in each location equals the measure of firms that settle there.[22] In turn, $\mathcal{U}(\ell) = L(\ell)(u(\ell) + \kappa(1 - u(\ell)))$ is the effective measure of searchers per unit of land in ℓ, impacted by the endogenous $L(\ell)$. An important implication is that firms' and workers' meeting rates, $(\lambda^F(\ell), \lambda^U(\ell), \lambda^E(\ell))$, as well as the unemployment rate, $u(\ell)$, can vary across locations. These location-specific meeting rates create congestion, which is an additional channel that affects the costs of competition and thus firm sorting.

The residential housing (or rental) market features exogenous supply, $h(\ell)$, in each location. Workers have Cobb-Douglas preferences over the final good and housing. We denote the share of income that is spent on housing (the final good) by ω $(1 - \omega)$. The income of employed

[21] Introducing mobility of the unemployed (as opposed to the employed) preserves the structure of our model and, moreover, employed workers are less mobile empirically: ∼90% of them are hired from a 100 km radius around the firm.

[22] To see this, note that under positive sorting, each ℓ is chosen by a single p, where we assume that for each p, there is a continuum of firms i s.t. $0 \leq i \leq q(p)$ with Lebesgue measure (i.e., a continuum of mass $Q'(p) = q(p)$). Under the equilibrium matching $p = \mu(\ell)$, so the mass of firms in location ℓ is $Q'(\mu(\ell)) = q(\mu(\ell))\mu'(\ell)$. Combined with the fact that in any ℓ the measure of firms equals the measure of vacancies, we have $\mathcal{V}(\ell)r(\ell) = q(\mu(\ell))\mu'(\ell)$ and thus $\mathcal{V}(\ell) = 1$.

workers is wage $w(y, \ell)$ and that of unemployed workers is benefit $w^U(\ell)$, financed via taxes on homeowners' income, τ.[23] Further, the government budget needs to balance

$$\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell), \tag{18}$$

where $d(\ell)$ is the housing price in ℓ, which adjusts in order to achieve housing market clearing:

$$h(\ell) = \omega \frac{w^U(\ell)}{d(\ell)} u(\ell)L(\ell) + \omega \frac{\mathbb{E}[w(y,\ell)|\ell]}{d(\ell)}(1 - u(\ell))L(\ell). \tag{19}$$

Housing supply $h(\ell)$ equals housing demand from both unemployed and employed workers.

The population size in each ℓ, $L(\ell)$, is pinned down by the fact that in equilibrium, each worker must be indifferent between any two locations—i.e., the value of search is equalized across space,

$$V^U(\ell') = V^U(\ell'') \qquad \forall \ell' \neq \ell'',$$

where $V^U(\ell)$, compared with (1) in the baseline model, reflects the fact that high local house prices, $d(\ell)$, low job-finding rates, $\lambda^E(\ell)$, and high separation rates, $\delta(\ell)$, render job search in location ℓ less attractive. In contrast, favorable local amenities, $B(\ell)$, render it more attractive:

$$\rho V^U(\ell) = B(\ell)d(\ell)^{-\omega} \left(z(\underline{y}, A(\ell)) + \lambda^E(\ell) \left[\int_{z(\underline{y},A(\ell))}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right] \right). \tag{20}$$

If location ℓ', for instance, has a better wage distribution than location ℓ'' (causing a temporary imbalance $V^U(\ell') > V^U(\ell'')$), workers will move into ℓ'. This puts downward pressure on market tightness (and thus workers' meeting rates) and upward pressure on housing prices in ℓ' until the difference in the locations' attractiveness is arbitraged away. Thus, a second source of congestion—beyond labor market congestion—stems from the residential housing market.

Importantly, despite these additions to the model, conditions similar to those in Section 3.1 guarantee the positive sorting of firms to locations; see Proposition 7 (Appendix C). The value that determines firms' location choices, $\overline{J}(p, \ell)$, is analogous to the baseline model with one key difference: Meeting rates are now endogenous. As a consequence, local competition has *two* components. It depends not only on local firm composition, Γ_ℓ (as before), but also on local labor market congestion, captured by $(\lambda^E(\ell), \lambda^F(\ell))$. Both components now affect how the firm size elasticity in (14) varies across space.

[23] The indirect utility of unemployed workers from consuming the final good and housing is given by $w^U(\ell)/d(\ell)^\omega$, where $d(\ell)$ is the housing price in ℓ. Thus, the flow utility of unemployed workers is given by $b(\ell) = B(\ell)w^U(\ell)/d(\ell)^\omega + \tilde{b}(\ell)$, where amenity $B(\ell)$ scales the consumption utility. We interpret $\tilde{b}$ as a non-monetary (possibly negative) utility component that stems from stigma. In practice, function $\tilde{b}$ gives us flexibility to satisfy Assumption 1, so that $w^R(\ell) = z(\underline{y}, A(\ell))$ for all ℓ.

Functional Forms. As for local labor markets, we assume that meeting rates are based on a Cobb-Douglas matching function $M(\mathcal{V}(\ell),\mathcal{U}(\ell)) = \mathcal{A}\sqrt{\mathcal{V}(\ell)\mathcal{U}(\ell)}$, where $\mathcal{A}$ is the overall matching efficiency. Recall that $\mathcal{U}(\ell) = L(\ell)(u(\ell) + \kappa(1 - u(\ell)))$. Parameter κ is the relative matching efficiency of employed workers, i.e., $\lambda^E(\ell) = \kappa\lambda^U(\ell)$.

As far as production is concerned, we assume that production function z is multiplicative and that the ex post productivity distribution is Pareto; that is,

$$z(y, A(\ell)) = A(\ell)y$$
$$\Gamma(y \mid p) = 1 - y^{-\frac{1}{p}}.$$

Based on this Pareto specification, firms with ex ante higher firm productivity p draw their ex post productivity y from a stochastically better distribution, in line with our theory.[24]

5.2 Identification

Our model is parameterized by a location ranking $[\underline{\ell}, \overline{\ell}]$, local TFP $A(\ell)$, local amenities $B(\ell)$, local separation rates $\delta(\ell)$, labor market parameters $(\kappa, \mathcal{A})$ and parameters of the housing market $(\omega, \tau, h(\ell))$. We must also identify the extent of firm sorting across space, captured by $\mu(\ell)$ (which in equilibrium equals p, the parameter guiding the Pareto's tail behavior).[25]

The key step in our identification problem is separating the effect of firm sorting $\mu(\ell)$ from local productivity $A(\ell)$. Intuitively, are locations prosperous because of high fundamental productivity or because of an advantageous firm composition? Under the assumptions of Pareto productivity and multiplicative technology, our model allows us to separately identify $\mu(\ell)$ and $A(\ell)$ using the average local labor share, $LS(\ell)$, and firm value added, $\mathbb{E}[z(y, A(\ell))|\ell]$:

$$LS(\ell) = 1 - \mu(\ell) \tag{21}$$
$$\mathbb{E}[z(y, A(\ell))|\ell] = A(\ell)(1 - \mu(\ell))^{-1}, \tag{22}$$

where the expectation is taken over ex post firm productivity distribution Γ_ℓ (for a similar labor share formula under the Pareto assumption in an economy with single labor market, see Gouin-Bonenfant, 2020). These equations illustrate how firm sorting and location productivity can be separated. Consider, first, the negative relationship between the labor share and firm productivity, which in the multiplicative class of z is independent of $A(\ell)$. Higher

[24] Note that we normalize the scale parameter, $\underline{y} = 1$. We also used two methods to investigate whether the Pareto assumption is justified (available on request): First, as a proxy for firm productivity, we use (residualized) AKM firm fixed effects and check that the tails of their local distributions are log-linear. Second, if ex post firm productivity is Pareto distributed in our model, then value added per employee, z, is as well—something we also check based on tail behavior.

[25] Given matching function μ, we can identify Q using $\mu(\ell) = Q^{-1}(R(\ell))$ (where we assume R is given; see below).

ex ante firm productivity $\mu(\ell) = p$ leads to higher ex post productivity of firms and thereby to a stochastically better employment distribution. Since more productive firms have more monopsony power and thus lower labor shares, locations in which employment is concentrated in top firms have lower average labor share.[26] Variation in local labor shares across space $LS(\cdot)$ thus identifies the extent of firm sorting $\mu(\cdot)$.

Conditional on local firm composition $\mu(\ell)$, location fundamental $A(\ell)$ can then be identified from average local value added. Specifically, the variation in value added across locations that is *not* accounted for by firm sorting must be driven by differences in location TFP.

To identify the parameters of the labor market, including relative search efficiency κ and matching efficiency $\mathcal{A}$, as well as job-separation rate $\delta(\ell)$, we exploit information on job-finding rates and local unemployment. First, the relative matching efficiency of employed workers κ is identified from their job-finding rate, relative to that of unemployed workers. Second, we identify the local job-separation rate using the steady-state formula for unemployment, as well as data on local unemployment and job-finding rates:

$$\delta(\ell) = \lambda^U(\ell)\frac{u(\ell)}{1-u(\ell)}. \tag{23}$$

Finally, the overall matching efficiency, $\mathcal{A}$, is identified from a mix of the job-finding rate of the unemployed, $\lambda^U(\ell)$, the job-destruction rate, $\delta(\ell)$, and the average firm size, $\bar{l}(\ell)$, in *any* ℓ,

$$\mathcal{A} = \sqrt{\lambda^U(\ell)(\delta(\ell) + \kappa\lambda^U(\ell))\bar{l}(\ell)}. \tag{24}$$

To identify the parameters of the housing market, we use the expenditure share of residential housing to pin down ω and the replacement rate of the unemployed to obtain tax rate τ for residential homeowners. And based on observed house prices—along with government budget constraint (18) and housing market clearing (19)—we can infer housing supply $h(\ell)$.

Last, we identify the amenity schedule B using the indifference condition whereby workers' value of search is equalized across space, given by (20) when imposing the normalization $\rho V^U = 1$. We now summarize this discussion:

Proposition 5 (Identification)**.** *Under the assumed functional forms (summarized in Assumption 3, Appendix F), the model is identified.*

[26] To bolster this intuition for why more productive firms have a lower labor share, we derive in Appendix I.2 (paragraph *Sufficient Conditions for* $\partial LS(\ell)/\partial\ell < 0$) a sufficient condition under which the firm-level labor share is decreasing in y for each ℓ. The pillar of this condition is that $\gamma_\ell(\cdot)$ is sufficiently decreasing—a property of many distributions, including (but not at all restricted to) the Pareto distribution—which implies that high-y firms have fewer competitors around them.

We provide more details on the presented derivations in the proof; see Appendix F.

5.3 Estimation: Strategy and Results

For estimation, we rely on regional data from the official records of the German Federal Statistical Office, which we aggregate at commuting-zone level; see Appendix D.1 for details. Specifically, we use employment, value added, and labor compensation, as well as unemployment rates, number of establishments, GDP, and population. In turn, for model validation, we use the worker- and firm-level data from the FDZ as in Section 4. The time unit is 1 month.

The identification argument provides us with a concrete estimation protocol that we will closely follow. Our implementation proceeds in seven steps. First, as in Section 4, we specify our location unit as *Commuting Zone* and we rank the 257 CZs based on their log GDP per capita—a commonly used measure of economic prosperity, which in our model is increasing in ℓ. Local log GDP per capita will be our discretized support $\{\ell_1, \ell_2, ..., \ell_{257}\}$ of land distribution R. To reflect differences in the number of firms across locations in the data (so that for aggregation we can appropriately weigh each CZ in the model), we then assign to each $\ell_j, j \in \{1, 2, ..., 257\}$ a probability mass $r(\ell_j)$ equal to its share of firms in Germany. Examples of the highest ranked CZs are Frankfurt, Munich, and Wolfsburg (all in West Germany); among the lowest ranked, we have Goerlitz (East) and Mansfeld-Südharz (rural East).

Second, we use (21) to obtain $\mu(\ell)$ from the observed labor share in ℓ. Because our model is stylized (e.g., it lacks noise in the firm-location matching process), we smooth any measurement error in the data moments before feeding them into the model. Specifically, we linearly fit each variable we target in estimation as a function of ℓ. Since the labor share is *decreasing* in ℓ (Figure 5, top left)—which we believe is a novel fact—we obtain an *increasing* matching function μ (top right).[27] This implies positive sorting between firms and locations, so that high-ℓ locations are characterized by stochastically better firm distributions.

Third, we obtain the overall matching efficiency, $\mathcal{A}$, from the Germany-wide observed matching rate, separation rate, and average firm size, using (24) (see Table A.7, Appendix G.1). To obtain the relative matching efficiency of employed workers κ, we take into account only those EE moves in the data that are associated with wage gains (59.7%) and set $\kappa = 0.597 \cdot \frac{\lambda^E}{\lambda^U}$, based on Germany-wide job-finding rates (λ^E, λ^U); see Table A.7.

Fourth, to pin down the local separation rates from (23), we use local unemployment and job-finding rates. To avoid using noisy CZ-specific job-finding rates from a small sample in

[27] The size of the dots in Figure 5 is proportional to the size of the CZ, as measured by its number of establishments.

the FDZ data, we infer $\lambda^U(\ell)$ in *each* ℓ from the average firm size $\bar{l}(\ell)$ provided by the German Federal Statistical Office (Figure 5, second row, left).[28] We then derive $\lambda^E(\ell) = \kappa\lambda^U(\ell)$ and $\lambda^F(\ell) = \lambda^U(\ell)/\theta(\ell)$. Based on the increasing $\bar{l}(\cdot)$, we obtain a slightly increasing $\lambda^U(\cdot)$, which implies higher meeting rates for workers and lower meeting rates for firms in high-ℓ locations (second row, right). Furthermore, an observed unemployment rate that is strongly decreasing in ℓ (Figure 5, third row, left), along with a fairly stable job-finding rate, translates into job-separation rates that are lower in more prosperous locations (Figure 5, third row, right).[29]

Fifth, we estimate location TFP based on the average value added per worker across locations using (22), except that we weigh each firm type by its employment.[30] Since average value added per worker is strongly increasing in ℓ (Figure 5, bottom left), we obtain an increasing A-schedule, even after controlling for firm sorting through μ (Figure 5, bottom right). To better understand the determinants of local TFP, we project the estimated A's on several location factors. We find that high local TFP is associated with a low cooperate tax rate, a high share of college educated workers, and the quality of infrastructure; see Table A.8, Appendix G.1.

Sixth, to pin down the parameters that govern residential housing markets, we target the average rent-to-income ratio of main tenant households (and obtain $\omega = 0.272$) and an average replacement rate of 60%, which implies a proportional tax rate on residential landlords of $\tau = 0.164$ (Table A.7, Appendix G.1). Finally, we pin down local housing supply $h(\ell)$, using observed location-specific rental rates $d(\ell)$; see Figure A.3 (bottom) in Appendix G.1.

Last, given $(\mu(\ell), A(\ell), \lambda^E(\ell), \delta(\ell), d(\ell))$ for each ℓ, we use (20) to back out amenity schedule B, which ensures that unemployed workers are indifferent between all locations. The top panel of Figure A.3 (Appendix G.1) shows that amenities are decreasing in the location index. Thus, even though residential housing is more expensive in high-ℓ locations (Figure A.3, bottom left), this force is not strong enough to dissuade workers from settling in those locations with high TFP *and* better firms, which calls for particularly low amenities in these places.

Importantly, at no point of the estimation do we impose PAM. Given the estimation output, we verify that the value of firm p of settling in ℓ, $\overline{J}(\ell, p)$, is supermodular in (p, ℓ), which verifies that the positive sorting of firms into locations is indeed *optimal* in the model.

[28] Solving (24) for $\lambda^U(\ell)$ while taking $\delta(\ell) = \lambda^U(\ell)u(\ell)/(1-u(\ell))$ into account gives $\lambda^U(\ell) = \mathcal{A}\big(\bar{l}(\ell) \cdot \big(\kappa + \frac{u(\ell)}{1-u(\ell)}\big)\big)^{-\frac{1}{2}}$.

[29] As an over-identification test, Figure A.2 shows that $\lambda^U(\cdot)$ and $\delta(\cdot)$, implied by our estimation, fit the data well.

[30] Instead of applying (22), we apply its weighted version $A(\ell) = \mathbb{E}_{g_\ell}[z(y, A(\ell))|\ell]/\big(\int y g_\ell(y)\, dy\big)$, where we observe average value added per employee, $\mathbb{E}_{g_\ell}[z(y, A(\ell))|\ell]$, in the data; and where we compute $\int y g_\ell(y)\, dy$ in the model, taking density g_ℓ based on (13) into account, which depends on $(\mu(\ell), \lambda^U(\ell), \delta(\ell))$ (all objects that we pinned down above). Note that the bottom left panel of Figure 5 is based on a linear fit of *log* (not level) value added to ℓ, which is why the displayed fit of value added against ℓ is not linear.

Figure 5: Model Fit of Targeted Moments (left) and Estimated Parameters (right)

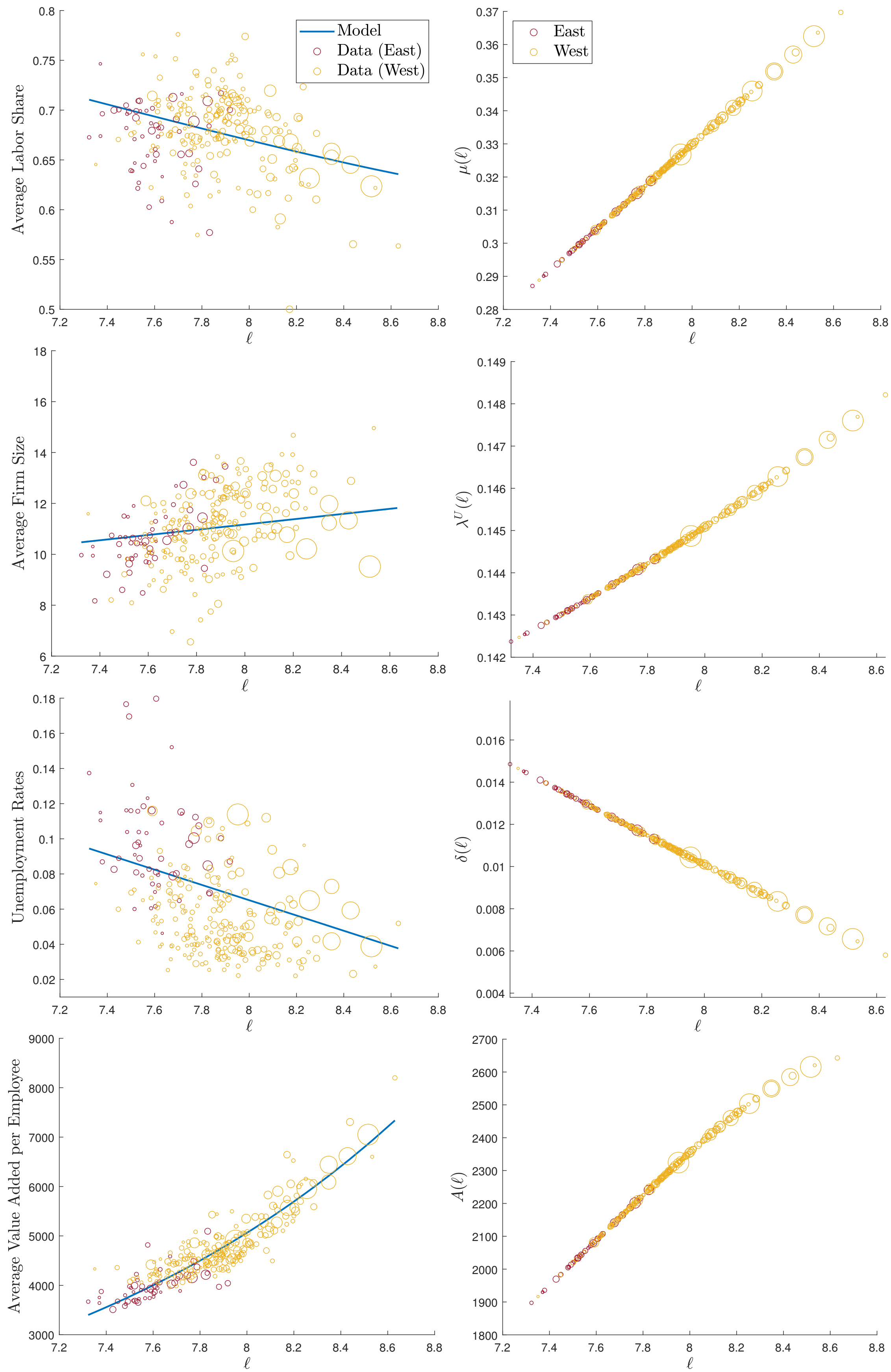
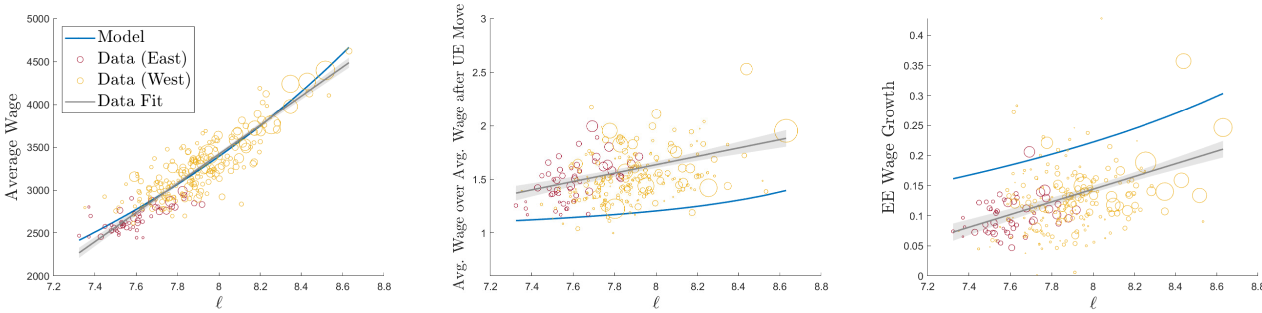


Figure 6: Model Fit: Non-Targeted Moments



Notes: Data sources: The left panel is based on local wages from the German Federal Statistical Office; the middle and right panels are based on worker-level wages from the LIAB-BHP. The right panel shows β_ℓ^{EE} from regression (17). These local (CZ-level) statistics are weighted by the number of workers, the number of wage observations, and the number of EE moves within the CZ, respectively, indicated by different sizes of the dots. 95% confidence intervals are displayed in gray.

5.4 Model Validation

Given our estimation approach, we fit the targeted data series of local labor shares, firm size, unemployment rates, and value added per worker by construction (Figure 5, left column). Importantly, our model also performs quite well regarding several non-targeted features of the data related to worker inequality and beyond.

In Figure 6, we first show the model's performance regarding the CZ's average wage, left panel. Second, and more importantly, our model captures quite well both the features of spatial wage inequality and of job ladder heterogeneity from our empirical evidence (Figure 4 of Section 4.2). In the middle panel of Figure 6, we again show the average wage in ℓ relative to the average wage of UE movers. Our model underestimates the level of this statistic, since only OJS and job ladder heterogeneity across space drive this ratio above 1, while in the data other factors—such as human capital accumulation on the job—also fuel this wage ratio. However, given our focus on spatial inequality, we stress that our model reproduces the observed spatial heterogeneity in this ratio. Finally, as shown in the right panel, our model also replicates the heterogeneity in EE wage growth across space, generated by steeper job ladders and stochastically better employment distributions in high-ℓ locations. We overestimate the level of wage growth but, reassuringly, we closely match the slope across locations.

We further validate the model using three additional features of the data: (i) firm size distribution across ℓ, as measured by the employment concentration in the largest 25% of firms in each ℓ; (ii) within-location wage dispersion, as measured by the 75/25-wage gap for each ℓ; and (iii) land price schedule k. Figure A.4 (Appendix G.1) shows that the model fits relatively well how these non-targeted features of the data vary across space, which is most important for our purposes.

The fact that our model fits well the spatial heterogeneity across commuting zones in a variety of dimensions makes us confident that it is a suitable measurement tool for spatial inequality in Germany and for decomposing this inequality into its driving forces.

6 The Drivers of Spatial Inequality

We now use our estimated model to decompose and understand the drivers of spatial inequality, focusing on firm sorting, on-the-job search, and spatial frictions. We concentrate on the East-West divide in Germany.[31] Our estimation revealed large discrepancies between East and West Germany in a variety of dimensions: The East is not only disadvantaged because of poor economic fundamentals (compare red and yellow CZs in Figure 5, bottom right) but also because workers lack access to the most productive firms (red and yellow CZs in Figure 5, top right). We now investigate to what extent these forces shape spatial wage inequality.

6.1 The German East-West Divide Through the Lens of Our Model

More than two decades after the Berlin wall came down, inequality between East and West Germany is still substantial; monthly nominal wages in the West exceed those in the East by 28% (Table 1). In comparison, our estimated model predicts a 22% gap, slightly underestimating this non-targeted feature of the data. Table A.9, Appendix G.2, reports spatial inequality in value added and underscores significant disparities between East and West Germany beyond pay.

Table 1: Monthly Wages (in €) and West-East Inequality

	Data	Model
Monthly Wage, West	3491.13	3480.72
Monthly Wage, East	2731.63	2845.38
West-East Gap	27.8%	22.3%

Notes: Data source: German Federal Statistical Office. Rows 1 and 2 report the average wage in each region and row 3 computes the percentage difference between rows 1 and 2.

6.2 The Role of Firm Sorting, On-the-Job Search and Spatial Frictions

To isolate the role of firm sorting, on-the-job search, and spatial frictions in the West-East German wage gap, we consider several counterfactual exercises. In each, we unpack the mechanism behind the resulting inequality changes by analyzing how the different components of

[31] Online Appendix OC.3 also analyzes wage gaps between urban and rural districts.

the spatial wage premium (16)—location differences in wage intercept, slope, and employment composition—are affected. Throughout the counterfactuals, we maintain Assumption 1. As a result, the overall firm distribution stays the same as in the baseline model.[32]

The Role of Firm Sorting. To assess the role of firm sorting in spatial inequality, we allocate firms randomly across locations and let the local labor market equilibrium play out. Comparison of this scenario with the baseline model tells us how much of the West-East wage gap is because better firms settle in the West. Appendix H.1 contains the technical details of this exercise.

Table 2, column 2, shows that in the absence of firm sorting the West-East wage gap would only be 18.8%, compared with 22.3% in the baseline model with sorting. Thus, spatial firm sorting can account for approximately 17% of the West-East wage gap.[33] Our conclusion is that the East is hurt not only because of poor economic fundamentals (low A) but this disadvantage is amplified by the fact that low-productivity firms cluster in those locations.

To put this into perspective, the impact of variation in location fundamentals $A(\cdot)$ on spatial inequality is almost three times as large as that of firm sorting. More specifically, when we shut down spatial differences in local TFP, $A = \mathbb{E}[A(\ell)]$, while keeping firm sorting at its baseline level, the West-East wage gap narrows to 11.8%.

Table 2: West-East Wage Inequality—Counterfactual Models

	Baseline Model	No Sorting	No OJS	No Spatial Frictions
West-East Gap	22.3%	18.8%	10.9%	1.2%

To better understand *why* inequality is lower without firm sorting, we zoom into our decomposition of the spatial wage premium given by (16). For illustration, we compare two locations at opposite ends of the spectrum of the local TFP distribution: Wolfsburg, the top West location, and Mansfeld-Suedharz, the bottom East location.

First, as shown in the left panel of Figure 7, the random matching of firms to locations reduces job ladder differences across space. While the job ladder was considerably steeper in

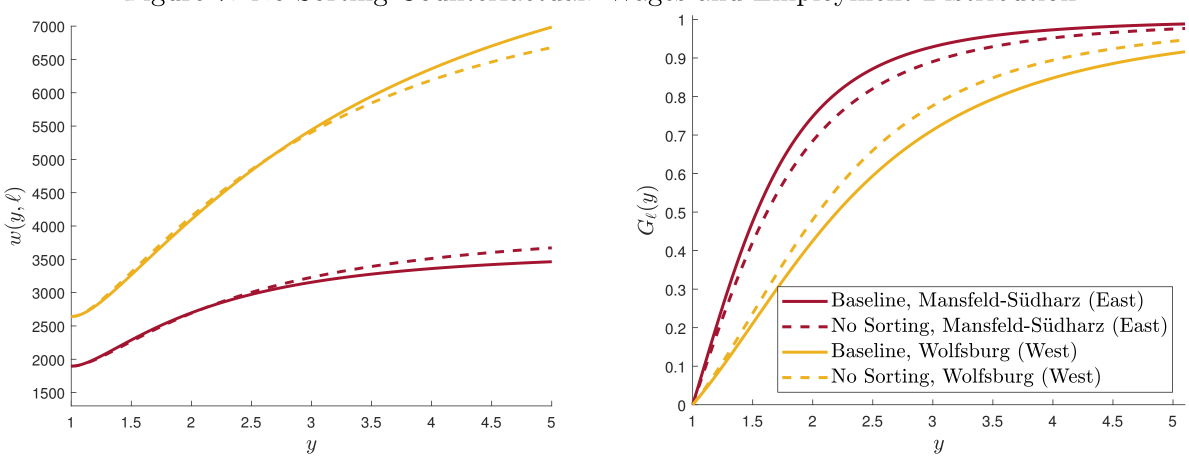
[32]By not taking into account changes in firm selection in these counterfactuals, we neglect a potentially interesting margin of adjustment. However, this approach allows us to analyze all counterfactuals in a coherent and tractable manner. Moreover, under this approach we obtain a conservative estimate of inequality reduction in each of the counterfactuals. Consider, e.g., our counterfactual on the role of firm sorting. There, Assumption 1 dampens the effect of firm sorting on spatial inequality: (i) our identification of tail parameter $1/\mu(\ell)$ of the firm productivity distribution is unaffected by this assumption; (ii) if, in addition, the selection margin operated so that $\underline{y}(\ell)$ can vary with ℓ, the estimated firm sorting would be even *stronger* than what we currently obtained, due to an increasing $\underline{y}(\cdot)$-schedule; and (iii) taking an increasing $\underline{y}(\cdot)$-schedule into account would flatten the estimated A-schedule, further amplifying the (relative) importance of firm sorting in spatial inequality. We ensure Assumption 1 in the counterfactuals, i.e., $w^R(\ell) = A(\ell)\underline{y}$ for all ℓ, by adjusting $\tilde{b}(\ell)$.

[33]Note that $17\% \sim (22.3 - 18.8)/22.3$.

the top West location (red solid) compared with the bottom East location (yellow solid) at baseline, this differential steepness shrinks when firm sorting is absent (dashed lines).

Second, the right panel highlights the role of spatial differences in employment composition. In the baseline model with firm sorting, the employment distribution in the West first-order stochastically dominates that in the East (compare the solid cdf's). Therefore, in the West, employment is clustered at more productive firms that pay higher wages. Importantly, shutting down firm sorting reduces this difference and thereby inequality. However, due to labor reallocation, the FOSD is not entirely undone by this counterfactual. Without firm sorting the West loses some its economic appeal, which is why workers reallocate to the East. As a result, workers' matching rates increase in the West, propelling workers to the top of the local job ladder faster compared with the East. Labor mobility thus alleviates the initial negative impact on the Western employment distribution, which mitigates the effect of firm sorting on spatial inequality.[34]

Figure 7: No-Sorting Counterfactual: Wages and Employment Distribution



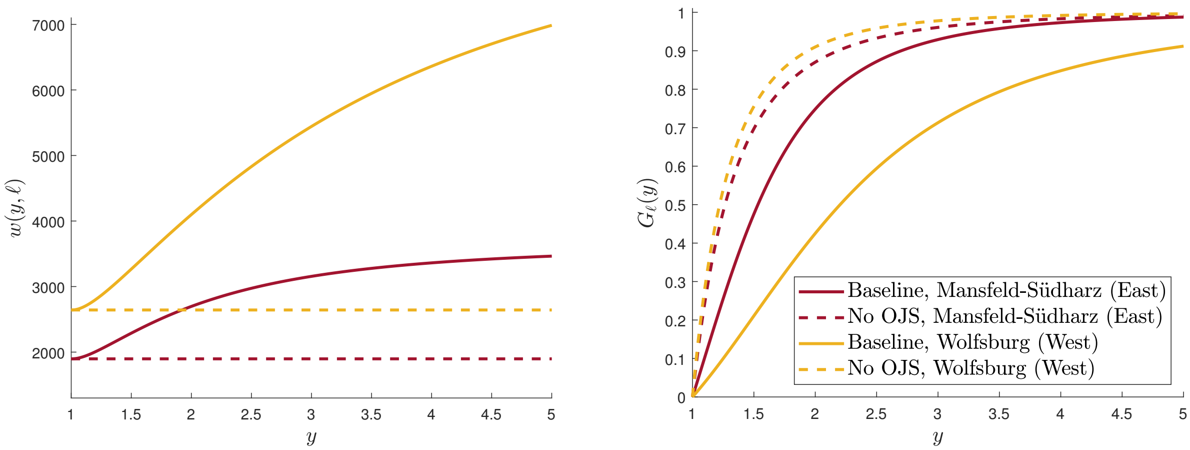
The Role of On-the-Job Search. In our theory, on-the-job search (and the associated job ladders) plays a crucial role in how firm sorting shapes inequality. To assess the quantitative importance of this feature, we eliminate on-the-job search by setting the search efficiency of employed workers to zero ($\kappa = 0$). See Appendix H.2 for details.

Table 2, column 3, shows that OJS is an important source of spatial wage inequality. Without it, the West-East gap would only be 10.9%, versus 22.3% in the baseline model.

The main change compared with our model with OJS search is that the sorting of firms to locations becomes largely *negative*, whereby the most productive firms settle in low-

[34] Without worker mobility, the resulting West-East wage gap would be 18.5% in this counterfactual, compared with 18.8% with mobility (Table 2). Note, however, that even without labor mobility (and thus fixed $\lambda^U(\cdot)$ at baseline level), the Western employment distribution would still dominate that in the East especially because $\delta(\cdot)$ is decreasing.

Figure 8: No On-the-Job-Search Counterfactual: Wages and Employment Distribution



productivity locations. The reason is that job ladders are flattened everywhere, which reduces the relative economic attractiveness of the West and leads to a large inflow of workers into low-ℓ (East) locations. This drives up firms' meeting rates λ^F in the East, which flips the modularity property of $\bar{J}(p, \ell)$ to submodular and induces negative sorting.

Figure 8, which is the analogue of Figure 7, highlights the two margins through which this counterfactual reduces the West-East German wage premium. First, the left panel shows that local job ladders collapse everywhere—the classical Diamond paradox, whereby all workers receive their reservation wage in the absence of OJS—which hurts especially the labor market prospects of workers in the West. Second, as shown in the right panel, due to the switch from positive to negative firm sorting, the employment distribution in the East now dominates that in the West. This highlights the fact that firm sorting only amplifies spatial inequality if it is *positive*. In contrast, negative firm sorting—which materializes in response to low enough κ, that is, if OJS is weak—dampens spatial inequality.

This counterfactual emphasizes once more the importance of allowing for endogenous firm location choices: Even though there is positive sorting between firms and locations in our estimated model, this need not be the case when the economic environment changes.

The Role of Spatial Frictions. In our final counterfactual, we elicit the role of spatial hiring frictions in spatial inequality. We eliminate the hiring friction firms face in the baseline model in which they were restricted to hire on-the-job searchers only from their local labor market. To do so, we integrate the labor market, so that the economy consists of a single job ladder with all firms hiring from everywhere. Firms are effectively characterized by the productivity index $z = A(\ell)y$ and workers climb the global z-job ladder, facing no geographic constraints regarding which firms can recruit them. This is an economy that offers a remote work option, which effectively decouples the place of residence from the place of work. Note that this counterfactual does not change the incentives of firms to sort across space; i.e., PAM is still in place. Technical details on the implementation are in Appendix H.3.

Table 2, column 4, shows the effect of labor market integration on the West-East wage gap. Without spatial hiring frictions, the spatial gap shrinks to 1.2% and essentially disappears.

Our decomposition of the spatial wage premium (16) sheds light on why this premium has become so small, with the average firm in the West paying almost the same wage as the average firm in the East. Although local job ladders in the two regions become more similar (Figure A.6, left, in Appendix H.3), the main driver is shifts in employment composition. While positive firm sorting still pushes toward stochastically better employment distributions G_ℓ in the West (fueling spatial inequality), this factor is counteracted by composition shifts that improve employment distributions in the East (dampening inequality). These composition shifts are due to the differential positioning of local firms on the global job ladder. In high-ℓ (West) locations, firms at the bottom of the local job ladder were disproportionally hurt under labor market segmentation because they faced severe competition from highly productive firms in their location. But—due to a high location fundamental A that increases their z on the global job ladder—they are globally competitive under labor market integration. They gain employment relative to more productive firms in their locations, so weight is shifting toward *less* productive firms, which deteriorates G_ℓ in the West.

The opposite is true for labor markets in the East. Due to positive firm sorting, the worst firms in those locations were shielded from tough competition under segmentation. Under labor market integration, however, they lose employment, which shifts weight toward more productive firms in their locations and improves G_ℓ. Indeed, Figure A.6 in the Appendix shows that G_ℓ in the East now stochastically dominates the employment distribution in the West, which gives rise to the stark reduction of spatial wage inequality documented in Table 2.[35]

7 Robustness

In this section, we summarize several checks we perform to highlight the robustness of our results; details are in Appendix I and Online Appendix OC.

[35]To put this into perspective, we also consider the counterfactual of eliminating the amenity differences across space (i.e., a constant B), which can be interpreted as shutting down the mobility frictions of workers. Doing so also reduces the West-East wage gap (to 15%), but the effect is quantitatively much smaller compared with eliminating the firms' hiring frictions.

Positive Sorting of Firms to Locations in the Data. An important ingredient of our estimation is the negative empirical relationship between local labor share and local GDP per capita. This—in combination with our model feature whereby locations with more productive firms have lower labor shares—lets us infer that there is positive sorting between firms and locations, i.e., productive firms sort into productive places with high GDP.

First, we demonstrate that the negative empirical relationship between log GDP per capita and local labor shares is robust to controlling for the local industry composition (column 2), local average firm size (column 3) and local population size (column 4); see Table A.11 (Appendix I.1). If anything, the negative relationship between labor share and GDP per capita becomes more pronounced when including these controls.

Second, we demonstrate in the model that the negative relationship between local labor shares and local firm productivity in (21)—which allows us to infer an upward-sloping matching function μ from the local labor shares that are decreasing in GDP—does *not* hinge on the assumption of Pareto-distributed firm productivity. More generally, this negative relationship between local labor shares and local firm productivity arises if firm-level labor shares are decreasing in y, because then locations with more productive firms have more mass on low-labor-share (i.e., high-y) firms. Note that decreasing firm-level labor shares tend to arise under firm productivity distributions with decreasing density; see Appendix I.2 for technical details. For example, if firm ex post productivity is (truncated) log-normally distributed with p shifting the mean productivity across space, we would also back out positive firm sorting from our data on local labor shares (see Figure A.7, Appendix I.3). By contrast, if spatial sorting is such that some locations have firm productivity that is simply a scaled version of others, then our identification strategy based on labor share variation would not detect it. These *neutral* shifts give rise to neither labor share variation nor spatial differences in job ladders or returns to EE flows (Appendix I.4). Because, empirically, prosperous locations have both lower labor shares and higher wage growth from switching jobs, we are confident that neutral productivity shifts are not a dominant feature of the data.

Estimation Conditional on Industry. One concern is that different industries operate under different technologies with different labor and capital intensities, which renders firms heterogeneous not only in their productivity but also in their labor intensity—a feature that is absent in our model. If industries sort across space, this could drive spatial labor share differences even in the absence of firm sorting. We address this concern in two ways.

First, we re-estimate our model after controlling for industry in the labor share data. We find that firm sorting accounts for an even larger share—27%—of the West-East wage gap (see Online Appendix OC.1 for details). This is consistent with our regression analysis above, which revealed that the within-industry relationship between labor shares and log GDP per capita is steeper than the unconditional one (see Table A.11, column 2).

Second, our model may be a better description of tradable industries such as manufacturing, where the local costumer base—a feature absent from our model—is unlikely to contribute to the attractiveness of a location. We therefore repeat our estimation using data on the manufacturing sector only; see Online Appendix OC.1. The results again remain similar to our baseline estimation, with firm sorting accounting for 24% of the West-East wage gap.

Estimation on Alternative Data Sources (FDZ). Our estimation heavily relies on regional statistics from the German Federal Statistical Office. As these are official records, we view them as more reliable than any regional aggregation we could perform on a smaller sample of firms in the FDZ. As a consequence, however, our analysis draws from two different data sources since the worker-level analysis (e.g., for model validation) requires individual-level data from the FDZ. For robustness, we therefore also perform our estimation using only data from the FDZ. The results remain very similar; see Online Appendix OC.2.

Other Dimensions of Spatial Inequality: The Urban Wage Premium. We complement our analysis of East-West inequality with an investigation of urban-rural wage gaps (Online Appendix OC.3). The role of firm sorting in the urban-rural wage gap is quantitatively similar ($\sim 19\%$) to that in the East-West divide (Table O.5). Moreover, OJS search is an important driver of the urban wage premium, giving rise to steeper job ladders in cities.

Positive Sorting of Firms to Locations in the Model. We generalize our model by allowing for *endogenous productivity spillovers*. Instead of assuming exogenous differences in A, we assume that locations are ex ante identical but differ ex post—after firms make location choices—in productivity, depending on the endogenous composition of firms; i.e., $A(\ell) = \tilde{A}(\Gamma_\ell)$, with $\tilde{A}' < 0$. This captures the idea of positive spillover effects from productive firms onto all firms in a location. In Proposition O3, Online Appendix OC.4.1, we show that if the (endogenous) location productivity advantage—along with its impact on firms' marginal productivity—is large enough relative to the cost of competition, then firms with high-p would indeed settle into high-ℓ locations, similar to the baseline model.

Finally, we endogenize vacancy creation instead of assuming that each firm has one open vacancy at all times. In Proposition O5, Online Appendix OC.4.2, we show that also in this case, the trade-off between productivity and competition determines firm sorting choices.

8 Conclusion

In this paper, we argue that the endogenous sorting of firms across local labor markets is an important contributor to spatial differences in economic performance. If productive areas are able to attract the most productive firms, non-productive labor markets are not only hurt through inferior location fundamentals but also lack access to productive employers.

We study firms' location decisions in a model with spatially segmented labor markets and on-the-job search. Our theory highlights the fact that firms face a fundamental trade-off when deciding which labor market to enter. Holding the distribution of competing firms fixed, productive locations are naturally more attractive. However, holding the productivity of a location fixed, being surrounded by more productive competitors exposes firms to poaching risk because they may lose employees quickly and have a hard time poaching workers from other firms. The degree of firm sorting in equilibrium thus depends on the balance of these forces.

We characterize this trade-off analytically and provide sufficient conditions for positive sorting—i.e., an allocation in which more productive firms settle in more productive locations. We show that positive sorting emerges as the unique equilibrium outcome if firm and location productivity are complements and labor market frictions are sufficiently large.

Using administrative data from Germany, we estimate our model to assess the degree of firm sorting in the data and to quantify its role for spatial inequality. Our estimates show that firm sorting is positive, with more productive firms sorting into more productive locations, which tend to be in West Germany. Quantitatively, firm sorting can account for 17%-27% of the West-East wage gap. Workers in the East are therefore not only harmed by poor economic fundamentals but also because they lack access to productive firms.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis: 1999, 'High Wage Workers and High Wage Firms'. *Econometrica* **67**(2), 251–333.

Ahlfeldt, G. M., S. Heblich, and T. Seidel: 2022, 'Micro-geographic property price and rent indices'. *Regional Science and Urban Economics*.

Bagger, J. and R. Lentz: 2019, 'An Empirical Model of Wage Dispersion with Sorting'. *Review of Economic Studies* **86**(1), 153–190.

Baum-Snow, N. and R. Pavan: 2011, 'Understanding the City Size Wage Gap'. *The Review of Economic Studies* **79**(1), 88–127.

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

Bilal, A.: 2022, 'The Geography of Unemployment'. NBER Working Paper 29269.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler: 2022, 'How Much Should we Trust Estimates of Firm Effects and Worker Sorting?'. *Journal of Labor Economics*.

Bruns, B.: 2019, 'Changes in Workplace Heterogeneity and How They Widen the Gender Wage Gap'. *American Economic Journal: Applied Economics* **11**(2), 74–113.

Burdett, K. and D. T. Mortensen: 1998, 'Wage Differentials, Employer Size, and Unemployment'. *International Economic Review* **39**(2), 257–273.

Card, D., J. Heining, and P. Kline: 2013, 'Workplace Heterogeneity and the Rise of West German Wage Inequality*'. *The Quarterly Journal of Economics* **128**(3), 967–1015.

Combes, P.-P., G. Duranton, and L. Gobillon: 2008, 'Spatial wage disparities: Sorting matters!'. *Journal of Urban Economics* **63**(2), 723–742.

Combes, P.-P., G. Duranton, L. Gobillon, D. Puga, and S. Roux: 2012, 'The productivity advantages of large cities: Distinguishing agglomeration from firm selection'. *Econometrica* **80**(6), 2543–2594.

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum: 2022, 'Matching in Cities'. *Journal of the European Economic Association*.

De La Roca, J. and D. Puga: 2017, 'Learning by Working in Big Cities'. *The Review of Economic Studies* **84**(1 (298)), 106–142.

Duranton, G. and D. Puga: 2004, 'Chapter 48 - Micro-Foundations of Urban Agglomeration Economies'. In: J. V. Henderson and J.-F. Thisse (eds.): *Cities and Geography*, Vol. 4 of *Handbook of Regional and Urban Economics*. Elsevier, pp. 2063–2117.

Gaubert, C.: 2018, 'Firm sorting and agglomeration'. *American Economic Review* **108**(11), 3117–53.

Glaeser, E. L. and D. C. Maré: 2001, 'Cities and Skills'. *Journal of Labor Economics* **19**(2), 316–342.

Gouin-Bonenfant, E.: 2020, 'Productivity Dispersion, Between-Firm Competition, and the Labor Share'. Working Paper.

Gould, E. D.: 2007, 'Cities, Workers, and Wages: A Structural Analysis of the Urban Wage Premium'. *The Review of Economic Studies* **74**(2), 477–506.

Heise, S. and T. Porzio: 2022, 'Labor Misallocation Across Firms and Regions'. NBER Working Paper 30298.

Milgrom, P. and C. Shannon: 1994, 'Monotone Comparative Statics'. *Econometrica* **62**(1), 157–180.

Moretti, E.: 2011, 'Local labor markets'. In: *Handbook of labor economics*, Vol. 4. Elsevier, pp. 1237–1313.

Moscarini, G. and F. Postel-Vinay: 2018, 'The cyclical job ladder'. *Annual Review of Economics* **10**, 165–188.

Nelson, K., D. Fredriksson, T. Korpi, W. Korpi, J. Palme, and O. Sjöberg: 2020, 'The social policy indicators (SPIN) database'. *International Journal of Social Welfare* **29**(3), 285–289.

Oberfield, E., E. Rossi-Hansberg, P.-D. Sarte, and N. Trachter: 2022, 'Plants in Space'. NBER Working Paper 27303.

Postel-Vinay, F. and J.-M. Robin: 2002, 'Equilibrium Wage Dispersion with Worker and Employer Heterogeneity'. *Econometrica* **70**(6), 2295–2350.

Ruf, K., L. Schmidtlein, S. Seth, H. Stüber, and M. Umkehrer: 2021a, 'Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975 – 2019.'. FDZ-Datenreport, 06/2021 (en), Nuremberg. DOI: 10.5164/IAB.FDZD.2106.en.v1.

Ruf, K., L. Schmidtlein, S. Seth, H. Stüber, M. Umkehrer, S. Grießemer, and S. Kaimer: 2021b, 'Linked-Employer-Employee-Data of the IAB (LIAB): LIAB longitudinal model 1975-2019, version 1'. Research Data Centre of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB). DOI: 10.5164/IAB.LIABLM7519.de.en.v1.

Villani, C.: 2009, *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften. Springer London, Limited.

Appendix

A Derivations

A.1 Alternative Formulation of Wage-Posting Problem

Firms' wage-posting problem (4) has an alternative formulation:

$$\tilde{J}(y,\ell) \equiv \max_{w \geq w^R(\ell)} h(w,\ell)J(y,w,\ell) = \max_{w \geq w^R(\ell)} \underbrace{\frac{\lambda^F \delta}{\delta + \lambda^E(1-F_\ell(w))}}_{=h(w,\ell)} \underbrace{\frac{z(y,A(\ell)) - w}{\rho + \delta + \lambda^E(1-F_\ell(w))}}_{=J(y,w,\ell)} \tag{A.1}$$

where $h(w,\ell)$ is the hiring rate of a firm posting w in location ℓ, and $J(y,w,\ell)$ is firm y's discounted flow profit when posting w in that location.[36] Using firm size expression (A.2) (Appendix A.2), we obtain (4).

A.2 Firm Size

As explained in Footnote 6, we can interpret the model's firm size as the product of the hiring rate and the expected duration of a match, which coincides with expression (3):

$$\begin{aligned}
l(y,\ell) &= \underbrace{\lambda^F\left(\frac{\lambda^U u(\ell)}{\lambda^U u(\ell) + \lambda^E(1-u(\ell))} + \frac{\lambda^E(1-u(\ell))}{\lambda^U u(\ell) + \lambda^E(1-u(\ell))}G_\ell(y)\right)}_{\text{Hiring Rate } h(y,\ell)} \underbrace{\frac{1}{\rho + \delta + \lambda^E(1-\Gamma_\ell(y))}}_{\text{Expected Match Duration}} \\
&= \lambda^F\left(\frac{\lambda^U \delta}{\lambda^U\delta + \lambda^E\lambda^U} + \frac{\lambda^E\lambda^U}{\lambda^U\delta + \lambda^E\lambda^U}\frac{\delta}{\delta + \lambda^E(1-\Gamma_\ell(y))}\Gamma_\ell(y)\right)\frac{1}{\rho + \delta + \lambda^E(1-\Gamma_\ell(y))} \\
&= \lambda^F \frac{\delta}{\delta + \lambda^E(1-\Gamma_\ell(y))}\frac{1}{\rho + \delta + \lambda^E(1-\Gamma_\ell(y))} \qquad (A.2) \\
\Rightarrow \quad l(y,\ell) &= \lambda^F \frac{\delta}{[\delta + \lambda^E(1-\Gamma_\ell(y))]^2} \quad \text{if } \rho \to 0.
\end{aligned}$$

Note that the matching rates of firms and workers need to be consistent with each other, that is: $\lambda^F = \lambda^U u + \lambda^E(1-u) = \lambda^U \frac{\delta}{\delta+\lambda^U} + \lambda^E \frac{\lambda^U}{\delta+\lambda^U} = \frac{\delta+\lambda^E}{\delta+\lambda^U}\lambda^U$. Plugging this into our definition of firm size above, we obtain $l(y,\ell) = \frac{\lambda^U(\delta+\lambda^E)}{\delta+\lambda^U}\frac{\delta}{[\delta+\lambda^E(1-\Gamma_\ell(y))]^2}$, which—when the measure of vacancies and workers coincide in each ℓ—is equivalent to the definition of firm size in Burdett and Mortensen (1998), who define it as

$$\frac{(1-u)g_\ell(y)}{1\cdot\gamma_\ell(y)} = \frac{\lambda^U}{\delta+\lambda^U}\frac{g_\ell(y)}{\gamma_\ell(y)} = \frac{\lambda^U}{\delta+\lambda^U}\frac{\delta(\delta+\lambda^E)}{(1+\lambda^E(1-\Gamma_\ell(y)))^2}.$$

[36] The hiring rate of firm y in location ℓ is $h(w,\ell) \equiv \lambda^F\left(\frac{\lambda^U u(\ell)}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))} + \frac{\lambda^E(1-u(\ell))}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))}E_\ell(w)\right)$, considering that a firm meets workers at rate λ^F from two pools: unemployment $u(\ell)$ (they will always accept the job), and employment $1-u(\ell)$ (they will accept if the new wage is higher than their current one). We denote the steady-state employment distribution by E_ℓ, where $E_\ell(w) = \delta\frac{F_\ell(w)}{\delta+\lambda^E(1-F_\ell(w))}$ (see (12) and (13)), so that $h(w,\ell)$ reduces to the expression in (A.1).

A.3 Wage Posting

Consider the firm's expected profits from employing workers (4). By the envelope theorem:

$$\frac{\partial \tilde{J}(y,\ell)}{\partial y} = l(w(y,\ell))\frac{\partial z(y,A(\ell))}{\partial y}.$$

And so,

$$\tilde{J}(y,\ell) = (z(y,A(\ell)) - w(y,\ell))l(w(y,\ell)) = \int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y} l(w(t,\ell))dt + \tilde{J}(\underline{y},\ell)$$

$$\Leftrightarrow \quad w(y,\ell) = z(y,A(\ell)) - \int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y}\frac{l(w(t,\ell))}{l(w(y,\ell))}dt - \frac{\tilde{J}(\underline{y},\ell)}{l(w(y,\ell))}. \tag{A.3}$$

Then:

$$\begin{aligned} w(y,\ell) = z(y,A(\ell)) &- \left[\delta + \lambda^E(1-\Gamma_\ell(y))\right]\left[\rho + \delta + \lambda^E(1-\Gamma_\ell(y))\right] \\ &\times \left\{\int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{[\delta + \lambda^E(1-\Gamma_\ell(t)] \cdot [\rho+\delta+\lambda^E(1-\Gamma_\ell(t))]}dt\right\} \\ &- \left[\delta + \lambda^E(1-\Gamma_\ell(y))\right] \cdot \left[\rho + \delta + \lambda^E(1-\Gamma_\ell(y))\right]\frac{\tilde{J}(\underline{y},\ell)}{\lambda^F\delta}. \end{aligned} \tag{A.4}$$

Plugging (A.4) into $\tilde{J}$, we obtain:

$$\tilde{J}(y,\ell) = \lambda^F\delta \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{[\delta + \lambda^E(1-\Gamma_\ell(t)] \cdot [\rho+\delta+\lambda^E(1-\Gamma_\ell(t))]}dt + \tilde{J}(\underline{y},\ell),$$

where $\tilde{J}(\underline{y};\ell) = l(w(\underline{y},\ell))(z(\underline{y},A(\ell)) - w^R(\ell)) = \frac{\lambda^F\delta}{\delta+\lambda^E}\frac{1}{\rho+\delta+\lambda^E}(z(\underline{y},A(\ell)) - w^R(\ell))$.

Imposing Assumption 1.2 (zero profits of the least productive firm type in each location, $\tilde{J}(\underline{y},\ell) = 0$) and $\rho = 0$, we obtain wage function (5) from (A.4).

A.4 Location Choice: Firm's FOC

The FOC of problem (7) is given by

$$\begin{aligned}\delta\lambda^F \int_{\underline{y}}^{\overline{y}} \Bigg(&\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}[\delta + \lambda^E(1-\Gamma_\ell(y))]^2}{[\delta+\lambda^E(1-\Gamma_\ell(y))]^4} \\ &- \frac{\frac{\partial z(y,A(\ell))}{\partial y}2\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]\lambda^E\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right)}{[\delta+\lambda^E(1-\Gamma_\ell(y))]^4}\Bigg)(1-\Gamma(y|p))\,dy = \frac{\partial k(\ell)}{\partial \ell}.\end{aligned}$$

Using the expression of firm size $l(y,\ell)$ from (3) and rearranging, this becomes (8).

A.5 Land Price Schedule

Using integration by parts and Assumption 1.2 (i.e., zero profits of $\underline{y}$ in all ℓ, implying $\tilde{J}(\underline{y}, \ell) = 0$) problem (7) can be expressed as

$$\max_{\ell} \int \frac{\partial \tilde{J}(y,\ell)}{\partial y}(1 - \Gamma(y|p))dy - k(\ell).$$

The FOC reads

$$\int \frac{\partial^2 \tilde{J}(y,\ell)}{\partial y \partial \ell}(1 - \Gamma(y|p))dy = \frac{\partial k(\ell)}{\partial \ell}.$$

Solving this differential equation, when evaluated at the equilibrium assignment, yields land price schedule $k(\cdot)$.

For the case with pure sorting given by matching function μ, solving for $k(\ell)$ yields:

$$k(\ell) = \bar{k} + \int_{\underline{\ell}}^{\ell} \int_{\underline{y}}^{\overline{y}} \frac{\partial \tilde{J}(y,\hat{\ell})}{\partial y \partial \ell} \left(1 - \Gamma(y|\mu(\hat{\ell}))\right) dy d\hat{\ell},$$

where $\overline{k}$ is a constant of integration. We anchor k by choosing $\bar{k}$ such that the landowner whose land commands the lowest price in equilibrium obtains zero.

A.6 Land Market Clearing

We can derive $R(\ell) = Q(\mu(\ell))$ from general land market clearing condition (10),

$$\begin{aligned}
R(\ell) &= \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\overline{p}} m_\ell(\tilde{\ell}|\tilde{p}) q(\tilde{p}) d\tilde{p} d\tilde{\ell} \\
&= \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\overline{p}} \frac{m(\tilde{\ell},\tilde{p})}{q(\tilde{p})} \frac{r(\tilde{\ell})}{r(\tilde{\ell})} q(\tilde{p}) d\tilde{p} d\tilde{\ell} \\
&= \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\overline{p}} \frac{r(\tilde{\ell})}{q(\tilde{p})} q(\tilde{p}) dM_p(\tilde{p}|\tilde{\ell}) d\tilde{\ell} \\
&= \int_{\underline{\ell}}^{\ell} \mu'(\tilde{\ell}) q(\mu(\tilde{\ell})) d\tilde{\ell} \\
&= Q(\mu(\ell)),
\end{aligned}$$

where, to go from line 3 to line 4, we use the fact that under positive sorting $M_p(p|\ell)$ is a Dirac measure, i.e., for each ℓ it puts positive mass only at $p = \mu(\ell)$, and conjecture $\mu'(\ell) = r(\ell)/q(\mu(\ell))$, which then indeed materializes.

B Baseline Model: Proofs and Additional Results

B.1 Proof of Proposition 2

Apply integration by parts to (6) to obtain

$$\begin{aligned}\overline{J}(p,\ell) &= \delta\lambda^F\left(\int_{\underline{y}}^{y}\frac{\frac{\partial z(t,A(\ell))}{\partial y}}{\left[\delta+\lambda^E(1-\Gamma_\ell(t))\right]^2}dt\Gamma(y|p)\Big|_{\underline{y}}^{\overline{y}}-\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]^2}\Gamma(y|p)dy\right)-k(\ell)\\ &= \delta\lambda^F\left(\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(t,A(\ell))}{\partial y}}{\left[\delta+\lambda^E(1-\Gamma_\ell(t))\right]^2}dt+\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]^2}(-\Gamma(y|p))dy\right)-k(\ell)\\ &= \delta\lambda^F\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]^2}(1-\Gamma(y|p))dy-k(\ell).\end{aligned}$$

To assess the conditions under which $\overline{J}(p,\ell)$ is supermodular in (p,ℓ), which is sufficient for the single-crossing property of $\overline{J}(p,\ell)$ in (p,ℓ), we differentiate wrt (p,ℓ):

$$\begin{aligned}\frac{\partial^2\overline{J}(p,\ell)}{\partial p\partial\ell}=\delta\lambda^F\int_{\underline{y}}^{\overline{y}}&\left(\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial\ell}\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]^2}{\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]^4}\right.\\ &\left.+\frac{\frac{\partial z(y,A(\ell))}{\partial y}2\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]\lambda^E\frac{\partial\Gamma_\ell}{\partial\ell}}{\left[\delta+\lambda^E(1-\Gamma_\ell(y))\right]^4}\right)\left(-\frac{\partial\Gamma(y|p)}{\partial p}\right)dy.\end{aligned}$$

In order for this expression to be (strictly) positive, it suffices that the integrand is positive for all $y\in[\underline{y},\overline{y}]$ and strictly so for some set of y of positive measure. In turn, for this it is sufficient that (recall that we assume $\frac{\partial\Gamma(y|p)}{\partial p}<0$ for all $y\in(\underline{y},\overline{y})$):

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial\ell}}{\frac{\partial z(y,A(\ell))}{\partial y}}>\frac{2\lambda^E}{\delta+\lambda^E(1-\Gamma_\ell(y))}\left(-\frac{\partial\Gamma_\ell(y)}{\partial\ell}\right).$$

Under positive sorting, there is a unique way of matching up firms' ex ante types with locations such that the land market clears, $Q(\mu(\ell))=R(\ell)$; see Appendix A.6. Further, based on our assumption of strictly positive densities (r,q), this assignment is one-to-one: μ is strictly increasing, where the firm type p assigned to location ℓ is given by $p=\mu(\ell)=Q^{-1}(R(\ell))$. Positive sorting then implies that the endogenous firm distribution in location ℓ is given by $\Gamma_\ell(y)=\Gamma(y|Q^{-1}(R(\ell)))$ (see also Footnote 12). Hence, $\frac{\partial\Gamma_\ell(y)}{\partial\ell}=\frac{\partial\Gamma}{\partial p}\frac{r(\ell)}{q(\mu(\ell))}$ and so to guarantee supermodularity of $\overline{J}(p,\ell)$ in (p,ℓ), we need to ensure that

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\lambda^E}{\delta+\lambda^E(1-\Gamma(y|Q^{-1}(R(\ell)))}\left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\right)\frac{r(\ell)}{q(Q^{-1}(R(\ell)))},$$

which is a condition in terms of primitives. To specify bounds that make this condition hold uniformly in (ℓ, y), let

$$\varepsilon^P \equiv \min_{\ell,y} \frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}}$$
$$t^P \equiv \max_{\ell,y}\left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\right)\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}.$$

Note that under our assumptions and the premise of the proposition, ε^P exists: It is strictly positive and bounded.

In turn, t^P exists (and it is also strictly positive and bounded) since we assume that $\Gamma(y|p)$ is continuously differentiable in p where both p and y are defined over compact sets, and that cdf's Q and R are continuously differentiable on the intervals $[\underline{p}, \overline{p}]$ and $[\underline{\ell}, \overline{\ell}]$, respectively, with strictly positive densities (q, r).

A sufficient condition for $\bar{J}$ to be supermodular in (ℓ, p) is therefore

$$\varepsilon^P > 2\varphi^E t^P.$$

So, equilibrium sorting is PAM, either if ε^P is sufficiently large or φ^E is sufficiently small. □

B.2 Proof of Proposition 3

We want to show that a fixed point in Γ_ℓ exists and we will do so by construction. Suppose the conditions of Proposition 2 hold, i.e., there is PAM of firms to locations.

Consider an assignment $\mu(\ell) = Q^{-1}(R(\ell))$, which yields a unique firm distribution across locations $\Gamma_\ell = \Gamma(y|\mu(\ell))$ and a unique wage function (5). We will show that the pair (μ, k) is the unique (up to a constant of integration) Walrasian equilibrium of the land market, where

$$k(\ell) = \overline{k} + \delta\lambda^F \int_{\underline{\ell}}^{\ell}\int_{\underline{y}}^{\overline{y}} \frac{\partial\left(\frac{\frac{\partial z(y,A(\hat{\ell}))}{\partial y}}{\left[\delta+\lambda(1-\Gamma_{\hat{\ell}}(y))\right]^2}\right)}{\partial \ell}(1-\Gamma(y|\mu(\hat{\ell})))dyd\hat{\ell}$$

is the land price schedule supporting assignment μ; see Appendix A.5.

By construction, μ clears the land market. To see that it is also *globally* optimal, we analyze firms' optimal behavior. Consider a firm with attribute p. It solves (7), i.e.,

$$\max_{\ell} \overline{J}(p,\ell) = \delta\lambda^F \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{\left[\delta+\lambda(1-\Gamma_\ell(y))\right]^2}(1-\Gamma(y|p))dy - \delta\lambda^F \int_{\underline{\ell}}^{\ell}\int_{\underline{y}}^{\overline{y}} \frac{\partial\left(\frac{\frac{\partial z(y,A(\hat{\ell}))}{\partial y}}{\left[\delta+\lambda(1-\Gamma_{\hat{\ell}}(y))\right]^2}\right)}{\partial \ell}(1-\Gamma(y|\mu(\hat{\ell})))dyd\hat{\ell} - \overline{k}.$$

To reduce notation, we define

$$\mathcal{J}(p,\ell) := \delta\lambda^F \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{\left[\delta+\lambda(1-\Gamma_\ell(y))\right]^2}(1-\Gamma(y|p))dy,$$

which is supermodular in (p,ℓ) under the conditions specified in Proposition 2. Firm p thus solves

$$\max_{\ell} \qquad \mathcal{J}(p,\ell) - \int_{\underline{\ell}}^{\ell} \frac{\partial\mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell}d\hat{\ell} - \overline{k},$$

with solution $p=\mu(\ell)$. To show that $\mu(\ell)$ is a *global* optimum, note that for any $\ell' < \ell$

$$\mathcal{J}(p,\ell) - \int_{\underline{\ell}}^{\ell} \frac{\partial\mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell}d\hat{\ell} \geq \mathcal{J}(p,\ell') - \int_{\underline{\ell}}^{\ell'} \frac{\partial\mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell}d\hat{\ell}$$

if and only if

$$\mathcal{J}(p,\ell) - \mathcal{J}(p,\ell') \geq \int_{\ell'}^{\ell} \frac{\partial\mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell}d\hat{\ell}. \tag{A.5}$$

Since $p=\mu(\ell)$ and since $\mathcal{J}(p,\ell)-\mathcal{J}(p,\ell') = \int_{\ell'}^{\ell}\frac{\partial\mathcal{J}(p,\hat{\ell})}{\partial \ell}d\hat{\ell}$, it follows that (A.5) is equivalent to

$$\int_{\ell'}^{\ell} \frac{\partial\mathcal{J}(\mu(\ell),\hat{\ell})}{\partial \ell}d\hat{\ell} \geq \int_{\ell'}^{\ell} \frac{\partial\mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell}d\hat{\ell}$$

and this holds due to the (strict) supermodularity of $\mathcal{J}(p,\ell)$ and $\mu(\ell) \geq \mu(\hat{\ell})$ for all $\hat{\ell} \in [\ell',\ell]$, which also holds strictly if $\hat{\ell} \neq \ell$. Hence, firm p strictly prefers ℓ over $\hat{\ell} < \ell$. A similar argument holds for $\hat{\ell} > \ell$, and hence choosing ℓ is the unique global optimum for p. Since p was arbitrary, all firm types behave optimally. We have shown that the optimal μ (and thus Γ_ℓ) coincides with the postulated μ (and thus Γ_ℓ) from above, i.e., we have constructed an equilibrium. Note that all land is occupied, and that, for each ℓ, land (owner) ℓ obtains $k(\ell) \geq 0$.

To see that this equilibrium is unique, we first note that under our assumptions, Theorem 10.28 in Villani (2009) implies that there exists a unique optimal assignment μ, which is deterministic. Second, the uniqueness of $k(\ell)$ (up to a constant of integration) then follows from Remarks 10.29 and 10.30 in Villani (2009). □

B.3 Proof of Proposition 4

We show that under the conditions of the proposition, (16) is positive, due to all of its three components being positive.

(i) follows directly from the assumptions that z is strictly increasing in A for all y, and that A is strictly increasing in ℓ.

(ii) follows from analyzing the cross-partial derivative of the wage function:

$$\frac{\partial^2 w(y,\ell)}{\partial y \partial \ell} = 2\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)\frac{\lambda^E}{\delta}\gamma_\ell(y)\int_{\underline{y}}^{y} \frac{\frac{\partial^2 z(t,A(\ell))}{\partial \ell \partial y}}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^2} - \frac{\frac{\partial z(t,A(\ell))}{\partial y} 2\frac{\lambda^E}{\delta}\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right)}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^3} dt$$
$$+2\left(\left(\frac{\lambda^E}{\delta}\right)^2 \frac{\partial \Gamma_\ell(y)}{\partial y}\left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell}\right) + \frac{\lambda^E}{\delta}\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)\frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}\right)\int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^2} dt.$$

The first line is positive under our conditions for PAM, which render the integrand positive for all y. But the second line is ambiguous unless we impose further assumptions. Denote

$$Z(t,\ell) := \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^2}.$$

Then, for all (y,ℓ)

$$\frac{\partial^2 w(y,\ell)}{\partial y \partial \ell} \geq 0,$$

if:

$$\int_{\underline{y}}^{y} \frac{\partial Z(t,\ell)}{\partial \ell} dt \geq -\frac{\left(\frac{\lambda^E}{\delta}\right)^2 \frac{\partial \Gamma_\ell(y)}{\partial y}\left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell}\right) + \frac{\lambda^E}{\delta}\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)\frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}}{\frac{\lambda^E}{\delta}\gamma_\ell(y)\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)} \int_{\underline{y}}^{y} Z(t,\ell) dt. \quad \text{(A.6)}$$

First, note that the (weak) inequality holds for $y = \underline{y}$.

Second, consider $y > \underline{y}$. We obtain the following sufficient condition for (A.6):

$$\frac{\int_{\underline{y}}^{y} \frac{\partial Z(t,\ell)}{\partial \ell} dt}{\int_{\underline{y}}^{y} Z(t,\ell) dt} \geq -\frac{\frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}}{\gamma_\ell(y)} \quad \forall (y,\ell),$$

which follows since under PAM $-\left(\left(\frac{\lambda^E}{\delta}\right)^2 \frac{\partial \Gamma_\ell(y)}{\partial y}\left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell}\right)\right)/\left(\frac{\lambda^E}{\delta}\gamma_\ell(y)\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)\right) \leq 0$ in the first term on the RHS of (A.6).

The condition on primitives ensuring this is given by

$$\min_{y,\ell}\left(\frac{\int_{\underline{y}}^{y}\frac{\partial Z(t,\ell)}{\partial \ell}dt}{\int_{\underline{y}}^{y}Z(t,\ell)dt}\right)\geq \max_{y,\ell}\left(-\frac{\frac{\partial^2\Gamma_\ell(y)}{\partial y\partial \ell}}{\gamma_\ell(y)}\right). \tag{A.7}$$

The maximum on the RHS is well-defined since it is taken over a continuous function on a compact set; moreover, the RHS is positive since $\partial^2\Gamma_\ell(y)/\partial y\partial\ell$ changes its sign in y and so the maximum is achieved at a positive value. We therefore need to assume that the minimum on the LHS is (sufficiently) positive. Specifically, we need to rule out that the minimum on the LHS is zero at $y=\underline{y}$. To this end, we use L'Hospital's rule, and obtain

$$\lim_{y\to\underline{y}}\frac{\int_{\underline{y}}^{y}\frac{\partial Z(t,\ell)}{\partial \ell}dt}{\int_{\underline{y}}^{y}Z(t,\ell)dt}=\lim_{y\to\underline{y}}\frac{\frac{\partial Z(y,\ell)}{\partial\ell}}{Z(y,\ell)}=\frac{\frac{\partial Z(\underline{y},\ell)}{\partial\ell}}{Z(\underline{y},\ell)}>0,$$

which is strictly positive under our sufficient condition for PAM (Proposition 2). Therefore, (A.7) is sufficient for w to be supermodular in (y,ℓ), which we can further unpack as:

$$\min_{\ell,y}\frac{\int_{\underline{y}}^{y}\frac{\partial^2 z(t,A(\ell))}{\partial\ell\partial y}dt}{\int_{\underline{y}}^{y}\frac{\partial z(t,A(\ell))}{\partial y}dt}\geq\left(1+\frac{\lambda^E}{\delta}\right)^2\left(\max_{y,\ell}\left(-\frac{\frac{\partial^2\Gamma_\ell(y)}{\partial y\partial\ell}}{\gamma_\ell(y)}\right)+2\frac{\lambda^E}{\delta}\max_{y,\ell}\left(\frac{\int_{\underline{y}}^{y}\frac{\partial z(t,A(\ell))}{\partial y}\left(-\frac{\partial\Gamma_\ell(t)}{\partial\ell}\right)dt}{\int_{\underline{y}}^{y}\frac{\partial z(t,A(\ell))}{\partial y}dt}\right)\right). \tag{A.8}$$

Note that the first term on the RHS has a well-defined maximum since we are maximizing a continuous function over a compact set and $\gamma_\ell>0$ due to our assumption that $\gamma>0$. In turn, regarding the second term, we need to rule out that it is infinite at $y=\underline{y}$. To this end, we use L'Hospital's rule, and obtain

$$\lim_{y\to\underline{y}}\frac{\int_{\underline{y}}^{y}\frac{\partial z(t,A(\ell))}{\partial y}\left(-\frac{\partial\Gamma_\ell(y)}{\partial\ell}\right)dt}{\int_{\underline{y}}^{y}\frac{\partial z(t,A(\ell))}{\partial y}dt}=\lim_{y\to\underline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}\left(-\frac{\partial\Gamma_\ell(y)}{\partial\ell}\right)}{\frac{\partial z(y,A(\ell))}{\partial y}}=-\frac{\partial\Gamma_\ell(\underline{y})}{\partial\ell}=0.$$

Thus, w is supermodular in (y,ℓ) if (A.8) holds, i.e., if the complementarities of z in (y,ℓ) on the LHS of (A.8) are sufficiently large.

(iii) follows directly from positive sorting, $\partial\Gamma_\ell(y)/\partial\ell\leq 0$, and the relationship between Γ_ℓ and G_ℓ in (13). $\square$

B.4 Proposition and Proof: Global vs. Local Rank

We state the formal result on the behavior of $D(y)$ that we described in the text. We maintain the following regularity assumption.

Assumption 2. *Both $\gamma(\underline{y}|p)$ and $\gamma(\overline{y}|p)$ are not constant in p.*

We can then show the following results.

Proposition 6 (Firm Sorting and the Difference between Global and Local Productivity Ranks)**.** *Suppose Assumption 2 holds.*

1. *If there is no spatial firm sorting, $\Gamma_{\ell'} = \Gamma_{\ell''}$ for all $\ell' \neq \ell''$, then $D(y) = 0$ for all $y \in [\underline{y}, \overline{y}]$.*
2. *If there is spatial firm sorting, $\Gamma_{\ell'} \neq \Gamma_{\ell''}$ for almost all $\ell' \neq \ell''$, then $D(y) = 0$ for $y = \{\underline{y}, \overline{y}\}$; in turn, there exists a firm type $y^* \in (\underline{y}, \overline{y})$ such that for all $y < y^*$, $D(y) < 0$, and a type $y^{**} \in (\underline{y}, \overline{y})$ with $y^{**} \geq y^*$ such that for all $y > y^{**}$, $D(y) > 0$.*

Proof. Recall that, under pure monotone sorting (PAM or NAM), we define:

$$D(y) := \int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) r(\ell) d\ell - \int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) \frac{\gamma(y|\mu(\ell)) r(\ell)}{\int_{\underline{\ell}}^{\overline{\ell}} \gamma(y|\mu(\hat{\ell})) r(\hat{\ell}) d\hat{\ell}} d\ell.$$

Our definition of local rank reflects the *average* local rank of any given firm y: $\int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) n_\ell(\ell|y) d\ell$, where $n_\ell(\ell|y)$ is defined as the (endogenous) location density conditional on y,

$$n_\ell(\ell|y) := \frac{n(\ell, y)}{n(y)} \underbrace{=}_{\text{PAM/NAM}} \frac{\gamma(y \mid \mu(\ell)) q(\mu(\ell)) \mu'(\ell)}{\int_{\underline{\ell}}^{\overline{\ell}} \gamma(y \mid \mu(\hat{\ell})) q(\mu(\hat{\ell})) \mu'(\hat{\ell}) d\hat{\ell}} = \frac{\gamma(y|\mu(\ell)) r(\ell)}{\int_{\underline{\ell}}^{\overline{\ell}} \gamma(y|\mu(\hat{\ell})) r(\hat{\ell}) d\hat{\ell}},$$

and where $n(\ell, y) := \gamma(y, \mu(\ell)) \mu'(\ell) = \gamma(y|\mu(\ell)) q(\mu(\ell)) \mu'(\ell)$ is the joint pdf of (ℓ, y) with corresponding marginal pdf, $n(y) := \int_{\underline{\ell}}^{\overline{\ell}} n(\ell, y) d\ell = \int_{\underline{\ell}}^{\overline{\ell}} \gamma(y|\mu(\ell)) q(\mu(\ell)) \mu'(\ell) d\ell$; in turn, $\gamma(y, p)$ is the pdf corresponding to the joint cdf $\Gamma(y, p)$.

Part 1. follows from the premise of no sorting, i.e., $\Gamma_{\ell'}(y) = \Gamma_{\ell''}(y) = \Gamma(y), \forall \ell', \ell'' \in [\underline{\ell}, \overline{\ell}]$, in which case

$$D(y) = \Gamma(y) \left(\int_{\underline{\ell}}^{\overline{\ell}} r(\ell) d\ell - \int_{\underline{\ell}}^{\overline{\ell}} n_\ell(\ell|y) d\ell \right) = 0.$$

Part 2., first statement, i.e. $D(\underline{y}) = D(\overline{y}) = 0$, also follows straight from the definition of D.

Part 2., second statement, follows from examining the slope of D at $y = \{\underline{y}, \overline{y}\}$.

Differentiate D wrt y to obtain

$$D'(y) = \int \gamma(y|\mu(\ell)))r(\ell)d\ell$$
$$- \left\{ \frac{\left(\int \left(\gamma(y|\mu(\ell))^2 + \Gamma_\ell(y)\frac{\partial\gamma(y|\mu(\ell))}{\partial y}\right) r(\ell)d\ell\right) \left(\int \gamma(y|\mu(\ell))r(\ell)d\ell\right)}{\left(\int \gamma(y|\mu(\ell))r(\ell)d\ell\right)^2} - \frac{\left(\int \Gamma_\ell(y)\gamma(y|\mu(\ell))r(\ell)d\ell\right) \left(\int \frac{\partial\gamma(y|\mu(\ell))}{\partial y} r(\ell)d\ell\right)}{\left(\int \gamma(y|\mu(\ell))r(\ell)d\ell\right)^2} \right\}$$

Evaluate this expression at $y = \{\underline{y}, \overline{y}\}$

$$D'(y)\big|_{y=\underline{y}} = \frac{\left(\int \gamma(\underline{y}|\mu(\ell))r(\ell)d\ell\right)^2 - \left(\int \gamma(\underline{y}|\mu(\ell))^2 r(\ell)d\ell\right)}{\int \gamma(\underline{y}|\mu(\ell))r(\ell)d\ell} = \frac{-\mathrm{Var}_r[\gamma(\underline{y}|\mu(\ell))]}{\int \gamma(\underline{y}|\mu(\ell))r(\ell)d\ell}$$
$$D'(y)\big|_{y=\overline{y}} = \frac{\left(\int \gamma(\overline{y}|\mu(\ell))r(\ell)d\ell\right)^2 - \left(\int \gamma(\overline{y}|\mu(\ell))^2 r(\ell)d\ell\right)}{\int \gamma(\overline{y}|\mu(\ell))r(\ell)d\ell} = \frac{-\mathrm{Var}_r[\gamma(\overline{y}|\mu(\ell))]}{\int \gamma(\overline{y}|\mu(\ell))r(\ell)d\ell},$$

where Var_r is our notation for the variance of a random variable, taking land distribution r into account. Both expressions are *strictly negative* if $\mathrm{Var}_r[\gamma(\underline{y}|\mu(\ell))] > 0$ and $\mathrm{Var}_r[\gamma(\overline{y}|\mu(\ell))] > 0$, which is the case under Assumption 2.

Since D starts at zero and first decreases, it is strictly negative for small $y > \underline{y}$; and since it ends at zero in a decreasing manner, it must be that for high $y < \overline{y}$ it is strictly positive. Hence, there must be at least one $y^* \in (\underline{y}, \overline{y})$ such that $D(y^*) = 0$ and at that point D crosses zero from below. If this interior crossing is unique, then $y^* = y^{**}$. In turn, if D has several interior zeros, then the first one, y^*, and the last one, $y^{**} > y^*$, share this 'crossing-from-below' property, proving the claim. □

Example with Unique Interior Crossing of $D(\cdot)$. We parameterize our model as follows:

$$R(\ell) = \frac{\ell - a}{b - a}, \quad b > a > 0$$
$$Q(p) = \frac{p - a}{b - a}$$
$$\Gamma(y|p) = y^p, \quad y \in [0, 1], p > 0$$

where $\ell \in [a, b]$ and $p \in [a, b]$. Thus, under PAM, $\mu(\ell) = \ell$ and $\Gamma_\ell(y|\mu(\ell)) = y^{\mu(\ell)} = y^\ell$. If $a = 1$ and $b = 2$, we can solve for the zeros of D in closed form, giving the unique interior $y^* = 0.5$. The shape of D thus resembles the one in Figure 2. Note that this example does not satisfy Assumption 2 for $\gamma(\underline{y}|p)$, which however is only sufficient (not necessary) for the result.

C Quantitative Model: Labor Mobility & Residential Housing

The firms' location choice problem has the same structure as in the baseline model, only that in (7) they now take into account that the meetings rates vary across locations.

From the firms' perspective, congestion—which can be measured by market tightness—is decreasing in the endogenous population size. If the local population is large, then market tightness is small and firm meeting rate $\lambda^F(\ell)$ is high, benefitting firms. In addition, competition that stems from poaching risk is mitigated in places with a large population: The job arrival rate for employed workers, $\lambda^E(\ell)$, decreases as the population gets larger and so the probability that firms retain workers rises.

Thus, an important question is how population size varies with ℓ. When agents conjecture that there is positive sorting between firms and locations, high-ℓ locations are more attractive (due to a stochastically better wage distribution), and draw in more workers. Labor market congestion therefore benefits firms in high-ℓ locations, $\partial\lambda^F/\partial\ell > 0$ and $\partial\lambda^E/\partial\ell < 0$, alleviating the competition channel and strengthening their desire to settle there (although this mechanism is mitigated by congestion in the residential housing market, which prevents a massive inflow of workers into high-ℓ locations). As a result, positive sorting materializes more easily than in the baseline model with exogenous meeting rates that are constant across space.

We now state our result on firm sorting under labor mobility formally. To do so, denote the minimum of the first term on the RHS of (14) (over ℓ, y) by ε^P. Assume that the labor market matching function is given by $M(\mathcal{V}(\ell),\mathcal{U}(\ell)) = \mathcal{A}\sqrt{\mathcal{V}(\ell)\mathcal{U}(\ell)}$ and that workers' flow utility function over housing and consumption is Cobb Douglas with share parameters ω and $1-\omega$, respectively. For illustration, assume that the exogenous functions $B(\cdot)$ and $\delta(\cdot)$ do not vary with ℓ (i.e., $B(\ell) = B = 1$ and $\delta(\ell) = \delta$).

Proposition 7. *If (i) z is strictly supermodular and either the productivity gains from sorting into higher ℓ, ε^P, are sufficiently large or the competition forces, $1/\delta$, are sufficiently small, and (ii) local housing supply is elastic and proportional to labor income, $h(\ell) \propto \mathbb{E}[w(y,\ell)|\ell]$, then there is positive sorting of firms p to locations ℓ.*

Proof of Proposition 7. The expected value of firm p in location ℓ is given by

$$\overline{J}(p,\ell) = \lambda^F(\ell)\delta \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(t))]^2} dt d\Gamma(y|p) - k(\ell),$$

where $\lambda^F(\cdot)$ is an endogenous function and where we will denote more compactly:

$$\hat{J}(p,\ell) := \delta \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(t))]^2} dt d\Gamma(y|p).$$

We can then compute the cross-partial derivative of $\overline{J}$ as

$$\frac{\partial^2 \bar{J}(p,\ell)}{\partial \ell \partial p} = \frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p}\lambda^F(\ell) + \frac{\partial \hat{J}(p,\ell)}{\partial p}\frac{\partial \lambda^F(\ell)}{\partial \ell}. \tag{A.9}$$

We apply integration by parts to $\hat{J}(p,\ell)$ to obtain

$$\hat{J}(p,\ell) = \delta \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|p))dy,$$

and then compute its derivatives:

$$\frac{\partial}{\partial p}\hat{J}(p,\ell) = \delta \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^2}\left(-\frac{\partial}{\partial p}\Gamma(y|p)\right) dy$$

$$\frac{\partial}{\partial \ell}\hat{J}(p,\ell) = \delta \int_{\underline{y}}^{\overline{y}} \left(\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^3} - \frac{\frac{\partial z(y,A(\ell))}{\partial y} 2\left(\lambda^E(\ell)\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right) + \frac{\partial \lambda^E(\ell)}{\partial \ell}(1-\Gamma_\ell(y))\right)}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^3}\right)(1-\Gamma(y|p))\, dy$$

$$\frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p} = \delta \int_{\underline{y}}^{\overline{y}} \left(\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^3} - \frac{\frac{\partial z(y,A(\ell))}{\partial y} 2\left(\lambda^E(\ell)\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right) + \frac{\partial \lambda^E(\ell)}{\partial \ell}(1-\Gamma_\ell(y))\right)}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^3}\right)\left(-\frac{\partial \Gamma(y|p)}{\partial p}\right) dy.$$

Plugging these derivatives into (A.9), we can write (A.9) as a single integral. Then, a sufficient condition for (A.9) to be positive (i.e., a sufficient condition for $\overline{J}(p,\ell)$ to be supermodular in (p,ℓ)) is that this integrand is positive for all $y \in [\underline{y},\overline{y}]$ and strictly so for a set of y of positive measure. Using $-\frac{\partial \Gamma(y|p)}{\partial p} \geq 0$, we then obtain the following sufficient condition for PAM:

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\left(\lambda^E(\ell)\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right) + \frac{\partial \lambda^E(\ell)}{\partial \ell}(1-\Gamma_\ell(y))\right)}{\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))} - \frac{\frac{\partial \lambda^F(\ell)}{\partial \ell}}{\lambda^F(\ell)}. \tag{A.10}$$

Define ε^P as the minimum of the LHS (as in the baseline model). It is strictly positive under our assumptions and the premise. Under labor mobility, the RHS depends on endogenous market tightness $\theta(\ell)$ through meeting rates $(\lambda^F(\ell), \lambda^E(\ell))$. Thus, the sufficient conditions for PAM from the baseline model are not readily applicable. Instead, we argue that the RHS is bounded. Thus, (A.10) holds for a large enough ε^P, made precise below.

To see this, first note that θ is *decreasing* in ℓ. Equivalently, λ^E is *decreasing* in ℓ and λ^F is *increasing* in ℓ. This follows from the fact that the value of unemployment is increasing in ℓ for a *fixed* λ^U (and thus a fixed $\lambda^E = \kappa\lambda^U$). Recall the value of unemployment in this extension:

$$\rho V^U(\ell) = d(\ell)^{-\omega}\left(z(\underline{y}, A(\ell)) + \lambda^E(\ell)\left[\int_{z(\underline{y},A(\ell))}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta + \lambda^E(\ell)(1 - F_\ell(t))}dt\right]\right).$$

Using the premise that $h(\ell) \propto \mathbb{E}[w(y, \ell)]$, government budget constraint (18) and housing market clearing (19), as well as (A.14), we obtain the following housing price:

$$d(\ell) \propto \frac{\omega}{1 - \omega\tau}(1 - u(\ell))L(\ell) = \frac{\omega}{1 - \omega\tau}\frac{\mathcal{A}^2}{(\delta + \kappa\lambda^U(\ell))\lambda^U(\ell)}.$$

The value of search can therefore be written as

$$\rho V^U(\ell) \propto \left(\frac{\omega}{1 - \omega\tau}\frac{\mathcal{A}^2}{(\delta + \lambda^E(\ell))\lambda^U(\ell)}\right)^{-\omega}\left(z(\underline{y}, A(\ell)) + \lambda^E(\ell)\left[\int_{z(\underline{y},A(\ell))}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta + \lambda^E(\ell)(1 - F_\ell(t))}dt\right]\right).$$

As the wage is strictly increasing in firm productivity and thus in their local rank, we can express this value as a function of the firm's rank in the local productivity distribution, $\mathcal{R}$, instead of the firm's wage rank. Set $t = w(\mathcal{R}, \ell)$. Using $F_\ell(t) = \mathcal{R}$, a change of variables yields:

$$\rho V^U(\ell) \propto \left(\frac{\omega}{1 - \omega\tau}\frac{\mathcal{A}^2}{(\delta + \lambda^E(\ell))\lambda^U(\ell)}\right)^{-\omega}\left(z(\underline{y}, A(\ell)) + \lambda^E(\ell)\left[\int_0^1 \frac{1 - \mathcal{R}}{\delta + \lambda^E(\ell)(1 - \mathcal{R})}\frac{\partial w(\mathcal{R}, \ell)}{\partial \mathcal{R}}d\mathcal{R}\right]\right). \tag{A.11}$$

Fixing, for the moment, $\lambda^U(\ell) = \lambda^U$ and thus also $\lambda^E = \kappa\lambda^U$, we can differentiate $\rho V^U(\ell)$:

$$\left.\frac{\partial \rho V^U}{\partial \ell}\right|_{\lambda^U(\ell) = \lambda^U} \propto d^{-\omega}\left(\frac{\partial z}{\partial A}\frac{\partial A(\ell)}{\partial \ell} + \lambda^E\int_0^1 \frac{1 - \mathcal{R}}{\delta + \lambda^E(1 - \mathcal{R})}\frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \mathcal{R}\partial \ell}d\mathcal{R}\right).$$

Denote firm y's local productivity rank by $\mathcal{R} = \Gamma_\ell(y)$. We apply a change of variables to wage function (5) (with $\Gamma_\ell(t) = x,\ \gamma_\ell(t)dt = dx$) and take the cross-partial derivative wrt $(\mathcal{R}, \ell)$:

$$w(\mathcal{R},\ell) = z(\Gamma_\ell^{-1}(\mathcal{R}), A(\ell)) - [\delta + \lambda^E(1-\mathcal{R})]^2 \int_0^{\mathcal{R}} \frac{\frac{\partial z(\Gamma_\ell^{-1}(x),A(\ell))}{\partial y}}{[\delta+\lambda^E(1-x)]^2} \frac{1}{\gamma_\ell(\Gamma_\ell^{-1}(x))} dx$$

$$\frac{\partial^2 w(\mathcal{R},\ell)}{\partial \mathcal{R} \partial \ell} = 2\frac{\lambda^E}{\delta}\left(1+\frac{\lambda^E}{\delta}(1-\mathcal{R})\right)\frac{\partial}{\partial \ell}\int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^2} dt. \tag{A.12}$$

Suppose that $\Gamma_\ell^{-1}(\mathcal{R})$ is increasing in ℓ (which is true if $\frac{\partial}{\partial \ell}\Gamma_\ell \leq 0$). In addition, suppose that, for any given λ^E such that $\underline{\lambda}^E \leq \lambda^E \leq \overline{\lambda}^E$ with $\underline{\lambda}^E = \min_\ell \lambda^E(\ell)$ and $\overline{\lambda}^E = \max_\ell \lambda^E(\ell)$, the integrand of (A.12), $\frac{\partial z(y,A(\ell))}{\partial y}/(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y)))^2$, is also increasing in ℓ. Both of these statements are true under the sufficient conditions for PAM that we provide below. Thus, the wage function is supermodular in $(\mathcal{R},\ell)$. This ensures that $\frac{\partial \rho V^U}{\partial \ell}\big|_{\lambda^U(\ell)=\lambda^U} > 0$.

This, together with the fact that V^U is increasing in λ^U (and thus λ^E), implies that for the equilibrium indifference condition of unemployed workers to hold (i.e., the value of unemployment, V^U, is equalized across ℓ), it must be that λ^U (and thus λ^E) is decreasing in ℓ, and so θ is decreasing in ℓ. Thus, λ^F is increasing in ℓ, rendering the second and third term on the RHS in (A.10) negative.

It then suffices that the first (positive) term on the RHS of (A.10) is bounded. Note that $\lambda^E(\cdot)$ is implicitly defined by (A.11), where, in equilibrium, V^U is a number that no longer depends on ℓ. All functions in this expression $(z, \partial w/\partial \mathcal{R})$ are continuous in ℓ (see (A.12)) where $\ell \in [\underline{\ell}, \overline{\ell}]$, and thus λ^E inherits this property.

The first term on the RHS of (A.10) is thus bounded from above by

$$\frac{2\overline{\lambda}^E\left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)}{\delta}.$$

Define

$$\tilde{t}^P \equiv \overline{\lambda}^E\left(\max_{y,\ell}\left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)\right),$$

which is positive and well-defined given that $\Gamma(y|p)$ is continuously differentiable in p, where both p and y are defined over compact sets, and cdf's Q and R are continuously differentiable on the intervals $[\underline{p}, \overline{p}]$ and $[\underline{\ell}, \overline{\ell}]$ with strictly positive densities (q, r).

Then, PAM obtains if ε^P is large enough or if $1/\delta$ is small enough, that is, $\varepsilon^P > 2\frac{1}{\delta}\tilde{t}^P$, since this ensures that (i) inequality (A.10) holds; and thereby (ii) $\Gamma_\ell^{-1}(\mathcal{R})$ is increasing in ℓ and the integrand of (A.12) is increasing in ℓ (which is equivalent to the condition that makes the integrand of $\frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p}$ positive), both of which we had postulated above. $\square$

D Data and Sample Restrictions

D.1 Administrative Regional-Level Data from the GFSO

Data Description. We obtain regional-level data from the German Federal Statistical Office (GFSO). To be consistent with our sample from the FDZ below, we focus on the years 2010-2017. We obtain district-level data (for 401 districts) for all years and aggregate them to the commuting-zone level (there are 257 CZs), using a crosswalk provided by the Federal Office for Building and Regional Planning of Germany (*Bundesinstitut für Bau-, Stadt- und Raumforschung*—BBSR). Finally, we take (simple) averages across years to obtain one value for each variable per commuting zone. If applicable, we adjust the variables to the monthly level, again for consistency with our FDZ sample.

Defining Important Variables.

Labor Compensation. Total *labor compensation* in a commuting zone at year t is defined as

$$\text{Labor Comp}_t = \frac{\text{Total hours worked by total workforce}_t}{\text{Total hours worked by employees}_t} \times \text{Comp of Employees}_t,$$

and *compensation of employees* consists of gross wages and salaries as well as employers' actual and imputed social contributions. We divide by 12 to obtain the monthly statistic.

Value Added per Worker. The monthly gross *value added per worker* is calculated as the ratio of (annual) gross value added and total employment, divided by 12.

Labor Share. We construct the local *labor share* as the ratio between *labor compensation* and gross value added in each commuting zone.

Average Wage. The *average monthly wage* of a commuting zone is defined by total (annual) *labor compensation* divided by total employment, divided by 12.

Average Firm Size. We define *average firm size* of a commuting zone by the total number of employees over the total number of establishments.

GDP per Capita. We take the ratio of (annual) GDP and population in each commuting zone; then divide by 12 to get the monthly figure. GDP corresponds to the gross value added of all sectors of the economy plus taxes on products, but excluding subsidies on products.

Unemployment Rate. We first use unemployment rates and number of unemployed workers at the district level to obtain the number of people who are in the labor force in each district. We then sum by commuting zone the number of unemployed workers as well as the number of people in the labor force and divide them to obtain the local *unemployment rate.*

Rent-to-Income Ratio. We use the Germany-wide *rent-to-income ratio* of the main tenant household.

Trade Tax Rate. The trade tax (*Gewerbesteuer*) is levied on the adjusted profit of corporations. It is a combination of base rate (universal to all municipalities, 3.5%) and a municipal tax rate (which is a multiplier to the base rate and at the discretion of each municipality). We focus on the municipal tax rate and refer to it as *trade tax rate.* We first aggregate municipal tax rates to the district level and then to the CZ level using population weights.

Share of Employees with a Degree. We take the ratio of employees with an academic degree and all employees subject to social security contributions at the place of work.

Net Business Registration Intensity. We define *net business registration intensity* at the CZ level as the balance between business registrations and de-registrations per 1,000 inhabitants.

D.2 Administrative Worker- and Firm-Level Data from the FDZ

Data Description. We draw from the Linked Employer-Employee Data (LIAB) provided by the Research Data Centre (FDZ) of the German Federal Employment Agency at the Institute for Employment Research. For more information, see Ruf et al. (2021a) and Ruf et al. (2021b).

The LIAB data links information on establishments from the IAB Establishment Panel (EP), a representative annual establishment survey, with information on all individuals employed at those establishments.[37] Surveyed establishments (the 'panel cases') are followed between 2009-2016 and we observe individual-level information for *all* their employees. This individual-level data, which includes workers' gender, education, full-time employment status, gross daily wages and work district, is assembled from official social security records. Moreover—from the Establishment History Panel 7518 (BHP), which we describe below—we observe basic information for *any* employer in those workers' entire employment history between 1975 and 2018. We call this dataset LIAB-BHP. While the original data is in a spell format, we transform it into a monthly panel.[38] We augment these datasets with the BHP, a 50% random sample of all German establishments with at least one employee subject to social security as of June 30 in any given year. In addition to standard information like total employment, average wages, sales or costs of inputs, it contains total inflows and outflows of workers at the establishment level.

[37] Because we only observe data at the level of the establishment, we use 'establishments' and 'firms' interchangeably.

[38] If a new spell starts in the middle of a month, we assign the month to the longest spell within the month.

Sample Restrictions. Our baseline sample pools the years 2010-2017.[39] We focus on full-time employees. We drop establishments with less than 5 employees and establishments whose mean real daily wage across the sample period is lower than 15 Euros, measured in 2015 euros (this wage restriction is based on Card et al. (2013)). We use various datasets from FDZ: LIAB, LIAB-BHP, BHP and EP.

Defining Important Variables.

Monthly Real Wage. To compute an individual's *monthly real wage*, we multiply daily wages by 30, and deflate these nominal wages using the German CPI (Table 61111-0001 in the GENESIS database of the Federal Statistical Office). The CPI base year is 2015. Data Source: Establishment History Panel (BHP).

Value Added per Full-Time Employee. We measure value added at the firm-level as the difference between sales and input costs as reported in the Establishment Panel, divided by the number of full-time employees. See also Bruns (2019). We deflate these variables using the same CPI as above. Data Source: Establishment Panel (EP).

Employment-to-Employment (EE) Transition. We say a worker made an EE move in month t in any of the following scenarios: (i) if they were employed at some establishment in month $t-1$ and are employed at a different establishment in month t; (ii) if they are employed at some establishment in month $t-3$ (or $t-2$) and disappear from the sample during months $t-2$ and $t-1$ (or only $t-1$) without claiming unemployment benefits, and are employed again at a different establishment in month t. In this second scenario, we consider it likely that the new job was already lined up when the worker left the previous one. Data Source: Linked Employer-Employee Data (LIAB).

Unemployment-to-Employment (UE) Transition. A worker made a UE move in month t if they were unemployed—that is, collecting unemployment benefits—in month $t-1$ and are employed at some establishment in t. Data Source: Linked Employer-Employee Data (LIAB).

Employment-to-Unemployment (EU) Transition. A worker made an EU move in month t if they were employed at some establishment in month $t-1$ and are (officially) unemployed in t or permanently disappear from the sample (we exclude December 2017 from this count, since it is the last month in our panel). Data Source: Linked Employer-Employee Data (LIAB).

Labor Market Transition Rates. In our regression analysis, we construct measures of work-

[39] The reason is that for some parts of the empirical analysis we use the establishment-level fixed effects, provided by the FDZ for all establishments in the LIAB-BHP. These are estimated by Card et al. (2013) for the period 2010-2017 using the methodology developed in Abowd et al. (1999).

ers' monthly transition rates from the data: We proxy the contact rate of employed workers λ^E by the realized EE transition rate in the data. For the contact rate λ^U and the rate of job destruction δ—since in the model, unemployed workers accept all offers and separations to unemployment are exogenous—they are equal to the realized rates. Specifically, in each t:

$$\begin{aligned}
\lambda_t^E &= \frac{\#\text{ Employed workers in } t-1 \text{ working in another firm in } t}{\#\text{ Employed workers in } t-1} \\
\lambda_t^U &= \frac{\#\text{ Unemployed workers in } t-1 \text{ who are employed in } t}{\#\text{ Unemployed workers in } t-1} \\
\delta_t &= \frac{\#\text{ Employed workers in } t-1 \text{ who are unemployed in } t}{\#\text{ Employed workers in } t-1}.
\end{aligned}$$

We measure these flows at the monthly frequency in each local labor market and then take the average over years 2010-2017 to obtain one number per local labor market. Data Source: Linked Employer-Employee Data (LIAB).

Firm Productivity y. Measuring y empirically is complicated by the fact that firms—according our theory—are sorted spatially and, thus, their output z and wages w depend not only on their productivity y but also on location productivity $A(\ell)$. Therefore, we cannot readily rank firms by either of these observable (wage or value added) statistics. In our structural estimation below, we show that we can identify firm productivity separately from location productivity. In turn, to obtain a measure of y for our reduced-form exercises, we implement the following procedure. Our starting point is the firm fixed effects from a standard two-way fixed effects wage regression, provided by the FDZ for all establishments in the LIAB-BHP. These are estimated by Card et al. (2013) for the period 2010-2017 using the methodology developed in Abowd et al. (1999) (henceforth, AKM). Using these fixed effects, we control for the effect of worker sorting across firms or locations, which is not present in our theory. But we still have to disentangle the effect of y from that of $A(\ell)$. To control for location productivity $A(\ell)$, we regress firm-level fixed effects on the average value added per worker in their location, and consider the residual of this regression as a proxy for y.

Poaching Share. To measure job flows and poaching at the firm level, we follow Moscarini and Postel-Vinay (2018) and Bagger and Lentz (2019) and measure firms' *poaching shares*, which we define as the ratio of EE inflows relative to all inflows.[40] Given our focus on *local* labor markets, we also compute firms' share of EE inflows and UE inflows that are local, i.e., from within the same commuting zone. Data Source: LIAB-BHP.

[40]In the terminology of Moscarini and Postel-Vinay (2018) and Bagger and Lentz (2019), this object refers to the poaching inflow share and the poaching index, respectively.

D.3 Variables from Other Data Sources

Residential Housing Prices. We use residential rental rates predicted for the centroids of postal codes (provided to us by Gabriel Ahlfeldt based on Ahlfeldt et al. (2022)) and aggregate them to the commuting-zone level. The model counterpart is $d(\ell)$ for each CZ ℓ.

Replacement Rate. We use the unemployment insurance net replacement rate. This variable is based on data from the Out-of-Work Benefits Dataset (OUTWB), provided as part of the Social Policy Indicator (SPIN) database (Nelson et al., 2020). Depending on household composition and earnings, replacement rates vary and we take 60% as a reference point.

Commercial Real Estate Prices. We use price data (EUR/m^2) for commercial properties 2012/13 from the German Real Estate Association (*Deutscher Immobilienverband*). We aggregate prices from the city to the commuting-zone level. The model counterpart is $k(\ell)$.

Distance to Highway. Distance to highway is proxied by the area-weighted average car driving time to the next federal motorway junction in minutes. We obtain this variable from the German Federal Office for Building and Regional Planning. Data is available only for 2020.

D.4 Defining Locations

Local Labor Markets. We consider 257 commuting zones (*Arbeitsmarktregionen*)—our local labor markets. These are defined for the year 2017 by the Federal Office for Building and Regional Planning of Germany (*Bundesinstitut für Bau-, Stadt- und Raumforschung*—BBSR).

East-West. We categorize commuting zones into East or West Germany based on whether the districts they consist of belong to Eastern or Western states. Many commuting zones contain more than one district; however, there are no commuting zones containing districts from both East and West Germany. We omit Berlin from our sample.

Urban-Rural. We categorize commuting zones into Rural or Urban based on their districts. To classify a district as Urban or Rural, we use the classification provided by the BBSR for the year 2018 (we use the 2017 definition of commuting zones and the 2018 definition of Urban/Rural because the 2017 definition of Urban/Rural has more than two categories, e.g., 'Mostly Rural', which would require more choices on our end). When a CZ is formed by districts that are all rural, we classify the CZ as Rural. When a CZ has at least one district that is urban, we classify it as Urban (note that there are only 27 commuting zones that have both urban and rural districts, the remaining 230 CZs are unambiguous).

E Empirical Analysis

E.1 Local Labor Markets

In Table A.1, we report aspects of the cross-sectional distribution of economic outcomes across local labor markets in Germany. In Table A.2, we give information on firms' poaching behavior, both at the firm level (Panel 1) and at the local level (Panel 2).

Table A.1: Spatial Heterogeneity: Distribution of Key Statistics

	Mean	S.D.	P10	P25	P50	P75	P90
Average Wages	3,132.72	400.87	2,617.73	2,849.02	3,092.29	3,361.76	3,661.95
Average Value Added	4,640.04	687.55	3,911.31	4,204.93	4,525.91	4,871.12	5,518.41
Average Firm Size	11.13	1.68	9.28	10.07	11.02	12.14	12.99
Share Emp. Top 10%	0.56	0.06	0.49	0.52	0.55	0.59	0.63
Population Density	292.20	421.55	83.29	109.94	164.73	272.41	589.32
Population	317,149	420,183	92,979	127,139	190,745	325,078	596,007

Notes: Data source: German Federal Statistical Office for all variables except 'share of employment of the largest 10% of firms' (Share emp. top 10%), which we compute from the BHP (using full-time employees only). Displayed statistics are computed at the commuting-zone level, and so the number of observations is 257. *Mean* (*S.D.*) is the average (standard deviation) of each variable across 257 commuting zones. *P10-P90* are the percentiles of their distributions. Wages and value added are reported at the monthly level, in 2015 Euros. See Appendix D.1 for more details on how the displayed variables are defined.

Table A.2: On-the-Job Search and Local Labor Markets

	Mean	S.D.	P10	P25	P50	P75	P90
Firm level (N = 5,895)							
Poaching Share	0.51	0.12	0.35	0.44	0.52	0.60	0.63
Share of local EE	0.70	0.17	0.47	0.62	0.73	0.81	0.88
Share of local UE	0.55	0.22	0.31	0.42	0.54	0.70	0.83
Commuting-zone level (N = 257)							
Poaching Share	0.49	0.05	0.42	0.46	0.49	0.53	0.55
Share of local EE	0.69	0.09	0.57	0.64	0.70	0.76	0.79
Share of local UE	0.58	0.11	0.45	0.52	0.60	0.66	0.69

Notes: Data source: LIAB-BHP, restricted to panel cases. In Panel A (Panel B) we report the statistics at the firm level (commuting-zone level). To aggregate the firm-level outcomes to the commuting-zone level, we weigh firms by total employment. The commuting-zone level statistics are weighed by the number of establishments in that location. EE and UE flows as well as Poaching Share are defined in Appendix D.2. Share of 'local' EE or UE transitions means that we divide worker transitions within a given commuting zone by total transitions to firms in that commuting zone.

E.2 Local Job Ladders and Firm-Level Outcomes: Empirical Evidence

We provide direct evidence on the importance of firms' local competitiveness for their outomes. Our theory predicts that there is a systematic relationship between firms' local competitiveness, captured by firms' *local rank* $\Gamma_\ell(y)$, and several firm-level outcomes. As we show in Lemma O1 (Online Appendix OB), firm-level wages w, net poaching shares n^E and size l are all increasing in $\Gamma_\ell(y)$. We now assess these predictions empirically by considering the specification

$$\ln \mathcal{D}_{f,\ell} = \alpha + \beta \Gamma_\ell(y_f) + X'_{f,\ell}\gamma + \epsilon_{f,\ell}, \tag{A.13}$$

where y_f is productivity y of firm f, $\mathcal{D}_{f,\ell}$ denotes the dependent variable (either $n^E(y_f, \ell)$—see (O.2) in Online Appendix OB for the definition—or $l(y_f, \ell)$ or $\ln w(y_f, \ell)$); $\Gamma_\ell(y_f)$ is the local productivity rank of firm f in location ℓ; and $X_{f,\ell}$ is a vector of additional controls.

In Panel A of Table A.3, we show the estimation results of (A.13) for firm-level net poaching share $n^E(y_f, \ell)$. Consistent with our theory, there is a strong positive correlation between a firm's net poaching share and its local productivity rank (column 1). In columns 2 and 3, we control for additional firm-level characteristics and a labor-market fixed effect. In column 4, we predict the net poaching share by the firm's *global* rank (instead of its local rank). Again, there is a positive relationship owing to the positive correlation between global and local ranks. In the last column, however, which controls for both the local and global rank, the global rank loses much of its predictive power—in line with our theory.

Panel B of Table A.3 focuses on the determinants of firm size $l(y_f, \ell)$. The structure is identical to Panel A. The first three columns again show a strong positive relationship between a firm's size and its local rank. In particular, as in the case of the net poaching share, the relationship between firm size and the local rank vs. the global rank is similar (columns 3-4). But column 5 shows that the local, not the global, rank is the main predictor of firm size.

Finally, in Panel C, we study firm-level (log) wages $\ln w(y_f, \ell)$. Again, we find a strong positive correlation with the local productivity rank (columns 1 and 2), which also holds true within local labor markets (column 3). The last column shows that the distinction between the local and global productivity rank is less clear cut. Our theory explains why. Consider two firms with the same global rank (same y) who are in two different markets $\ell' < \ell''$. The firm in market ℓ' has a higher local rank (due to overall worse firm composition in that market under PAM), but given $\ell' < \ell''$, it also faces $A(\ell') < A(\ell'')$. Thus, conditional on the firm's global rank, its local rank is *negatively* related to local wages due to this TFP effect.

Table A.3: The Local Rank and Firm-Level Outcomes

	(1)	(2)	(3)	(4)	(5)
Panel A: Net poaching share ($N = 368,331$)					
Local Rank	0.25***	0.18***	0.18***		0.13***
	(0.01)	(0.00)	(0.00)		(0.01)
Global Rank				0.17***	0.05***
				(0.00)	(0.01)
Panel B: Normalized log full-time employees ($N = 393,567$)					
Local Rank	2.96***	2.37***	2.30***		2.35***
	(0.11)	(0.08)	(0.06)		(0.26)
Global Rank				2.22***	0.02
				(0.07)	(0.27)
Panel C: Log real wage of full-time employees ($N = 393,567$)					
Local Rank	1.19***	0.94***	0.94***		0.04
	(0.02)	(0.02)	(0.02)		(0.04)
Global Rank				0.97***	0.94***
				(0.01)	(0.04)
Controls	N	Y	Y	Y	Y
Location FE	N	N	Y	N	N

Notes: Data source: LIAB-BHP. Observations are at the firm level. Standard errors clustered by commuting zone are in parentheses. Global (local) ranks describe the rank of each firm in the global (local) distribution of firm productivity, proxied by the residualized AKM firm fixed effects (see Appendix D.2 for details). Regressions (except Panel B) are weighted by mean (across years) number of full-time employees. Firm size in panel B is measured relative to the firm size of the 5th percentile in the local firm size distribution. All panels control for the CZ-level unemployment rate and 3-digit industry fixed effects. Panel C additionally controls for the CZ-level log real value added and the mean share of full-time and marginal employees.

E.3 Spatial Wage Inequality

Real Spatial Wage Inequality versus Nominal Spatial Wage Inequality.

Table A.4: Spatial Inequality (Monthly, €): Real versus Nominal

	German CPI		Local CPI	
	Wage	Value Added	Wage	Value Added
West-East Inequality				
West	3491.13	5237.02	3704.85	5552.25
East	2731.63	4045.24	3122.56	4624.49
West/East	1.28	1.30	1.19	1.20
Urban-Rural Inequality				
Urban	3510.01	5270.60	3701.94	5552.52
Rural	2984.37	4429.01	3372.72	5007.15
Urban/Rural	1.18	1.19	1.10	1.11

Notes: Data source: German Federal Statistical Office. With some abuse, we denote by 'Nominal' those variables that are deflated using the *Germany-wide* CPI in 2015; and by 'Real' we denote the variables that are deflated using the *Local* CPI, i.e., using commuting zone-level price deflators (computed from district-level price deflators from BBSR).

Wage Variance Decomposition.

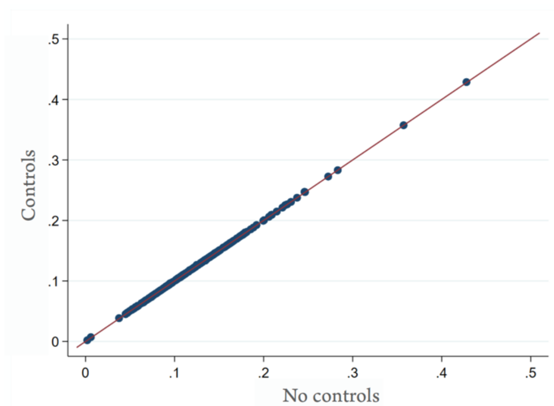
Table A.5: Variance Decomposition of Log Wages

	Total	Location FE	Residual
Variance	0.341	0.071	0.270
Share in %		20.90	79.10

Notes: Data Source: LIAB. N =72,632,961. We run a regression of individual-level log wages on location (CZ) fixed effects and compute the variance of the estimated fixed effects and the variance of the residuals (first row). We then compute the shares of these variances in the total wage variance (second row).

Wage Growth due to EE Transitions: Robustness. We provide some robustness on regression (17), by showcasing how the coefficients of interest change if we include additional controls: age, education and gender. In Figure A.1 we plot, on the horizontal axis, the demeaned estimated coefficients based on (17) (labeled 'No Controls') and, on the vertical axis, the demeaned estimated coefficients based on an augmented regression that controls for indicator variables for age bins, education, and gender (labeled 'Controls'). Since the coefficients of the baseline and augmented regression line up along the 45 degree line, our results from Figure 4 do not appear to be driven by omitting these variables from the baseline regression.

Figure A.1: Wage Growth due to EE Transitions: Robustness



Decomposition of Life-Time Earnings. In this exercise, we study the impact of heterogeneous job ladders across space on spatial inequality in life-time earnings. We compare life-time earnings in two regions, 'rich' and 'poor' locations (where 'rich' and 'poor' locations refer to the top and bottom 25% commuting zones in terms of GDP per capita). We focus on a single cohort of workers in each region: They are 25-30 years old in 2002, and we follow them over 15 years, from 2002 to 2017. We restrict the sample to those workers who remain in the region where we first observe them.

First, we measure the average starting wages of workers in rich and poor locations, before they begin climbing the job ladder. Second, we compute average wages of workers in each region after 15 years. Third, we decompose the total average wage growth within regions into three parts: (i) the average wage growth of workers who never changed jobs nor experienced unemployment for more than four months (i.e., the 'stayers'), (ii) the average wage growth of workers who changed jobs at least once and did not experience unemployment for more than four months (i.e., the 'EE movers'), and (iii) the average wage growth of workers who have been unemployed at least once for more than four months (which we call the 'unemployed'). Average wage growth of a region is equal to the weighted average of wage growth in these three categories, with the weights being equal to the number of workers in each category.

This decomposition of average wage growth allows us to assess the contribution of heterogeneous job ladders across space to spatial wage inequality as follows. We compute wage growth in rich locations imposing the (counterfactual) wage growth of EE movers from poor locations, while keeping the number of EE movers fixed. This way, we get a measure of life-time income inequality across space keeping job ladders *the same* across regions.

The results are in Table A.6. If poor and rich regions had the same job ladder, spatial inequality in life-time income would be 24% percent lower than under the heterogeneous job ladders that we see in the data, i.e., the rich region would be characterized by 41.6% higher wages than the poor one, instead of the observed 54.8%.

Table A.6: Decomposition of Life-Time Earnings in Top and Bottom 25% of Local Labor Markets

	Top 25%	Bottom 25%
Wage Growth (total)	0.916	0.633
Wage Growth of Stayers	0.854	0.642
Wage Growth of Unemployed	0.680	0.566
Wage Growth of EE Movers	1.052	0.655
Starting Wage	3036.18	2301.67
Wage after 15 Years	5817.34	3758.63
After-15-years Spatial Wage Inequality (data)	1.548	
After-15-years Spatial Wage Inequality under the Same Job Ladder (counterfactual)	1.416	
Contribution of Job Ladder Differences to Spatial Wage Inequality	0.241	

Notes: Data source: LIAB. Top and bottom 25% of local labor markets (CZs) are categorized based on GDP per capita. The last row reports the percentage difference between the actual (row 7) and counterfactual (row 8) spatial wage inequality 24% $\sim$ (54.8−41.6)/54.8.

F Identification

We prove identification of our model under the following assumption:

Assumption 3. *We assume the following functional forms and normalizations:*

1. *The labor market matching function is given by* $M(\mathcal{V}(\ell), \mathcal{U}(\ell)) = \mathcal{A}\sqrt{\mathcal{V}(\ell)\mathcal{U}(\ell)}$.
2. *Workers' flow utility function over housing and consumption is Cobb Douglas with share parameters* ω *and* $1-\omega$.
3. *The ex post firm productivity distribution is given by* $\Gamma(y \mid p) = 1 - y^{-\frac{1}{p}}$.
4. *The production function is given by* $z(y, A(\ell)) = yA(\ell)$.
5. R *is given.*
6. *Normalize* $\rho V^U = 1$.

Proof of Proposition 5. We need to identify the ranking of locations $[\underline{\ell}, \overline{\ell}]$; functions $(Q,\ A,\ B)$; the tail parameters p of the ex post productivity distribution; the separation rate schedule δ; the relative matching efficiency κ and the overall efficiency of the matching function $\mathcal{A}$; as well as the parameters pertaining to the housing market (ω, τ, h).

First, we can assign $\ell \in [\underline{\ell}, \overline{\ell}]$ to each location, based on any observed statistic that—according to our model—is increasing in ℓ.

Second, $\mu(\ell)$ (and thus $p = \mu(\ell)$) can be obtained from a location's labor share, $LS(\ell) = 1 - \mu(\ell)$. That is, under goods' market clearing, aggregate output in ℓ equals aggregate wages plus profits and land prices in equilibrium

$$\int_{\underline{y}}^{\overline{y}} z(y, A(\ell)) l(y,\ell) d\Gamma(y|(\mu(\ell)) = \int_{\underline{y}}^{\overline{y}} w(y,\ell) l(y,\ell) d\Gamma(y|(\mu(\ell)) + \varphi^F \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{A(\ell)}{[1+\varphi^E(1-\Gamma(t|\ell))]^2} dt d\Gamma(y|(\mu(\ell))$$
$$= \int_{\underline{y}}^{\overline{y}} w(y,\ell) l(y,\ell) d\Gamma(y|(\mu(\ell)) + \varphi^F \int_{\underline{y}}^{\overline{y}} \frac{A(\ell)}{[1+\varphi^E(1-\Gamma(y|(\mu(\ell)))]^2} (1-\Gamma(y|(\mu(\ell)))dy,$$

where we use integration by parts for the second line. Thus, the labor share is given by:

$$LS(\ell) := \frac{\int_{\underline{y}}^{\overline{y}} w(y,\ell) l(y,\ell) d\Gamma(y|(\mu(\ell))}{\int_{\underline{y}}^{\overline{y}} z(y, A(\ell)) l(y,\ell) d\Gamma(y|(\mu(\ell))} = 1 - \frac{\varphi^F \int_{\underline{y}}^{\overline{y}} \frac{A(\ell)}{[1+\varphi^E(1-\Gamma(y|(\mu(\ell)))]^2} (1-\Gamma(y|(\mu(\ell)))dy}{\int_{\underline{y}}^{\overline{y}} z(y, A(\ell)) l(y,\ell) d\Gamma(y|(\mu(\ell))}.$$

At the same time, aggregate output can be expressed as follows, using that Γ is Pareto and firm size expression (3):

$$\int_{\underline{y}}^{\overline{y}} z(y, A(\ell))l(y, \ell)d\Gamma(y|(\mu(\ell)) = \int_{\underline{y}}^{\overline{y}} A(\ell)y \cdot l(y, \ell)\frac{1}{\mu(\ell)}y^{-\frac{1}{\mu(\ell)}-1}dy$$
$$= \frac{1}{\mu(\ell)}\varphi^F \int_{\underline{y}}^{\overline{y}} \frac{A(\ell)}{[1+\varphi^E(1-\Gamma(y|(\mu(\ell)))]^2}(1-\Gamma(y|(\mu(\ell)))dy$$

Plugging aggregate output back into $LS(\ell)$ above, we obtain $LS(\ell) = 1 - \mu(\ell)$. If μ satisfies PAM, Q is identified from $\mu(\ell) = Q^{-1}(R(\ell))$, for a given R.

Third, we obtain κ as described in the text. In turn, equation (24), which allows us to back out the overall matching efficiency, is derived as follows. First, note that:

$$\begin{aligned}
\lambda^E(\ell) &= \frac{M(\mathcal{V}(\ell), \mathcal{U}(\ell))}{(u(\ell)+\kappa(1-u(\ell)))L(\ell)}\kappa \\
&= M(\mathcal{V}(\ell), \mathcal{U}(\ell))\frac{\kappa(1-u(\ell))}{u(\ell)+\kappa(1-u(\ell))}\frac{1}{(1-u(\ell))L(\ell)} \\
&= \mathcal{A}\mathcal{V}(\ell)^{\frac{1}{2}}\big(u(\ell)+\kappa(1-u(\ell))L(\ell)\big)^{\frac{1}{2}}\frac{\lambda^E(\ell)}{\delta(\ell)+\lambda^E(\ell)}\frac{\delta(\ell)+\lambda^U(\ell)}{\lambda^U(\ell)}\frac{1}{L(\ell)} \\
&= \mathcal{A}\left(\frac{\delta(\ell)+\lambda^E(\ell)}{\delta(\ell)+\lambda^U(\ell)}\right)^{\frac{1}{2}}\frac{\lambda^E(\ell)}{\delta(\ell)+\lambda^E(\ell)}\frac{\delta(\ell)+\lambda^U(\ell)}{\lambda^U(\ell)}L(\ell)^{-\frac{1}{2}} \\
\Rightarrow \quad L(\ell) &= \mathcal{A}^2\frac{\delta(\ell)+\lambda^U(\ell)}{\delta(\ell)+\kappa\lambda^U(\ell)}\left(\frac{1}{\lambda^U(\ell)}\right)^2. \qquad \text{(A.14)}
\end{aligned}$$

Next, note that average firm size in location ℓ is given by $\bar{l}(\ell) = (1-u(\ell))L(\ell)$, and thus,

$$L(\ell) = \left(1 + \frac{\delta(\ell)}{\lambda^U(\ell)}\right)\bar{l}(\ell). \qquad \text{(A.15)}$$

Equalizing (A.14) and (A.15), and solving for $\mathcal{A}$ gives (24) in the text, where we treat $(\lambda^U(\ell), u(\ell), \bar{l}(\ell))$ as observed for all ℓ.

Fourth, we obtain $\delta(\ell)$ in each ℓ from local unemployment and job-finding rates, see (23).

Fifth, we obtain the A-schedule from how average value added varies across space:

$$\mathbb{E}[z(y, A(\ell))|\ell] = A(\ell)\mathbb{E}[y|\ell] = A(\ell)\tfrac{1}{1-\mu(\ell)}$$
$$\Rightarrow A(\ell) = (1-\mu(\ell))\mathbb{E}[z(y, A(\ell))|\ell].$$

Sixth, regarding the housing market parameters, we treat $(u(\ell), L(\ell), d(\ell), \mathbb{E}[w(y, \ell)|\ell], \mathcal{R})$ as observed for all ℓ (where $\mathcal{R}$ is the economy-wide replacement rate of the unemployed) and obtain $(\omega, \tau, h(\cdot), w^U(\cdot))$ from a system of four equations:

We obtain ω from the expenditure share of housing in any location ℓ (for given $(h(\ell), w^U(\ell))$,

$$\omega = \frac{h(\ell)d(\ell)}{w^U(\ell)u(\ell)L(\ell) + \mathbb{E}[w(y,\ell)|\ell](1-u(\ell))L(\ell)}.$$

Then, given ω, we obtain the tax rate on home owners, τ, using government budget constraint (18), housing market clearing (19), and replacement rate $\mathcal{R}$, which satisfies

$$\mathcal{R}\sum_{\ell}\mathbb{E}[w(y,\ell)|\ell]\frac{(1-u(\ell))L(\ell)}{\sum_{\hat{\ell}}(1-u(\hat{\ell}))L(\hat{\ell})} = \sum_{\ell} w^U(\ell)\frac{u(\ell)L(\ell)}{\sum_{\hat{\ell}}u(\hat{\ell})L(\hat{\ell})},$$

where the RHS is the aggregate unemployment benefit. Note that

$$\sum_{\ell} w^U(\ell)\frac{u(\ell)L(\ell)}{\sum_{\hat{\ell}}u(\hat{\ell})L(\hat{\ell})} = \tau\sum_{\ell}\frac{d(\ell)h(\ell)}{\sum_{\hat{\ell}}u(\hat{\ell})L(\hat{\ell})} = \frac{\omega\tau}{1-\tau\omega}\sum_{\ell}\frac{\mathbb{E}[w(y,\ell)|\ell](1-u(\ell))L(\ell)}{\sum_{\hat{\ell}}u(\hat{\ell})L(\hat{\ell})},$$

where the first equality uses government budget constraint (18) and the second one uses a combination of (18) and housing market clearing (19), which gives $\mathbb{E}[w(y,\ell)|\ell](1-u(\ell))L(\ell) = (1-\omega\tau)d(\ell)h(\ell)$. Equalizing the last two equations and solving for τ gives:

$$\tau = \frac{1}{\omega}\frac{\mathcal{R}}{\frac{\sum_{\hat{\ell}}(1-u(\hat{\ell}))L(\hat{\ell})}{\sum_{\hat{\ell}}u(\hat{\ell})L(\hat{\ell})} + \mathcal{R}}.$$

Further, we can combine government budget constraint (18) and housing market clearing (19) to obtain housing supply function h using observed rental rates d as well as (τ, ω)

$$h(\ell) = \frac{\omega\mathbb{E}[w(y,\ell)|\ell](1-u(\ell))L(\ell)}{(1-\tau\omega)d(\ell)} \quad \forall\ell.$$

Then, given (τ, h), we obtain unemployment income schedule w^U from government budget constraint (18), which holds for each ℓ.

Last, to identify amenity schedule B, our starting point is the value of unemployment:

$$\rho V^U(\ell) = d(\ell)^{-\omega}B(\ell)w^U(\ell) + \tilde{b}(\ell) + d(\ell)^{-\omega}B(\ell)\lambda^U(\ell)\left[\int_{w^R(\ell)}^{\overline{w}}\frac{1-F_\ell(t)}{\delta(\ell)+\lambda^E(\ell)(1-F_\ell(t))}dt\right],$$

which is the same as in the baseline model, except that unemployed workers receive unemployment benefit $w^U(\ell)$ and enjoy local amenity $B(\ell)$, but suffer from unemployment stigma, captured by $\tilde{b}(\ell)$. Next, as before, reservation wage w^R is implicitly defined by a condition that equalizes the value of unemployment with the value of holding a job:

$$d(\ell)^{-\omega}B(\ell)w^R(\ell) = d(\ell)^{-\omega}B(\ell)w^U(\ell) + \tilde{b}(\ell)$$
$$+ \quad d(\ell)^{-\omega}B(\ell)(\lambda^U(\ell) - \lambda^E(\ell))\left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))}dt\right],$$

where, to satisfy Assumption 1, we now set $\tilde{b}(\ell)$ such that $w^R(\ell) = z(\underline{y}, A(\ell))$:

$$\tilde{b}(\ell) = d(\ell)^{-\omega}B(\ell)\left(z(\underline{y}, A(\ell)) - w^U(\ell) - (\lambda^U(\ell) - \lambda^E(\ell))\left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))}dt\right]\right). \quad \text{(A.16)}$$

Plug $\tilde{b}(\ell)$ back into V^U above, and use a change of variable (to re-express F_ℓ using Γ_ℓ, where $\Gamma_\ell = \Gamma(y|\mu(\ell))$) and make use of the Pareto assumption on Γ to obtain

$$\rho V^U = d(\ell)^{-\omega}B(\ell)A(\ell)\left(1 + 2\left(\lambda^E(\ell)\right)^2 \int_1^\infty y^{-\frac{1}{\mu(\ell)}} \frac{1}{\mu(\ell)} y^{-\frac{1}{\mu(\ell)}-1} \int_1^y \frac{dt}{\left[\delta(\ell) + \lambda^E(\ell)t^{-\frac{1}{\mu(\ell)}}\right]^2}dy\right),$$

which allows us to back out $B(\ell)$ for each ℓ, given the normalization $\rho V^U = 1$ and given (A, μ) (obtained above) as well as observed rental rates d and transition rates (λ^E, δ). □

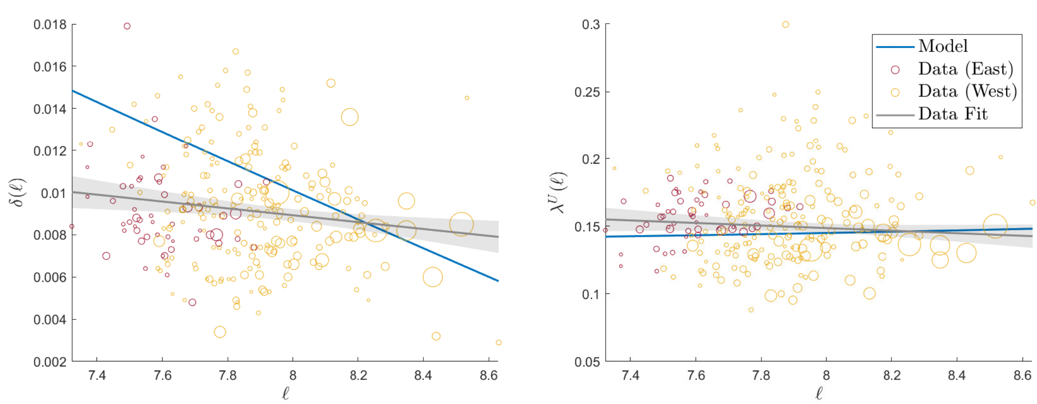
G Estimation

G.1 Estimation Results

Table A.7: Calibrated Parameters

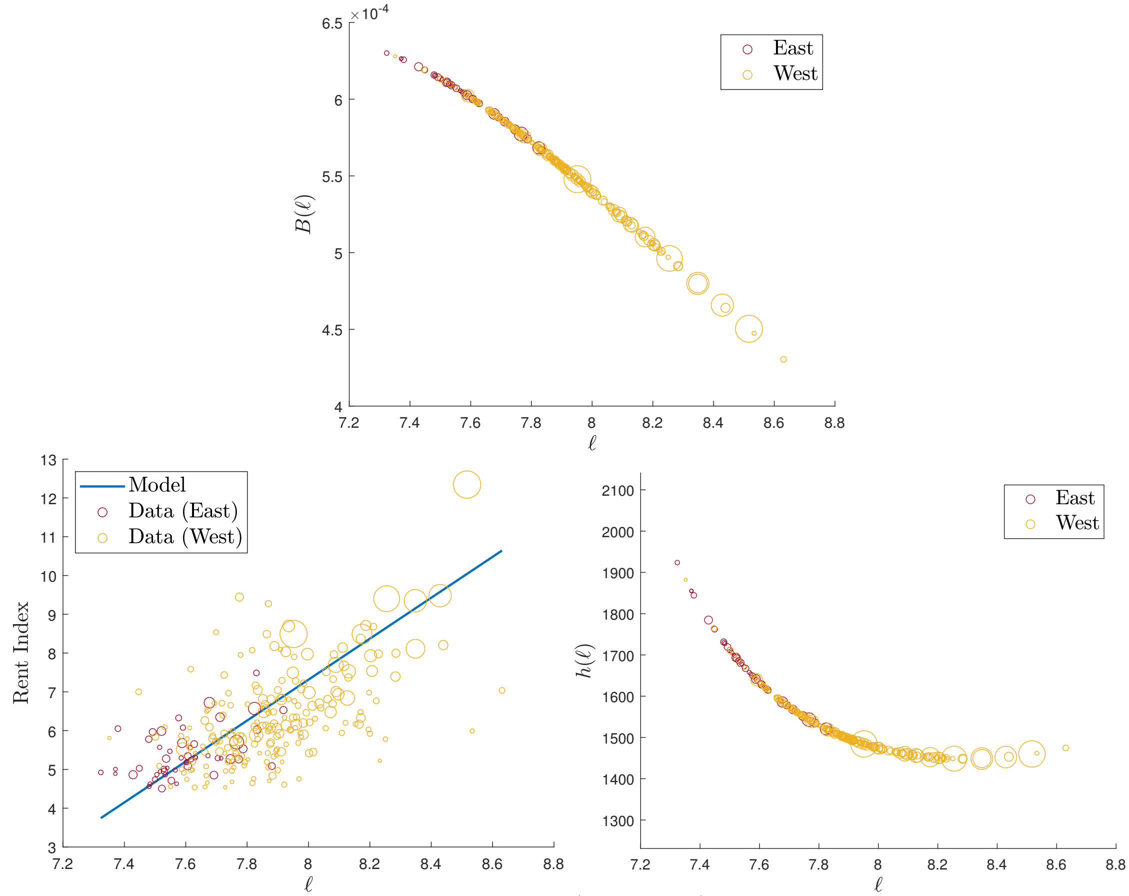
Parameter	Value	Calibration
κ	0.253	monthly UE and EE transition rate (LIAB)
$\mathcal{A}$	0.276	monthly UE transition rate (LIAB) and average firm size (GFSO)
ω	0.272	rent-to-income of main tenant households (GFSO)
τ	0.164	replacement rate of unemployed workers (SPIN)

Figure A.2: Over-Identification: δ and λ^U in Data and Model



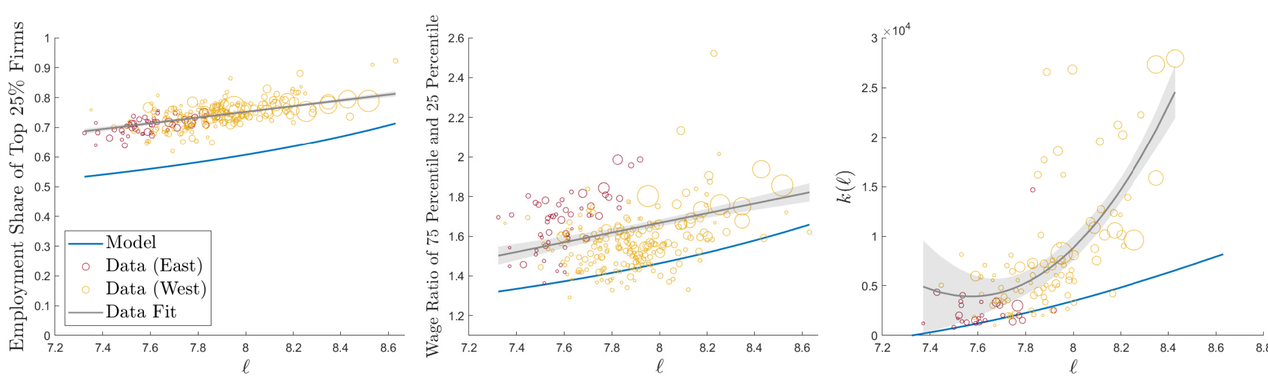
Notes: Data Sources: LIAB. For details on how these variables are constructed, see Appendix D.2. Observations are weighted by the number of establishments in each CZ, indicated by different sizes of the dots. 95% confidence intervals are displayed in gray.

Figure A.3: Additional Parameter Estimates: Location Preference Schedule (top); Housing Supply (bottom right) obtained from Residential Rents (bottom left)



Notes: Data Source: The residential rent index (bottom left) was constructed by Ahlfeldt et al. (2022), see Appendix D.3. Observations are weighted by the number of establishments in each CZ, indicated by different sizes of the dots.

Figure A.4: Model Fit: Additional Non-Targeted Moments



Notes: Data Sources: Left panel is based on BHP; middle panel is based on firm-level wages of full-time employees from BHP (and the percentiles are taken from the wage distribution that weighs observations by the number of full-time employees of each firm); right panel is based on price data for commercial properties 2012/13 from the German Real Estate Association; see Appendices D.2 and D.3 for details. Observations are weighted by the number of establishments in each CZ, indicated by different sizes of the dots. 95% confidence intervals are displayed in gray.

Table A.8: Determinants of Local TFP $A(\ell)$

	(1)	(2)	(3)	(4)	(5)
Trade Tax	0.133*				-0.141**
	(0.060)				(0.047)
Distance to Highway		-0.054***			-0.039***
		(0.012)			(0.009)
% of Employees with a College Degree			0.851***		0.674***
			(0.115)		(0.116)
Net Business Registration Intensity				0.039**	0.016
				(0.013)	(0.013)
Observations	257	257	257	257	257
Adjusted R^2	0.033	0.151	0.416	0.245	0.499

Notes: Data Source: German Federal Statistical Office and Federal Office for Building and Regional Planning. All regressions are run at the commuting-zone level and weighted by the number of establishments in each CZ. Data is averaged across years for the period 2010-2017. Dependent variable in columns (1)-(5) is 'Log Local TFP, log $A(\ell)$' obtained from our estimation for each ℓ; see Section 5.3. See Appendices D.1 and D.3 for the definition of the independent variables.

G.2 East-West Comparison: Value Added

West-East Inequality in Value Added. In Table A.9, we report the West-East gap in value added, both in the data and estimated model. Note that our model does not perfectly match the data moment because we target the *fitted line* of local value added per worker (against ℓ) instead of the exact data values.

Table A.9: West-East Inequality: Monthly Value Added per Worker (in €)

	Data	Model
Monthly Value Added per Worker, West	5237.02	5227.19
Monthly Value Added per Worker, East	4045.24	4122.54
West-East Gap	29.5%	26.8%

Notes: Data source: German Federal Statistical Office. We sum (monthly) value added across all locations in East Germany and across all locations in West Germany; and similarly for employees. We then take the ratio to obtain *Value Added per Worker* for each region. See Appendix D.1 for more details.

Counterfactuals.

Table A.10: West-East Value Added Inequality: Counterfactual Models

	Model	No Sorting	No OJS	No Spatial Frictions
West-East Gap	27%	19%	7%	2%

Notes: This table is the analogue of Table 2 but for West-East inequality in value added (not wages).

H Counterfactuals: Technical Details

H.1 The Role of Firm Sorting

We adjust $\tilde{b}(\ell)$ so that the reservation wage in each ℓ remains the same as in the baseline model, i.e., $w^R(\ell) = A(\ell)\underline{y}$, see (A.16). Moreover, we keep the estimated schedules $(A(\ell), B(\ell), h(\ell))$ from the baseline model in place. Without spatial firm sorting, F_ℓ (and thus Γ_ℓ), (λ^U, λ^E) and d all differ from the baseline model.

First, since the wage function in each ℓ is still strictly increasing in y, we have $F_\ell(w(y,\ell)) = \Gamma_\ell(y)$. But here $\Gamma_\ell(y) = \Gamma(y)$, which follows from the premise of random matching, i.e., the ex post productivity distribution is the same across locations.

Second, as unemployed workers are freely mobile across regions, we calculate $\lambda^E(\ell)$ for each ℓ to equalize the value of search while adjusting house price $d(\ell)$ such that the housing market clears in each ℓ, given the estimated $(A(\ell), B(\ell), h(\ell))$ from the baseline model:

$$\rho V^U = d(\ell)^{-\omega} B(\ell) A(\ell) \left[1 + 2(\lambda^E(\ell))^2 \int_1^\infty (1-\Gamma(y))\gamma(y) \int_1^y \frac{1}{[\delta(\ell) + \lambda^E(\ell)(1-\Gamma(t))]^2} dt dy \right]$$
$$d(\ell)h(\ell) = \frac{\omega}{1-\tau\omega} \mathbb{E}[w(y,\ell)|\ell](1-u(\ell))L(\ell)$$

where Γ is the country-wide productivity distribution of firms (no longer ℓ-specific). Note that compared to baseline, we need to determine a new value of search, ρV^U, to calculate $\lambda^E(\ell)$. We choose ρV^U to guarantee the same total population size as in the baseline economy, $\bar{L} = \int L(\ell) dR(\ell)$. In practice, we solve a fixed point in ρV^U so that it satisfies both welfare equalization of workers as well as this population constraint. Once we determine $\lambda^E(\ell)$ for each ℓ, we can compute $\lambda^U(\ell) = \lambda^E(\ell)/\kappa$.

H.2 The Role of On-the-Job Search

When reducing the intensity of OJS κ (e.g., $\kappa = 0$ as in the text), the modularity properties of $\bar{J}$ may change, so we need to re-solve for the sorting decision of firms. The population size in each location (and thus worker and firm meeting rates) depends on the firm composition in each ℓ, but at the same time impacts firms' sorting choices. We therefore need to solve for a fixed point in firm allocation μ (and thus in Γ_ℓ):

Fix κ and postulate a matching function μ. Given this μ, we first obtain Γ_ℓ, and then find meeting rate λ^U and housing price d (both as a function of $(\ell; \kappa, \Gamma_\ell, \rho V^U)$) so that—given the estimated schedules $(A(\ell), B(\ell), h(\ell))$ from the baseline model—the value of search for unemployed workers is equalized across space and local housing market clearing holds:

$$\rho V^U = d(\ell)^{-\omega} B(\ell) A(\ell) \left[1 + 2(\kappa\lambda^U(\ell))^2 \int_1^\infty (1-\Gamma_\ell(y))\gamma_\ell(y) \int_1^y \frac{1}{[\delta(\ell) + \kappa\lambda^U(\ell)(1-\Gamma_\ell(t))]^2} dtdy \right]$$

$$d(\ell)h(\ell) = \frac{\omega}{1-\tau\omega} \mathbb{E}[w(y,\ell)|\ell](1-u(\ell))L(\ell),$$

where the value of search is again calculated assuming $w^R(\ell) = A(\ell)\underline{y}$, supported by adjusting $\tilde{b}(\ell)$;[41] and where we set the new value of search, ρV^U, to achieve consistency with the total population size from the baseline economy, $\bar{L} = \int L(\ell)dR(\ell)$. Based on unemployed workers' welfare equalization, we obtain $\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)$ and therefore $\lambda^E(\ell;\kappa,\Gamma_\ell,\rho V^U) = \kappa\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)$. With $\lambda^U(\ell)$ for each ℓ in hand, we can also compute $\lambda^F(\ell;\kappa,\Gamma_\ell,\rho V^U) = \mathcal{A}^{\frac{1}{\alpha}}(\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U))^{1-\frac{1}{\alpha}}$, as well as the *match value* of a firm type p and location ℓ,

$$\bar{J}(p,\ell) + k(\ell) = \delta(\ell)\lambda^F(\ell;\kappa,\Gamma_\ell,\rho V^U)A(\ell) \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{1}{[\delta(\ell) + \kappa\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)(1-\Gamma_\ell(t))]^2} dtd\Gamma(y|p).$$

To find the optimal allocation $\hat{\Gamma}_\ell$, we maximize the sum of this value across (p,ℓ)-pairs, subject to market clearing in the land market, using a linear program. If $\Gamma_\ell = \hat{\Gamma}_\ell$ for all ℓ, we have found the equilibrium. If $\Gamma_\ell \neq \hat{\Gamma}_\ell$ for at least one ℓ, we use $\hat{\Gamma}_\ell$ as a new starting point and repeat the same steps, until convergence.

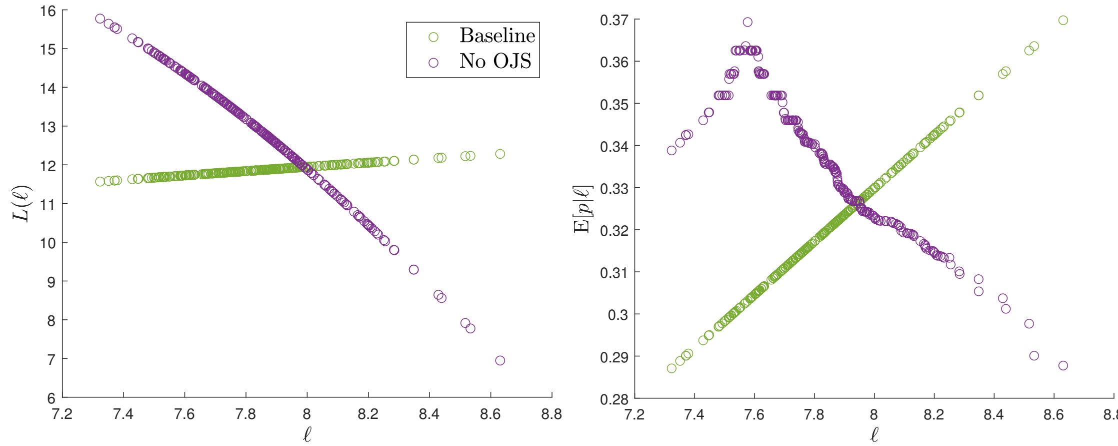
Figure A.5 below demonstrates the equilibrium properties when shutting down OJS in this counterfactual ($\kappa = 0$). In the left panel, the population gradient in the location index flips its sign from positive (green—baseline) to negative (purple—counterfactual), as there is a large outflow of workers from high-ℓ places due to their loss of economic appeal. As a result, low-ℓ locations become attractive to firms, especially to productive ones, since filling vacancies has become easier. This turns the supermodular $\bar{J}$ from the baseline model into a (mostly) submodular $\bar{J}$, inducing negative firm sorting for a wide range of locations (right panel).

H.3 The Role of Spatial Frictions

When the labor market is integrated the economy has a single job ladder and the model is similar to the basic wage posting model with firm productivity z and economy-wide productivity distribution $\tilde{\Gamma}(z) = \int \Gamma\left(\frac{z}{A(\ell)}\middle|\mu(\ell)\right)dR(\ell)$. Employed workers accept a job offer if the new wage is higher than the current one and the wage function is strictly increasing in z, so that the wage cdf is $F(w(z)) = \tilde{\Gamma}(z)$. The employment distribution becomes $\tilde{G}(z) = \delta\frac{\tilde{\Gamma}(z)}{\delta+\lambda^E\left(1-\tilde{\Gamma}(z)\right)}$; and, in terms of the local employment distribution, G_ℓ is no longer

[41] In particular, $\tilde{b}(\ell)$ is defined as in the "No-Sorting" case based on (A.16) but using F_ℓ (which we can compute based on the postulated Γ_ℓ) and (λ^U, λ^E) obtained above.

Figure A.5: Equilibrium under Active OJS (Baseline, $\kappa > 0$) and No OJS (Counterfactual, $\kappa = 0$)



given by (13) but by $G_\ell(y) = \left(\int_{\underline{y}}^{y} \frac{\tilde{g}(A(\ell)y')}{\tilde{\gamma}(A(\ell)y')}\gamma_\ell(y')dy'\right)/\left(\int_{\underline{y}}^{\overline{y}} \frac{\tilde{g}(A(\ell)y')}{\tilde{\gamma}(A(\ell)y')}\gamma_\ell(y')dy'\right)$. We keep the estimated schedules $(A(\ell), B(\ell), h(\ell))$ from the baseline model in place.

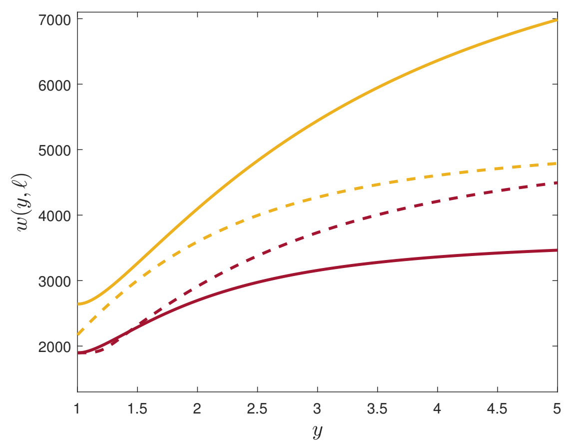
Note that $(\delta, \lambda^U, \lambda^E, \tilde{b})$ are now all constant across ℓ: There are no differences in the job-separation rate, as we compute the economy-wide δ from the average location-specific separation rates, $\delta = \int \delta(\ell)\frac{L(\ell)}{\overline{L}}dR(\ell)$ (where $\delta(\ell)$ and $L(\ell)$ are taken from the baseline model); further, all workers—irrespective of their residence location—have the same chances to find jobs, so there are economy-wide meeting rates for employed workers λ^E and unemployed workers λ^U. We determine these rates using the total population size $\overline{L}$ from the baseline model and $\bar{L} = \mathcal{A}^{\frac{1}{\alpha}} \frac{\delta+\lambda^U}{\delta+\kappa\lambda^U}\left(\lambda^U\right)^{-\frac{1}{\alpha}}$, which we derive from average firm size $\bar{l}(\ell) = (1-u)\overline{L}$ and (24). Similar to the previous counterfactual exercises, we adjust the unemployment flow benefit $\tilde{b}(\ell) = \tilde{b}$ (only that here it is independent of ℓ) so that $w^R(\ell) = w^R = A(\underline{\ell})\underline{y}$, i.e., in an integrated labor market the reservation wage is also determined economy wide, not location specific. Further, to equalize the value of search across all locations

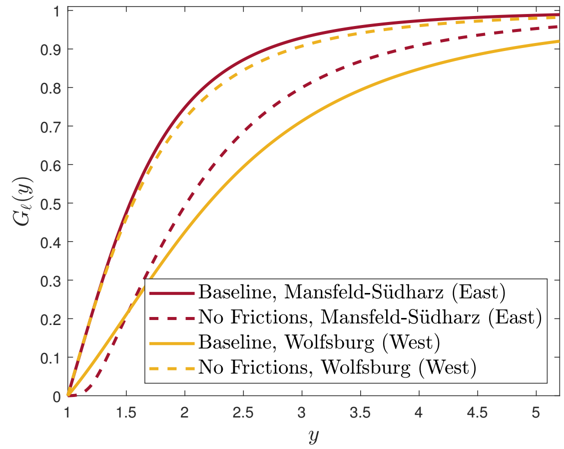
$$\rho V^U(\ell) = d(\ell)^{-\omega}B(\ell)w^U + \tilde{b} + d(\ell)^{-\omega}B(\ell)\lambda^U\left[\int_{w^R}^{\overline{w}} \frac{1-F(t)}{\delta+\lambda^E(1-F(t))}dt\right],$$

despite differences in local amenities $B(\ell)$ (as estimated from the baseline model), housing prices d need to adjust so that the 'real' value of local amenity $d(\ell)^{-\omega}B(\ell)$ is the same everywhere. As in the other counterfactuals, we choose the value of ρV^U to be consistent with the aggregate population size $\overline{L}$. Finally, to make the obtained house price schedule d consistent with local housing market clearing, population size $L(\cdot)$ adjusts so that (19) holds.

Figure A.6 displays the components of decomposition (16) for the baseline model (solid) and the counterfactual without spatial hiring frictions (dashed).

Figure A.6: No-Spatial-Frictions Counterfactual: Wages and Employment Distribution





I PAM of Firms and Locations: Robustness

I.1 Negative Relation of Local Labor Share and GDP per Capita

Table A.11: Labor Shares—Robustness

	(1)	(2)	(3)	(4)	(5)
log(GDP per capita)	-0.0625***	-0.1010***	-0.0992***	-0.0794***	-0.1147***
	(0.0087)	(0.0211)	(0.0130)	(0.0086)	(0.0236)
Share of Employment in Industry	N	Y	N	N	Y
Establishment Size	N	N	Y	N	Y
log(Population Density)	N	N	N	Y	Y
N	257	257	257	257	257

Notes: Data Source: German Federal Statistical Office. All regressions are run at the commuting-zone level (CZ) and weighted by number of establishments in each CZ. Data is averaged across years for the period 2010-2017. Column (1) is our baseline; columns (2)-(5) add controls. When controlling for local employment shares by industry ('Share of Employment in Industry'), we take the following industries into account: agriculture; mining and electricity, gas, water supply; manufacturing; construction; trade, transportation, information and communication; finance, insurance, real estate; and public and other, education, health. 'Establishment size' refers to controls that contain the share of establishments (out of the total number of establishments in a CZ) that have 0-9, 10-49, 50-249, and 250+ employees.

I.2 Negative Relation of Labor Share and Firm Productivity: General Case

We investigate the relationship between local labor share and local firm productivity beyond the specific functional forms assumed in our quantitative setting (Section 5.1).

We are especially interested in two properties. First, given the empirical observation that local labor shares are decreasing in ℓ, can we infer positive spatial sorting by firms (i.e., is PAM necessary for $\partial\,\mathrm{LS}(\ell)/\partial\ell < 0$)? Second, we are interested in the intuition for why PAM is a force behind the decreasing local labor share function (i.e., we want to unpack the sufficient conditions for $\partial\,\mathrm{LS}(\ell)/\partial\ell < 0$).

Denote the firm-level labor share by $Ls(y,\ell) := w(y,\ell)/z(y,A(\ell))$ and let the value-added-weighted employment density be given by $\tilde{g}_\ell(y) := \frac{z(y,A(\ell))g_\ell(y)}{\int_{\underline{y}}^{\overline{y}} z(y',A(\ell))g_\ell(y')dy'}$, with corresponding cdf $\tilde{G}_\ell(y)$. The local labor share in each ℓ is then

$$\begin{aligned}\mathrm{LS}(\ell) &= \frac{\int_{\underline{y}}^{\overline{y}} w(y,\ell)g_\ell(y)dy}{\int_{\underline{y}}^{\overline{y}} z(y,A(\ell))g_\ell(y)dy} \\ &= \int_{\underline{y}}^{\overline{y}} \frac{w(y,\ell)}{z(y,A(\ell))}\frac{z(y,A(\ell))g_\ell(y)}{\int_{\underline{y}}^{\overline{y}} z\left(y',A(\ell)\right)g\left(y',\ell\right)dy'}dy \\ &= \int_{\underline{y}}^{\overline{y}} Ls(y,\ell)\tilde{g}_\ell(y)dy.\end{aligned}$$

Thus, for the local labor share to be decreasing in ℓ, the following must hold:

$$\frac{\partial LS(\ell)}{\partial \ell} = \frac{\partial Ls(\underline{y},\ell)}{\partial \ell} + \int_{\underline{y}}^{\overline{y}} \frac{\partial \frac{\partial Ls(y,\ell)}{\partial y}}{\partial \ell}(1-\tilde{G}_\ell(y)) + \frac{\partial Ls(y,\ell)}{\partial y}\left(-\frac{\partial \tilde{G}_\ell(y)}{\partial \ell}\right)dy < 0, \qquad \text{(A.17)}$$

where the first term is independent of ℓ.

Necessary Conditions for $\partial LS(\ell)/\partial\ell < 0$. If $\partial LS(\ell)/\partial \ell < 0$ *for all decreasing* firm-level labor shares $Ls(\cdot,\ell)$ (i.e., also for those that are supermodular, $\partial^2 Ls(y,\ell)/\partial y\partial\ell \geq 0$), then it must be that $\partial\tilde{G}_\ell/\partial\ell \leq 0$ on some set of y of positive measure. This, in turn, is a strong indication of positive sorting of firms across space, $\mu' > 0$.[42]

Sufficient Conditions for $\partial LS(\ell)/\partial\ell < 0$. We obtain $\partial LS(\ell)/\partial\ell < 0$ if the integrand of (A.17) is negative, i.e., if the firm-level labor share is (i) submodular, $\partial^2 Ls(y,\ell)/\partial y\partial\ell < 0$; and (ii) decreasing in y, $\partial Ls(y,\ell)/\partial y < 0$, which renders the second term negative if positive firm sorting across space ensures that $\partial\tilde{G}_\ell/\partial\ell \leq 0$. Whereas condition (i) is difficult to guarantee in general, we investigate (ii) more closely. First, note that $\partial\tilde{G}_\ell/\partial\ell \leq 0$ if there is

[42] To see this, first note that for any $\ell' < \ell''$ and $y < \overline{y}$, $\tilde{G}_{\ell''} \leq \tilde{G}_{\ell'}$ if

$$\begin{aligned}&\frac{\int_{\underline{y}}^{y} g_{\ell''}(\tilde{y})z(\tilde{y},A(\ell''))d\tilde{y}}{\int_{\underline{y}}^{\overline{y}} g_{\ell''}(\tilde{y})z(\tilde{y},A(\ell''))d\tilde{y}} \leq \frac{\int_{\underline{y}}^{y} g_{\ell'}(\tilde{y})z(\tilde{y},A(\ell'))d\tilde{y}}{\int_{\underline{y}}^{\overline{y}} g_{\ell'}(\tilde{y})z(\tilde{y},A(\ell'))d\tilde{y}} \\ \Leftrightarrow \quad &\int_{\underline{y}}^{y} g_{\ell''}(\tilde{y})z(\tilde{y},A(\ell''))d\tilde{y}\int_{\underline{y}}^{\overline{y}} g_{\ell'}(\tilde{y})z(\tilde{y},A(\ell'))d\tilde{y} \leq \int_{\underline{y}}^{y} g_{\ell'}(\tilde{y})z(\tilde{y},A(\ell'))d\tilde{y}\int_{\underline{y}}^{\overline{y}} g_{\ell''}(\tilde{y})z(\tilde{y},A(\ell''))d\tilde{y},\end{aligned}$$

i.e., if $\int_{\underline{y}}^{y} g_\ell z(y,A(\ell))$ is log-supermodular in (y,ℓ), which is guaranteed if both g_ℓ and z are log-supermodular, with the latter being satisfied, for instance, under the multiplicative z. Cross-differentiating $\log(g_\ell)$ shows that g_ℓ is log-supermodular if

$$\min_{\ell,y} \frac{\partial^2\gamma_\ell}{\partial y\partial\ell}\gamma_\ell - \frac{\partial\gamma_\ell}{\partial\ell}\frac{\partial\gamma_\ell}{\partial y} > C \quad \Leftrightarrow \quad \min_{\ell,p,y}\left(\frac{\partial^2\gamma}{\partial p\partial y}\gamma - \frac{\partial\gamma}{\partial y}\frac{\partial\gamma}{\partial p}\right)\mu' > C$$

where C is a positive and bounded constant. This condition is satisfied if there is positive sorting of firms across space, $\mu' > 0$, and if γ is sufficiently log-supermodular in (p,y).

positive sorting of firms to locations, $\mu' > 0$, and if γ is sufficiently log-supermodular in (p, y) (see Footnote 42), which is similar but slightly stronger than our assumption in the baseline model that p shifts $\Gamma(\cdot|p)$ in the FOSD sense. Second, the feature that the firm-level labor share is decreasing in y requires additional assumptions on the density γ_ℓ, which is what we will turn to next.

We provide sufficient conditions, under which the firm-level labor share,

$$Ls(y,\ell) = \frac{w(y,\ell)}{z(y,A(\ell))} = 1 - \frac{\left[\delta(\ell) + \lambda^E(\ell)(1-\Gamma_\ell(y))\right]^2 \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{\left[\delta(\ell)+\lambda^E(\ell)(1-\Gamma_\ell(t))\right]^2} dt}{z(y,A(\ell))},$$

is decreasing in firm productivity y for all locations ℓ. Differentiation and some algebra yield:

$$\frac{\partial Ls(y,\ell)}{\partial y} = (1 - Ls(y,\ell)) \frac{2\lambda^E(\ell)\gamma_\ell(y)}{\delta(\ell) + \lambda^E(\ell)(1-\Gamma_\ell(y))} - Ls(y,\ell) \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))}. \tag{A.18}$$

We want to show that under the following assumptions this expression is negative and thus the firm-level labor share is decreasing in y for all ℓ:

1. $\gamma_\ell(\cdot)$ is decreasing such that $\forall y$:

$$\max_y \frac{\partial \gamma_\ell(y)}{\partial y} < \min_y \left(-\frac{\lambda^E(\ell)}{\delta(\ell) + \lambda^E(\ell)(1-\Gamma_\ell(y))} \gamma_\ell(y)^2 \right);$$

2. $z(\cdot, A(\ell))$ is weakly log-convex in y.

We proceed by contradiction. Suppose that $\frac{\partial Ls(y,\ell)}{\partial y} > 0$ for some $y \in [\underline{y}, \overline{y}]$, that is, (A.18) is strictly positive at y.

First, note that, at $y = \underline{y}$, (A.18) is negative, since the first term is zero while the second one is strictly positive,

$$\underbrace{\underbrace{(1 - Ls(\underline{y},\ell))}_{=0} \frac{2\lambda^E(\ell)\gamma_\ell(\underline{y})}{\delta(\ell) + \lambda^E(\ell)(1-\Gamma_\ell(\underline{y}))}}_{:=LHS(\underline{y})} < \underbrace{Ls(\underline{y},\ell) \frac{\frac{\partial z(\underline{y},A(\ell))}{\partial y}}{z(\underline{y},A(\ell))}}_{:=RHS(\underline{y})}.$$

Second, in order for $\frac{\partial Ls(y,\ell)}{\partial y} > 0$ for some $y \in (\underline{y}, \overline{y})$ on a set $[y-\epsilon, y+\epsilon], \epsilon > 0$, it must be that $RHS(y)$ is *smaller* than $LHS(y)$, i.e., $RHS(\cdot)$ must cross $LHS(\cdot)$ at least once and the first crossing at point $\hat{y} \in (\underline{y}, \overline{y})$ is such that $RHS(\hat{y})$ crosses $LHS(\hat{y})$ *from above*. That is, there exists $\hat{y} \in (\underline{y}, \overline{y})$ such that the slope of $RHS(\hat{y})$ is smaller than the slope of $LHS(\hat{y})$. We now investigate these slopes.

The slope of $LHS(y)$ is given by:

$$-\frac{\partial Ls(y,\ell)}{\partial y}\frac{2\lambda^E(\ell)\gamma_\ell(y)}{\delta(\ell)+\lambda^E(\ell)(1-\Gamma_\ell(y))}+(1-Ls(y,\ell))\frac{2\lambda^E(\ell)\left(\frac{\partial\gamma_\ell(y)}{\partial y}(\delta(\ell)+\lambda^E(\ell)(1-\Gamma_\ell(y)))+\lambda^E\gamma_\ell(y)^2\right)}{(\delta(\ell)+\lambda^E(\ell)(1-\Gamma_\ell(y)))^2},$$

whose second term is *negative* under the 1. assumption above.

Next, the slope of $RHS(y)$ is given by:

$$\frac{\partial Ls(y,\ell)}{\partial y}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))}+Ls(y,\ell)\frac{\partial\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))}}{\partial y},$$

where the second term is *positive* under the 2. assumption from above.

Consider the first crossing of $RHS(\cdot)$ and $LHS(\cdot)$, i.e., the point $\hat{y}\in(\underline{y},\overline{y})$ at which $RHS(\cdot)$ crosses $LHS(\cdot)$ *from above*. First note that at crossing $\hat{y}$, $\partial Ls(\hat{y},\ell)/\partial y=0$ by construction. Moreover, by the 1. assumption, $LHS(\hat{y})$ has a *negative* slope, while by the 2. assumption $RHS(\hat{y})$ has a *positive* slope. Therefore, the slope of $RHS(\cdot)$ is *larger* than the slope of $LHS(\cdot)$ at $\hat{y}$, and so $RHS(\cdot)$ cannot cross $LHS(\cdot)$ from above at $\hat{y}$—a contradiction.

As a result, under the specified assumptions (particularly that density γ_ℓ is sufficiently decreasing), the firm-level labor share, $Ls(\cdot,\ell)$, is decreasing in firm productivity y, for each ℓ.

I.3 Negative Relation of Labor Share and Firm Productivity: Log-Normal

We now demonstrate that the property of the negative relationship between local labor share and local firm productivity from our quantitative model does not hinge on the Pareto assumption on $\Gamma(y|p)$. To do so, we provide simulations for a common alternative functional form of firm productivity. Specifically, we assume that $\Gamma(y|p)$ follows a log-normal distribution, whereby $\log(y)$ follows a normal distribution with parameters $(p,0.5^2)$ and, in equilibrium, $(\mu(\ell),0.5^2)$. That is, firms with higher ex ante type (higher μ) have higher mean productivity ex post. In line with the discussion above, we left-truncate this productivity distribution such that its density is sufficiently decreasing for all μ (we truncate all productivity distributions at the median productivity of the worst location in order to maintain a common $\underline{y}$). Figure A.7 shows that, as in the Pareto case, there is a negative relationship between local labor shares and firm ex ante productivity under the log-normal assumption. As a result, we back out an *increasing* matching function $\mu(\cdot)$ (indicating PAM between firms and locations) from the decreasing local labor share function $LS(\cdot)$ observed in the data.

To gain intuition for the negative relationship between local labor share and local firm productivity, Figure A.8 plots for two locations—the top and bottom CZ—firm-level labor shares $Ls(\cdot, \ell)$ (left panel) and the cdf of (weighted) employment (right panel), informing the two terms $\frac{\partial Ls(y,\ell)}{\partial y}\left(-\frac{\tilde{G}_\ell(y)}{\partial \ell}\right)$ in the integrand of (A.17): In each location, firm-level labor shares decrease in productivity y; moreover, the weighted employment distribution of the top location is stochastically better (under PAM). Thus, a decreasing empirical local labor share function $LS(\cdot)$ calls—under decreasing firm-level labor shares $Ls(\cdot, \ell)$—for PAM so that in high ℓ more employment is concentrated in top firms with low labor shares.

Figure A.7: Identifying Firm Sorting from Local Labor Shares: Truncated Log-Normal $\Gamma(y|p)$

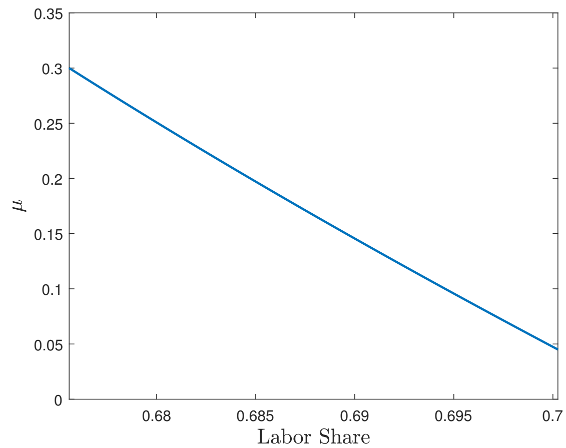
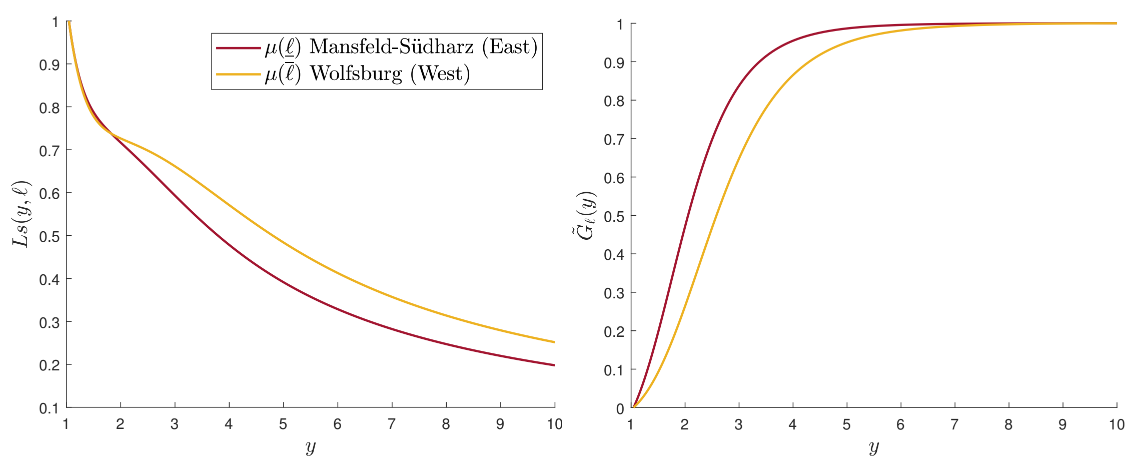


Figure A.8: Firm-Level Labor Shares and Employment Distribution: Truncated Log-Normal $\Gamma(y|p)$



I.4 No Relation of Labor Share and Firm Productivity: Neutral Productivity Shifts

In contrast to the discussed productivity shifts across locations that stem from changes in the tail parameter of the Pareto distribution or in the mean of the log-normal distribution, there are other distributional shifts—which we call *neutral* shifts—that would not translate into labor share differences across locations. These are shifts under which firm productivity in one location is simply the scaled firm productivity of another location. One example of neutral productivity shifts is a scenario in which regions differ in the scale (and not the tail) parameter of the Pareto distribution. We now formalize this discussion and show that neutral productivity shifts impact neither local labor shares nor EE wage growth.

Suppose $y \sim \Gamma_\ell, y' \sim \Gamma_{\ell'}$ and $\frac{y'}{a} \sim \Gamma_\ell$ for some constant $a > 0$, that is, productivity in location ℓ' is a scaled version of productivity in ℓ, $y' = ay$. Then, $\Gamma_{\ell'}(y') = \mathbb{P}\big[Y' \leq y'\big] = \mathbb{P}\big[\frac{Y'}{a} \leq \frac{y'}{a}\big] = \Gamma_\ell(\frac{y'}{a})$ and $\gamma_{\ell'}(y') = \frac{1}{a}\gamma_\ell(\frac{y'}{a})$. We also assume $z(y, A(\ell)) = A(\ell)y$.

Local Labor Shares. We can express the wage in location ℓ' as:

$$
\begin{aligned}
w(y',\ell') &= A(\ell')y' - (1+\varphi^E(1-\Gamma_{\ell'}(y')))^2 \int_{\underline{y}'}^{y'} \frac{A(\ell')}{[1+\varphi^E(1-\Gamma_{\ell'}(t'))]^2} dt' \\
&= A(\ell')y' - (1+\varphi^E(1-\Gamma_{\ell}(y'/a)))^2 \int_{\underline{y}'}^{y'} \frac{A(\ell')}{[1+\varphi^E(1-\Gamma_{\ell}(t'/a))]^2} dt' \\
&= aA(\ell')y'/a - a(1+\varphi^E(1-\Gamma_{\ell}(y'/a)))^2 \int_{\underline{y}}^{y'/a} \frac{A(\ell')}{[1+\varphi^E(1-\Gamma_{\ell}(t))]^2} dt \quad \text{(change of var. } t' = at) \\
&= aw(y'/a,\ell)\frac{A(\ell')}{A(\ell)}.
\end{aligned}
$$

Recall that $l(y,\ell)$ is a function of $\Gamma_\ell(y)$ and from our observations, $l(y'/a,\ell) = l(y',\ell')$. Then:

$$
\begin{aligned}
\text{LS}(\ell') = \frac{\int_{\underline{y}'}^{\overline{y}'} w(y',\ell')l(y',\ell')\gamma_{\ell'}(y')dy'}{\int_{\underline{y}'}^{\overline{y}'} z(y',A(\ell'))l(y',\ell')\gamma_{\ell'}(y')dy'} &= \frac{\int_{\underline{y}}^{\overline{y}} w(ay,\ell')l(ay,\ell')\gamma_{\ell'}(ay)ady}{\int_{\underline{y}}^{\overline{y}} z(ay,A(\ell'))l(ay,\ell')\gamma_{\ell'}(ay)ady} \quad \text{(change of var. } y' = ay) \\
&= \frac{\int_{\underline{y}}^{\overline{y}} w(ay,\ell')l(y,\ell)\gamma_{\ell}(y)dy}{\int_{\underline{y}}^{\overline{y}} z(ay,A(\ell'))l(y,\ell)\gamma_{\ell}(y)dy} \\
&= \frac{\int_{\underline{y}}^{\overline{y}} aw(y,\ell)\frac{A(\ell')}{A(\ell)}l(y,\ell)\gamma_{\ell}(y)dy}{\int_{\underline{y}}^{\overline{y}} z(y,A(\ell))a\frac{A(\ell')}{A(\ell)}l(y,\ell)\gamma_{\ell}(y)dy} = \text{LS}(\ell).
\end{aligned}
$$

Local EE Wage Growth: Let $y(\cdot)$ be a quantile function s.t. $\Gamma_\ell(y(s)) = s$. Then, $s = \Gamma_{\ell'}(y'(s)) = \Gamma_\ell\left(\frac{y'(s)}{a}\right)$, i.e. $\frac{y'(s)}{a} = y(s)$. Consider two quantiles $s_1 < s_2$. Then:

$$
\frac{w(y'(s_2),\ell')}{w(y'(s_1),\ell')} = \frac{w(y'(s_2)/a,\ell)}{w(y'(s_1)/a,\ell)} = \frac{w(y(s_2),\ell)}{w(y(s_1),\ell)}.
$$

Thus, the wage ratio of two firms with ranks s_2 and s_1 is the same in ℓ and ℓ', which implies that EE wage growth is the same across locations: A job-to-job transition from a firm at productivity quantile s_1 to a firm at productivity quantile s_2 leads to the same wage gain in *all* locations.