Firm Sorting and Spatial Inequality*

Ilse Lindenlaub
Yale University and NBER

Ryungha Oh
Yale University

Michael Peters
Yale University and NBER

April 2023

Abstract

We study the importance of spatial firm sorting for inequality both between and within local labor markets. 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 provide conditions under which there is positive firm sorting, with more productive firms settling in more productive locations. We show that positive firm sorting increases both the mean and the dispersion of wages in productive markets relative to less productive ones. The main mechanism is that firm sorting steepens the job ladder in productive places. We estimate our model using administrative data from Germany and identify firm sorting from a novel fact: Average local labor shares are lower in productive locations, which indicates a higher concentration of top firms with strong monopsony power. We infer that there is positive sorting of firms across space. Quantitatively, firm sorting can account for at least 16% of the spatial variation in mean wages and at least 38% of the variation in within-location wage dispersion.

*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, Boston University, the NBER Summer Institute 2022, SED 2022, EEA 2022 and the NYU-Columbia Trade Day 2023 for useful comments. We are grateful to Hector Chade for comments on the paper and 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 from 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. 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: Wages in urban areas are almost 20% higher, and the West still has 28% higher wages than the East. In addition, richer local labor markets are characterized by more wage dispersion. The phenomenon of a spatial divide has often been attributed to the spatial sorting of more productive workers to cities or to the productivity advantages of areas with an increasingly larger population.

In this paper, we argue that firm sorting across locations plays an important role in explaining these patterns of spatial inequality. We develop a theory of how heterogeneous firms sort across space: Firms first decide where to produce and then compete for workers in a frictional local labor market that is characterized by job search by both unemployed and employed workers. We show that spatial firm sorting impacts the local distribution of wages and hence inequality. If more productive firms sort into more productive locations, the local job ladder in these locations steepens. This amplifies spatial wage inequality across and within locations. We use our framework to shed new light on spatial wage disparities in Germany. Our quantitative findings indicate that the poorest local labor markets in Germany have significantly lower wages and less wage dispersion because they lack highly productive firms. Firm sorting can account for 16%-34% of the spatial variation in mean wages and for 38%-66% of the variation in within-location wage dispersion.

A key feature of our model is that the sorting of heterogeneous firms to heterogeneous locations is determined in equilibrium. Once firms settle in a local labor market, they hire workers subject to search frictions and produce output that depends on both firm and location productivity. The second crucial feature of our model is that in each local labor market, not only unemployed but also employed workers search for jobs; they all aim to climb their local job ladder toward increasingly higher wages. Importantly, the shape of the local job ladder is endogenous and determined by the firms that settle there.

In this setting, 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, these 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, makes it difficult to poach and retain workers and thereby curbs firm size. Hence, firms' location decisions balance two considerations: local productivity and local competitiveness. To our knowledge, this is the first model that integrates on-the-job search with firms' location choices and highlights this novel trade-off. This may be surprising in light of evidence that firms routinely hire from other firms within their local labor market. For example, in Germany, around two-thirds of hires come from the firm's commuting zone, and a substantial share of firms' hires (around 50%) are poached from other firms.

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 also show that under the conditions for monotone sorting, an equilibrium exists and is unique.

Our theory makes precise predictions why firm sorting affects spatial inequality. First, positive firm sorting intensifies labor market *competition* for workers in productive locations, which steepens their wage (job) ladders. Second, it leads to a stochastically better employment *composition* whereby workers in productive locations are frequently hired by productive firms that pay higher wages. Both factors amplify the spatial wage premium and wage dispersion across firms in localities that are more productive to start with.

These predictions are in line with evidence from Germany. Using administrative data, we show that the top quartile of local labor markets in terms of their GDP per capita have 37% higher average wages and 18% higher wage dispersion compared with the bottom quartile. Importantly, we provide direct evidence for the main mechanism of our model that generates these features: Wage growth from an employment-to-employment (EE) transition is more than twice as large in the richest compared with the poorest locations, which indicates that job ladders are indeed steeper in prosperous places. Moreover, spatial heterogeneity in job ladders matters for spatial inequality: A statistical decomposition of spatial differences in lifetime earnings suggests that around 25% of the spatial earnings gap is due to these heterogeneous job ladders, which underscores the role of on-the-job search in understanding spatial disparities.

Our model generates these heterogeneous job ladders through spatial firm sorting. To provide more direct evidence for this, we exploit the implication of our theory whereby firm sorting affects local labor shares. This is because a concentration of highly productive firms—who have a high degree of monopsony power and low labor shares—leads to a low average labor share in their location. We document the novel fact that the local labor share is *decreasing* in local GDP per capita, which implies that there is *positive sorting* between firms and locations.

To quantitatively assess the importance of spatial firm sorting for spatial inequality, we estimate our model using administrative data from Germany. A central 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 spatial variation in local labor shares identifies the extent of firm sorting. Our finding that local labor shares are *lower* in locations with high local GDP per capita calls for *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 pins down the fundamental productivity of each location.

We quantify the equilibrium impact of firm sorting on spatial inequality in a counterfactual that matches firms *randomly* to locations. Firm sorting can account for at least 16% of the wage gap between the poorest and richest locations. Hence, poor locations are not only disadvantaged because of inferior economic fundamentals, but this weakness is amplified by the fact that low-productivity firms tend to cluster there. Moreover, firm sorting affects spatial differences in local wage inequality, accounting for at least 38% of the spatial gap in within-location wage dispersion. Firm sorting therefore plays an important role in spatial inequality in Germany.

Related Literature. Our project merges two strands of the literature that so far 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 ($\sim$10%-40%).[1] Despite this evidence on the importance of firms and search for cross-sectional inequality, 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 is the first to conduct such an analysis both theoretically and quantitatively.

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); Spain (De La Roca and Puga, 2017); France (Combes et al., 2008); and Germany (Bamford, 2021; Dauth et al., 2022). In turn, Schmutz and Sidibé (2018) and Heise and Porzio (2022) analyze spatial wage gaps using a job ladder model and focus on *worker* mobility, mobility costs, and preference frictions as the main sources of inequality. In contrast to our work, what these papers do not account for is how the *spatial sorting of firms* drives spatial inequality by endogenously shaping local job ladders.[2]

A smaller but growing literature analyzes firms' location choices but differs from ours in focus and modeling choices. Combes et al. (2012) use firm-level data to disentangle firm selection from agglomeration economies in the productivity advantage of cities. Behrens et al. (2014) combine worker sorting, firm selection, and agglomeration economies to rationalize the urban wage premium. Both papers feature frictionless labor markets, as does Gaubert (2018), who 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. A single wage prevails in any local labor market, and so there is no relationship between firm sorting, spatial differences in EE returns, and within/across-location inequality. Bilal (2022) is the first paper that analyzes the effect of firms' location choices in a model that features labor market frictions. Given his focus on spatial unemployment differences, he abstracts from on-the-job search and thus from the spatial heterogeneity of job ladders and EE returns; we argue that these are key determinants of spatial inequality. Our work complements these studies by drawing out the implications of firm sorting for wage inequality.

[1] Using a structural model, Postel-Vinay and Robin (2002) find a contribution of firm heterogeneity to wage dispersion of around 30% and a contribution of search frictions of around 40% in France. Bagger and Lentz (2019) find a contribution of firms of 18% and a contribution of search of 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 (e.g., Card et al. (2013) for Germany).

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

2 The Model

2.1 Environment

Time is continuous, the horizon infinite 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 $A(\ell)$ 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 utility $b(\ell)$ 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 $p \in [\underline{p}, \bar{p}]$, distributed with a continuously differentiable cdf Q, with 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 $\Gamma(\cdot|p)$, where Γ is continuously differentiable in both y and p. We assume that the corresponding density, $\gamma(\cdot|p)$, satisfies the strict monotone likelihood ratio property in (y, p). This implies that $\partial\Gamma(y|p)/\partial p < 0$ for all $y \in (\underline{y}, \overline{y})$, i.e., 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 who yields a positive profit. Firm y in location ℓ produces output $z(y, A(\ell))$ per worker hired. We assume that z is twice continuously differentiable 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, which allows them to vary across space.

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.

To simplify our analytical arguments, we impose the following assumption.

Assumption 1.

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

An implication of the common support assumption from part 1, which is supported by the evidence[3], is that locations inhabited by firms with higher ex ante productivity p have an ex post productivity distribution $\Gamma(y|p)$ that puts more mass on highly productive firms. In turn, part 2 will be guaranteed by making sure that the output of firm $\underline{y}$ 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. In our quantitative analysis, we show that our results are robust to relaxing Assumption 1.

2.2 Equilibrium

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

Workers. Workers face a single decision: whether to accept a job offer, both when employed and unemployed. We discuss this job acceptance decision briefly since it is standard.

Consider first a worker who is employed at wage w. The value of being employed at wage

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

w in location ℓ, $V^E(w,\ell)$, solves the recursive equation

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

where F_ℓ is the endogenous wage-offer distribution in location ℓ with support $[\underline{w}(\ell),\overline{w}(\ell)]$ and $V^U(\ell)$ denotes the value of unemployment, given by

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

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, which pins down $w^R(\ell)$ for each ℓ from a worker who is indifferent between accepting and rejecting a job,

$$V^E(w^R(\ell),\ell) = V^U(\ell). \tag{2}$$

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 backward.

Wage Posting. When posting a wage, a firm in location ℓ trades off profit per worker against firm size, which is given by (see Appendix A.2)[4]

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

Firms that are higher ranked in local wage distribution F_ℓ are larger—since they poach more and are being poached less—than 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 in a FOSD sense, since the firm faces fiercer competition. Importantly, the (relative) EE rate, φ^E, governs the extent to which firm size depends on local competition. If labor market frictions are severe and EE flows are rare, $\varphi^E \to 0$, then the competition channel is mitigated and, in the limit, firm size is independent of the local wage distribution.

Each firm posts the wage that maximizes its net present discounted value of profits, which

[4] Firm size can be derived as the hiring rate times the expected match duration or, as in Burdett and Mortensen (1998), as the measure of workers employed at y divided by the measure of firms with y.

can be expressed as per-worker profit times size (see Appendix A.1),[5]

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

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

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 strictly 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),

$$w(y,\ell) = z(y,A(\ell)) - \int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y} \frac{l(t,\ell)}{l(y,\ell)} dt, \tag{5}$$

except that here there is one such wage function in each location ℓ. Firms have monopsony power due to search frictions. This creates a wedge between the firm's wage $w(y,\ell)$ and its marginal product $z(y,\ell)$, which depends on the local labor market competition "from below": If the competitive pressure surrounding firm $t < y$ relative to firm y—captured by a relatively small size of t, $l(t,\ell)$, relative to that of firm y, $l(y,\ell)$—is strong, it is difficult for firm y to poach workers which bids up its wage. The extent of local labor market competition is determined by the local productivity distribution, Γ_ℓ, which affects firm y's size relative to its lower ranked competitors (see (3)) and thereby wages. Pinning down Γ_ℓ is what 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. Each firm p considers the expected value from settling in location ℓ, which is the expected present discounted value of profits net of the price of land, $k(\ell)$:

$$\overline{J}(p,\ell) = \int \tilde{J}(y,\ell) d\Gamma(y|p) - k(\ell) = \int_{\underline{y}}^{\overline{y}} \frac{\partial z(y,A(\ell))}{\partial y} l(y,\ell)\,(1-\Gamma(y|p))dy - k(\ell),$$

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

which we derived using $\tilde{J}(y,\ell)$ from (4) and wage function (5) (and integration by parts). 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

$$\max_{\ell} \overline{J}(p,\ell). \tag{6}$$

The solution to (6) describes firms' location decisions and is at the center of our analysis. The FOC of this problem highlights firms' fundamental location choice trade-off:[6]

$$\int_{\underline{y}}^{\overline{y}} \left(\frac{\partial \ln\left(\frac{\partial z(y,A(\ell))}{\partial y}\right)}{\partial \ell} + \frac{\partial \ln l(y,\ell)}{\partial \ell} \right) \frac{\partial z(y,A(\ell))}{\partial y} l(y,\ell)\left(1-\Gamma(y|p)\right) dy = \frac{\partial k(\ell)}{\partial \ell}, \tag{7}$$

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 (7) 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 z and thus firm profits per employee (first term in brackets on the LHS). 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 l (second term in brackets). 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, they command higher land prices, $\partial k(\ell)/\partial \ell > 0$.

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,

$$\Gamma_\ell(y) = \int_{\underline{p}}^{\overline{p}} \Gamma(y|p) m_p(p|\ell) dp \qquad \forall y \in [\underline{y},\overline{y}], \tag{8}$$

where we define by $m(\ell,p)$ the endogenous joint matching density between (ℓ,p) with conditional density $m_p(p|\ell)$ (and also $m_\ell(\ell|p)$).[7] In addition, the FOC pins down the land price schedule, $k(\cdot)$, that sustains this allocation. This is obtained by solving (7) for k, when evaluated at the equilibrium assignment (see Appendix A.4). The land price in location ℓ is given by the cumulative marginal contributions of land to the match surplus (between firms and land) in locations that are weakly less productive than ℓ.[8]

[6] We proceed as if Γ_ℓ (and thus $l(y,\ell)$) is continuously differentiable in ℓ, which will be the case under pure sorting below.

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

[8] In this competitive land market, firms that maximize expected profits and landowners who maximize land prices will

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

$$R(\ell) = \int_{\underline{\ell}}^{\ell} \int_{\underline{p}}^{\overline{p}} m_\ell(\tilde{\ell}|\tilde{p}) q(\tilde{p}) d\tilde{p} d\tilde{\ell}, \tag{9}$$

which ensures that the mass of land with quality below ℓ equals the mass of firms settling in those locations. Thus, 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,

$$\int_{\underline{y}}^{\overline{y}} z(y, A(\ell)) l(y,\ell) d\Gamma_\ell(y) = \int_{\underline{y}}^{\overline{y}} w(y,\ell) l(y,\ell) d\Gamma_\ell(y) + \bar{J}(\mu(\ell),\ell) + k(\ell). \tag{10}$$

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):

$$\delta(1 - u(\ell)) = u(\ell)\lambda^U \quad \Rightarrow \quad u(\ell) = \frac{1}{1 + \varphi^U}. \tag{11}$$

Second, the inflow into and outflow out of employment in firms with productivity below y must balance (for all y), taking into account the optimal job acceptance decisions of employed workers. This determines the cdf of employment in location ℓ, denoted by G_ℓ:

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

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 $\lambda^E(1 - \Gamma_\ell(y))$). Local employment distribution G_ℓ reflects the local firm productivity distribution, Γ_ℓ, but is stochastically better as long as there is an active job ladder, $\varphi^E > 0$.

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

result in the same allocation of firms to locations, which is why we detail only one side's decision: the one by firms.

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* (11) *and* (12) *hold, pinning down* $u(\ell)$ *and* G_ℓ*;*
5. *Good market clearing* (10) *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.[9]

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,* $\mu'(\ell) > 0$.

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 (8) to

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

so that high-ℓ locations have ex post productivity distributions that are stochastically better.[10]

[9] For completeness, we analyze the allocation with negative assortative matching (NAM) in Online Appendix OA-1.1.

[10] 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 (8) becomes

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

To obtain *sufficient* conditions for positive sorting, recall that firm p chooses location ℓ to maximize $\overline{J}(p,\ell)$ given in (6). 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 the strict supermodularity of $\overline{J}(p,\ell)$ in (p,ℓ) is sufficient for the strict single-crossing property. Thus, complementarities of $\overline{J}(p,\ell)$ in (p,ℓ) lead to positive sorting, which echoes familiar insights from the literature on 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 (7)) 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:[11]

$$\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell} > 0 \qquad \text{if} \qquad \underbrace{\frac{\partial \ln \left(\frac{\partial z(y,A(\ell))}{\partial y} \right)}{\partial \ell}}_{\text{Productivity Gains}} + \underbrace{\frac{\partial \ln l(y,\ell)}{\partial \ell}}_{\text{Competition}} > 0. \tag{13}$$

Whereas the local productivity gains from settling into high-ℓ locations are positive if production technology z is supermodular (first term in (13)), the local competition effect is negative under positive sorting since productive firms cluster in the best locations (second term in (13)). Positive sorting thus emerges if the productivity benefits that boost profits per worker outweigh the costs from competition that translate into lower expected firm size.

Productivity gains are large if productivity differences across space are large (the A-schedule is steep) and when complementarities of z in $(y, A(\ell))$ are strong. Note that for the multiplicative technology $z(y,A) = Ay$—the functional form used in our quantitative analysis—these gains become $\partial \ln A(\ell)/\partial \ell$ and thus only depend on the spatial variation in local TFP.

In turn, the cost of local competition depresses firm size and is captured by

$$\frac{\partial \ln l(y,\ell)}{\partial \ell} = -\frac{2\varphi^E}{1+\varphi^E\left(1-\Gamma_\ell(y)\right)} \left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell} \right) \leq 0. \tag{14}$$

It depends on two forces: first, on how endogenous firm distribution Γ_ℓ varies across space, and second, on the degree of labor market frictions $\varphi^E = \lambda^E/\delta$ that determines the impact of

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

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 in that case 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 predominantly results 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.

To guarantee the supermodularity of $\overline{J}(p,\ell)$ in (p,ℓ) in terms of *primitives*, we use (13) and (14) to obtain the following sufficient condition for PAM:

Proposition 1 (Spatial Sorting of Firms)**.** *If z is strictly supermodular, and either the productivity gains from sorting into higher ℓ 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 productivity gains" and "sufficiently small φ^E" precise.

Under the conditions of Proposition 1, 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, $\partial\Gamma_\ell/\partial\ell = \left(\partial\Gamma(y|\mu(\ell))/\partial p\right)\mu'(\ell) \leq 0$.

Note that, in the absence of on-the-job search, $\varphi^E = 0$, $\partial \ln l(y,\ell)/\partial\ell = 0$ and complementarities in production are enough to sustain positive sorting. It may be surprising at first sight that larger labor market frictions (lower φ^E) facilitate sorting. What is important to realize, however, is that in the frictionless case, $\varphi^E \to \infty$, a winner-takes-all allocation takes hold whereby the most productive firm attracts all workers in a given local labor market. Ex ante, this discourages firms from collocating with productive peers, which prevents positive sorting.

Generalizations. We generalize Proposition 1 in three ways. First, we can allow for *endogenous productivity spillovers*. Instead of assuming exogenous differences in A, we assume that A depends 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 O2, Online Appendix OA-2.1, we show that if the (endogenous) location productivity advantage is large enough relative to the cost of competition, firms with high-p would indeed settle into high-ℓ locations, similar to the baseline model. Second, we *endogenize vacancy creation*. In Proposition O4, Online Appendix OA-2.2, we show that also in this case, the trade-off between productivity and competition determines firm sorting choices. Finally, we can allow for *labor mobility* and *residential housing* (Proposition 8, Appendix C).

3.2 Existence and Uniqueness

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

Proposition 2 (Existence & Uniqueness)**.** *Assume that the conditions from Proposition 1 hold; then a unique equilibrium (up to a constant of integration in land price function k) exists.*

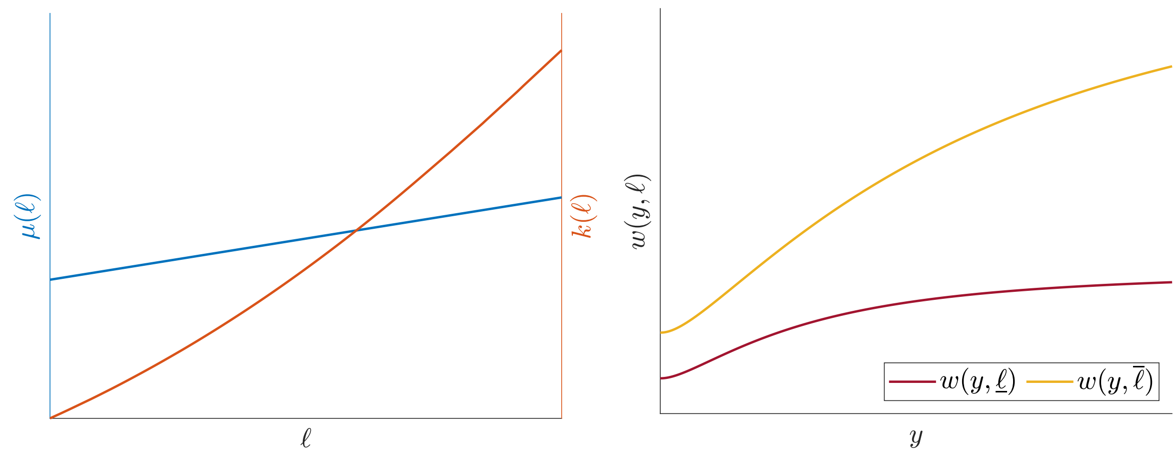
The proof is in Appendix B.2. We show the existence of a fixed point in Γ_ℓ by construction. In turn, uniqueness arises because, under the conditions on primitives stated in Proposition 1, the impact of endogenous firm composition on the firms' value function leaves the complementarity properties of $\overline{J}$ 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. The 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 and steepness ensure 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 Γ_ℓ. Both aspects are important determinants of spatial inequality.

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.

3.4 Spatial Firm Sorting and Spatial Wage Inequality

We now turn to the main focus of this paper: the implications of firm sorting for spatial wage inequality. Spatial firm sorting affects local wage distributions through two channels: It impacts local labor market *competition*, and thereby local job ladders; and it alters the employment distribution through a *composition* effect. Through these mechanisms, spatial sorting amplifies spatial wage inequality both within and across locations.

To simplify exposition, we will state all propositions in this section for the case of multiplicative technology $z(y, A(\ell)) = yA(\ell)$.[12] In this case, wage schedule (5) reads

$$w(y,\ell) = A(\ell)\left(y - \int_{\underline{y}}^{y} \frac{l(t,\ell)}{l(y,\ell)}dt\right). \tag{15}$$

This wage function is log-additive in location TFP $A(\ell)$ and a term that captures firm y's competition through its relative firm size, which in turn is driven by local productivity distribution Γ_ℓ and the extent of labor market frictions φ_E (see (3)).

Spatial Firm Sorting and Spatial Wage Inequality. Consider first inequality *across* locations, which we measure by the *spatial wage premium*. It compares the average wages of more productive locations with the wage of the least productive one, $\mathbb{E}[w(y,\ell)|\ell]/\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]$. To illustrate the drivers of spatial inequality, we consider how this measure varies as we increase ℓ, where the derivative of the spatial wage premium wrt ℓ is equal in sign ($\stackrel{\text{S}}{=}$) to

[12] The proofs in this section are contained in Appendix B.3-B.6. All results can be easily stated for general z, except for Proposition 4 and Corollary 1, in which multiplicative z avoids ambiguity.

$$\frac{\partial \frac{\mathbb{E}[w(y,\ell)|\ell]}{\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]}}{\partial \ell} \overset{\text{S}}{=} \underbrace{\underline{y}\frac{\partial A(\ell)}{\partial \ell}}_{\substack{\text{Intercept of} \\ \text{Job Ladder}}} + \int_{\underline{y}}^{\overline{y}} \Bigg(\underbrace{\frac{\partial^2 w(y,\ell)}{\partial y \partial \ell}}_{\substack{\text{Spatial Variation in} \\ \text{Steepness of Job Ladder}}} (1 - G_\ell(y)) + \underbrace{\frac{\partial w(y,\ell)}{\partial y}}_{\substack{\text{Job} \\ \text{Ladder}}} \underbrace{\left(-\frac{\partial G_\ell(y)}{\partial \ell} \right)}_{\substack{\text{Spatial Variation in} \\ \text{Employment Composition}}} \Bigg) dy. \quad (16)$$

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, which alters labor market *competition* in ℓ as well as the local employment *composition*. These differences drive the three factors that underlie cross-location wage inequality. The first two are displayed in the right panel of Figure 1: First, higher-ℓ locations have a higher intercept of the wage function and thus job ladder, due to higher location TFP. Second, higher-ℓ locations have steeper wage functions (i.e., a *steeper job ladder*), which stems from both the productive complementarities in (y, ℓ) along with a sufficiently increasing TFP A; and positive firm sorting: Tougher competition for workers among highly productive firms pushes up wages. In addition, there is an important composition effect due to positive sorting: Higher-ℓ locations have better firms and hence a stochastically better employment distribution G_ℓ. More employment is clustered at the upper part of the wage schedule, where wages are higher.

Proposition 3 (Firm Sorting & Between-Location Wage Inequality)**.** *Suppose the elasticity of local TFP, $\partial \ln A(\ell)/\partial \ell$, is sufficiently large, which satisfies the conditions from Proposition 1 and thus renders positive firm sorting across space. Then:*

(i) Local job ladder $w(\cdot, \ell)$ is steeper in high-ℓ locations, $\partial^2 w(y,\ell)/\partial y \partial \ell > 0$.

(ii) Spatial wage premium $\mathbb{E}[w(y,\ell)|\ell]/\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]$ is increasing in ℓ.

Anticipating our empirical measurement below, we note that the wage gradient with respect to firm's local productivity *rank* $\mathcal{R} := \Gamma_\ell(y)$ also steepens in ℓ (see the Remark, Appendix B.3).

In turn, to analyze how *within-location wage inequality* varies across space, we focus on the variation in the local *wage range*, $w(\overline{y}, \ell)/w(\underline{y}, \ell)$, derived using (15):

$$\frac{\partial \frac{w(\overline{y},\ell)}{w(\underline{y},\ell)}}{\partial \ell} \overset{\text{S}}{=} -\int_{\underline{y}}^{\overline{y}} \underbrace{\frac{\partial \frac{l(t,\ell)}{l(\overline{y},\ell)}}{\partial \ell}}_{\text{Competition}} dt = \int_{\underline{y}}^{\overline{y}} \frac{2\varphi^E}{(1+\varphi^E(1-\Gamma_\ell(t)))^3} \left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell} \right) dt \quad (17)$$

Wage dispersion is *larger* in high-ℓ locations since, under positive firm sorting, they have a cluster of productive firms that induces stronger competition for workers (reflected in lower firm size for all $y < \overline{y}$) and puts upward pressure on the highest wages. The variation in spatial wage dispersion is independent of local TFP and only depends on firm sorting. In the absence of firm sorting, $\partial\Gamma_\ell(y)/\partial\ell = 0$, within-location inequality would be equalized across space.

Proposition 4 (Firm Sorting & Within-Location Wage Inequality)**.** *If there is positive firm sorting across space, then wage dispersion, $w(\overline{y},\ell)/w(\underline{y},\ell)$, is increasing in ℓ.*

Importantly, firm sorting affects between- and within-location inequality only if there is OJS. Without OJS, $\varphi^E = 0$, local job ladders collapse everywhere and all workers receive their local reservation wage—a spatial Diamond paradox. So the effect of firm sorting on both competition (which affects the steepness of job ladders) and employment composition along the local job ladder is mute.

Detecting Spatial Firm Sorting. Given the importance of spatial firm sorting for spatial inequality, a natural question is how to detect it in the data. We show that firm sorting has distinct implications for the spatial variation in local labor shares and local productivity dispersion—all outcomes that are, in principle, observable.

We first show that under positive firm sorting, high-ℓ locations have a *lower* labor share, defined as the weighted average of firm-level labor shares $Ls(y,\ell) := w(y,\ell)/z(y,A(\ell))$ in each ℓ

$$LS(\ell) \equiv \int_{\underline{y}}^{\overline{y}} Ls(y,\ell)\tilde{g}_\ell(y)dy,$$

where $\tilde{g}_\ell$ is the value-added weighted employment distribution (see Appendix B.5 for the derivation). The spatial gradient of the local labor share is then given by

$$\frac{\partial LS(\ell)}{\partial \ell} = \int_{\underline{y}}^{\overline{y}} \Bigg(\underbrace{\frac{\partial^2 Ls(y,\ell)}{\partial y \partial \ell}}_{\substack{\text{Spatial Variation in} \\ \text{Firm Monopsony Power}}} (1-\tilde{G}_\ell(y)) + \underbrace{\frac{\partial Ls(y,\ell)}{\partial y}}_{\substack{\text{Heterogeneity in Firm} \\ \text{Monopsony Power}}} \underbrace{\left(-\frac{\partial \tilde{G}_\ell(y)}{\partial \ell}\right)}_{\substack{\text{Spatial Variation in} \\ \text{Employment Composition}}} \Bigg) dy, \tag{18}$$

where $\tilde{G}_\ell$ is the cdf corresponding to density $\tilde{g}_\ell$. The local labor share is decreasing in ℓ if firm-level labor share $Ls(y,\ell)$ is submodular in (y,ℓ) and decreasing in firm productivity y; and if positive firm sorting triggers a composition effect that translates into stochastically better employment distributions $\tilde{G}_\ell$ in high-ℓ places. The intuition why positive firm sorting along with these properties of the *firm*-level labor share leads to lower average labor shares in high-ℓ locations is straightforward. Under positive sorting, high-ℓ places tend to have employment distributions that are tilted toward productive firms. If, additionally, more productive firms have more monopsony power and thus lower labor shares $Ls(y,\ell)$, locations in which employment is concentrated in top firms tend to have a lower average labor share. In turn, submodularity of the firm-level labor share means that productive firms in high-ℓ locations have especially strong monopsony power and thus particularly low labor shares, which further lowers the average labor share in their location.

Proposition 5 (Firm Sorting & Local Labor Shares)**.** *If there is positive firm sorting across space and φ^E is sufficiently small, then the local labor share, $LS(\cdot)$, is decreasing in ℓ.*

Positive firm sorting leads to a decreasing local labor share if sufficiently strong labor market frictions ensure that firm-level labor shares are decreasing in productivity (under this condition, productivity schedule $z(\cdot, A)$ increases faster in y than wage schedule $w(\cdot, \ell)$).

In our quantitative setting below, we will assume $\Gamma(y|p) \sim Pareto(1, 1/p)$. Because the density of the Pareto distribution is decreasing, more productive firms in each ℓ face less labor market competition, which translates into more monopsony power and lower labor shares (circumventing the need for the condition on φ^E in Proposition 5).[13] Indeed, in this case, the local labor share *only* depends on the Pareto tail coefficient, p, of the local productivity distribution. High-ℓ locations have a thicker tail and therefore more mass on firms with high monopsony power, which reduces the local labor share.

Corollary 1 (Firm Sorting & Local Labor Shares: Pareto)**.** *Suppose* $\Gamma(y|p) \sim Pareto(1, 1/p)$. *If and only if there is positive firm sorting across space, then the local labor share, given by* $LS(\ell) = 1 - \mu(\ell)$, *is decreasing in* ℓ.

Corollary 1 plays an important role in our quantitative analysis below because it allows us to infer firm sorting, captured by $\mu(\ell)$, directly from local labor shares.

We supplement this result on how to detect firm sorting in two ways. First, we show that positive firm sorting also implies that high-ℓ locations have more productivity dispersion, captured by the quantile ratio $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$ (where $t', t'' \in (0, 1)$ and $t'' > t'$), which is increasing in ℓ (Proposition 7, Appendix B.7). This result applies to the class of productivity distributions $\Gamma(y|p)$ in which stochastic dominance wrt p is more pronounced for higher y. In Corollary 2 (Appendix B.7), we show that the Pareto assumption satisfies this distributional requirement and renders positive sorting not only sufficient but also necessary for the result. Second, we show that firm sorting has testable implications for the relationship between the local and the global (economy-wide) productivity rank of firms (Proposition O5, Online Appendix OA-3).

[13] Another example of a distribution with decreasing density that renders a negative relationship between local labor share and local firm productivity is the (truncated) log-normal distribution; see Online Appendix OA-5.1.

4 Descriptive Evidence

Turning to our empirical application, we first assess our model's *qualitative* predictions on firm sorting and wage inequality, both within and across locations. In Section 6, we estimate our model to highlight the *quantitative* implications of firm sorting for spatial inequality.

4.1 Data and Measurement

We base our analysis on 257 commuting zones (CZ)—our local labor markets—which are similar to US commuting zones. We order them by local GDP per capita—a commonly used measure of economic prosperity—which in our model is increasing in productivity index ℓ.

We use three main data sources: (i) regional data from the German Federal Statistical Office on GDP per capita, labor compensation, value added, and unemployment rates for each CZ; (ii) worker-level panel data from linked employer-employee data in Germany (LIAB) provided by the Research Data Centre (FDZ) of the German Federal Employment Agency, which are based on workers' social security records and contain information on wages and worker flows; and (iii) firm-level data from the Establishment History Panel (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.[14] From the BHP, we construct firm-level average wages—the relevant concept of wages through the lens of our theory—to assess wage inequality within and across local labor markets. We deflate wages using a nationwide CPI.[15] We complement our main data sources with information on firm-level sales from the German Establishment Panel (EP). Throughout, we focus on the period 2010-2017.

In Appendix D, we describe these data sources in more detail, define important variables, and report some basic statistics, which show that average wages, firm size, and value added differ substantially across commuting zones (see Table A.1, Appendix D.5).

4.2 Spatial Firm Sorting and Spatial Wage Inequality: Evidence

We have shown that, in theory, firm sorting amplifies wage inequality between and across labor markets (Section 3.4). An important mechanism through which this occurs is that firm sorting steepens the local job ladder in more productive locations. Our theory also suggests that we can identify firm sorting in the data based on spatial variation in local labor shares. We now provide empirical evidence along all these dimensions.

[14]The data are at establishment level, but we use the terms "firm" and "establishment" interchangeably.

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

Local Job Ladders and Local EE Wage Growth. We first provide evidence that local job ladders differ systematically across space, consistent with positive spatial sorting of firms (Proposition 3(i)). For a given location ℓ, we measure the local job ladder by the local wage as a function of firms' local productivity *rank* $\mathcal{R} := \Gamma_\ell(y)$ or $w(\mathcal{R}, \ell)$.

In Figure 2, left panel, we plot local job ladders for four groups of labor markets that are ordered by their local GDP per capita. Specifically, we first compute the local job ladder in each location based on the quantiles of the firm-level wage distribution (corresponding to $w(\mathcal{R}, \ell)$ in our model since $\mathcal{R} = \Gamma_\ell(y) = F_\ell(w(y, \ell))$). We then take the average of all job ladders within each GDP quartile. To focus on the differential *steepness* of these local job ladders, we normalize the bottom wage of each local ladder to unity. We find that local job ladders differ meaningfully across space: More productive locations are characterized by steeper job ladders. Whereas in rich locations top wages exceed bottom wages by a factor of 3.5, poor locations only see a rise in wages by a factor of less than 3.

The fact that job ladders are steeper in rich locations suggests that the returns to EE transitions should be higher in those labor markets. To assess this claim, we run the following regression:

$$\frac{w_{i\ell,t} - w_{i\ell,t-1}}{w_{i\ell,t-1}} = \sum_{\ell=1}^{257} \beta_\ell + \sum_{\ell=1}^{257} \beta_\ell^{EE} \, \mathrm{EE}_{i\ell,t} + \sum_{\ell=1}^{257} \beta_\ell^{EXT} \, \mathrm{EXT}_{i\ell,t} + \varepsilon_{i\ell,t}, \tag{19}$$

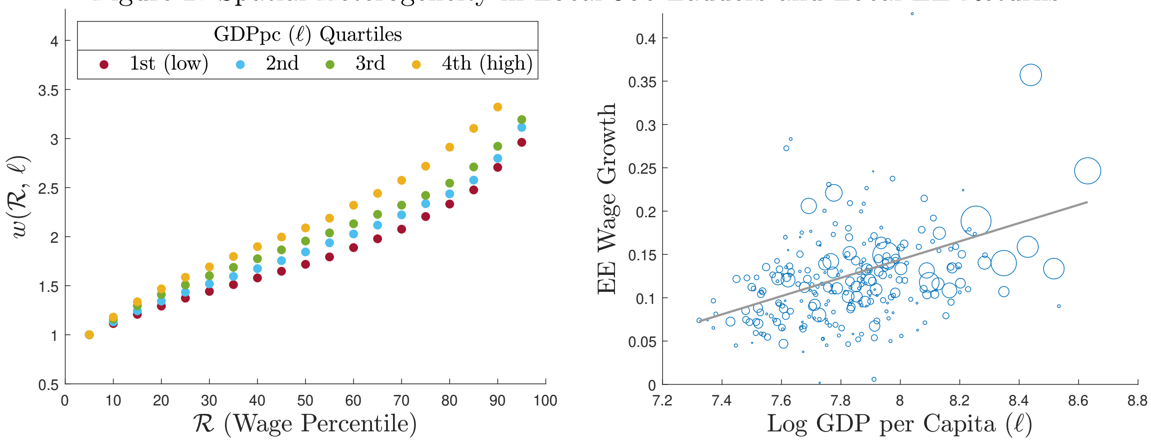
where $w_{i\ell,t}$ is the wage of individual i in CZ ℓ and month t, $EE_{i\ell,t}$ indicates whether individual i made an EE move to a job in CZ ℓ between months t and $t-1$, and $EXT_{i\ell,t}$ indicates an EE transition to a job in CZ ℓ from a job outside of ℓ. The coefficients $(\beta_\ell, \beta_\ell^{EE}, \beta_\ell^{EXT})$ are CZ fixed effects and CZ-specific returns to EE moves from within and outside the CZ, respectively.

In the right panel of Figure 2, we plot β_ℓ^{EE}—the effect of an EE transition *within* local labor market ℓ on wage growth—against ℓ. Consistent with steeper local job ladders, EE returns are substantially higher in productive places: 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 as in the poorest location.

It is important to note that the high EE returns in rich places are *not* driven by spatial differences in worker composition along observable dimensions. We show in Appendix E.1 (Figure A.1) that our estimates are virtually unchanged when we control for individuals' age, gender, and education.

To assess the quantitative impact of this spatial variation in EE returns on spatial wage

Figure 2: Spatial Heterogeneity in Local Job Ladders and Local EE Returns



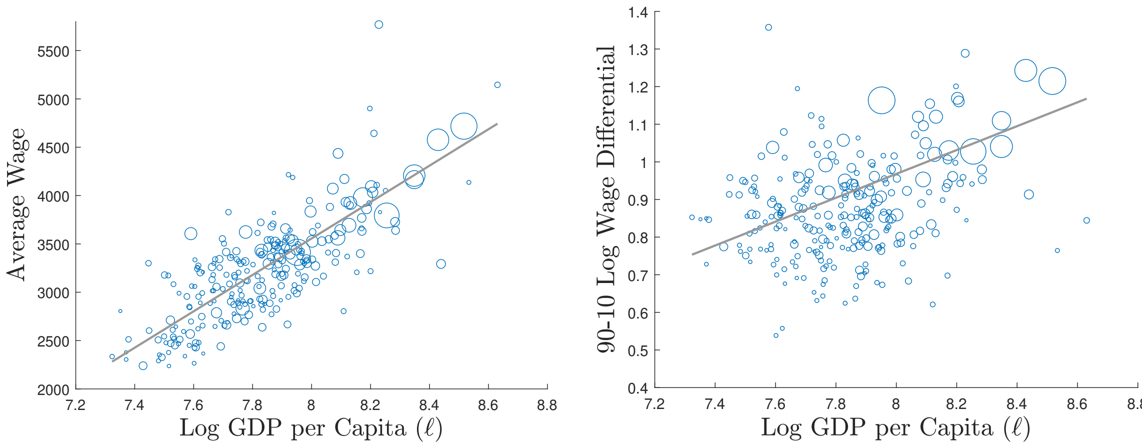
Notes: Data sources: BHP and LIAB. In the left panel, we display wage quantiles relative to the 5% quantile for four groups of CZs, ordered by quartiles of local GDPpc. In the right panel, we plot the location-specific return to an EE-transition, that is, β_ℓ^{EE} across ℓ based on (19). The size of the markers indicates the number of EE moves within each CZ.

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 more than 25% 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 Appendix E.1 for details on this exercise.

Spatial Wage Inequality. Figure 3 assesses the extent of spatial wage inequality in the data. The left panel shows that the most productive locations have substantially higher (by almost 50%) average wages than the least productive ones. In turn, we use the difference between the 90th and 10th quantiles of log wages within each CZ to measure the wage range within locations. The right panel shows that wage inequality within locations is substantially higher in productive locations: The 90-10 gap in log wages differs by about 0.4 log points between poor and rich labor markets.

Our theory implies that positive firm sorting fuels these patterns of spatial wage inequality through two channels: (i) by intensifying labor market *competition* in high-ℓ locations and (ii) through a *composition* effect whereby employment in high-ℓ places is concentrated in more productive firms that pay higher wages. More productive locations then have higher average wages, above and beyond the direct TFP effect through A (see decomposition (16) and Proposition 3), and more wage dispersion (see (17) and Proposition 4).

Figure 3: Wage Inequality Across and Within Locations



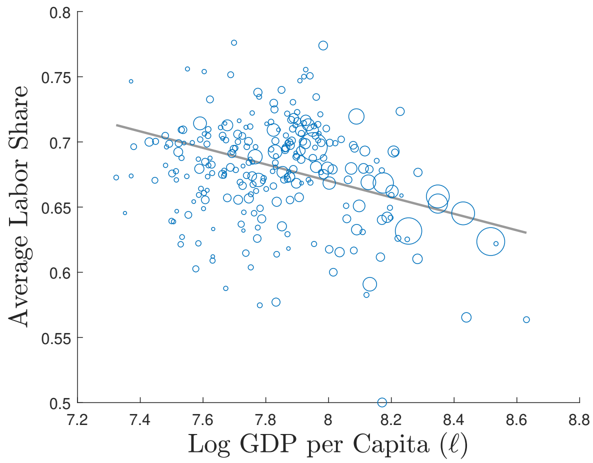
Notes: Data source: BHP. The left panel shows a scatter plot between average local wages and local log GDPpc. The right panel shows a scatter plot between the local difference in the 90th and 10th quantiles of log wages and local log GDPpc. Wages are measured at the firm level. The size of the markers indicates the size of the CZ (i.e., the number of firms in each CZ).

Detecting Spatial Firm Sorting. The spatial job ladder heterogeneity and inequality documented in Figures 2 and 3 are consistent with spatial firm sorting but only provide indirect evidence for it. Obtaining direct evidence is challenging because our dataset—like many others—does not contain a direct measure of physical firm productivity and does not allow us to identify firm movers. We therefore use our model predictions to obtain evidence for spatial firm sorting and identify it from variation in local labor shares (Proposition 5 and Corollary 1). As stated in Corollary 1, if firm productivity is Pareto distributed and technology multiplicative, local labor shares are *entirely* pinned down by the tail parameter of the local productivity distribution. Specifically, if the local labor share is decreasing in local productivity index ℓ, there is positive sorting of firms across space.

In Figure 4, we display the correlation between local labor shares and GDP per capita across 257 labor markets in Germany. Richer locations indeed have lower labor shares—which we believe is a new fact—and this spatial variation is economically meaningful: The labor share is about 6% lower in rich compared with poor labor markets. We deduce that firm sorting is positive, with stochastically better firm distributions in more productive locations. The intuition for why a negative correlation between local labor shares and local GDP indicates positive firm sorting is that within each labor market, more productive firms have more monopsony power and thus lower *firm-level* labor shares. Rich locations with low aggregate labor shares must therefore have a higher concentration of employment in top firms with low labor share—a composition effect implied by positive firm sorting.

The link between local labor shares and the spatial allocation of firms plays an important

Figure 4: Local Labor Shares Across Space



Notes: Data source: German Federal Statistical Office. The figure shows a scatter plot between local log GDPpc and local labor shares, defined as the ratio between *labor compensation* and *gross value added* in each CZ (see Appendix D.1 for details). The size of the markers reflects the number of firms in each CZ.

role in our quantitative analysis below because we use it as a calibration target to identify firm sorting. In Appendix E.2, we supplement this evidence of positive firm sorting, using spatial variation in firm sales per worker. Specifically, we document that the within-location dispersion in sales per worker is larger in productive labor markets, which—based on Corollary 2 in Appendix B.7—indicates positive firm sorting across space.[16] In addition, we show that the relationship between the difference in firms' global and local productivity ranks and firm productivity is consistent with spatial firm sorting (Online Appendix OA-3). However, given that we observe sales for only a small subset of firms, these results, while qualitatively in line with positive firm sorting, are imprecisely estimated. As a consequence, our preferred evidence for spatial firm sorting stems from local labor shares, based on regional data from the official records of the German Federal Statistical Office.

Alternative Explanations for Our Facts. Figures 2, 3, and 4 show that local labor markets in Germany systematically differ: Prosperous locations have higher wages, more within-location wage inequality, steeper job ladders, higher EE returns, and lower labor shares. We have demonstrated that all of these patterns are natural implications of our model, in which firms positively sort into segmented local labor markets that feature search frictions and OJS. We now consider several alternative mechanisms and argue that they are unlikely to be the driving force behind these facts.

In each column of Table 1, we regress one of our outcomes—local log wage, local 90-10

[16] The predictions of Corollaries 1 and 2 hinge on the assumption that local productivity is Pareto distributed. In Appendix E.2, we analyze tail variation in local productivity distributions and show that this assumption is supported by the data.

log wage differential, local EE return, and local labor share—on log GDP per capita. Row 1 shows the unconditional relationships from Figures 2 to 4, which is our baseline.

A first concern might be that these patterns do not reflect firm sorting but are driven by regional differences in the industrial composition (Gaubert, 2018); our model abstracts from this aspect. For example, the negative correlation between local GDP per capita and local labor shares could be driven by the greater importance of capital-intensive industries in prosperous locations. In row 2 of Table 1, we show that this is not the case. When controlling for regional differences in the industrial composition through sectoral employment shares, the relationship between the local labor share and local GDP per capita becomes even more pronounced (and the relationships between the other three variables and local GDP remain similar).

Second, a natural question is to what extent the spatial sorting of workers (rather than of firms) or agglomeration economies can account for these patterns. In particular, wage differences across locations (column 1) could be driven by the fact that highly productive workers settle in more prosperous places (Heise and Porzio, 2022); wage differences within locations (column 2) could reflect a higher variance of skills in more productive locations (Eeckhout et al., 2014); and differences in EE returns (column 3) could reflect better match quality if rich and dense locations benefit from increasing returns to matching (Dauth et al., 2022). In row 3, we therefore control for the worker composition. In columns 1 and 2 we measure firm-level wages as the residual after controlling for the employees' education; in columns 3 and 4 we control for the skill composition of the local workforce. Furthermore, in row 4 we control for population density. Again, we find that our baseline results are qualitatively robust to these controls.

Finally, the higher mean and greater dispersion of wages in more productive places may be due to the accumulation of more valuable experience (Baum-Snow and Pavan, 2011; De La Roca and Puga, 2017). We therefore investigate the role of the location-specific experience premia of job "stayers" (i.e., of workers who do not change jobs), captured by β_ℓ in regression (19). In stark contrast to the EE returns (β_ℓ^{EE}), experience premia (β_ℓ) are both relatively small (on the order of 0.3% per month) and vary little across space; see Figure A.2 (Appendix E.1).[17] This suggests that the spatial heterogeneity in dynamic wage profiles emphasized by De La Roca and Puga (2017) may in part stem from the spatial variation in EE returns (Figure 2, right panel), which is the focus of our paper but has so far not received much attention.

[17]This is in line with Bowlus and Neuman (2006), who use earnings data from the U.S. Census (1940–2000) to show that the wage growth of stayers is a small fraction of overall wage growth, which suggests that job mobility is more important.

Table 1: Wage Inequality, EE Returns, and Labor Shares Across Local Labor Markets

	(1) Log Wage	(2) 90-10 Log Wage Diff.	(3) EE Return	(4) Labor share
Log GDPpc [No Controls]	0.5350***	0.3172***	0.1054***	-0.0632***
	(0.0271)	(0.0459)	(0.0309)	(0.0088)
Log GDPpc [Control: Industry]	0.4433***	0.1685***	0.1728*	-0.1024***
	(0.0519)	(0.0479)	(0.0775)	(0.0212)
Log GDPpc [Control: Worker Composition]	0.2600***	0.3789***	0.1417**	-0.1130***
	(0.0272)	(0.0566)	(0.0447)	(0.0134)
Log GDPpc [Control: Pop. Density]	0.4723***	0.2029***	0.1109**	-0.0799***
	(0.0418)	(0.0601)	(0.0380)	(0.0086)
N	257	257	257	257

Notes: Data Sources: German Federal Statistical Office, BHP, and LIAB. We run regressions of the form $out_\ell = \beta \ln GDP_\ell + x'_\ell \gamma + \epsilon_\ell$, where GDP_ℓ denotes GDP per capita in CZ ℓ and out_ℓ denotes one of the outcomes in the different columns of the table. The table reports estimates of β and the associated standard error in parentheses. In row 2, we control for local industrial employment shares in seven industries (agriculture; mining and utilities; manufacturing; construction; trade and transportation; professional services; public administration and health). In row 3, columns 1 and 2, we measure firm-level average wages as the residual after controlling for the education level of the firms' employees. In row 3, columns 3 and 4, we control for workers' skill levels at the commuting-zone level. In row 4, we control for local log population density.

5 Estimation

To evaluate the quantitative importance of firm sorting for spatial wage inequality, we structurally 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.[18] 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 will ensure that unemployed workers' value of search is the same across space. Last, we allow job separation rates δ to vary (exogenously) across locations to rationalize the spatial variation in unemployment rates that we observe.[19] In Section 7, we show robustness of our quantitative results to additional changes: introducing imperfect labor mobility; dispensing with Assumption 1 so that there can be firm selection at the lower end of the local productivity distribution; and controlling for local capital intensity via industry.

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

[19] Importantly, in the estimated model positive firm sorting is also optimal if δ is constant in ℓ.

By allowing for spatial mobility among the unemployed, our model endogenizes local population size $L(\ell)$, and thereby also local meeting rates of workers ($\lambda^U(\ell)$, $\lambda^E(\ell)$) and firms $\lambda^F(\ell)$. We assume that in each location there is a labor market matching function with constant returns to scale, so that 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.[20] 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 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 workers is wage $w(y, \ell)$ and that of unemployed workers is benefit $w^U(\ell)$, financed via taxes on homeowners' income, τ.[21] Further, the government budget needs to balance, $\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell)$, where $d(\ell)$ is the housing price in ℓ. It adjusts to clear the housing market, balancing housing demand from unemployed and employed workers with housing supply $h(\ell)$.

The population size in each ℓ, $L(\ell)$, is pinned down by the fact that in equilibrium, workers 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:

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

[20] 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 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 $\mathcal{V}(\ell) = 1$.

[21] 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.

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 our baseline model guarantee the positive sorting of firms to locations; see Proposition 8 (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 (13) varies across space.

Functional Forms. As for local labor markets, we assume that worker-firm meetings 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,

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

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.[22]

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 spatial firm sorting, captured by $\mu(\ell)$ (which in equilibrium equals p, the (inverse) Pareto tail of ℓ's productivity distribution).[23]

The key step in our identification strategy 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? As already highlighted in

[22] Note that we normalize the scale parameter, $\underline{y} = 1$. We also investigated whether the Pareto assumption is justified. If ex post firm productivity is Pareto distributed in our model, then output, z, is as well. We checked empirically that the tails of the local distributions of firm sales are log-linear.

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

Corollary 1, 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, $\mathbb{E}[z(y, A(\ell))|\ell]$,

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

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 a single labor market, see Gouin-Bonenfant, 2020). Variation in local labor shares across space $LS(\cdot)$ thus identifies the extent of firm sorting $\mu(\cdot)$.[24] 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, 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 can identify the local job-separation rate, $\delta(\ell)$, using the steady-state formula for unemployment, as well as data on local unemployment and job-finding rates:

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

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,

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

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 the government budget constraint and housing market clearing—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:

[24] Pareto is not the only distribution that renders a negative relationship between local labor share and local firm productivity. See Online Appendix OA-5.1 for an example, in which local firm productivity is log-normally distributed.

Proposition 6 (Identification)**.** *Under the assumed functional forms (summarized in Assumption 2, Appendix F), the model is identified.*

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, and GDP. In turn, for model validation, we use 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 follow closely. Our implementation proceeds in seven steps. First, we again rank the 257 CZs based on their log GDP per capita (our proxy for productivity index ℓ). Local log GDP per capita will be our discretized support $\{\ell_1, \ell_2, ..., \ell_{257}\}$ of the model's land distribution R. To reflect differences in the number of firms across locations in the data, we 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), we obtain an *increasing* matching function μ (top right).[25] This implies positive sorting between firms and locations.

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.5, 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.5.

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

[25] 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 the (endogenous) job finding rate $\lambda^U(\ell)$ in *each* ℓ from the average firm size $\bar{l}(\ell)$ provided by the German Federal Statistical Office (Figure 5, second row, left).[26] We then compute $\lambda^E(\ell) = \kappa\lambda^U(\ell)$ and $\lambda^F(\ell) = \lambda^U(\ell)/\theta(\ell)$. Since average firm size is increasing in ℓ, 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).

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.[27] 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 corporate tax rate, a high share of college-educated workers, and the quality of infrastructure; see Table A.6, 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.5, Appendix G.1). Finally, we pin down local housing supply $h(\ell)$ using observed location-specific rental rates $d(\ell)$; see Figure A.6 (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.6 (Appendix G.1) shows that amenities are decreasing in the location index. Thus, even though residential housing is more expensive in high-ℓ places, 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.

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

[27] 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 (12) into account, which depends on $(\mu(\ell), \lambda^U(\ell), \delta(\ell))$—all objects that we pinned down above.

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

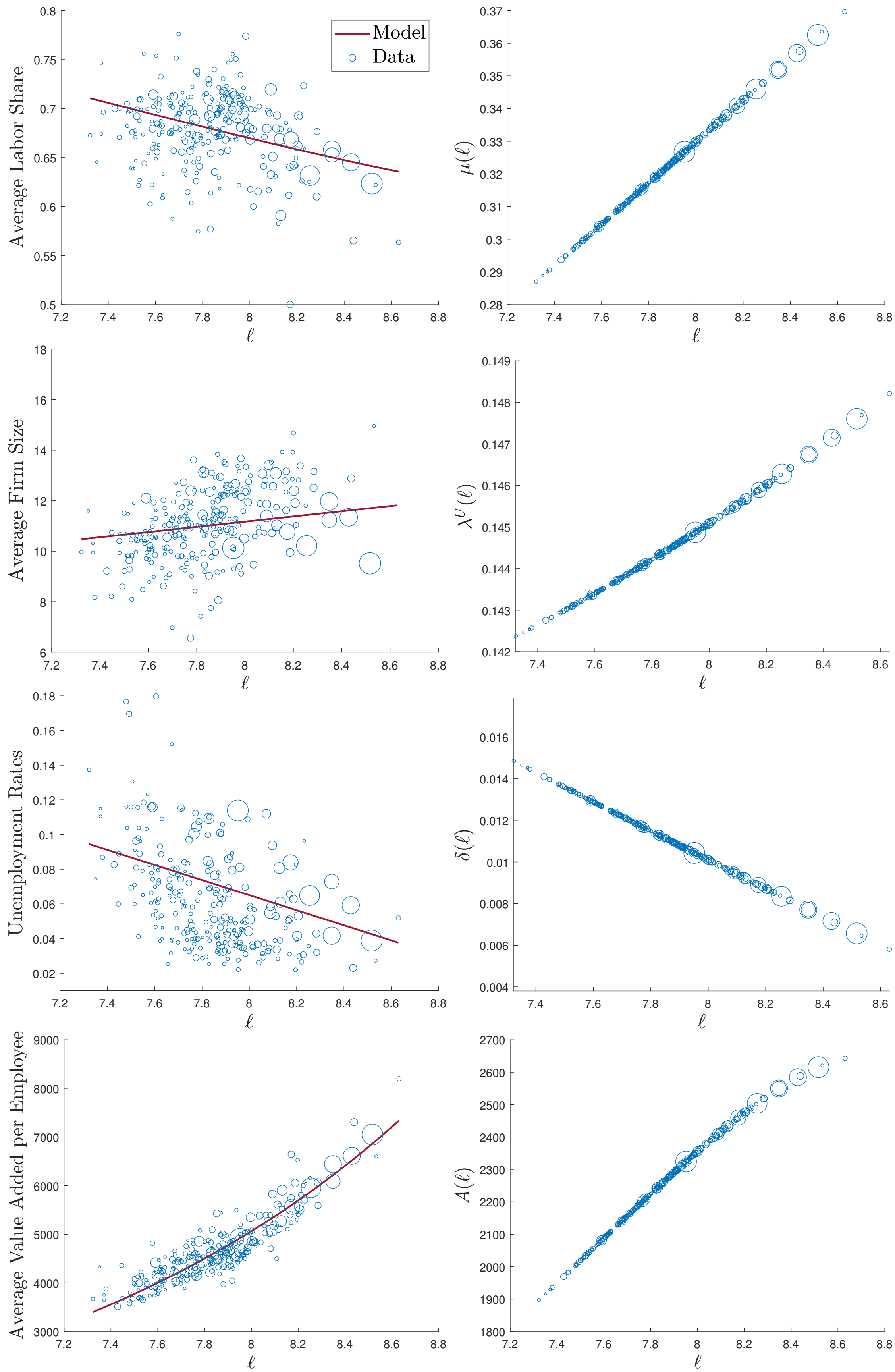
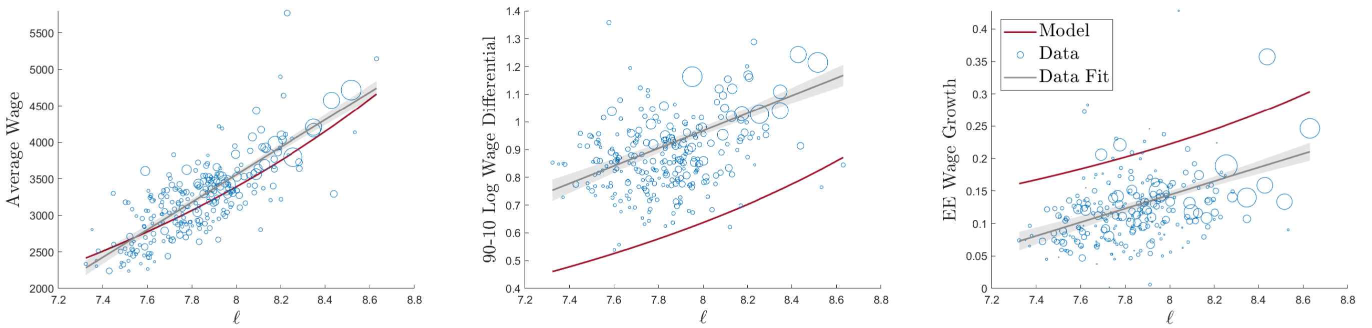


Figure 6: Model Fit: Non-Targeted Moments



Notes: Data sources: The left and middle panels are based on firm-level wages of full-time (FT) employees from BHP (and the percentiles are from the wage distribution that weighs observations by the number of FT employees of each firm); and the right panel is based on worker-level wages from the LIAB. The right panel shows β_ℓ^{EE} from regression (19). These local statistics are weighted by the number of firms (left and middle panel) and the number of EE moves within the CZ (right panel), indicated by different marker sizes. 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 that are related to worker inequality and beyond.

In Figure 6, we show that our model replicates other important features of our empirical analysis, contained in Figures 2 and 3. In the left panel, we display the average local wage across space. Because our model perfectly matches value added per capita across ℓ, it is expected that it also fits spatial differences in average wages. Second, and more importantly, our model captures quite well both spatial differences in wage inequality within locations and spatial job ladder heterogeneity. In the middle panel of Figure 6, we display local 90-10 log wage gaps across space, which is the model counterpart of the right panel of Figure 3. Although our model slightly underestimates the level of within-location inequality, it almost perfectly rationalizes the spatial variation in this measure. The right panel shows that 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—the empirical counterpart to the right panel of Figure 2. We overestimate the level of EE wage growth, but, reassuringly, closely match the slope across locations.

We further validate the model along the following dimensions (see Appendix G.1). First, as an overidentification check, we show that our model matches relatively well the empirically observed job-finding rate among the unemployed, $\lambda^U(\cdot)$, and the rate of job loss, $\delta(\cdot)$ (Figure A.4). Second, our model replicates the fact that employment is more concentrated among top firms in rich labor markets. Specifically, in both the data and the model, the employment share among the largest 25% of firms is increasing in ℓ (Figure A.5, left panel). Finally, our model is consistent with the fact that commercial land prices are higher in rich labor markets, even though the increase is more pronounced in the data (Figure A.5, right panel).

The fact that our model fits well the spatial heterogeneity across CZs in a variety of dimensions suggests that it is a suitable tool for quantifying the relationship between firm sorting and spatial inequality.

6 Spatial Firm Sorting: Quantitative Implications

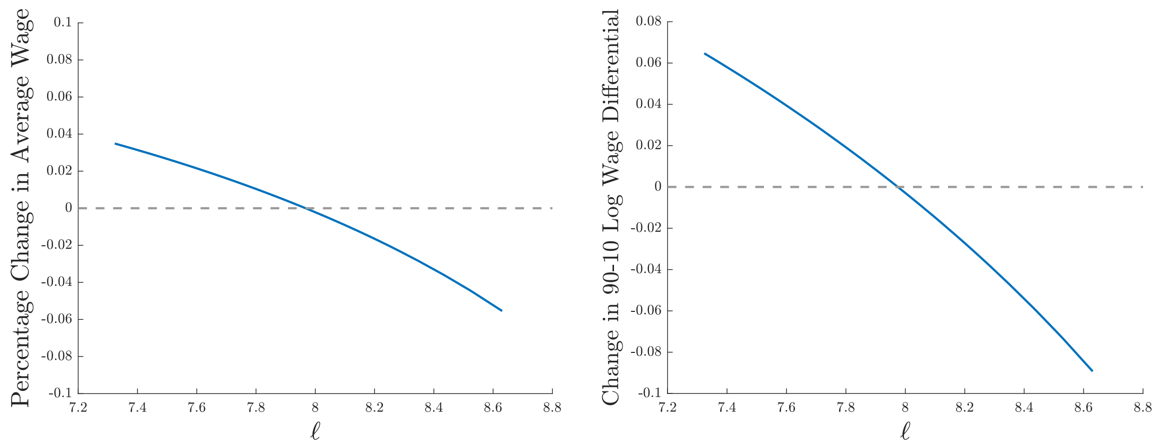
We use our estimated model to quantify the role of spatial firm sorting in spatial inequality. We also show that the presence of on-the-job search and spatial hiring frictions are central for our results.

6.1 The Implications of Spatial Firm Sorting for Spatial Inequality

Our estimation reveals large discrepancies between poor and rich labor markets in a variety of dimensions: Unproductive labor markets are not only disadvantaged because of poor economic fundamentals but also because workers lack access to the most productive firms. To assess to what extent this scarcity of productive employers shapes spatial wage inequality, both across and within labor markets, we allocate firms randomly across locations and let the local labor market equilibrium play out. Hence, workers reallocate across space, job-finding rates adjust via local market tightness, and local wage schedules change. Appendix H.1 contains technical details on this exercise. Throughout, we maintain Assumption 1. As a result, the economy-wide firm distribution stays the same as in the baseline model. In Section 7, we explicitly allow for firm selection as robustness.

We start with the implications of sorting for wage inequality *across* German labor markets. Firm sorting helps productive locations because it amplifies their advantage from a higher local TFP A. Hence, spatial inequality would be lower in the absence of sorting. The left panel of Figure 7, which displays the change in average local wages in the absence of sorting relative to the baseline equilibrium, indicates the quantitative importance of this channel. Without sorting, wages would be 6% lower in the richest labor markets and around 4% higher in the poorest locations. Table 2 summarizes the quantitative impact of firm sorting on spatial inequality. In the left panel, we begin with the spatial wage premium between poor and rich locations in Germany (i.e., between locations in the bottom and top quartiles of the

Figure 7: The Effect of Spatial Firm Sorting on Spatial Wage Inequality



Notes: The left panel shows the percentage change in average local wages if firm sorting is abolished. The right panel shows the percentage point change in the 90-10 difference in log wages within locations if firm sorting is abolished.

distribution of local GDP per capita). Average wages are 40% higher in rich labor markets. In the absence of firm sorting, this wage premium would decline to 34%. Spatial firm sorting thus accounts for $6/40 \approx 16\%$ of the observed spatial wage gap.

To put the quantitative role of sorting into perspective, we can shut down spatial differences in local TFP and set $A = \mathbb{E}[A(\ell)]$ in all locations while keeping firm sorting at its baseline level. Doing so would cut the spatial wage gap in half—that is, reduce it to 21%. Hence, the effect of firm sorting on spatial inequality is about one-third as large as the effect of innate differences in location TFP.

Table 2: The Effect of Spatial Firm Sorting on Spatial Wage Inequality

Across-Location Inequality *(Spatial Wage Premium)*			Within-Location Inequality *(90-10 Difference in log Wages)*		
Baseline	No Sorting	Contribution of Sorting	Baseline	No Sorting	Contribution of Sorting
40%	34%	16%	0.20	0.12	38%

Notes: Columns 1 and 2 contain the spatial wage premium between rich and poor labor markets in the baseline model and the counterfactual equilibrium without sorting. Column 3 reports the contribution of sorting as the percentage difference between columns 1 and 2. Columns 4-6 report the analogue for the 90-10 difference in log wages within labor markets. Throughout, we define poor (rich) location as the bottom (top) quartile of commuting zones ranked by their local GDP per capita.

We now turn to the implications of firm sorting for differences in wage inequality *within* German labor markets, which, as before, we measure by the 90-10 difference in local log wages. The right panel of Figure 7 shows how within-location inequality would change if firm sorting were removed: Wage inequality within rich labor markets would decline by almost 0.1 log points (because firm composition worsens), while inequality in poor localities would increase by 0.06 log points (because firm composition improves). These results suggest that firm sorting is a central determinant of the spatial variation in within-location inequality.

Indeed, Table 2 (right panel) shows that firm sorting accounts for 38% of the higher within-inequality in rich labor markets: Whereas the difference in within-location inequality between the top and bottom quartiles of German labor markets is 0.2 log points in our baseline model with sorting, this gap would drop to 0.12 log points in the absence of it. While differences in local TFP A do *not* affect within-location inequality (see (15)), spatial heterogeneity in job-destruction rates δ and job-finding rates λ does impact local job ladders and thereby account for the remaining spatial differences in within-inequality.

To better understand *why* changes in firm sorting affect inequality, we first 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 at the top and Mansfeld-Südharz at the bottom in terms of local GDP per capita.

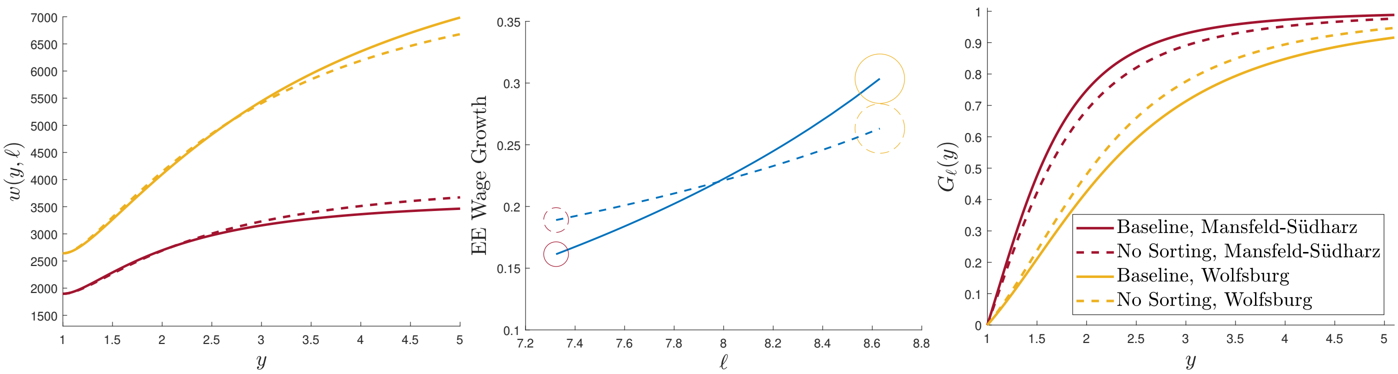
First, as shown in the left panel of Figure 8, random matching of firms to locations reduces job ladder differences across space. Replacing positive firm sorting with a random allocation improves firm productivity in bottom locations and deteriorates it in top ones. This alleviates labor market *competition* for workers in top locations while reinforcing it at the bottom. While the job ladder was considerably steeper in the top location (yellow solid) compared with the bottom location (red solid) at baseline, this differential steepness shrinks when firm sorting is absent (dashed lines).

These changes in local job ladders directly impact local returns from EE transitions. The middle panel of Figure 8 shows the effect of firm sorting on EE returns. Relative to our baseline estimation (solid line), spatial differences in EE returns are muted without sorting (dashed line). For our two locations, shown as red and yellow circles, the difference in EE returns falls by around 75%. This suggests that the observed spatial gap in EE wage growth is largely attributable to firm sorting.

Spatial convergence of local wage ladders and EE returns is an important mechanism behind the reduced wage inequality *across* space; and—by shrinking the wage range in top locations and widening it in bottom locations—this channel also explains the documented changes in *within*-location inequality.

Finally, the right panel of Figure 8 highlights the role of spatial differences in employment *composition*. In the baseline model with firm sorting, the employment distribution in Wolfsburg first-order stochastically dominates that of Mansfeld-Südharz (compare the solid cdf's). Therefore, in Wolfsburg, employment is clustered in more productive firms that pay higher

Figure 8: No-Sorting Counterfactual: Wages, EE Returns and Employment Distribution



Notes: The left panel shows the local job ladders in Wolfsburg (WB, yellow) and Mansfeld-Südharz (MS, red) with and without sorting. The middle panel reports wage growth due to EE moves with sorting (solid line) and without sorting (dashed line) as a function of ℓ. WB and MS are denoted by yellow and red circles, which are proportional to their relative size. The right panel shows employment distributions in WB and MS with and without sorting.

wages. Without firm sorting, the employment distributions across space converge.[28] This further contributes to a decline in wage inequality across space.

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

In our theory, firm sorting particularly affects spatial inequality by shaping the returns to on-the-job search (and the associated job ladders) within spatially segmented labor markets. To highlight the interaction between these features and spatial firm sorting, we show that firm sorting would have substantially *smaller* effects on inequality (both across and within locations) if there were less OJS or if local labor markets were integrated.

To that end, we consider two variations of our estimated model. First, we reduce the importance of OJS by lowering the search efficiency of employed workers, κ. Second, we integrate local labor markets 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 job ladder, facing no geographic constraints regarding which firms can recruit them. This is an economy that offers a remote work option, which decouples the place of residence from the place of work. In both cases—low OJS and no spatial frictions—the positive sorting of firms across space remains intact. Appendix H.2 and H.3 detail the implementation.

We start with the role of OJS in spatial inequality. The elasticity of across-location

[28]Due to labor reallocation, the FOSD in firm productivity distributions is not entirely undone by this counterfactual. Without firm sorting, top locations lose some of their economic appeal, which is why workers reallocate to poorer places. As a result, workers' matching rates in top places increase, which propels workers to the top of the local job ladder faster. Labor mobility thus mitigates the effect of firm sorting on spatial inequality.

inequality wrt to the poaching share—a proxy for the magnitude of OJS in the economy—is sizable: A 1% decline in the poaching share reduces the spatial wage premium of the most prosperous places by 0.6%. A reduction in OJS causes job ladders to flatten everywhere, but with a larger impact on well-off locations, in which competition for employed workers had been fierce. OJS is therefore an important amplifier of spatial inequality in our model.[29]

To analyze the *interaction* between OJS and firm sorting, we implement our random matching counterfactual in the environment with reduced OJS (implemented via a drop in κ, which—for illustration—corresponds to a 10% decline in the poaching share).[30] Comparing the impact of sorting on inequality in the baseline model (blue dashed lines) with its impact when OJS is muted (flatter green lines) in Figure 9 shows that firm sorting is now less important for inequality, both across and within locations. We conclude that OJS is an important reason *why* firm sorting affects inequality: If workers were to climb their local job ladder at a slower pace, spatial job ladder heterogeneity would be less important and the effect of firm sorting on spatial inequality would be mitigated.

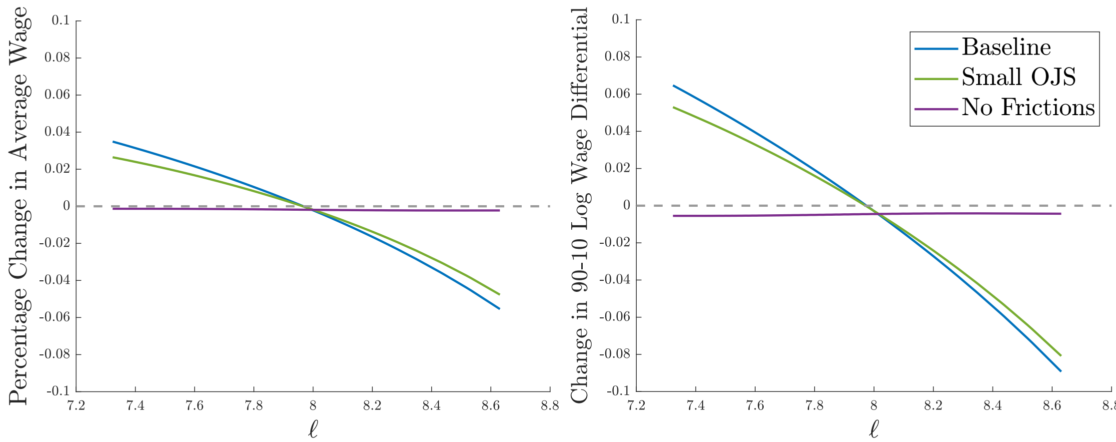
We next turn to the role of spatial hiring frictions in spatial inequality. Labor market integration leads to a decline in *across*-location inequality by 93%, which almost entirely eliminates the spatial wage premium. Our decomposition of the spatial wage premium (16) sheds light on the mechanism. Although local job ladders across regions become more similar, the main driver is shifts in employment composition that improve employment distributions in poor places and thereby dampen inequality (see Figure A.7, Appendix H.3). These composition shifts are due to the differential positioning of local firms on the global job ladder. In rich 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. The opposite is true for less prosperous labor markets. These composition shifts also substantially alter *within*-location inequality across space, which is now 6% *higher* in *low*-ℓ relative to high-ℓ locations.

To elicit the extent to which firm sorting *interacts* with spatial hiring frictions in shaping spatial inequality, we implement the random matching counterfactual in this setting of fully

[29]Interestingly, even though a reduction in OJS reduces inequality within all CZs, it *amplifies* spatial differences in within-location inequality. Under lower OJS, prosperous places lose some of their economic appeal and workers leave, which leads to a relative increase in job-finding rates and thereby within-location wage inequality in rich locations.

[30]A firm's poaching share is defined as the ratio of EE inflows relative to all worker inflows between two periods.

Figure 9: The Effect of Firm Sorting on Spatial Wage Inequality: Reduced OJS and No Spatial Frictions



Notes: The figure reports changes in average local wages and in the 90-10 difference of log wages within locations if firm sorting is abolished for three cases: (i) the baseline model (blue line), (ii) an economy with smaller OJS (i.e., the poaching share is 10% smaller than baseline, green line), and (iii) an economy with an integrated labor market (purple line).

integrated local labor markets. The flat purple lines in both panels of Figure 9 indicate that when hiring frictions are absent, eliminating firm sorting does *not* differentially affect wages across locations and fails to mitigate spatial inequality. We conclude that firm sorting matters for spatial inequality only if labor markets feature some degree of segmentation.

6.3 The Effect of Firm Sorting on East-West and Urban-Rural Gaps

A natural application of our model is to assess the extent to which spatial firm sorting is responsible for two well-known spatial wage gaps: the wage gap between East and West Germany and the gap between urban and rural areas. Almost three decades after the Berlin Wall fell, the West-East wage gap is 28%. In turn, the German urban-rural wage premium is 18%.

We show in Online Appendix OA-4 that spatial firm sorting can account for 17% of the West-East wage gap and for 19% of the urban-rural wage gap. Workers in the East and in rural Germany not only lose out because they work in places with poor economic fundamentals, but also because they do not have access to the most productive firms. The crucial channel through which more productive firms benefit workers in the West and in cities is by steepening the job ladders and rendering a "better" employment composition in these places.

7 Robustness

Our quantitative results are robust to a variety of different measurement and modeling choices. For all of these alternatives, we re-estimate our model and perform the *no-sorting counterfactual*, which allows us to compare the role of firm sorting in spatial inequality to our baseline model. We conclude that our baseline estimation provides a lower bound of the role of sorting in spatial inequality. Table 3 summarizes the results; Online Appendix OA-5 contains details.

Measurement. A first concern may be that different industries operate under different technologies with different labor and capital intensities, which renders firms heterogeneous not only in their productivity but also their labor intensity—a feature absent from 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 the local industry composition in the labor share data. As shown in the second column of Table 3, focusing on the within-industry variation, if anything, amplifies the role of sorting: Sorting now accounts for, respectively, 26% and 48% of inequality across and within locations.[31] 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. As shown in column 3 of Table 3, doing so strengthens the impact of firm sorting on spatial inequality, both between and within labor markets. The reason is that there is more regional variation in local labor shares *within* manufacturing. As a consequence, our estimation infers more spatial firm heterogeneity.

Further, note that our baseline estimation relies on regional statistics from the German Federal Statistical Office. Since 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 data. As a consequence, 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. As seen in column 4, this yields a more prominent role of firm sorting in spatial inequality.

Alternative Model Assumptions. We also probed the robustness of our results with respect to two model dimensions.

First, so far we have maintained Assumption 1, i.e., all firm types $y \in [\underline{y}, \overline{y}]$ are active in each market and the least productive firm makes zero profit. This assumption, ensured by adjusting the disutility from unemployment, ruled out spatial differences in firm selection at the lower end of local productivity distributions. In Online Appendix OA-5.5, we implement

[31] This is consistent with our regression analysis in Table 1, which revealed that the within-industry relationship between labor shares and log GDP per capita is steeper than the unconditional one.

a version of our model in which we first pin down the disutility of unemployment for all ℓ and then determine, for each labor market, endogenous productivity cutoff $\underline{y}(\ell)$, which implies that some firm types y with $\underline{y} \leq y < \underline{y}(\ell)$ choose to exit. In the estimated model, exit threshold $\underline{y}(\ell)$ turns out to be increasing in ℓ, which implies that firm sorting is stronger than in our baseline estimation. Thus, allowing for this form of firm selection amplifies the impact of sorting on inequality: Sorting now accounts for 22% of spatial wage gaps and 57% of differences in within-location inequality (column 5 of Table 3).

Second, our estimation was based on the simplifying assumption of full mobility of unemployed workers. Since imperfect mobility may be more accurate, we also estimate an extended model with location preference shocks that hamper worker mobility. We discipline workers' propensity to migrate by targeting a spatial labor supply elasticity that is consistent with the empirical literature; see Online Appendix OA-5.4. Column 6 shows that this does not change our quantitative conclusions about the role of sorting in within- and across-location inequality.

Table 3: Spatial Firm Sorting and Spatial Inequality: Robustness

	Baseline	Measurement Choices			Model Assumptions	
		Within Industry	Only Manufac.	FDZ Data	Selection	Imperfect Mobility
Spatial Wag Gap	16 %	26 %	22 %	34 %	22 %	16 %
90-10 Difference	38 %	48 %	52 %	66 %	57 %	39 %

Notes: This table reports the contribution of firm sorting to the spatial gap in average local wages (row 1) and the spatial gap of the 90-10 difference in log wages within local labor markets (row 2) for various measurement choices and specifications of our model.

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 by 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. At the center of our theory firms face the following 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. Importantly, if more productive firms sort into more productive locations, the local job ladder in these locations steepens and more employment is concentrated in top firms that pay high wages. This amplifies spatial wage inequality across and within locations.

Using administrative data from Germany, we estimate our model to assess the degree of firm sorting in the data and quantify its role for spatial inequality. We identify firm sorting from a novel fact: Local labor shares are lower in productive locations. This indicates that there is a higher concentration of top firms with strong monopsony power in productive places and thus positive sorting of firms across space. Quantitatively, firm sorting can account for 16%-34% of the spatial variation of mean wages and for 38%-66% of the variation in within-location wage dispersion. Workers in the least prosperous locations are not only harmed by poor economic fundamentals, but also because they lack access to productive firms.

References

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.

Bamford, I.: 2021, 'Monopsony Power, Spatial Equilibrium, and Minimum Wages'. Working Paper.

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.

Bowlus, A. J. and G. R. Neuman: 2006, 'Chapter 10 The Job Ladder'. In: H. Bunzel, B. J. Christensen, G. R. Neumann, and J.-M. Robin (eds.): *Structural Models of Wage and Employment Dynamics*, Vol. 275 of *Contributions to Economic Analysis*. Elsevier, pp. 217–235.

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.

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

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.

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.

Schmutz, B. and M. Sidibé: 2018, 'Frictional Labour Mobility'. *The Review of Economic Studies* **86**(4), 1779–1826.

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

Appendix

A Baseline Model: 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.[32] Using firm size expression (A.2) (Appendix A.2), we obtain (4).

A.2 Firm Size

As explained in Footnote 4, 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):

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

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 the measure of workers employed at firms of type y over the measure of firms of type y

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

[32] 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 (11) and (12)), 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 $\tilde{J}(\underline{y};\ell) = l(w(\underline{y},\ell))(z(\underline{y},A(\ell)) - w^R(\ell))$.

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

A.4 Land Price Schedule

Using integration by parts and Assumption 1.2 (i.e., zero profits of firm type $\underline{y}$ in all ℓ, implying $\tilde{J}(\underline{y},\ell) = 0$) problem (6) 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.

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.5 Land Market Clearing

We can derive market clearing under pure matching, $R(\ell) = Q(\mu(\ell))$, from general land market clearing condition (9),

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

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

B.1 Proof of Proposition 1

We use $\tilde{J}(y,\ell)$ from (4), wage function (5) (and integration by parts), to obtain $\bar{J}(p,\ell)$ as:

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

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):

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

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.5. Further, based on our assumption of strictly positive densities (r,q), this assignment is one-to-one: μ is strictly increasing, whereby 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 10). Hence, $\frac{\partial \Gamma_\ell(y)}{\partial \ell} = \frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p} \frac{r(\ell)}{q(\mu(\ell))}$ 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 the productivity gains, ε^P, are sufficiently large or if φ^E is sufficiently small. □

B.2 Proof of Proposition 2

We want to show that a fixed point in Γ_ℓ exists and we will do so by construction. Suppose the conditions of Proposition 1 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.4.

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 (6), i.e.,

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

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 1. 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 $\hat{\ell} \in [\ell',\ell]$. Moreover, it 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). $\square$

B.3 Proof of Proposition 3

Part (i). We analyze the cross-partial derivative of wage function (5):

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

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 ensuring this is given by

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

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, which is positive, is (sufficiently) large. 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 1). Therefore, (A.7) is sufficient for w to be supermodular in (y,ℓ), which we can further unpack as:

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

Using the multiplicative production function $z = yA$ and $\Gamma_\ell(y) = \Gamma(y|Q^{-1}(R(\ell)))$ with $\mu(\ell) = Q^{-1}(R(\ell))$, this becomes:

$$\min_{\ell,y} \frac{\frac{\partial A(\ell)}{\partial \ell}}{A(\ell)} \geq \left(1 + \frac{\lambda^E}{\delta}\right)^2 \left(\max_{y,\ell} \left(-\frac{\frac{\partial^2 \Gamma(y|Q^{-1}(R(\ell)))}{\partial y \partial p} \mu'(\ell)}{\gamma(y|Q^{-1}(R(\ell)))} \right) + 2\frac{\lambda^E}{\delta} \max_{y,\ell} \left(\frac{\int_{\underline{y}}^{y} \left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell}\right) dt}{y - \underline{y}} \right) \right). \quad \text{(A.9)}$$

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 since $\gamma > 0$ everywhere. In turn, regarding the second term, we need to rule out that it is infinite at $y = \underline{y}$. We use L'Hospital's rule, and obtain

$$\lim_{y \to \underline{y}} \frac{\int_{\underline{y}}^{y} \left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell}\right) dt}{y - \underline{y}} = \lim_{y \to \underline{y}} \frac{\left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell}\right)}{1} = -\frac{\partial \Gamma_\ell(\underline{y})}{\partial \ell} = 0.$$

Thus, w is supermodular in (y,ℓ) if (A.9) holds, i.e., if the elasticity of $A(\ell)$ wrt ℓ on the LHS of (A.9) is sufficiently large. Finally note that when (A.9) holds, the sufficient condition for PAM also holds (as stated in the premise of the proposition), which was given by:

$$\int_{\underline{y}}^{y} \frac{\frac{\partial^2 z(t,A(\ell))}{\partial \ell \partial y}}{(1 + \frac{\lambda^E}{\delta}(1 - \Gamma_\ell(t)))^2} - \frac{\frac{\partial z(t,A(\ell))}{\partial y} 2\frac{\lambda^E}{\delta}\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right)}{(1 + \frac{\lambda^E}{\delta}(1 - \Gamma_\ell(t)))^3} dt > 0$$

This holds if:

$$\int_{\underline{y}}^{y} \frac{\frac{\partial^2 z(t,A(\ell))}{\partial \ell \partial y}}{(1 + \frac{\lambda^E}{\delta})^2} - \frac{\partial z(t,A(\ell))}{\partial y} 2\frac{\lambda^E}{\delta} \left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right) dt > 0,$$

and under the assumed functional form for z, if

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

which is implied by (A.9) since the first term on the RHS of (A.9) is positive.

Part (ii). We show that under the conditions of the proposition, (16) is positive, due to all of its three components being positive. First, that the intercept is higher in high-ℓ locations follows directly from the assumption that A is strictly increasing in ℓ. Next, that the wage function is steeper in high-ℓ locations under the stated conditions (steep enough A which also guarantees PAM), $\partial^2 w/\partial y \partial \ell > 0$ follows from Part (i). Finally, due to positive sorting, $\partial \Gamma_\ell(y)/\partial \ell \leq 0$, which translates into first-order stochastic dominance of G_ℓ in ℓ, see (12). $\square$

Remark. We note that the conditions of the proposition are also sufficient for the result that wage schedule $w(\mathcal{R},\ell)$ is steeper in high-ℓ places, where $\mathcal{R} := \Gamma_\ell(y)$ is the productivity rank of firm type y. To see this, consider a firm in ℓ with rank $\mathcal{R} \in [0,1]$. Its wage is given by:

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

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

where we used a change of variable and $\gamma_\ell(\Gamma_\ell^{-1}(x))$ is the pdf at the xth quantile. Differentiate wrt $\mathcal{R}$:

$$\begin{aligned}\frac{\partial w(\mathcal{R},\ell)}{\partial \mathcal{R}} &= \frac{\partial z(\Gamma_\ell^{-1}(\mathcal{R}),A(\ell))}{\partial y}\frac{\partial \Gamma_\ell^{-1}(\mathcal{R})}{\partial \mathcal{R}} + 2\lambda^{\mathbb{E}}[\delta+\lambda^E(1-\mathcal{R})]\int_0^{\mathcal{R}} \frac{\frac{\partial z(A(\ell),\Gamma_\ell^{-1}(x))}{\partial y}}{[\delta+\lambda^E(1-x)]^2}\frac{1}{\gamma_\ell(\Gamma_\ell^{-1}(x))}dx - \frac{\frac{\partial z(\Gamma_\ell^{-1}(\mathcal{R})),A(\ell)}{\partial y}}{\gamma_\ell(\Gamma_\ell^{-1}(\mathcal{R}))}\\ &= 2\lambda^E[\delta+\lambda^E(1-\mathcal{R})]\int_0^{\mathcal{R}} \frac{\frac{\partial z(\Gamma_\ell^{-1}(x)),A(\ell)}{\partial y}}{[\delta+\lambda^E(1-x)]^2}\frac{1}{\gamma_\ell(\Gamma_\ell^{-1}(x))}dx\\ &= 2\lambda^E[\delta+\lambda^E(1-\mathcal{R})]\int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^2}dt\end{aligned}$$

Differentiate once more wrt ℓ:

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

which is positive since (i) $\Gamma_\ell^{-1}(\mathcal{R})$ is increasing in ℓ (given that under PAM $\frac{\partial}{\partial \ell}\Gamma_\ell \leq 0$); and (ii) under the sufficient conditions for PAM in Proposition 1, $\frac{\partial z(y,A(\ell))}{\partial y}/(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y)))^2$, is also increasing in ℓ.

B.4 Proof of Proposition 4

With multiplicative z, the wage range based on (15) simplifies to:

$$\frac{w(\overline{y},\ell)}{w(\underline{y},\ell)} = \frac{\overline{y} - \int_{\underline{y}}^{\overline{y}} \frac{l(t,\ell)}{l(\overline{y},\ell)} dt}{\underline{y}}.$$

This is increasing in ℓ if $\frac{l(\overline{y},\ell)}{l(t,\ell)}$ is increasing in ℓ for all $t \in [\underline{y}, \overline{y}]$, where

$$\frac{\partial \frac{l(\overline{y},\ell)}{l(t,\ell)}}{\partial \ell} = 2(1+\varphi^E(1-\Gamma_\ell(t)))\varphi^E\left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell}\right) > 0.$$

This expression is positive if there is positive firm sorting, $-\frac{\partial \Gamma_\ell(y)}{\partial \ell} = -\frac{\partial \Gamma(y|p)}{\partial p}\mu'(\ell) > 0$. □

B.5 Proof of Proposition 5

Preliminaries. 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

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

Thus, for the local labor share to be decreasing in ℓ, the following must hold (after applying integration by parts):

$$\frac{\partial LS(\ell)}{\partial \ell} = \frac{\partial Ls(\underline{y},\ell)}{\partial \ell} + \int_{\underline{y}}^{\overline{y}} \frac{\partial \frac{\partial Ls(y,\ell)}{\partial y}}{\partial \ell}(1-\tilde{G}_\ell(y)) + \frac{\partial Ls(y,\ell)}{\partial y}\left(-\frac{\partial \tilde{G}_\ell(y)}{\partial \ell}\right) dy < 0, \quad \text{(A.10)}$$

which coincides with (18) since the first term is independent of ℓ.

The proof proceeds in three steps. First, we show that for $\varphi^E \to 0$, the firm-level labor share is modular in (ℓ, y), which renders the second term in (A.10) zero. Second, we show that $\varphi^E \to 0$ guarantees that the firm-level labor share is downward sloping. Third, we show that PAM and $\varphi^E \to 0$ imply that $\tilde{G}_\ell$ is shifted by ℓ in the FOSD sense. Steps 2 and 3 ensure that the third term in (A.10) is negative. Thus, under PAM and large enough labor market frictions, i.e., if φ^E is small enough, $LS(\ell)$ is decreasing in ℓ. We now provide the details.

Step 1. We focus on the second term in (A.10) and show that $Ls(y, \ell)$ is modular in (y, ℓ) under the stated conditions. First, we spell out the condition we have to sign:

$$\begin{aligned}\frac{\partial^2 Ls(y,\ell)}{\partial y \partial \ell} &\overset{\text{S}}{=} z^2\left(\frac{\partial^2 w}{\partial y \partial \ell} z + \frac{\partial w}{\partial y}\frac{\partial z}{\partial \ell} - \frac{\partial w}{\partial \ell}\frac{\partial z}{\partial y} - w\frac{\partial^2 z}{\partial y \partial \ell}\right) - 2z\frac{\partial z}{\partial \ell}\left(\frac{\partial w}{\partial y} z - w\frac{\partial z}{\partial y}\right) \\ &= -zw\left(\frac{\partial^2 z}{\partial y \partial \ell} z - 2\frac{\partial z}{\partial y}\frac{\partial z}{\partial \ell}\right) + \frac{\partial^2 w}{\partial y \partial \ell} z^3 - z^2\frac{\partial w}{\partial y}\frac{\partial z}{\partial \ell} - z^2\frac{\partial w}{\partial \ell}\frac{\partial z}{\partial y}\end{aligned}$$

Note that for large labor market frictions ($\varphi^E \to 0$):

$$\begin{aligned}\left.\frac{\partial^2 w}{\partial y \partial \ell}\right|_{\varphi^E=0} &= 0 \\ \left.\frac{\partial w}{\partial y}\right|_{\varphi^E=0} &= 0 \\ \left.\frac{\partial w}{\partial \ell}\right|_{\varphi^E=0} &= \frac{\partial z(\underline{y}, A(\ell))}{\partial \ell} \\ \left. w\right|_{\varphi^E=0} &= z(\underline{y}, A(\ell)).\end{aligned}$$

Hence, under the multiplicative production function $z = yA$, we obtain[33]

$$\left.\frac{\partial^2 Ls(y,\ell)}{\partial y \partial \ell}\right|_{\varphi^E=0} = -z(y, A(\ell))z(\underline{y}, A(\ell))\left(\frac{\partial^2 z}{\partial y \partial \ell} z - 2\frac{\partial z}{\partial y}\frac{\partial z}{\partial \ell}\right) - z^2\frac{\partial z(\underline{y}, A(\ell))}{\partial \ell}\frac{\partial z(y, A(\ell))}{\partial y} = 0.$$

Step 2. As for the third term in (A.10), we show that under the premise, firm-level labor share,

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

is *decreasing* in firm productivity y in each location ℓ. Differentiation and some algebra yield:

$$\frac{\partial Ls(y,\ell)}{\partial y} = (1 - Ls(y,\ell))\frac{2\varphi^E \gamma_\ell(y)}{1 + \varphi^E(1-\Gamma_\ell(y))} - Ls(y,\ell)\frac{\frac{\partial z(y, A(\ell))}{\partial y}}{z(y, A(\ell))}. \tag{A.11}$$

[33]The proof can be easily adjusted to a general z that is sufficiently complementary in (y, A), so that $\left.\frac{\partial^2 Ls(y,\ell)}{\partial y \partial \ell}\right|_{\varphi^E=0} < 0$.

We will show that this expression is negative for sufficiently small φ^E. It suffices that

$$(1 - Ls(y,\ell))\frac{2\varphi^E\gamma_\ell(y)}{1+\varphi^E(1-\Gamma_\ell(y))} < Ls(y,\ell)\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))},\ \ \forall(y,\ell).$$

Then, it suffices that

$$2\varphi^E\gamma_\ell(y) < \left(1 - \frac{\left[1+\varphi^E(1-\Gamma_\ell(y))\right]^2 \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{[1+\varphi^E(1-\Gamma_\ell(t))]^2}dt}{z(y,A(\ell))}\right)\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))}.$$

or that:

$$2\varphi^E\gamma_\ell(y) < \left(1 - \frac{\left[1+\varphi^E\right]^2 (z(y,A(\ell)) - z(\underline{y},A(\ell))}{z(y,A(\ell))}\right)\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))}.$$

For (A.11) to be negative it is therefore sufficient that

$$(1+\varphi^E)^2\min_{\ell,y}\left(\frac{z(\underline{y},A(\ell)}{z(y,A(\ell))}\right) > 2\varphi^E\max_{\ell,y}\left(\gamma_\ell(y)\frac{z(y,A(\ell))}{\frac{\partial z(y,A(\ell))}{\partial y}}\right) + 2\varphi^E + (\varphi^E)^2 \tag{A.12}$$

i.e., it suffices that φ^E is small enough: The LHS is strictly positive for $\varphi^E = 0$ and increasing in φ^E. The RHS is zero for $\varphi^E = 0$, increasing in φ^E, and its slope in φ^E is steeper than that of the LHS. Then, by the Intermediate Value Theorem, there is an intersection of the two such that the RHS crosses the LHS from below. It follows that there exists a point $\tilde{\varphi}^E$ such that for $\varphi^E < \tilde{\varphi}^E$, (A.12) holds and so $\partial Ls(y,\ell)/\partial y < 0$.

Step 3. We now show that sufficiently large labor market frictions and PAM guarantee that ℓ shifts $\tilde{G}_\ell$ in a FOSD sense.

First note that for any $\ell' < \ell''$ and $y < \overline{y}$, $\tilde{G}_{\ell''} \leq \tilde{G}_{\ell'}$ iff

$$\frac{\int_{\underline{y}}^{y} g_{\ell''}(\tilde{y})z(\tilde{y},A(\ell''))d\tilde{y}}{\int_{\underline{y}}^{\overline{y}} g_{\ell''}(\tilde{y})z(\tilde{y},A(\ell''))d\tilde{y}} \leq \frac{\int_{\underline{y}}^{y} g_{\ell'}(\tilde{y})z(\tilde{y},A(\ell'))d\tilde{y}}{\int_{\underline{y}}^{\overline{y}} g_{\ell'}(\tilde{y})z(\tilde{y},A(\ell'))d\tilde{y}}$$

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

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. The latter is satisfied under multiplicative z. Regarding the former, note that

$$g_\ell(y) = \frac{\gamma_\ell(y)(1+\varphi^E)}{(1+\varphi^E(1-\Gamma_\ell(y)))^2}.$$

Cross-differentiating $\log(g_\ell)$ shows that g_ℓ is log-supermodular if for all (y,ℓ)

$$\frac{\partial^2 \log\gamma_\ell(y)}{\partial y\partial\ell} \geq -\frac{2\varphi^E}{(1+\varphi^E(1-\Gamma_\ell(y)))^2}\left(\frac{\partial^2\Gamma_\ell(y)}{\partial\ell\partial y}(1+\varphi^E(1-\Gamma_\ell(y)))+\varphi^E\frac{\partial\Gamma_\ell(y)}{\partial y}\frac{\partial\Gamma_\ell(y)}{\partial\ell}\right)$$

$$\frac{\partial^2 \log\gamma_\ell(y)}{\partial y\partial\ell} \geq -\left(\frac{2\varphi^E}{(1+\varphi^E(1-\Gamma_\ell(y)))}\frac{\partial^2\Gamma_\ell(y)}{\partial\ell\partial y}+\frac{2(\varphi^E)^2}{(1+\varphi^E(1-\Gamma_\ell(y)))^2}\frac{\partial\Gamma_\ell(y)}{\partial y}\frac{\partial\Gamma_\ell(y)}{\partial\ell}\right)$$

$$\frac{\partial^2 \log\gamma(y|p)}{\partial y\partial p}\mu'(\ell) \geq -\left(\frac{2\varphi^E}{(1+\varphi^E(1-\Gamma_\ell(y)))}\frac{\partial^2\Gamma(y|p)}{\partial p\partial y}\mu'(\ell)+\frac{2(\varphi^E)^2}{(1+\varphi^E(1-\Gamma_\ell(y)))^2}\frac{\partial\Gamma(y|p)}{\partial y}\frac{\partial\Gamma(y|p)}{\partial p}\mu'(\ell)\right)$$

$$\frac{\partial^2 \log\gamma(y|p)}{\partial y\partial p} \geq -\left(\frac{2\varphi^E}{(1+\varphi^E(1-\Gamma_\ell(y)))}\frac{\partial^2\Gamma(y|p)}{\partial p\partial y}+\frac{2(\varphi^E)^2}{(1+\varphi^E(1-\Gamma_\ell(y)))^2}\frac{\partial\Gamma(y|p)}{\partial y}\frac{\partial\Gamma(y|p)}{\partial p}\right),$$

where the last inequality follows under PAM, $\mu'(\ell)>0$. We want the following to hold for all (y,p):

$$\frac{\partial^2 \log\gamma(y|p)}{\partial y\partial p} \geq 2\varphi^E\max_{y,p}\left(-\frac{\partial^2\Gamma(y|p)}{\partial p\partial y}\right)+2(\varphi^E)^2\max_{y,p}\left(\frac{\partial\Gamma(y|p)}{\partial y}\left(-\frac{\partial\Gamma(y|p)}{\partial p}\right)\right), \tag{A.13}$$

where the maxima on the RHS are positive and well-defined as we maximize continuous functions over compact sets. The LHS is strictly positive for all (y,p) under our assumption that $\gamma(y|p)$ satisfies the strict monotone likelihood ratio property (which is equivalent to $\gamma(y|p)$ being log-supermodular in (y,p)). For (A.13) to hold, it therefore suffices that φ^E is small enough: The LHS is constant in φ^E and strictly positive; and the RHS is increasing in φ^E, starting at 0 and ending at ∞. Then, by the Intermediate Value Theorem, there is an intersection of the two and so there exists a point $\widehat{\varphi}^E$ such that for $\varphi^E<\widehat{\varphi}^E$, (A.13) holds strictly. It follows that g_ℓ is log-supermodular, implying that ℓ shifts $\tilde{G}_\ell$ in the FOSD sense, i.e., $\partial\tilde{G}_\ell/\partial\ell\leq 0$, provided that φ^E is small enough. As a result of Step 2 and Step 3, the third term of (A.10) is negative for sufficiently small φ^E.

We have shown that the integrand of (A.10) is strictly negative for $\varphi^E\to 0$, and so by continuity of LS in φ^E, there exists a neighborhood $(0,\overline{\varphi}^E)$ such that for $\varphi^E<\overline{\varphi}^E$, (A.10) is negative. $\square$

B.6 Proof of Corollary 1

Under goods' market clearing, aggregate output in ℓ equals aggregate wages plus profits and land prices in equilibrium

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

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

where we used integration by parts in 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 in the denominator can be expressed as follows, using that Γ is Pareto and firm size expression (3):

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

Plugging aggregate output back into $LS(\ell)$ above, we obtain $LS(\ell) = 1-\mu(\ell)$. Thus, $LS'(\ell) < 0$ if and only if $\mu'(\ell) > 0$. □

B.7 Proposition and Proof: Firm Sorting & Local Productivity Dispersion

Proposition 7 (Firm Sorting & Local Productivity Dispersion)**.** *If there is positive firm sorting across space, then the quantile ratio of local productivity, $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$, is increasing in ℓ, provided that the elasticity of $(-\Gamma_p/\Gamma_y)$ with respect to y exceeds 1.*

Proof. We provide conditions under which the quantile ratio of the productivity distribution

$$\frac{\Gamma_\ell^{-1}(t'')}{\Gamma_\ell^{-1}(t')} = \frac{\Gamma^{-1}(t'',\mu(\ell))}{\Gamma^{-1}(t',\mu(\ell))}$$

is increasing in ℓ, where $\Gamma^{-1}(t,\mu(\ell))$ is the t-th quantile, $t \in (0,1)$, pertaining to productivity distribution $\Gamma(y|\mu(\ell))$. To simplify notation, we define $\Psi(t,\mu(\ell)) \equiv \Gamma^{-1}(t,\mu(\ell))$, and so

$$\frac{\Psi(t'',\mu(\ell))}{\Psi(t',\mu(\ell))} = \frac{\Gamma^{-1}(t'',\mu(\ell))}{\Gamma^{-1}(t',\mu(\ell))}.$$

We aim to show under which conditions this ratio is increasing in ℓ or, stated differently, conditions under which $\Psi(t,\mu(\ell))$ is log-supermodular in (t,ℓ):

$$\mu'(\ell)\big(\Psi_{tp}\Psi - \Psi_t\Psi_p\big) \geq 0.$$

This holds if:

$$
\begin{aligned}
& \mu'(\ell)\Bigg(\bigg(\frac{\Gamma_{yy}\Gamma_p - \Gamma_{py}\Gamma_y}{\Gamma_y^2}\bigg)\Psi + \bigg(\frac{\Gamma_p}{\Gamma_y}\bigg)\Bigg) && \geq 0 \\
\Leftrightarrow \quad & \frac{\Gamma_{yy}\Gamma_p - \Gamma_{py}\Gamma_y}{\Gamma_y^2} && \geq \bigg(- \frac{\Gamma_p}{\Gamma_y}\bigg)\frac{1}{y} \\
\Leftrightarrow \quad & \frac{\partial\big(- \Gamma_p/\Gamma_y\big)}{\partial y}\frac{y}{\big(- \Gamma_p/\Gamma_y\big)} && \geq 1
\end{aligned}
$$

where to go from the first to the second line, we use PAM, $\mu'(\ell) > 0$ and $\Psi = y$. □

Corollary 2 (Firm Sorting & Local Productivity Dispersion: Pareto Case)**.** *If and only if there is positive firm sorting across space, then both the quantile ratio of local productivity,* $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$*, and the quantile difference of the log value added distribution,* $\Pi_\ell^{-1}(t'') - \Pi_\ell^{-1}(t')$ *are increasing in* ℓ *(where we denote by* $\Pi_\ell(z)$ *the cdf of log value added* $log(z)$*).*

Proof. In Proposition 7, we saw that the quantile ratio of productivity, $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$, is increasing in ℓ if

$$\bigg(\frac{\Gamma_{yy}\Gamma_p - \Gamma_{py}\Gamma_y}{\Gamma_y^2}\bigg)\Psi + \bigg(\frac{\Gamma_p}{\Gamma_y}\bigg) \geq 0.$$

If $y \sim Pareto(1, 1/p)$, i.e., $\Gamma(y|p) = 1 - (1/y)^{1/p}$, then this expression becomes

$$\bigg(\frac{\Gamma_{yy}\Gamma_p - \Gamma_{py}\Gamma_y}{\Gamma_y^2}\bigg)y + \bigg(\frac{\Gamma_p}{\Gamma_y}\bigg) = \frac{y}{p} > 0.$$

And therefore

$$\mu'(\ell)\bigg(\frac{\Gamma_{yy}\Gamma_p - \Gamma_{py}\Gamma_y}{\Gamma_y^2}\bigg)y + \bigg(\frac{\Gamma_p}{\Gamma_y}\bigg) = \mu'(\ell)\frac{y}{p} > 0$$

if and only if $\mu'(\ell) > 0$, proving the claim.

Further, regarding the claim about log value added, first note that if y is Pareto distributed as specified then log(z) follows an exponential distribution. To see this note that

$$\log z(y, A(\ell)) = \log(A(\ell)) + \log y,$$

where $\log y \sim exp(1/p)$ due to the assumption that the location parameter in y's Pareto distribution equals 1. Then, conditional on ℓ, $A(\ell)$ is a constant and so

$$\begin{aligned}\Pi_\ell(\tilde{z}) \equiv \mathbb{P}[\log(z) \leq \tilde{z}] &= \mathbb{P}[\log y \leq \tilde{z} - \log(A(\ell))] \\ &= 1 - e^{-(\tilde{z}-\log(A(\ell)))\frac{1}{\mu(\ell)}}.\end{aligned}$$

Then, the t-th quantile of the log value added distribution is given by,

$$\Pi_\ell^{-1}(t) = \log(A(\ell)) - \mu(\ell)\log(1-t),$$

and the difference of two quantiles corresponding to $t'' > t'$ is given by:

$$\begin{aligned}\Pi_\ell^{-1}(t'') - \Pi_\ell^{-1}(t') &= \log(A(\ell)) - \mu(\ell)\log(1-t'') - (\log(A(\ell)) - \mu(\ell)\log(1-t')), \\ &= \mu(\ell)(\log(1-t') - \log(1-t'')).\end{aligned}$$

It follows that

$$\frac{\partial(\Pi_\ell^{-1}(t'') - \Pi_\ell^{-1}(t'))}{\partial \ell} > 0 \quad \Leftrightarrow \quad \mu'(\ell) > 0.$$

□

C Quantitative Model: Labor Mobility & Residential Housing

The firms' location choice problem has the same structure as in the baseline model, only that in (6) they now take into account that their 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, benefiting 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, again denote the minimum of the first term on the LHS of (13) (over ℓ, y) by ε^P, i.e.,

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

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 $\delta(\ell) = \delta$); and that local housing supply is given by $h(\ell) = d(\ell)^\xi$, where ξ is the housing supply elasticity.

Proposition 8. *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) housing supply elasticity ξ is sufficiently large, then there is positive sorting of firms p to locations ℓ.*

Proof of Proposition 8. The expected value for firm p from settling 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

$$\frac{\partial^2 \bar{J}(p,\ell)}{\partial \ell \partial p} = \frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p} \lambda^F(\ell) + \frac{\partial \hat{J}(p,\ell)}{\partial p} \frac{\partial \lambda^F(\ell)}{\partial \ell}. \tag{A.14}$$

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:

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

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

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

Plugging these derivatives into (A.14), we can write (A.14) as a single integral. Then, a sufficient condition for (A.14) 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:

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

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.15) holds for a large enough ε^P, made precise below. We proceed in 3 steps.

Step 1. We first show that the value of unemployment is increasing in ℓ for a *fixed* λ^U (and thus a fixed $\lambda^E = \kappa\lambda^U$) if housing supply elasticity ξ is sufficiently large. We now unpack this statement.

Recall the value of unemployment in this extension of the model:

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

Using the government budget constraint,

$$\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell),$$

the housing market clearing condition,

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

the local population size (for a derivation, see (A.20) in Appendix F)

$$L(\ell) = \mathcal{A}^2\frac{\delta(\ell)+\lambda^U(\ell)}{\delta(\ell)+\kappa\lambda^U(\ell)}\left(\frac{1}{\lambda^U(\ell)}\right)^2,$$

as well as the postulated housing supply function, we obtain the following housing price:

$$d(\ell) = \left(\frac{\omega}{1-\omega\tau}\mathbb{E}[w(y,\ell)](1-u(\ell))L(\ell)\right)^{1/(1+\xi)}.$$

Denote by

$$\tilde{d}(\ell) \equiv \frac{\omega}{1-\omega\tau}\mathbb{E}[w(y,\ell)](1-u(\ell))L(\ell).$$

As the wage is strictly increasing in firm productivity and thus in firm's local rank, we can express the value of unemployment 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:

$$\rho V^U(\ell) = \tilde{d}(\ell)^{-\frac{\omega}{1+\xi}}\left(z(\underline{y}, A(\ell)) + \lambda^E(\ell)\left[\int_0^1 \frac{1-\mathcal{R}}{\delta+\lambda^E(\ell)(1-\mathcal{R})}\frac{\partial w(\mathcal{R},\ell)}{\partial \mathcal{R}}d\mathcal{R}\right]\right). \tag{A.16}$$

We now differentiate value (A.16) wrt ℓ for a fixed $\lambda^U(\ell) = \lambda^U$ (and thus fixed $\lambda^E = \kappa\lambda^U$) and obtain

$$\begin{aligned}\left.\frac{\partial \rho V^U}{\partial \ell}\right|_{\lambda^U(\ell)=\lambda^U} &= \tilde{d}(\ell)^{-\frac{\omega}{1+\xi}} \times \left(\frac{\partial z}{\partial A}\frac{\partial A(\ell)}{\partial \ell} + \lambda^E\int_0^1 \frac{1-\mathcal{R}}{\delta+\lambda^E(1-\mathcal{R})}\frac{\partial^2 w(\mathcal{R},\ell)}{\partial\ell\partial\mathcal{R}}d\mathcal{R}\right) \\ &\quad - \frac{\omega}{1+\xi}\tilde{d}^{-\frac{\omega}{1+\xi}-1}\frac{\partial\tilde{d}(\ell)}{\partial\ell} \times \left(z(\underline{y}, A(\ell)) + \lambda^E\left[\int_0^1 \frac{1-\mathcal{R}}{\delta+\lambda^E(1-\mathcal{R})}\frac{\partial w(\mathcal{R},\ell)}{\partial\mathcal{R}}d\mathcal{R}\right]\right).\end{aligned} \tag{A.17}$$

We will show that the first line is positive while the second line is negative. However, for large enough

ξ, the second line becomes sufficiently small, rendering the overall expression positive.

To see that the first line of (A.17) is positive under the premise, 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 $(\mathcal{R}, \ell)$:

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

$$\frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \mathcal{R} \partial \ell} = 2\frac{\lambda^E}{\delta}\left(1 + \frac{\lambda^E}{\delta}(1-\mathcal{R})\right) \frac{\partial}{\partial \ell} \int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})} \frac{\frac{\partial z(t, A(\ell))}{\partial y}}{(1 + \frac{\lambda^E}{\delta}(1 - \Gamma_\ell(t)))^2} dt. \tag{A.18}$$

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.18), $\frac{\partial z(y, A(\ell))}{\partial y}/(1 + \frac{\lambda^E}{\delta}(1 - \Gamma_\ell(y)))^2$, is also increasing in ℓ. Both of these statements are true under the sufficient conditions for PAM that we provide below, so that the wage function is supermodular in $(\mathcal{R}, \ell)$. This ensures that the first line of (A.17) is positive.

In turn, to see that the second line of (A.17) is negative note that

$$\left.\frac{\partial \tilde{d}(\ell)}{\partial \ell}\right|_{\lambda^U(\ell) = \lambda^U} = \frac{\omega}{1 - \omega\tau}(1-u)L \left.\frac{\partial \mathbb{E}[w(y, \ell)]}{\partial \ell}\right|_{\lambda^U(\ell) = \lambda^U} > 0.$$

But if the housing supply elasticity is large $\xi \to \infty$, the second line vanishes since, for fixed λ^U and λ^E,

$$\lim_{\xi \to \infty}\left(-\frac{\omega}{1+\xi}\tilde{d}^{-\frac{\omega}{1+\xi}-1}\frac{\partial \tilde{d}(\ell)}{\partial \ell}\right) = 0 \times \frac{\partial \tilde{d}(\ell)}{\partial \ell} = 0.$$

Importantly, the limit of the first line is positive and given by

$$\begin{aligned}\lim_{\xi \to \infty} \tilde{d}(\ell)^{-\omega/(1+\xi)} &\left(\frac{\partial z}{\partial A}\frac{\partial A(\ell)}{\partial \ell} + \lambda^E \int_0^1 \frac{1-\mathcal{R}}{\delta + \lambda^E(1-\mathcal{R})} \frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \ell \partial \mathcal{R}} d\mathcal{R}\right) \\ &= \left(\frac{\partial z}{\partial A}\frac{\partial A(\ell)}{\partial \ell} + \lambda^E \int_0^1 \frac{1-\mathcal{R}}{\delta + \lambda^E(1-\mathcal{R})} \frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \ell \partial \mathcal{R}} d\mathcal{R}\right).\end{aligned}$$

Thus, by continuity of V^U in ξ, there exists a $\widehat{\xi}$ such that for $\xi > \widehat{\xi}$, the positive effect stemming from the first line of (A.17) dominates the negative effect stemming from the second line, which renders (A.17) positive. As a result, the value of unemployment is increasing in ℓ for a *fixed* λ^U (and thus a fixed $\lambda^E = \kappa\lambda^U$) if housing supply elasticity ξ is sufficiently large.

Step 2. A similar argument shows that for large enough ξ, V^U is increasing in λ^U since its positive effect on $\left(\frac{\partial z}{\partial A}\frac{\partial A(\ell)}{\partial \ell} + \lambda^E \int_0^1 \frac{1-\mathcal{R}}{\delta + \lambda^E(1-\mathcal{R})} \frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \ell \partial \mathcal{R}} d\mathcal{R}\right)$ dominates its (ambiguous) effect on $\tilde{d}(\ell)^{-\omega/(1+\xi)}$. Denote the level of the housing supply elasticity for which this is (weakly) true by $\tilde{\xi}$, and so for $\xi > \tilde{\xi}$, V^U is increasing in λ^U. Going forward we assume that $\xi > \max\{\widehat{\xi}, \tilde{\xi}\}$, consistent with our premise that the housing supply elasticity is "large enough".

Step 3. This discussion implies that for the equilibrium indifference condition of searching 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 while λ^F is increasing in ℓ. This renders the second and third term on the RHS in (A.15) negative.

For (A.15) to hold, it then suffices that the first (positive) term on the RHS is bounded and "dominated" by the LHS. Note that $\lambda^E(\cdot)$ is implicitly defined by (A.16), where, in equilibrium, V^U is a number that no longer depends on ℓ. If there is PAM, $\mu'(\ell) > 0$, all functions in this expression $(\tilde{d}, z, \partial w/\partial\mathcal{R})$ are continuous in ℓ (see (A.18)) where $\ell \in [\underline{\ell}, \overline{\ell}]$, and thus λ^E inherits this property.

The first term on the RHS of (A.15) is thus bounded from above. Recall that $-\frac{\partial\Gamma_\ell(y)}{\partial\ell} = -\frac{\partial\Gamma(y|\mu(\ell))}{\partial\ell} = -\frac{\partial\Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}$ and 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 so that $\varepsilon^P > 2\frac{1}{\delta}\tilde{t}^P$. This ensures that (i) inequality (A.15) holds; and thereby that (ii) $\Gamma_\ell^{-1}(\mathcal{R})$ is differentiable and increasing in ℓ and the integrand of (A.18) is increasing in ℓ, all 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* in any given CZ is calculated as the ratio of (annual) gross value added and total employment in a location, 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 combines a 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. In each CZ, 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 use worker- and firm-level data provided by the Research Data Centre (FDZ) of the German Federal Employment Agency at the Institute for Employment Research.

We use three datasets from FDZ: LIAB (Linked Employer-Employee Data), BHP (Establishment History Panel 7518), and EP (Establishment Panel).

The LIAB data links annual information on establishments with information on all individuals employed at those establishments.[34] Surveyed establishments (the 'panel cases') are followed between 2009-2016 and we observe individual-level information for *all* their employees. This individual-level

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

data, which includes workers' gender, education, full-time employment status, gross daily wages and work district, is assembled from official social security records. For more information, see Ruf et al. (2021a) and Ruf et al. (2021b).

We augment these datasets with the Establishment History Panel 7518 (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 and average wages, it contains total inflows and outflows of workers at the establishment level.

Moreover, linking the BHP to the LIAB, we observe basic information for *any* employer in workers' entire employment history between 1975 and 2018. While the original data is in a spell format, we transform it into a monthly panel.[35]

We complement these main data sources with information on firm-level sales and costs of inputs from the Establishment Panel (EP). The EP is a nationally representative survey of about 15,000 firms that reports standard balance-sheet information on sales, inputs and employment, as well as information on a variety of survey questions on topics related to employment policy.

Sample Restrictions. Our baseline sample pools the years 2010-2017. 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)).

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 German 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

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

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 workers' 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. We proxy the productivity type y of a firm with its sales per worker, residualized against year and two-digit industry fixed effects. To avoid that our results are driven by outliers, we winsorize the distribution of sales per worker at the top and bottom 1%. Data Source: Establishment Panel (EP).

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.[36] 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: Linked Employer-Employee Data (LIAB).

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

[36] 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.

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).

For the supplementary empirical analysis of firm sorting, Appendix E.2, we consider a different, more aggregate definition of local labor markets (38 NUTS2 regions, defined by the European Union), because the number of observations per CZ is too small.

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.

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 categorization 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 out of 257 commuting zones that have both urban and rural districts).

D.5 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	3,133	401	2,616	2,849	3,093	3,364	3,662
Average Value Added	4,640	687	3,903	4,202	4,523	4,872	5,518
Average Firm Size	11	2	9	10	11	12	13
Share Emp. Top 10%	0.56	0.06	0.49	0.52	0.55	0.59	0.63
Population Density	292	422	83	110	165	272	589
Population	317,149	420,183	92,979	127,139	190,745	325,078	596,006

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 = 5,958)							
Poaching Share	0.51	0.13	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	0.56	0.22	0.31	0.42	0.54	0.70	0.83
Commuting-zone level (N = 252)							
Poaching Share	0.49	0.05	0.42	0.46	0.48	0.52	0.54
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, 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 Empirical Analysis

E.1 Spatial Wage Inequality

Real Spatial Wage Inequality versus Nominal Spatial Wage Inequality. In our main analysis, we use—consistent with our theory—nominal wages to measure inequality both within and across locations. In Table A.3 we report average wage premia after deflating wages with local CPIs. Doing so decreases spatial inequality by about one third.

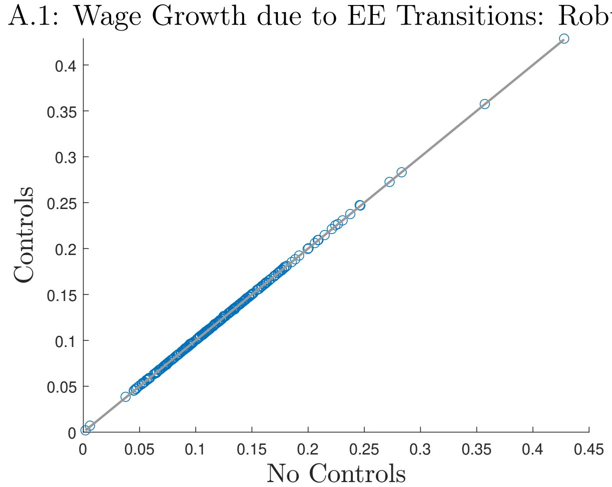
Table A.3: Spatial Inequality (Monthly, €): Real versus Nominal

	German CPI		Local CPI	
	Wage	Value Added	Wage	Value Added
Rich-Poor Inequality				
Rich	3784.6	5820.2	3899.6	5993.3
Poor	2755.2	4034.2	3135.2	4591.8
Rich/Poor	1.37	1.44	1.24	1.31
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). "Rich-Poor Inequality" refers to the comparison of the bottom and top quartile of CZs when grouped according to their GDP per capita.

Spatial Variation in EE Wage Growth: Robustness. We provide some robustness on regression (19), 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 β_ℓ^{EE} based on (19) (labeled 'No Controls') and, on the vertical axis, the demeaned estimated coefficients β_ℓ^{EE} 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 almost exactly along the 45 degree line, our results from Figure 2 (right panel) are not driven by the omission of these variables from the baseline regression.

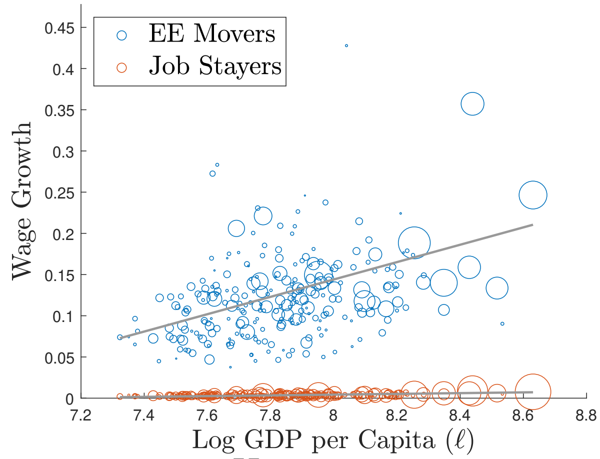
Figure A.1: Wage Growth due to EE Transitions: Robustness



Notes: Data source: LIAB. We plot, on the horizontal axis, the demeaned estimated coefficients β_ℓ^{EE} based on (19) (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').

Spatial Variation in Wage Growth: EE Movers versus Job Stayers. In Figure A.2, we compare the estimated wage growth of EE-transitions (β_ℓ^{EE} in regression (19)), shown in blue, with the average wage growth of stayers (β_ℓ in regression (19)), shown in orange. Figure A.2 shows that EE-return are sizable and that their spatial variation is much larger.

Figure A.2: Wage Growth of Job Movers and Job Stayers



Notes: Data source: LIAB. We plot coefficients β_ℓ^{EE} (denoted by "EE Movers") and β_ℓ (denoted by "Job Stayers") from regression (19) against ℓ. β_ℓ^{EE} is weighted by the number of EE moves within a commuting zone; β_ℓ is weighted by the number of stayers in each CZ.

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 again refer to the top and bottom 25% of 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 average starting wages of workers in rich and poor locations, before they climb 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 equals 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.4. If poor and rich regions had the same job ladder, spatial inequality in life-time income would be around 27% percent lower than under the heterogeneous job ladders that we see in the data, i.e., the rich region would be characterized by only 40.1% higher wages than the poor one, instead of the observed 54.8%.

Table A.4: 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 (Monthly, €)	3036.19	2301.67
Wage after 15 Years (Monthly, €)	5817.34	3758.63
After-15-years Spatial Wage Inequality (data)	1.548	
After-15-years Spatial Wage Inequality (counterfactual)	1.401	
Contribution of Job Ladder Differences to Spatial Wage Inequality	0.268	

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 27% $\sim (54.8 - 40.1)/54.8$.

E.2 Detecting Firm Sorting: Firm-Level Evidence

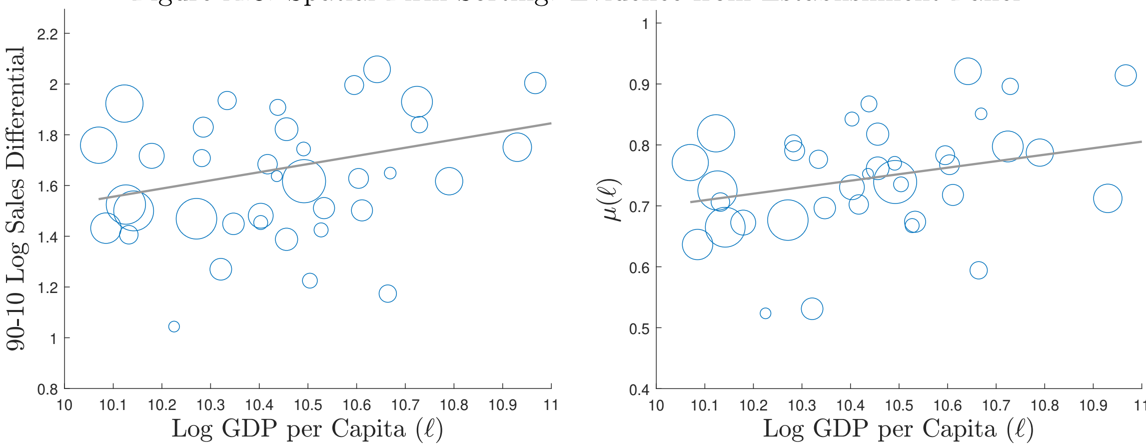
We provide additional evidence on positive firm sorting using firm-level productivity indicators. We consider this analysis as only supplementary to our evidence based on local labor shares since the firm-level productivity data is based on the Establishment Panel, which has a relatively small sample size (10,719 firms). This also implies that we cannot estimate the local productivity distribution at the level of 257 CZs but have to aggregate these data to the 38 NUTS2 regions. Moreover, because the Establishment Panel is a survey, the data is relatively noisy.

Within-Location Dispersion of Productivity and Sales. Corollary 2 (Appendix B.7) suggests a test of positive firm sorting based on how the local dispersion of (log) output per worker varies across space. If and only if sorting is positive, then high-ℓ locations are characterized by more dispersion in output per worker.

When assessing this prediction we measure output per worker at the firm level by sales per worker.[37] Figure A.3 (left) plots the difference of the 90% and 10% quantile of the distribution of log sales per worker. Based on Corollary 2, the positive relationship between sales dispersion and ℓ indicates positive firm sorting across space.

[37] We prefer sales per worker as our measure of z, because the data on intermediate inputs (and hence value added) are noisy. However, the results based on value added are very similar.

Figure A.3: Spatial Firm Sorting: Evidence from Establishment Panel



Notes: Data source: Establishment Panel. The left panel shows a scatter plot between the log difference of the 90th and 10th quantile of firm sales per worker against local log GDPpc. We compute the 90th and 10th percentiles using frequency weights, where we weigh each firm observation by the number of firms in the same size class (see footnote 38). In the right panel, we plot $-1/\beta_\ell$ from regression (A.19), where $z(A(\ell), y)$ in the dependent variable is measured as sales per worker. For each location ℓ, quantile k is taken from the local firm productivity distribution, where we use the same frequency weights as those in the left panel. Coefficient, β_ℓ, is weighted by the number of firms in each NUTS2 region. The size of the markers indicates the size of the region (number of firms in each NUTS2 region).

Pareto Tails of Firm Productivity. When assessing firm sorting based on the spatial variation in local labor shares or in the dispersion of sales per worker, we implicitly assume that firm productivity y in each ℓ follows a Pareto distribution with shape parameter $1/p$ (which in equilibrium becomes $1/\mu(\ell)$), see Corollaries 1 and 2. Positive sorting of firms across locations means that μ is increasing and thus richer locations have a thicker Pareto tail of the local productivity distribution. To assess this prediction, we proxy firm productivity by sales per worker and estimate the Pareto shape parameter at the NUTS2 regional level by implementing the following regression at the local level

$$\log(1 - \mathbb{P}[z(A(\ell), y) \leq k]) = \alpha_\ell + \beta_\ell \log(k) + \epsilon, \tag{A.19}$$

where $k = z^{(1)}, z^{(2)}, \cdots, z^{(n_\ell - 1)}, z^{(n_\ell)}$, and $(z^{(j)}, n_\ell)$ are the j-th order statistics and the number of firms in region ℓ, respectively.[38]

Under the assumptions of multiplicative technology and Pareto productivity distributions as well as the validity of our productivity proxy, regression coefficient β_ℓ captures the Pareto shape parameter $\frac{1}{\mu(\ell)}$. The R^2 of these regional Pareto regressions varies between 0.7 and 0.9, which suggests that the Pareto assumption is reasonable. Furthermore, the positive slope of the estimated $\mu(\ell)$ against (log) GDP per capita, as shown in the right panel of Figure A.3, is consistent with positive firm sorting across space.

In Online Appendix OA-3, we present additional evidence on spatial firm sorting. In particular,

[38]Note that the Establishment Panel samples firms based on firm size and industry across Germany. The sample is not representative at the regional level. To obtain a more representative empirical distribution of firm productivity, we weigh each observation with the local proportion of firms within the same size class, obtained from the German Federal Statistical Office, which provides the number of firms with fewer than 10, 10-50, 50-250, more than 250 employees at the district-year level.

we show that (globally) unproductive firms, that is firms with low y, are relatively more productive in their local labor market as they collocate with other, unproductive firms. By contrast, firms with a high y from an economy-wide point of view are relatively *un*productive in their local labor market, because the firms around them are, on average, also productive. In contrast to our "tests" of firm sorting discussed in this appendix, that insight does *not* rely on any parametric restrictions on local firm productivity distributions.

F Identification

We prove identification of our model under the following assumption:

Assumption 2. *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. *Land distribution* R *and housing expenditure share* ω *are observed.*
6. *Normalize the value of unemployment such that* $\rho V^U = 1$.

Proof of Proposition 6. 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 (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, $LS(\ell) = 1 - \mu(\ell)$, see Corollary 1.

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:

$$
\begin{aligned}
\lambda^E(\ell) &= \frac{M(\mathcal{V}(\ell),\mathcal{U}(\ell))}{(u(\ell)+\kappa(1-u(\ell)))L(\ell)}\kappa \\
&= M(\mathcal{V}(\ell),\mathcal{U}(\ell))\frac{\kappa(1-u(\ell))}{u(\ell)+\kappa(1-u(\ell))}\frac{1}{(1-u(\ell))L(\ell)} \\
&= \mathcal{A}\mathcal{V}(\ell)^{\frac{1}{2}}\big(u(\ell)+\kappa(1-u(\ell))L(\ell)\big)^{\frac{1}{2}}\frac{\lambda^E(\ell)}{\delta(\ell)+\lambda^E(\ell)}\frac{\delta(\ell)+\lambda^U(\ell)}{\lambda^U(\ell)}\frac{1}{L(\ell)} \\
\Rightarrow \quad L(\ell) &= \mathcal{A}^2\frac{\delta(\ell)+\lambda^U(\ell)}{\delta(\ell)+\kappa\lambda^U(\ell)}\left(\frac{1}{\lambda^U(\ell)}\right)^2. \qquad \text{(A.20)}
\end{aligned}
$$

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

$$L(\ell) = \left(1 + \frac{\delta(\ell)}{\lambda^U(\ell)}\right)\bar{l}(\ell). \tag{A.21}$$

Equalizing (A.20) and (A.21), 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, given $\mu(\ell)$, 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 $(u(\ell), L(\ell), d(\ell), \mathbb{E}[w(y, \ell)|\ell], \mathcal{T}, \omega)$ as observed for all ℓ (where $\mathcal{T}$ is the economy-wide replacement rate of the unemployed) and obtain $(\tau, h(\cdot), w^U(\cdot))$ from a system of three equations:

We use government budget constraint,

$$\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell), \tag{A.22}$$

housing market clearing,

$$h(\ell) = \omega\frac{w^U(\ell)}{d(\ell)}u(\ell)L(\ell) + \omega\frac{\mathbb{E}[w(y, \ell)|\ell]}{d(\ell)}(1 - u(\ell))L(\ell), \tag{A.23}$$

and an equation obtained from the definition of replacement rate $\mathcal{T}$

$$\tau = \frac{1}{\omega}\frac{\mathcal{T}}{\frac{\sum_{\hat{\ell}}(1-u(\hat{\ell}))L(\hat{\ell})}{\sum_{\hat{\ell}} u(\hat{\ell})L(\hat{\ell})} + \mathcal{T}}. \tag{A.24}$$

We computed (A.24) as follows. The replacement rate $\mathcal{T}$ satisfies

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

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 (A.22) and the second one uses a combination of (A.22) and housing market clearing (A.23), 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 yields (A.24).

Equation (A.24) pins down τ given the observed housing expenditure share and replacement rate

$(\omega, \mathcal{T})$. For each location ℓ and given τ, equations (A.23)–(A.22) are then a system of two equations and two unknowns $(w^U(\ell), h(\ell))$, which can be solved uniquely.

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

$$\rho V^U(\ell) = d(\ell)^{-\omega} B(\ell) w^U(\ell) + \tilde{b}(\ell) + d(\ell)^{-\omega} B(\ell) \lambda^U(\ell) \left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right], \quad \text{(A.25)}$$

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 holding a job with the value of unemployment:

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

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

$$\tilde{b}(\ell) = d(\ell)^{-\omega} B(\ell) \left(z(\underline{y}, A(\ell)) - w^U(\ell) - (\lambda^U(\ell) - \lambda^E(\ell)) \left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right] \right). \quad \text{(A.26)}$$

Plug $\tilde{b}(\ell)$ back into V^U in (A.25), and use a change of variable (to re-express F_ℓ using Γ_ℓ, where $\Gamma_\ell(y) = \Gamma(y|\mu(\ell))$) 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

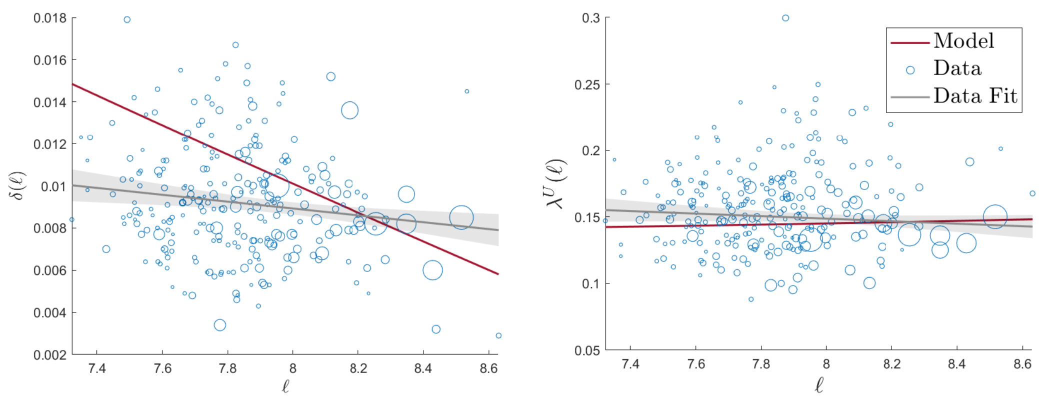
G.1 Estimation Results

Table A.5: 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)

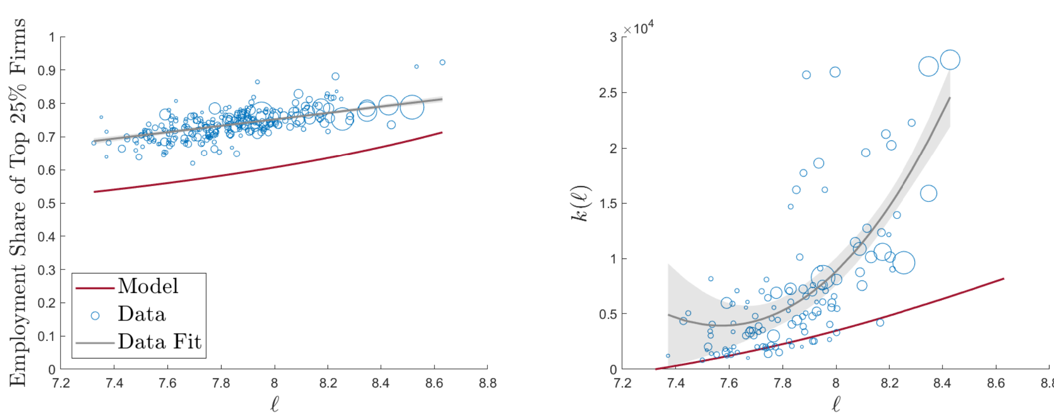
Notes: For data sources and variable definitions, see Appendix D. To obtain the relative matching efficiency of employed workers, κ, we use Germany-wide job-finding rates λ^E based on monthly EE transition rates, counting a transition only for those workers who move to a new firm with higher wage.

Figure A.4: Over-Identification: δ and λ^U in Data and Model



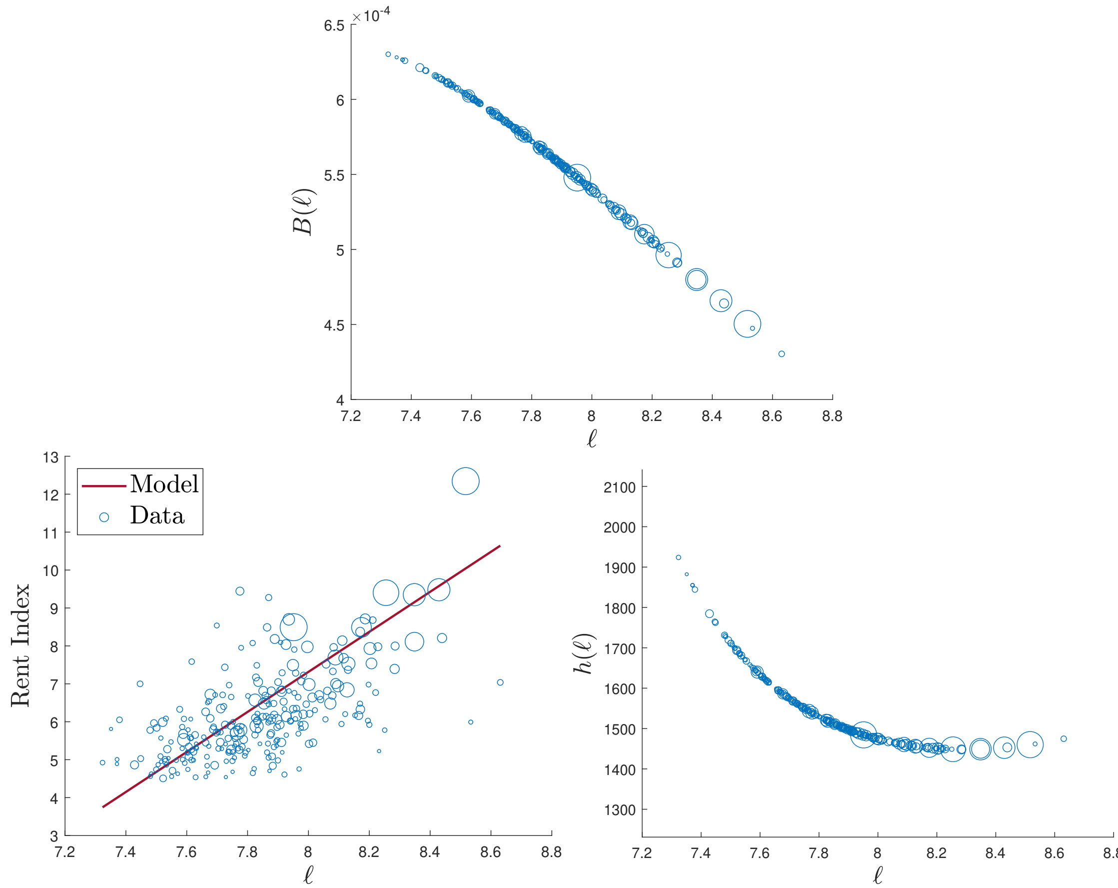
Notes: Data Sources: LIAB. For details on how these variables are constructed, see Appendix D.2. Observations are weighted by the number of establishments in each CZ, indicated by different sizes of the markers. 95% confidence intervals are displayed in gray.

Figure A.5: Model Fit: Additional Non-Targeted Moments



Notes: Data Sources: Left panel is based on BHP; 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 markers. 95% confidence intervals are displayed in gray.

Figure A.6: 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 markers.

Table A.6: Determinants of Local TFP $A(\ell)$

	(1)	(2)	(3)	(4)	(5)
Trade Tax	0.1298*				-0.0854
	(0.0600)				(0.0458)
Distance to Highway		-0.0536***			-0.0424***
		(0.0123)			(0.0093)
% of Employees with a College Degree			0.8765***		0.6466***
			(0.1423)		(0.1275)
Net Business Registration Intensity				0.0511***	0.0376***
				(0.0063)	(0.0064)
N	257	257	257	257	257
R^2	.0346	.153	.356	.274	.562

Notes: Data Sources: 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. The 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.

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.26). We also keep the estimated schedules (A, B, h) from the baseline model. But, 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 economy-wide productivity distribution of firms (no longer ℓ-specific). Note that compared to the 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 for a fixed point in ρV^U so that it satisfies both welfare equalization of workers and 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 lowering search efficiency of employed workers κ, we find meeting rate λ^U and housing price d given the estimated schedules (A, B, h) from the baseline model, so that the value of search for unemployed workers is equalized across space and local housing market clearing holds:

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

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

where the value of search is again calculated assuming $w^R(\ell) = A(\ell)\underline{y}$, supported by adjusting $\tilde{b}(\ell)$. We verify that the positive sorting of firms is optimal in this counterfactual equilibrium.

H.3 The Role of Spatial Hiring 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 given by (12) 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 $(A(\cdot), B(\cdot), h(\cdot))$ 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 $\tilde{l}(\ell) = (1-u)\overline{L}$ 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 at the level of the economy, and not location specific. Further, to equalize the value of search across all locations, given by

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

despite differences in local amenities $B(\ell)$ (as estimated from the baseline model), housing prices $d(\ell)$ 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 $d(\cdot)$ consistent with local housing market clearing, the local population size $L(\ell)$ adjusts so that housing market clearing holds.

Figure A.7 displays the components of decomposition (16) for the baseline model (solid) and the counterfactual without spatial hiring frictions (dashed).

Figure A.7: No-Spatial-Frictions Counterfactual: Wages and Employment Distribution