Firm Sorting and Spatial Inequality*

Ilse Lindenlaub
Yale University and NBER

Ryungha Oh
Yale University

Michael Peters
Yale University and NBER

October 2023

Abstract

We study the importance of spatial firm sorting for wage inequality both between and within local labor markets. We develop a novel model in which heterogeneous firms first choose a location and then hire workers in a frictional 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. Positive sorting—with productive firms settling in productive locations—emerges as the unique equilibrium if firm and location productivity are sufficient complements or labor market frictions are sufficiently large. Positive sorting steepens the job ladder in productive locations and, as a consequence, increases both their average wages and wage dispersion. We estimate our model using administrative data from Germany and identify firm sorting from a novel fact: Labor shares are lower in productive locations, which indicates a higher concentration of top firms with strong monopsony power. Positive firm sorting can account for at least 15% of the spatial variation in average wages and for 40% of the spatial variation in within-location wage dispersion.

Keywords: Firm Sorting, Inequality, Frictional Labor Markets, On-the-Job search, Monopsony Power, Labor Shares

*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 comments. We are grateful to Ariel Burstein, Hector Chade, Kyle Herkenhoff, Oleg Itskhoki, Rasmus Lentz, Guido Menzio, Simon Mongey and Aysegul Sahin for useful discussions and comments, 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. Recent literature highlights wage disparities not only across people but also across space. Indeed, local labor markets differ substantially in both their average wage and wage dispersion. Take Germany as an example: Wages in the most prosperous local labor markets are 40% higher than in the poorest ones. At the same time, local wage dispersion, measured by the 90-10 log wage gap, is 0.2 log points higher in affluent places. 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 wage inequality, complementing the explanations that rely on worker composition. 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 and find that the poorest local labor markets have significantly lower wages and less wage dispersion because they lack highly productive firms. Firm sorting can account for 15%-34% of the spatial variation in average wages and for 40%-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 basic 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, importantly, 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 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, 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 40% higher average wages and 20% 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 20% 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 evidence for it, we exploit another important 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.

Even though we do not model spatial worker heterogeneity and sorting explicitly, we recognize their importance for spatial wage differences. In our theory, we capture the variation in worker attributes across space through local TFP differences, which firms take into account when choosing their location. Under common functional form assumptions, however, local TFP—i.e., the local worker composition or other determinants of local productivity—does *not* impact several of our key outcomes that are shaped by firm sorting: the steepness of local job ladders, local EE returns, within-location wage dispersion and local labor shares. The documented spatial variation in these outcomes is therefore indicative of spatial firm sorting. In our empirical assessment of firm sorting and its implications, we nevertheless control for observed and unobserved worker heterogeneity and show that there is remaining spatial variation in our outcomes of interest, which we argue firm sorting can help explain.

To gauge the quantitative importance of firm sorting for spatial differences in local wage distributions, 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 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 15% 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 40% of the spatial gap in within-location wage dispersion. Firm sorting therefore plays an important role in spatial wage inequality in Germany. The presence of on-the-job search is crucial for this conclusion: When the rate of on-the-job search decreases by 10%, spatial firm sorting contributes 15% less to across-location wage inequality and 14% less to the spatial gap in within-location inequality.

Finally, to underscore the importance of *endogenous* firm sorting (as opposed to fixed spatial differences in firm productivity) we conduct a policy exercise that captures a reduction in place-based subsidies for a set of low-income regions in Germany. This policy leads to *re-sorting* of firms across locations with significant effects on spatial inequality.

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.

A key insight of the literature on labor search and frictional wage dispersion is that search frictions and, especially, on-the-job search are crucial drivers of wage inequality and allow for firm heterogeneity to impact wages (e.g., Diamond, 1971; Burdett and Mortensen, 1998; Bontemps et al., 2000; Postel-Vinay and Robin, 2002; Karahan et al., 2017, Bagger and Lentz, 2019). Two important findings are that, conditional on worker heterogeneity, firm heterogeneity accounts for a sizable share of the cross-sectional wage dispersion (∼15%-30%); and that search frictions and on-the-job search do as well (∼10%-40%).[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 feature 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; Behrens et al., 2014); 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), Heise and Porzio (2022) and Martellini (2022) analyze spatial wage gaps using a job ladder model and focus on *worker* sorting, mobility costs, and preference frictions or learning interactions as the main sources of inequality. In contrast to our work, these papers do not highlight how the *spatial sorting of firms* drives spatial inequality by *endogenously* shaping local job ladders.

A smaller but growing part of the urban literature analyzes firms' location choices but differs from ours in focus and modeling choices: We highlight that firm sorting affects wage inequality by

[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).

shaping local job ladders in frictional labor markets that feature on-the-job search. In contrast, Gaubert (2018) builds a frictionless model of spatial firm sorting 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 show that these are key determinants of spatial inequality. Our work complements these studies by drawing out the implications of firm sorting for wage inequality within and across locations.

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). Importantly, while we do not explicitly model worker heterogeneity across space, it is implicitly contained in local productivity $A(\ell)$—something we confirm in our estimated model—and thus contributes to a location's attractiveness. In this sense, more productive workers increase firms' average output and thus impact spatial firm sorting.

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.[2]

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 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 local population size and hence these meeting rates through endogenous labor mobility, 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 flow 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.

The model timing is such that firms first make a one-time location choice and then the continuous-time economy described above plays out in each local labor market.

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

[2]Alternatively, we could have obtained non-degenerate firm distributions in all ℓ by having firms with different p draw preference shocks over locations or by letting the matching process between p and ℓ be subject to search frictions.

[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.

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. While Assumption 1 gives us analytical tractability, we show that our quantitative results are robust to dropping it.

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 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), $V^E(\cdot, \ell)$ 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 a reservation wage strategy, pinning 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 w, a firm in location ℓ trades off profit per worker against firm size, which is given by (see Appendix SA.1.2)[4]

[4]Firm size can be derived as the hiring rate times the expected match duration or, as in Burdett and Mortensen (1998), as

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

Firms that are higher ranked in local wage distribution F_ℓ are larger since they poach more and are being poached less. 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, $\varphi^E \to 0$, 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 can be expressed as per-worker profit times size (see Appendix SA.1.1),[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, $l(w,\ell)$, but reduces flow profits, $z(y,A(\ell))-w$.

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 and denote firm size by $l(y,\ell) := l(w(y),\ell)$.

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; and that ℓ matters through both its effect on TFP $A(\ell)$ and firm size $l(y,\ell)$. Note that firms have monopsony power due to

the measure of workers employed at y divided by the measure of firms with y.

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

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 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—and hence monopsony power—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, $k(\ell)$:

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

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, for each firm of any type p the 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

[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.

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 SA.1.4). 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]

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 (where we cancel a multiplicative $r(\ell)$ on both sides),

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

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):

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

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

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:

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

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.

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. 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 (see Appendix SA.2.1 for the analysis of negative sorting).

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.[9]

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.[10]

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:

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

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

[9]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

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

[10]Below we check the case in which firms do not postulate PAM and show that it is not consistent with equilibrium.

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 (13) reduces to a comparison between local TFP and local competition.

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

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

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 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 any equilibrium features positive sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(R(\ell))$.*

The proof is in Appendix A.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.[11] 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 location. Ex ante, this discourages firms from collocating with productive peers, which prevents positive sorting.

Generalizations. We generalize our main Proposition 1 in various 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 in ℓ; i.e., $A(\ell) = \tilde{A}(\Gamma_\ell)$, where $\tilde{A}$ is increasing in the location's average firm productivity. Proposition SA2, Appendix SA.3.2, shows that, as in the baseline model, positive sorting arises if the (endogenous) location productivity advantage is large enough relative to the cost of competition. Second, we *endogenize vacancy creation*. In Proposition SA3, Appendix SA.3.3, we show that also in this case, the trade-off between productivity and competition determines firm sorting. Third, we *endogenize land distribution* R by allowing it to depend on land demand through price $k(\cdot)$—still the same outbidding and sorting logic as in the baseline model applies (Proposition SA4, Appendix SA.3.4). Finally, we can allow for *labor mobility* and *residential housing* (Proposition A1, Appendix B), endogenizing local population size, which will be our quantitative setting and further discussed below.

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.*

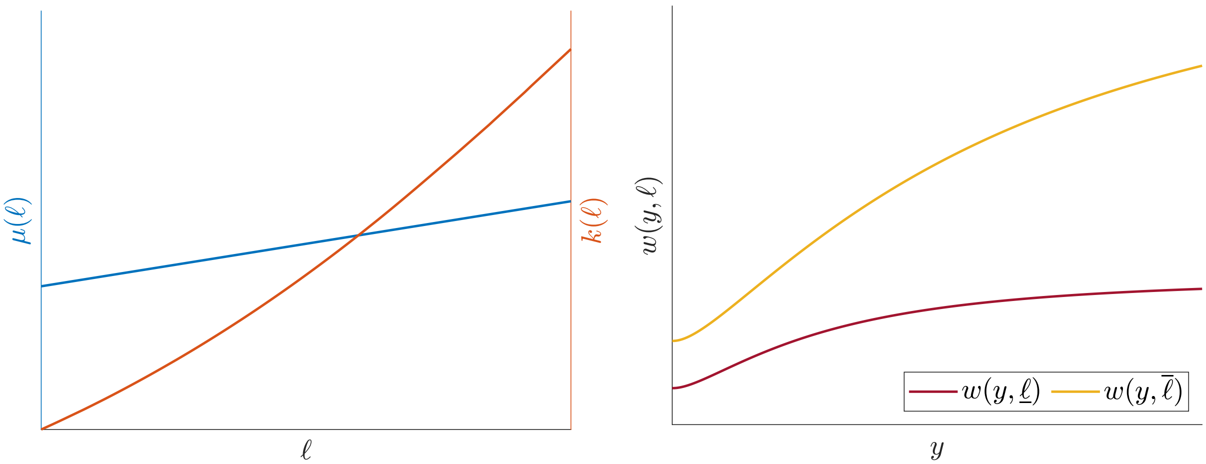
The proof is in Appendix A.2. We show the existence of a fixed point in Γ_ℓ by construction. In turn, uniqueness arises because, under the conditions on primitives stated in Proposition 1, the impact of endogenous firm composition on the firms' value function leaves the complementarity properties of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

[11] Even though frictions (small φ^E) facilitate positive sorting in our context, they alone are not sufficient: Some complementarities of z in (y, A) are needed, otherwise $\partial^2\overline{J}/\partial\ell\partial p$—see equation (A.2) in Appendix A.1—cannot be positive.

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 ($\ell = \bar{\ell}$, yellow) and a low-productivity ($\ell = \underline{\ell}$, red) location.

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, $\bar{\ell} > \underline{\ell}$. 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 wage inequality.

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 gaps in both average wages and wage dispersion.

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*, that is average wages of more productive locations relative to 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 ($\overset{\text{S}}{=}$) to

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

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, 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 that affects the degree of firm competition across different locations: Tougher poaching competition for workers among highly productive firms in high-ℓ locations pushes up the wage level at a faster rate. The third factor 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:*

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

(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.

Firm sorting not only affects the spatial gap in average wages but also in *within-location wage inequality*, captured by the spatial variation in the local *wage range*, $w(\overline{y},\ell)/w(\underline{y},\ell)$:

$$\frac{\partial \frac{w(\overline{y},\ell)}{w(\underline{y},\ell)}}{\partial \ell} \overset{\text{S}}{=} - \int_{\underline{y}}^{\overline{y}} \underbrace{\frac{\partial \frac{l(t,\ell)}{l(\overline{y},\ell)}}{\partial \ell}}_{\substack{\text{Spatial Variation in} \\ \text{Labor Market Competition}}} dt \;=\; \int_{\underline{y}}^{\overline{y}} \frac{2\varphi^E}{(1+\varphi^E(1-\Gamma_\ell(t)))^3} \left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell}\right) dt. \tag{17}$$

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. Note that 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. OJS therefore amplifies the effect of firm sorting on both competition (which affects the steepness of job ladders) and employment composition.

Spatial Firm Sorting and Local Labor Shares. 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, which are observable.

We demonstrate 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 A.5 for the derivation). The spatial gradient of the local labor share is then given by

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

where $\tilde{G}_\ell$ is the cdf corresponding to density $\tilde{g}_\ell$. The first force pushing toward declining local labor shares, $LS(\cdot)$, is that the *firm*-level labor share $Ls(\cdot, \ell)$ is decreasing in firm productivity y; and positive firm sorting triggers a composition effect that translates into stochastically better employment distributions $\tilde{G}_\ell$ in high-ℓ places. The intuition is straightforward. Under positive sorting, high-ℓ places tend to have a better firm *composition* and thus employment distributions that are tilted toward productive firms. If, additionally, more productive firms in any given ℓ face milder *competition* than less productive firms and thus have more monopsony power and lower labor shares $Ls(y, \ell)$, locations in which employment is concentrated in top firms tend to have a lower average labor share. A second force behind declining local labor shares in (18) can stem from a submodular firm-level labor share $Ls(y, \ell)$.[13]

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 $z(\cdot, A(\ell))$ increases faster in y than wage schedule $w(\cdot, \ell)$).

One way to circumvent the restriction on φ^E in Proposition 5 is to assume that the density of local firm productivity, $\gamma(\cdot|p)$, is sufficiently *decreasing*. In this case, more productive firms in any given ℓ face less labor market competition than low-productivity firms, which translates into more monopsony power and lower labor shares of top firms (despite the fact that they pay higher wages). Examples of commonly used firm distributions that have this property are the Pareto and log-normal distributions.[14] In our quantitative setting, we will assume $\Gamma(y|p) \sim Pareto(1, 1/p)$. Indeed, the local labor share then *only* depends on the Pareto tail coefficient, p, of the local productivity distribution (see Gouin-Bonenfant (2022) for a similar result under Pareto in a single-market economy). Under positive sorting, high-ℓ locations have a thicker tail and thus 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.*

[13] Submodularity of the firm-level labor share means that productive firms in high-ℓ locations have especially strong monopsony power and thus even lower labor shares, which would further lower the average labor share in their location. Note, however, that we are not relying on this second force in the proof of Proposition 5. Instead, we ensure that the first force is strong enough.

[14] In Appendix SA.8.1, we show that the assumption of (truncated) log-normal local firm productivity also renders a negative relationship between the local labor share and local firm productivity.

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.

In the Appendix, we supplement this result on how to detect firm sorting, using the spatial variation in local productivity dispersion (Proposition SA5 and Corollary SA1) or in the relationship between the local and the global productivity rank of firms (Proposition SA6).

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 Sections 5 and 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. For graphical illustrations, it is convenient to order locations by index ℓ—or, equivalently, by local TFP $A(\ell)$—which however are both unobserved. We therefore use local GDP per capita as a proxy for ℓ, which, albeit endogenous in our model, is increasing in $A(\ell)$. In our quantitative analysis, we draw on our estimated model to exactly decompose local GDP into A and endogenous firm sorting.

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) a worker-level panel 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.[15] 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.[16] 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 (using an average).

In Appendix C, we describe these data sources in more detail and define important variables. In Appendix SA.5, we report some basic statistics, which show that average wages, firm size, and value added differ substantially across commuting zones (see Table SA.2).

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

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

4.2 Spatial Firm Sorting and Spatial Wage Inequality: Evidence

In this section, we provide evidence that more productive firms in Germany sort into more productive locations. Obtaining direct evidence on firm sorting is challenging because our dataset—like many others—does not contain a clean measure of physical firm productivity and does not allow us to identify firm movers.[17] And deducing firm sorting indirectly from spatial variation in commonly observed outcomes, such as spatial inequality in wages or value added, is also difficult because these statistics conflate firm sorting with local TFP differences.

To provide evidence for spatial firm sorting, we therefore focus on various measures that, according to our theory, are only impacted by firm sorting. We proceed in two steps. First, to disentangle local firm productivity from location TFP, we use variation in local labor shares. Second, we show that our model's theoretical implications of firm sorting for spatial wage inequality are borne out empirically: Firm sorting provides a *unified* explanation for why prosperous places have not only higher wages, but especially also more within-region inequality, steeper job ladders and—as a natural implication—higher EE returns.

Local Labor Shares. Our model highlights that local labor shares identify the extent of spatial firm sorting (Proposition 5 and Corollary 1). As stated in Corollary 1, if firm productivity is Pareto distributed and technology multiplicative, local labor shares are pinned down by the tail parameter of the local productivity distribution, $LS(\ell) = 1 - \mu(\ell)$.[18] Specifically, there is positive sorting of firms across space if the local labor share is decreasing in ℓ (or, equivalently in productivity $A(\ell)$). The top left panel of Figure 2 shows that 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 CZs.[19,20]

The intuition for why a negative correlation between local labor shares and local GDP indicates positive firm sorting is that within each labor market of our model, more productive firms have more monopsony power and thus lower *firm-level* labor shares—something that resonates with

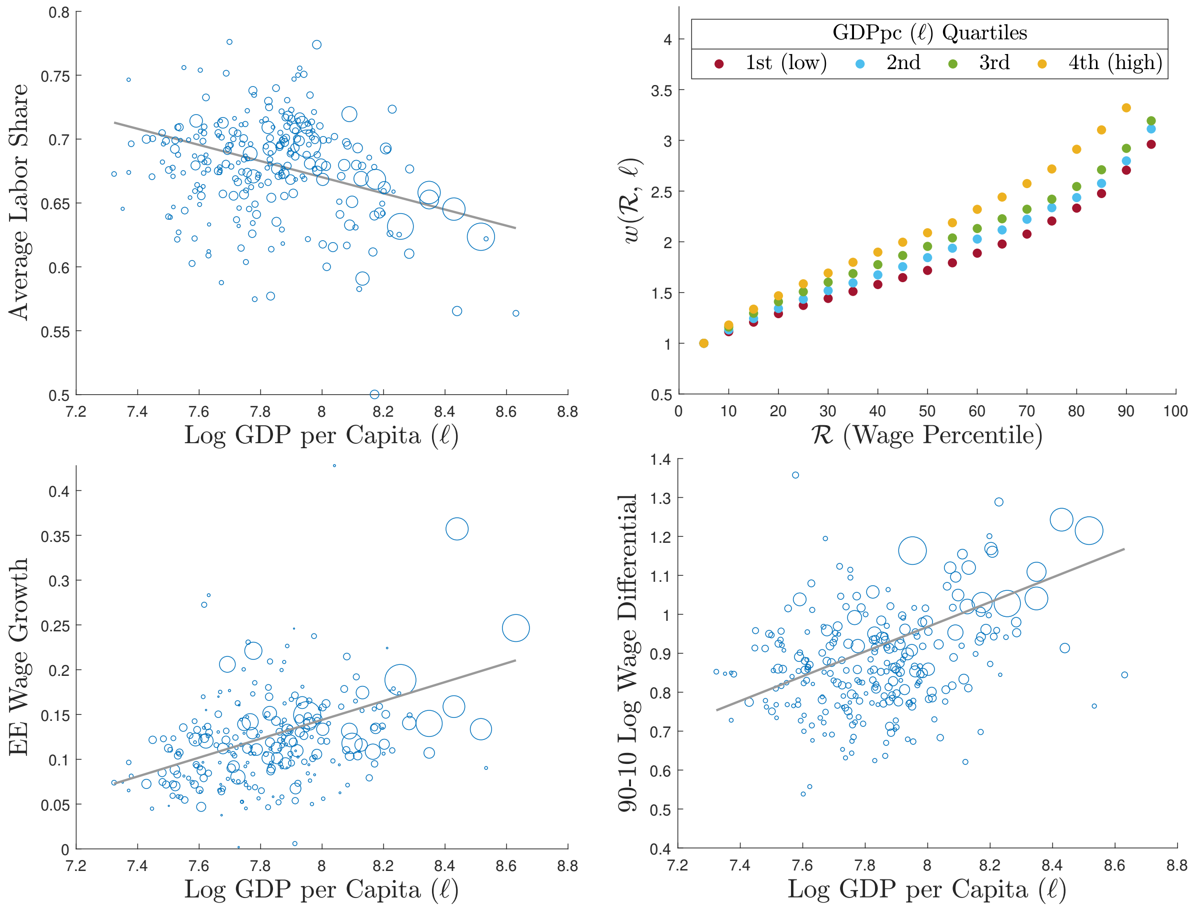
[17] We deliberately do not use AKM firm fixed effects as a proxy for firm productivity since they conflate firm with location productivity. Moreover, limited mobility bias tends to inflate the variance of these (FDZ-provided) firm fixed effects (e.g., Bonhomme et al., 2022), which makes them unsuitable for assessing local firm productivity *distributions*, especially in small CZs.

[18] Appendix SA.4.1 provides empirical support for the Pareto assumption based on tail variation in local productivity.

[19] Our finding, based on local labor shares, that places with high GDP per capita feature more monopsony power is not inconsistent with Hirsch et al. (2020) who find, based on firm-level labor supply elasticities, that large German cities feature less monopsony power: GDP per capita and population density are only weakly correlated in Germany ($corr$ =0.14).

[20] Appendix SA.4.1 supplements this evidence of positive firm sorting, using spatial variation in firm sales per worker. We document that the within-location dispersion in sales per worker is larger in productive labor markets, which—based on Corollary SA1, Appendix SA.4.1—indicates positive firm sorting across space. In addition, we show that the empirical relationship between the difference in firms' global and local productivity ranks and firm productivity is consistent with spatial firm sorting (Proposition SA6, Appendix SA.4.2). However, given that we observe sales for only a small subset of firms, our preferred evidence for spatial firm sorting stems from local labor shares, based on *regional* data from the German Federal Statistical Office.

Figure 2: Implications of Spatial Firm Sorting: Spatial Variation in Labor Shares, Job Ladders, EE Returns and Within-Location Wage Inequality



Notes: Data sources: BHP and LIAB. In the top left panel, we plot local labor shares against local log GDP pc. Local labor shares are defined as the ratio between labor compensation and gross value added in each CZ (see Appendix C.1 for details). In the top right panel, we display wage quantiles relative to the 5% quantile for four groups of CZs, ordered by quartiles of local GDPpc. In the bottom left panel, we plot location-specific EE returns, i.e., β_ℓ^{EE} based on (19), against log GDPpc. In the bottom right panel, we plot the local 90-10 log difference in firm-level wages against local GDP pc. The size of the markers either indicates the number of firms or the number of EE moves within each CZ.

empirical findings from a variety of countries (e.g., Lochner and Schulz, 2022 for Germany; Bontemps et al., 2000 and Cahuc et al., 2006 for France; Yeh et al., 2022 for the US).[21] Rich locations with low aggregate labor shares must then have a higher concentration of employment in top firms with low labor share—a composition effect implied by positive firm sorting.[22]

While our theory implies that local labor shares only depend on firm sorting μ, we now explore whether other mechanisms outside of our model can explain the documented labor share patterns. We summarize this discussion in the first column of Table 1, which reports in row 1 our baseline correlation between local labor shares and GDP per capita from Figure 2. A first concern

[21] Given the link between local labor shares and firm composition in our model, measuring local labor market power based on the labor share is natural in our context. Alternative models/data may suggest a different route, e.g., by estimating firm-level labor supply elasticities (Manning, 2003, Manning, 2011, Hirsch et al., 2020), which are then aggregated to the local level, or by computing local concentration indices (e.g., Berger et al., 2022, Berger et al., 2023 who build on Atkeson and Burstein, 2008).

[22] In support of this prediction, we show in Figure SA.1 that high-ℓ locations have thicker-tailed local sales distributions.

might be that this correlation does not reflect firm sorting but is driven by regional differences in the industrial composition (Gaubert, 2018). For example, the negative correlation between local GDP and local labor shares could be driven by the greater importance of capital-intensive industries in prosperous locations or industry-specific union power. However, when controlling for regional differences in the industrial composition through sectoral employment shares in row 2, the negative relationship between local labor shares and local GDP becomes even more pronounced.

In rows 3 and 4, we control for spatial differences in worker composition. If, for example, skilled workers have more bargaining power or skilled work is complimentary with capital, spatial worker sorting (e.g., Diamond, 2016; Heise and Porzio, 2022) rather than firm sorting could drive the negative relationship between labor shares and GDP per capita. But when we control for the observed skill composition of the local workforce (row 3) or for differences in unobserved skills via the local average AKM worker fixed effect provided by the FDZ (row 4), the negative relationship between local labor shares and GDP gets reinforced.

Finally, a potential concern may be that land—a factor of production our model omits—is more expensive in rich places, which could mechanically lower their local labor shares. In row 5, we control for local prices of commercial real estate, which does not affect our conclusion.

Local Job Ladders and Local EE Wage Growth. Our evidence suggests that firms sort positively across space and—based on our theory—this has important implications for the local wage structure, which we now assess using firm- and worker-level micro data.

One crucial channel through which positive firm sorting affects local wages is by steepening local job ladders in productive (high-ℓ) locations (Proposition 3(i)). In Figure 2, top right 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 ℓ based on the quantiles of the firm-level wage distribution (corresponding to $w(\mathcal{R}, \ell)$, where $\mathcal{R} := \Gamma_\ell(y)$ is the firms' local productivity *rank*).[23] We express all wages relative to the bottom wage of that location. This allows us to focus on the differential *steepness* of these local job ladders that is driven by firm sorting and independent of $A(\ell)$. We then average all job ladders within each GDP quartile.

Local job ladders differ meaningfully across space: Whereas in rich regions, 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.

A natural implication of this spatial job ladder heterogeneity is that EE wage growth is higher in prosperous locations. To assess this prediction, we run the following regression, which allows

[23]Importantly, as a corollary to Proposition 3(i), the wage gradient in firms' local productivity *rank* $\mathcal{R} := \Gamma_\ell(y)$ also steepens in ℓ if the sufficient conditions for positive firm sorting from Proposition 1 hold (see the Remark in Appendix A.3).

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

	(1) Labor share	(2) EE Return	(3) 90-10 Log Wage Diff.	(4) Log Wage
Log GDPpc [No Controls]	-0.0632***	0.1094***	0.3172***	0.5350***
	(0.0088)	(0.0313)	(0.0459)	(0.0271)
Log GDPpc [Control: Industry]	-0.1024***	0.1057***	0.2328***	0.5116***
	(0.0212)	(0.0315)	(0.0362)	(0.0253)
Log GDPpc [Control: Observed Comp.]	-0.1130***	0.1082***	0.1052***	0.3560***
	(0.0134)	(0.0314)	(0.0203)	(0.0330)
Log GDPpc [Control: Unobserved Comp.]	-0.0776***	0.1087***	0.2094***	0.1880***
	(0.0104)	(0.0317)	(0.0267)	(0.0301)
Log GDPpc [Control: Other]	-0.0997***	0.1152**	0.2029***	0.4723***
	(0.0183)	(0.0385)	(0.0601)	(0.0418)
N	257	257	257	257

Notes: Data Sources: German Federal Statistical Office, BHP, and LIAB. We report coefficient β from the regression $\Upsilon_\ell = \beta \ln GDPpc_\ell + \epsilon_\ell$, where $GDPpc_\ell$ denotes GDP per capita in CZ ℓ and Υ_ℓ denotes one of the outcomes in the different columns. The different rows refer to the inclusion of various controls at the regional level (column 1) or to the case in which we residualize firm/worker-level outcomes at the micro level before aggregating them to regional outcome Υ_ℓ (columns 2-4). In row 2, we account for industry impact: by controlling for local employment shares in seven industries (agriculture; mining and utilities; manufacturing; construction; trade and transportation; professional services; public administration and health) in column 1; by controlling for 1-digit industry fixed effects in (19) when estimating the LHS variable β_ℓ^{EE} in column 2; and by residualizing firm-level wages before computing the regional wage outcomes in columns 3 and 4. In row 3, we control for workers' observable skills/traits: through local shares of college graduates (column 1); by controlling for age/gender/education when estimating β_ℓ^{EE} from (19) (column 2); and by residualizing firm-level wages using firms' shares of college graduates (column 3 and 4). In row 4, we control for workers' unobserved skills: by controlling for workers' average local AKM fixed effects (provided by the FDZ) in column 1; or by including individual (not AKM) fixed effects when estimating (19) for column 2; in columns 3 and 4, we use the LIAB and compute workers' *residual* wages as $\tilde{w}_{i\ell,t} \equiv \ln w_{i\ell,t} - \alpha_i$, where $w_{i\ell,t}$ is the wage of worker i in location ℓ and month t, and α_i is the AKM worker fixed effect (from FDZ). We then compute local wage dispersion (column 3) and average wages (column 4) after taking out month fixed effects. In row 5, we control for local corporate real estate prices in column 1 and for log population density in columns 2-4. Due to limited availability of local land prices, the regression in row 5, column 1, has only 118 observation.

us to zero in on *within*-location moves in line with our model:

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

where $w_{i\ell,t}$ is the wage of individual i in CZ ℓ and month t, $EE_{i\ell,t}$ indicates whether 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. 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. Our coefficient of interest is β_ℓ^{EE}: the impact of an EE move *within* CZ ℓ on wage growth.

In the bottom left panel of Figure 2, we plot β_ℓ^{EE} against local GDP per capita. More prosperous locations have substantially higher EE returns: 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.[24] This finding is consistent with the spatial heterogeneity in dynamic wage profiles emphasized by De La Roca and Puga (2017) and suggests that such heterogeneity may in part reflect variation in EE returns across space.

[24]We compute an average EE return of 13%, similar to the results in Heise and Porzio (2022) who also use the LIAB.

To assess the quantitative impact of this spatial variation in EE returns on spatial wage inequality, we perform a statistical decomposition of the spatial wage gap in lifetime earnings (see also Baum-Snow and Pavan, 2011), considering the following drivers: spatial differences in wage growth due to EE moves, starting wages, wage growth during continuing job spells, and 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 find that 20% of the spatial earnings gap that emerges 15 years into workers' careers is due to differential EE wage growth. Importantly, these spatial differences in EE wage growth are not driven by heterogeneity in local EE transition rates (see Figure A.1, middle panel) but reflect higher EE returns in prosperous labor markets. Differential EE returns are more important for the lifetime earnings gap than differential wage growth of job stayers (accounting for 13%), with most of the remaining part stemming from regional differences in starting wages. See Appendix D.2 for details.[25]

Within our model, since wages are proportional to local TFP $A(\ell)$, the spatial variation in wage growth from EE moves, β_ℓ^{EE}, only reflects differences in firm sorting and is not directly impacted by $A(\ell)$ and its empirical determinants. In Table 1 column 2, we nevertheless investigate alternative mechanisms outside of our model. To do so we estimate β_ℓ^{EE} in (19) with different sets of RHS variables and then project it on local log GDP per capita. Row 1 contains our baseline specification from Figure 2: A rise in GDP per capita by one log point is associated with an increase in EE returns by 11 percentage points. We show in row 2 that this spatial correlation is not due to a high concentration of industries with faster EE wage growth in rich labor markets: When we estimate (19) while controlling for sector fixed effects, our main coefficient hardly changes.

An additional concern is that spatial worker sorting drives regional differences in EE returns, if, e.g., skilled workers predominantly settle in rich labor markets and experience faster EE wage growth; or if workers who accumulated more valuable experience in the past move to prosperous places and see their wages grow relative to workers who lack these dynamic human capital benefits (De La Roca and Puga, 2017). Row 3 shows that the correlation between local EE returns β_ℓ^{EE}—obtained from (19) when controlling for observable worker characteristics such as age, gender and education fixed effects—and GDP per capita is little affected. In row 4, we control for unobserved heterogeneity in individuals' wage *growth* by including an individual fixed effect in (19). Doing so leaves the correlation between local EE returns and GDP per capita unchanged.

Finally, row 5 controls for local population density when regressing β_ℓ^{EE} on local GDP per

[25] Baum-Snow and Pavan (2011) find that differential EE wage growth across small and large US cities accounts for 13% of the earning gap after 15 years, possibly lower than our estimate because EE returns are higher in Germany (Engbom, 2021) and because city size and GDP per capita—our statistic to rank locations—are only weakly correlated in Germany ($corr = 0.14$).

capita, which only has a mild impact on our coefficient of interest. This addresses the concern that higher EE returns in rich and dense locations may reflect better match quality (see, e.g., Dauth et al., 2022) if these places benefit from increasing returns to matching.

Spatial Wage Inequality Within Labor Markets. We now turn to the implications of spatial firm sorting for inequality within local labor markets, which we measure by the difference between the 90th and 10th quantiles of log firm wages within each CZ. Our evidence on steeper job ladders and higher EE wage growth in prosperous places suggests that wage inequality is more pronounced in these locations. This is in line with our model's predictions (Proposition 4): Within-location inequality is higher in high-ℓ locations if firm sorting is positive, *independent* of local productivity $A(\ell)$. The bottom right panel of Figure 2 indeed shows that the 90-10 gap in log wages is about 0.4 log points higher in the most prosperous labor markets.

We now entertain several explanations for this fact that are outside of our model. In column 3 of Table 1, we report the bivariate correlation between local GDP per capita and within-location wage dispersion, which we compute after residualizing firm wages in different ways.

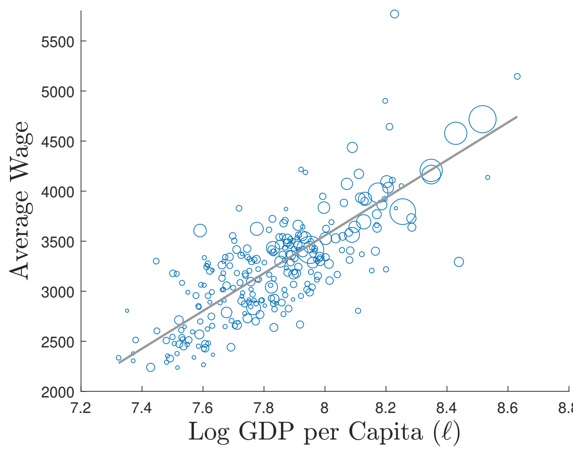
Row 2 shows that prosperous labor markets have more dispersed wages also within industries, which addresses the concern that industries may be systematically sorted across space and, e.g., differ in union coverage (Jäger et al., 2022) and thus the extent of wage compression.

Another valuable concern is that local wage dispersion mostly reflects the spatial sorting of workers (rather than firms). This may be the case if observed skills are more dispersed in productive locations (Eeckhout et al., 2014) or if the dispersion of unobserved ability is higher among skilled workers who cluster in those places (Card et al., 2023). In row 3, before computing local wage dispersion, we control for observed worker heterogeneity by residualizing firm-level wages with respect the share of firms' college graduates. The resulting correlation with local GDP is lower but still positive and significant. To address the concern about unobserved worker heterogeneity, we turn to worker-level (rather than firm-level) wages and purge them from their AKM worker fixed effects (see the *Notes* of Table 1 for details). Row 4 shows that the correlation between the local dispersion of residualized wages and local GDP remains positive and sizeable.

Finally, in row 5, we again consider the possibility that rich and dense places feature stronger worker-firm sorting, which would augment local wage inequality. Controlling for local population density, however, does not annul our baseline correlation from row 1.

Spatial Wage Inequality Across Labor Markets Finally, we turn to the implications of our model for wage inequality across locations (Proposition 3(ii)). Poor locations are not only

Figure 3: Implication of Spatial Firm Sorting: Spatial Variation in Wages



Notes: Data source: BHP. The figure shows a scatter plot between local average firm wages and local log GDPpc. The size of the markers indicates the size of the CZ (i.e., the number of firms in each CZ).

hurt by adverse fundamentals but also by the fact that unproductive firms settle there. Indeed, Figure 3 shows that spatial wage inequality in Germany is pervasive: Average monthly wages paid by firms vary between 2,500EUR in the poorest and 4,500EUR in the richest places.

While our model predicts that positive firm sorting across space is one contributing factor, spatial wage differences clearly also reflect variation in local productivity $A(\ell)$ and its determinants. To address this concern in our empirical analysis as good as possible, column 4 of Table 1 reports the correlation between local average wages and GDP per capita after controlling for firms' industry (row 2), observed and unobserved heterogeneity of workers (rows 3 and 4) and population density (row 5). The correlation between local wages and GDP per capita remains positive and sizeable, but we recognize that fully isolating the role of firm sorting in across-location wage inequality requires our structural model; see Section 5 below.

5 Estimation

To evaluate the quantitative importance of firm sorting for spatial wage inequality, we structurally estimate our model. To this end, we first enrich our model along four dimensions and then discuss our identification proof, estimation strategy, and results.

5.1 Bringing our Model to the Data

Setting. Our goal is to introduce minimal changes that render our model suitable for estimation while preserving its key mechanism and tight link to our analytical results. 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.[26] 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 with ℓ and scale individuals' real consumption utility. Last, we allow job separation rates δ to vary (exogenously) across locations to rationalize the observed spatial variation in unemployment.[27] Importantly, 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 distributions; and controlling for local capital intensity via industry.

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 ℓ there is a labor market matching function with constant returns to scale, so that meeting rates are determined by local market tightness, $\theta(\ell) = \mathcal{V}(\ell)/\mathcal{U}(\ell)$, where the measure of vacancies in each location, $\mathcal{V}(\ell)$, equals the measure of firms that settle there.[28] In turn, we let parameter κ be the relative matching efficiency of employed workers (i.e., $\lambda^E(\ell) = \kappa\lambda^U(\ell)$) so that $\mathcal{U}(\ell) = L(\ell)(u(\ell) + \kappa(1 - u(\ell)))$ is the effective measure of searchers in ℓ, impacted by the endogenous $L(\ell)$. An important implication is that firms' and workers' meeting rates, $(\lambda^F(\ell), \lambda^U(\ell), \lambda^E(\ell))$ 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—a second source of congestion—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.[29] 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

[26]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, where we assess the workers' location based on the last employer's location.

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

[28]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 $Q'(p) = q(p)$). In equilibrium $p = \mu(\ell)$, so the mass of firms in 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 $\mathcal{V}(\ell) = q(\mu(\ell))\mu'(\ell) = r(\ell)$, where the mass of vacancies *per unit of land in* ℓ is one.

[29]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.

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:

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

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

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 A1 (Appendix B). 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. As far as production is concerned, we assume that production function z is multiplicative and that the ex post productivity distribution is Pareto with tail parameter $1/p$:[30]

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

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.

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

[30]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.

$(\omega, \tau, h(\ell))$. 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).[31]

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 Corollary 1, our model allows us to separately identify $\mu(\ell)$ and $A(\ell)$ using the average local labor share, $LS(\ell)$, and firm value added, $\mathbb{E}[z(y, A(\ell))|\ell]$,

$$LS(\ell) = 1 - \mu(\ell) \tag{21}$$

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

where the expectation is taken over Γ_ℓ. Variation in local labor shares across space $LS(\cdot)$ thus identifies firm sorting $\mu(\cdot)$. Conditional on local firm composition $\mu(\ell)$, location productivity $A(\ell)$ can then be identified from average local value added: The spatial variation in value added that is *not* accounted for by firm sorting must be driven by differences in local 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:

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

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

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

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

Last, we identify the amenity schedule $B(\ell)$ 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:

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

Proposition 6 (Identification)**.** *Under the assumed functional forms (summarized in Assumption A1, Appendix E), the model is identified.*

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

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 C.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 for our empirical analysis 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, as in Section 4, we rank the 257 CZs based on their log GDP per capita. Local log GDP per capita will be our discretized support $\{\ell_1, \ell_2, ..., \ell_{257}\}$ of the model's land distribution R. We identify R from the number of firms in each CZ, i.e., 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 Munich and Wolfsburg (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 4, top left), we obtain an *increasing* matching function μ (top right).[32] 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.3, Appendix F). 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 $\kappa = 0.597 \cdot \frac{\lambda^E}{\lambda^U}$; see Table A.3.

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 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 4, second row, left).[33] We

[32] The size of the dots in Figure 4 is proportional to the size of the CZ, as measured by its number of establishments.
[33] Solving (24) for $\lambda^U(\ell)$ while taking $\delta(\ell) = \lambda^U(\ell)u(\ell)/(1-u(\ell))$ into account gives $\lambda^U(\ell) = \mathcal{A}\big(\bar{l}(\ell) \cdot \big(\kappa + \frac{u(\ell)}{1-u(\ell)}\big)\big)^{-\frac{1}{2}}$.

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 decreasing in ℓ (Figure 4, third row, left), along with a fairly stable job-finding rate, translates into job-separation rates that are lower in more prosperous locations (Figure 4, 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.[34] Since value added per worker is strongly increasing in ℓ (Figure 4, bottom left), we obtain an increasing A-schedule, even after controlling for firm sorting through μ (Figure 4, 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, the quality of infrastructure, and especially the skills of local workers; see Table A.4, Appendix F. This is consistent with our implicit assumption in the baseline model that spatial worker heterogeneity shapes $A(\ell)$ and is therefore taken into account by firms when making their location choices.

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.3, Appendix F). Finally, we pin down local housing supply $h(\ell)$ using observed location-specific rental rates $d(\ell)$; see Figure A.3 in Appendix F.

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 left panel of Figure A.3 (Appendix F) 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, while we stipulate that the empirical firm sorting can be captured by a function μ, 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) (in line with Proposition 1), which verifies that the positive sorting of firms into locations is indeed *optimal* in the estimated model.

[34] 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 4: Model Fit of Targeted Moments (left) and Estimated Parameters (right)

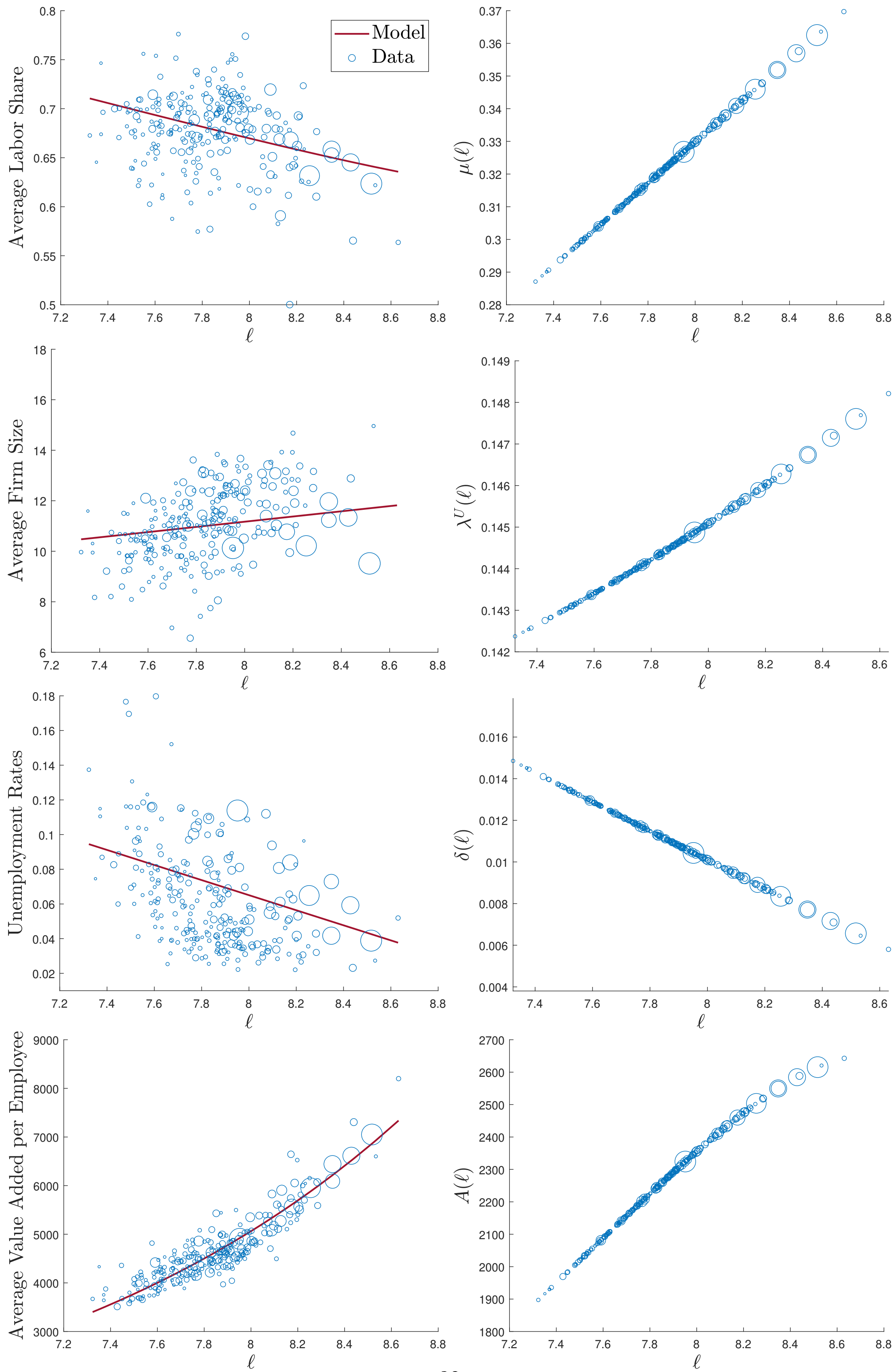
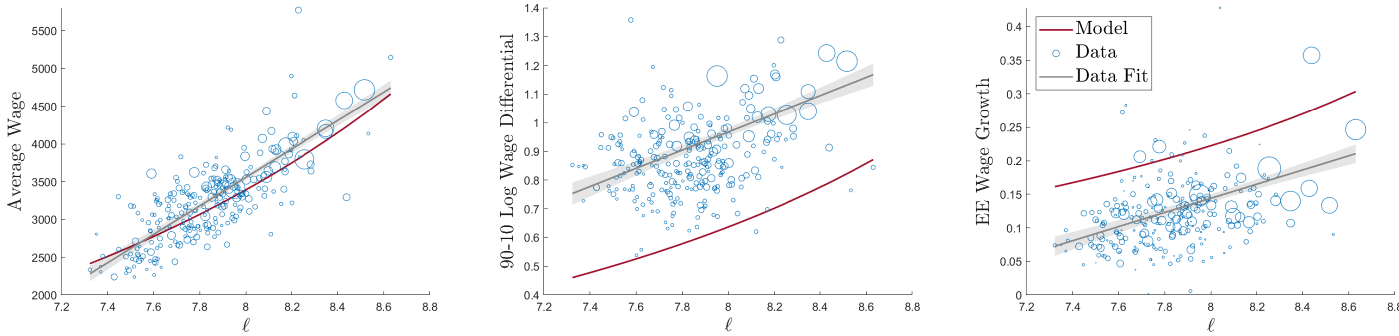


Figure 5: Model Fit: Non-Targeted Moments



Notes: Data sources: The left and middle panels are based on firm-level wages of full-time (FT) employees from BHP; and the right panel is based on worker-level wages from the LIAB. The right panel (data) 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 4, left column). Importantly, despite its parsimony, our model also performs reasonably well regarding several non-targeted features of the data that are related to worker inequality and beyond.

In Figure 5, we confront our model with the findings from our empirical analysis (Section 4). 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 5, we display local 90-10 log wage gaps across space, which is the model counterpart of the bottom right panel of Figure 2. Although our model underestimates the level of within-location inequality, it almost perfectly rationalizes its spatial variation. 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 theoretical counterpart to the bottom left panel of Figure 2. We overestimate the level of EE wage growth, but, reassuringly, closely match the variation across space.

We further validate the model along the following dimensions (see Appendix F). First, we show that our model matches relatively well the decreasing rate of job loss, $\delta(\cdot)$, as well as the flat EE transition rate, $\mathbb{P}(EE|\cdot)$, and UE transition rate, $\lambda^U(\cdot)$ (Figure A.1). 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.2, 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.2, right panel).

6 Spatial Firm Sorting: Quantitative Implications

We use our estimated model to quantify the role of spatial firm sorting in spatial wage disparities. We then show that the presence of OJS and spatial hiring frictions are central for our results.

6.1 The Implications of Spatial Firm Sorting for Spatial Wage Inequality

Our estimation implies that 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 consider a counterfactual in which we allocate firms randomly to locations and let the equilibrium play out. Hence, workers reallocate across space, job-finding rates adjust via local market tightness, and local wage schedules change. Appendix SA.6.1 contains technical details on this exercise.[35]

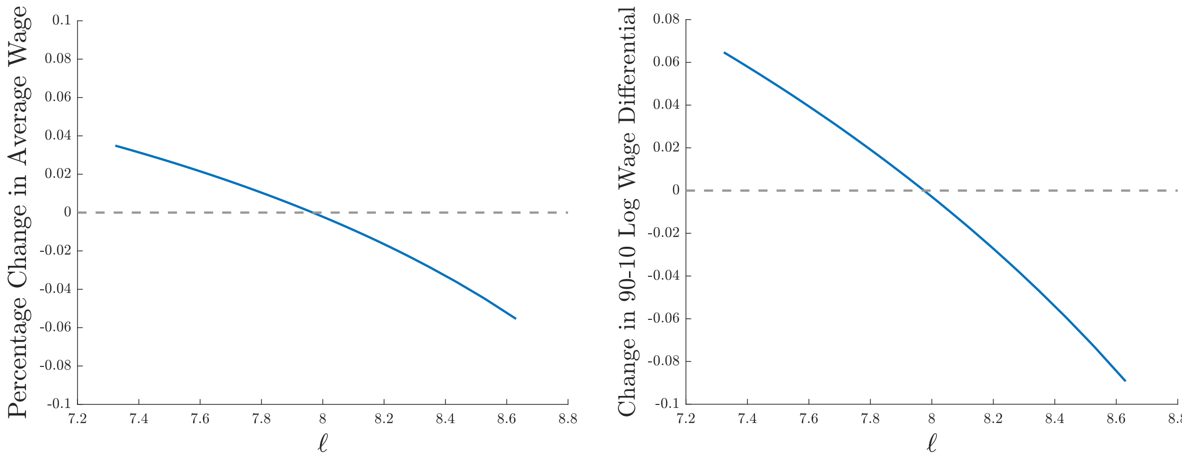
We start with the implications of sorting for wage inequality *across* local labor markets. Firm sorting helps productive locations because it amplifies their advantage from a higher local TFP A. Hence, spatial inequality would be lower in the absence of sorting. The left panel of Figure 6, which displays 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. This implies that the spatial wage premium between poor and rich locations in Germany (i.e., between locations in the bottom and top quartiles of the distribution of local GDP per capita) drops from 40% to 34% in the absence of firm sorting (Table 2, left panel). Spatial firm sorting thus accounts for $6/40 = 15\%$ of the observed spatial wage gap.[36]

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 differences in location TFP (much of which is driven by spatial worker heterogeneity, see also Table A.4, Appendix F).

[35] 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.

[36] We focus on the comparison between the richest and poorest CZs in the text, but show in Appendix SA.7 that spatial firm sorting can also account for 17% of the West-East wage gap and for 19% of the urban-rural wage gap.

Figure 6: 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.

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%	15%	0.20	0.12	40%

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 spatial gap in the 90-10 difference of log wages within labor markets, where we first assess the gap between the 90th and 10th quantiles of log wages in each location and then compute the difference of this measure across poor and rich locations. Throughout, we define poor (rich) locations 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 spatial 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 6 shows how within-location inequality would change if firm sorting were removed: Wage inequality would decline by almost 0.1 log points within productive labor markets and increase by 0.06 log points in unproductive places. 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 40% 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.[37]

To better understand *why* changes in firm sorting affect wage distributions across space, 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.

[37]While differences in local TFP A do *not* affect the spatial gap in within-location inequality (see (17)), spatial heterogeneity in job-destruction rates δ and job-finding rates λ does impact local job ladders and account for the remaining spatial variation.

First, as shown in Figure 7 (left panel), 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 moves. The middle panel of Figure 7 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 7 highlights the role of spatial differences in employment *composition*. In the baseline model with firm sorting, the employment distribution in Wolfsburg strongly 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 wages. Without firm sorting, the employment distributions across space converge.[38] This further contributes to a decline in wage inequality across space.

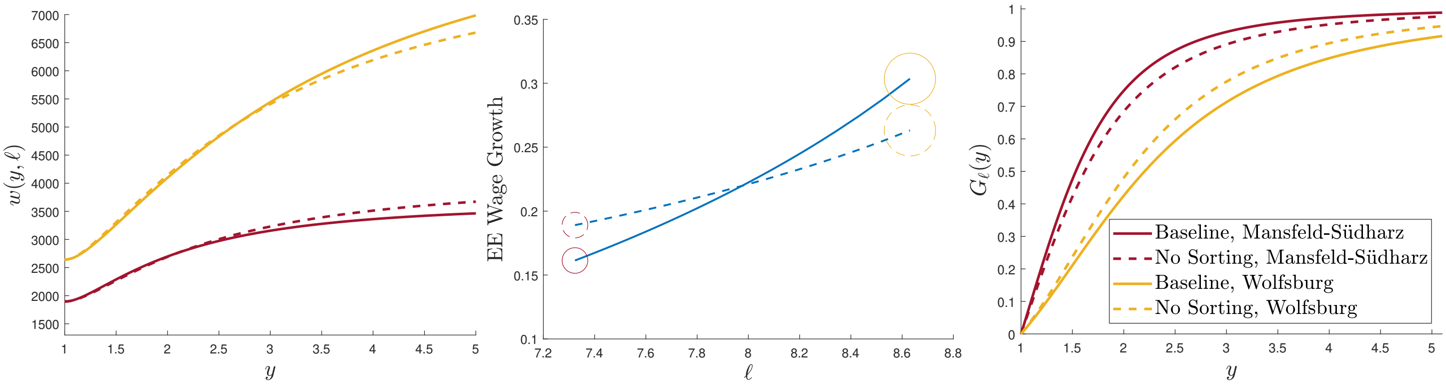
6.2 The Role of On-the-Job Search and Spatial Frictions

In our theory, firm sorting affects spatial inequality especially by shaping the returns to on-the-job search (and the associated job ladders) within spatially segmented labor markets. This suggests 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 show that this is indeed the case, 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

[38] 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 induces workers to 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.

Figure 7: 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.

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 SA.6.2 and SA.6.3 detail the implementation.

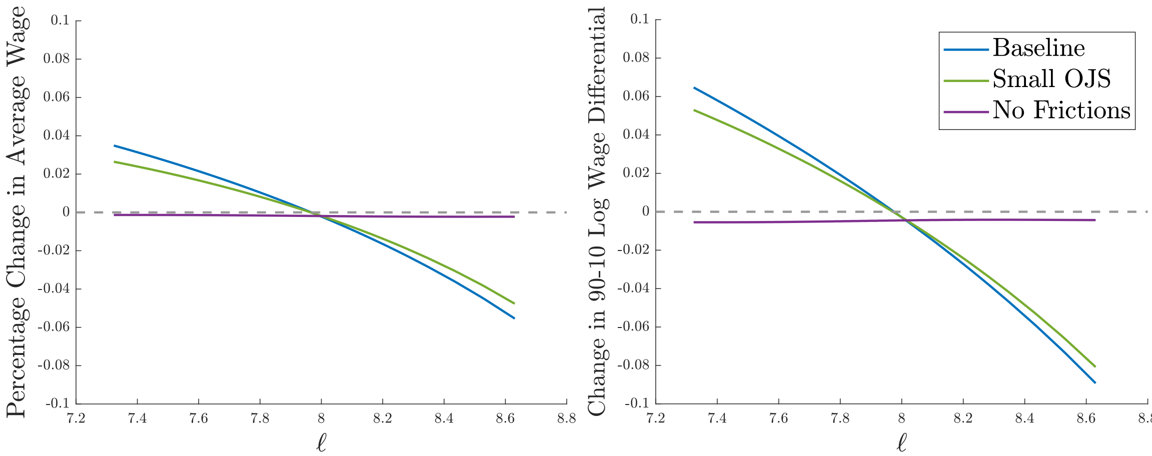
We start with the role of OJS in spatial wage inequality. The elasticity of across-location inequality wrt to the poaching share—a proxy for the magnitude of OJS in the economy—is sizable:[39] A 1% decline in the poaching share reduces the spatial wage premium of the most prosperous places by 0.6%. A 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.[40]

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). Comparing the impact of sorting on wage inequality in the baseline model (blue lines) with its impact when OJS is muted (flatter green lines) in Figure 8 shows that firm sorting is now less important for inequality, both across and within locations—an effect that is quantitatively meaningful: When the poaching share decreases by 10%, spatial firm sorting contributes 15% less to across-location inequality and 14% less to the spatial gap in within-location inequality. We conclude that OJS is an important reason

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

[40] 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.

Figure 8: 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).

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 wage disparities 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 (see Figure SA.4, Appendix SA.6.3). 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 the distribution of wages *within* locations: Wage dispersion across firms is now 6% *higher* in *low*-ℓ relative to high-ℓ locations.

To study the *interaction* of firm sorting and spatial hiring frictions, we implement the random matching counterfactual in this setting of fully integrated local labor markets. The flat purple lines in both panels of Figure 8 indicate that firm sorting shape wages across locations only if labor markets feature some degree of segmentation.

6.3 Firm Sorting and Place-Based Policies

In a final exercise, we use our model to examine the role of place-based policies in spatial wage disparities, which highlights the importance of considering firms' location choices as *endogenous*. In 1969, the German Federal Ministry of Economics established an initiative for the improvement of local economies (GRW), which provides discretionary investment grants to economically disadvantaged regions to equalize living standards across space. Since GRW covers a wide range of CZs during our sample period, its impact is reflected in the estimated local productivity, $A(\ell)$.

We consider a reduction in these place-based subsidies in some structurally weak CZs (Appendix SA.6.4 gives details). For illustration, we decrease local TFP $A(\ell)$ by 10% in locations $\{\ell_5, ..., \ell_{33}\}$, which all received GRW subsidies during the time period of our study (Figure 9, left).

This policy alters the relative attractiveness of our local labor markets and thereby affects firm's location choices: The composition of firms worsens in locations $\{\ell_5, ..., \ell_{33}\}$, whose subsidy is reduced, and improves in the originally least developed CZs, $\{\ell_1, ..., \ell_4\}$, which have become relatively more attractive (Figure 9, middle panel). Quantitatively, the average productivity of firms in the "treated" regions declines by 0.13%. By contrast, firm productivity in the initially least developed locations increases by 2.1%.

In the right panel of Figure 9, we display the implications for local wages in this counterfactual equilibrium. Consider first locations $\{\ell_1, ..., \ell_4\}$, which indirectly benefit from the policy shift because they are able to attract better firms: Both average wages (shown in green) and wage dispersion (shown in purple) increase in response to the change in firm sorting.

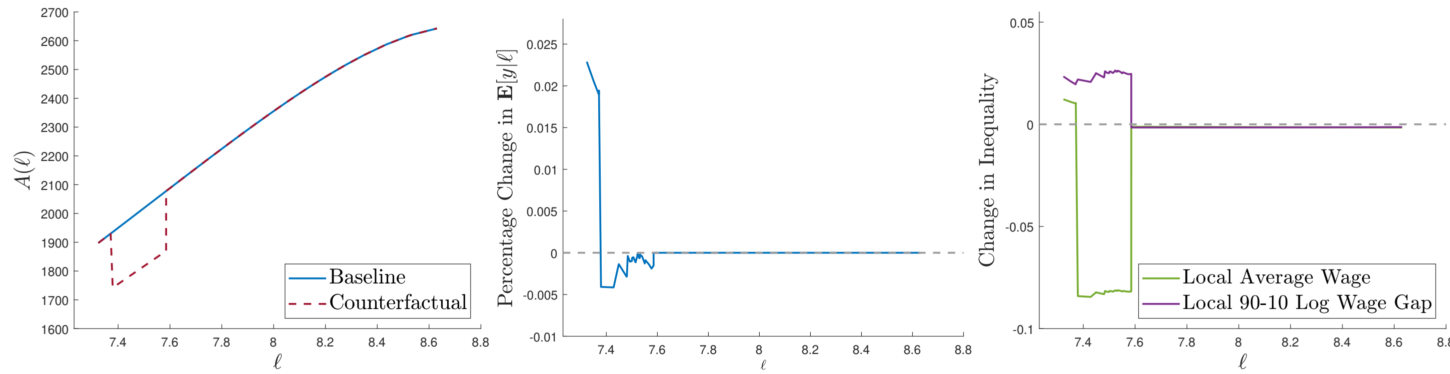
The equilibrium impacts on the treated locations $\{\ell_5, ..., \ell_{33}\}$ are subtle. The direct productivity effect of this policy reduces local wages but the overall decline is less than the subsidy reduction of 10%. The reason is that, like productive firms, workers *also* leave these labor markets. This increases the meeting rates of the remaining workers and counteracts the initial wage decline. Higher worker meeting rates also explain the increase in local wage dispersion.

This exercise highlights the subtle effects of place-based policies on treated and non-treated areas due to equilibrium effects that are triggered by endogenous firm sorting.

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

Figure 9: The Impact of Place-Based Policies on Local TFP, Firm Sorting and Spatial Wage Inequality



Notes: The left panel shows changes in local TFP $A(\ell)$ induced by the counterfactual policy. The middle panel shows the percentage changes of average firm productivity across ℓ induced by the counterfactual. The right panel plots percentage changes in local average log wages (green) and changes in the local 90-10 log wage gap (purple) across regions.

inequality. Table 3 summarizes the results; Appendix SA.8 contains details.

Measurement. A first concern may be that different industries operate under different technologies with different labor and capital intensities. 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.[41] Second, our model may be a better description of tradable industries, 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 stronger spatial firm sorting.

Further, in our baseline estimation we relied 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.

[41] 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.

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 a flexible b-schedule, ruled out spatial differences in firm selection at the lower end of local productivity distributions. In Appendix SA.8.5, we implement a version of our model without this assumption. In that model, each labor market has an endogenous productivity cutoff $\underline{y}(\ell)$, and all firms with $y < \underline{y}(\ell)$ choose to exit. In the estimated model, exit threshold $\underline{y}(\ell)$ is 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. As an alternative, we estimate a model with location preference shocks that hamper labor mobility. We discipline workers' propensity to migrate by targeting a spatial labor supply elasticity that is consistent with the empirical literature; see Appendix SA.8.4. Column 6 shows that this does not change our quantitative conclusions about the role of sorting in within- and across-location inequality.

Table 3: The Effect of Spatial Firm Sorting and Spatial Wage Inequality: Robustness

	Baseline	Measurement Choices			Model Assumptions	
		Within Industry	Only Manufac.	FDZ Data	Selection	Imperfect Mobility
Spatial Wag Gap	15 %	26 %	22 %	34 %	22 %	16 %
90-10 Difference	40 %	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. For both measures, we compare 'poor' (bottom quartile of CZs in terms of GDPpc) and 'rich' (top quartile) locations.

8 Conclusion

In this paper, we study the link between spatial firm sorting and spatial wage inequality. We analyze 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 and makes it difficult to poach workers from other firms.

We characterize this trade-off analytically and show that positive firm sorting—i.e., an allocation in which productive firms settle in productive locations—emerges as the unique equilibrium if firm and location productivity are sufficient complements or labor market frictions are sufficiently large. Our theory makes rich predictions on how firm sorting impacts local wage distributions. If productive firms sort into productive locations, the local job ladder in these places steepens, EE returns are higher and more employment is concentrated in top firms that pay high wages. This amplifies the spatial gaps in both the mean and dispersion of wages.

Using administrative data from Germany, we first provide empirical support for these predictions. We then estimate our model to assess the degree of firm sorting in the data and quantify its role for wage disparities across and within locations: Firm sorting can account for 15%-34% of the spatial variation of mean wages and for 40%-66% of the spatial differences 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*.

Allen, T. and D. Donaldson: 2020, 'Persistence and Path Dependence in the Spatial Economy'. NBER Working Paper 28059.

Atkeson, A. and A. Burstein: 2008, 'Pricing-to-Market, Trade Costs, and International Relative Prices'. *American Economic Review* **98**(5), 1998–2031.

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.

Berger, D., K. Herkenhoff, and S. Mongey: 2022, 'Labor Market Power'. *American Economic Review* **112**(4), 1147–93.

Berger, D. W., K. F. Herkenhoff, A. R. Kostøl, and S. Mongey: 2023, 'An Anatomy of Monopsony: Search Frictions, Amenities, and Bargaining in Concentrated Markets'. In: *NBER Macroeconomics Annual 2023, volume 38*, NBER Chapters. National Bureau of Economic Research, Inc.

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

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

Bontemps, C., J.-M. Robin, and G. J. Van Den Berg: 2000, 'Equilibrium Search with Continuous Productivity Dispersion: Theory and Nonparametric Estimation'. *International Economic Review* **41**(2), 305–358.

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.

Cahuc, P., F. Postel-Vinay, and J.-M. Robin: 2006, 'Wage Bargaining with On-the-Job Search: Theory and Evidence'. *Econometrica* **74**(2), 323–364.

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.

Card, D., J. Rothstein, and M. Yi: 2023, 'Location, Location, Location'. Working Paper.

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.

Diamond, P. A.: 1971, 'A Model of Price Adjustment'. *Journal of Economic Theory* **3**(2), 156–168.

Diamond, R.: 2016, 'The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000'. *American Economic Review* **106**(3), 479–524.

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.

Engbom, N.: 2021, 'Labor Market Fluidity and Human Capital Accumulation'. Working Paper.

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, É.: 2022, 'Productivity Dispersion, Between-Firm Competition, and the Labor Share'. *Econometrica* **90**(6), 2755–2793.

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.

Hirsch, B., E. J. Jahn, A. Manning, and M. Oberfichtner: 2020, 'The Urban Wage Premium in Imperfect Labor Markets'. *Journal of Human Resources*.

Jäger, S., S. Noy, and B. Schoefer: 2022, 'The German Model of Industrial Relations: Balancing Flexibility and Collective Action'. *Journal of Economic Perspectives* **36**(4), 53–80.

Karahan, F., R. Michaels, B. Pugsley, A. Şahin, and R. Schuh: 2017, 'Do Job-to-Job Transitions Drive Wage Fluctuations over the Business Cycle?'. *American Economic Review* **107**(5), 353–57.

Lochner, B. and B. Schulz: 2022, 'Firm Productivity, Wages, and Sorting'. *Journal of Labor Economics* **forthcoming**.

Manning, A.: 2003, *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton University Press.

Manning, A.: 2011, 'Chapter 11 - Imperfect Competition in the Labor Market'. Vol. 4 of *Handbook of Labor Economics*. Elsevier, pp. 973–1041.

Martellini, P.: 2022, 'Local Labor Markets and Aggregate Productivity'. Working Paper.

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.

Yeh, C., C. Macaluso, and B. Hershbein: 2022, 'Monopsony in the US Labor Market'. *American Economic Review* **112**(7), 2099–2138.

Firm Sorting and Spatial Inequality: Appendix

- FOR ONLINE PUBLICATION -

A Baseline Model: Proofs

Mostly for expositional reasons, we maintain the assumption $\rho = 0$ (as stated in footnote 5).

A.1 Proof of Proposition 1

We proceed in three steps. In Step 1, we show that under the premise of the proposition, an assignment with positive sorting is among the optimal ones. In Step 2, we show that any optimal assignment $\mu(\cdot)$ is strictly increasing, i.e., satisfies positive sorting. In Step 3, we show that given our assumptions on firm productivity and land distributions, there exists a unique optimal assignment that is strictly increasing and thus satisfies positive sorting. Taken together, any equilibrium features positive sorting.

Step 1. *An assignment with positive sorting satisfies firms' optimality.* To this end, we show that, if firms conjecture that there is positive sorting, then value $\overline{J}(p, \ell)$ is indeed globally supermodular in (p, ℓ) under the premise, inducing positive sorting based on firms' optimal choices.

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

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

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):

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

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 for all (y, ℓ) (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 (the conjecture of) positive sorting, land market clearing is given by $Q(\mu(\ell)) = R(\ell)$; see Appendix SA.1.5. Further, based on our assumption of strictly positive densities (r, q), this PAM assignment is one-to-one, i.e., μ is strictly increasing, whereby the firm type p assigned to location ℓ is $p = \mu(\ell) = Q^{-1}(R(\ell))$. 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 9). 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

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\lambda^E}{\delta + \lambda^E(1 - \Gamma(y|Q^{-1}(R(\ell)))} \left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p} \right) \frac{r(\ell)}{q(Q^{-1}(R(\ell)))}, \quad \text{(A.3)}$$

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

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

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, PAM is optimal, either if the productivity gains, ε^P, are sufficiently large or if φ^E is sufficiently small.

Step 2. *Any optimal assignment is strictly increasing and thus satisfies positive sorting.* To show this, we proceed by contradiction. That is, suppose there exists $\{\ell', \ell''\} \in [\underline{\ell}, \overline{\ell}]$ with $\ell' < \ell''$ such that the assigned firm types satisfy $\mu(\ell') > \mu(\ell'')$, where $\mu(\ell') = p''$, $\mu(\ell'') = p'$ and $p' < p''$, i.e., there exists a negatively assorted pair of agents. For this negatively assorted pair to be consistent with the optimal choices of firms p' and p'', it must be that if firms conjecture this assignment then their objective function $\overline{J}(p,\ell) - k(\ell)$ satisfies decreasing differences when evaluated at this pair, i.e., $\overline{J}(p',\ell') + \overline{J}(p'',\ell'') < \overline{J}(p',\ell'') + \overline{J}(p'',\ell')$. Otherwise there exists a blocking pair, which renders the assignment non-optimal. This requirement of decreasing differences translates into the requirement that the following expression, based on (A.1), be positive:

$$\int_{\underline{y}}^{\overline{y}} \underbrace{\left(\frac{\frac{\partial z(y,A(\ell''))}{\partial y}}{\left[\delta + \lambda^E(1-\Gamma_{\ell''}(y))\right]^2} - \frac{\frac{\partial z(y,A(\ell'))}{\partial y}}{\left[\delta + \lambda^E(1-\Gamma_{\ell'}(y))\right]^2} \right)}_{>0} \underbrace{\left(1-\Gamma(y|p') - (1-\Gamma(y|p''))\right)}_{<0} dy.$$

Note that the first term of the integrand is strictly positive since: (i) $\frac{\partial z(y,A(\ell''))}{\partial y} > \frac{\partial z(y,A(\ell'))}{\partial y}$ due to strict monotonicity of A and strict supermodularity of z; and (ii) $\left[\delta + \lambda^E(1-\Gamma_{\ell''}(y))\right]^2 < \left[\delta + \lambda^E(1-\Gamma_{\ell'}(y))\right]^2$ based on the conjecture of negative sorting, which implies that the firm productivity distribution in ℓ'' is stochastically worse than in ℓ'. In turn, the second term of the integrand is negative under our assumption on Γ. Thus, the integral is *negative*, i.e., $\overline{J}(p,\ell)$ satisfies *increasing* differences when evaluated at this pair of agents, which implies that negative sorting violates firm optimality: Any optimal assignment must be strictly increasing.

Step 3. *There is a unique optimal assignment with positive sorting, given by $\mu(\ell) = Q^{-1}(R(\ell))$.* We conclude from Step 2 that, under the premise, the only conjecture about sorting that is consistent with firm optimality is that of positive sorting; and we have shown in Step 1 that in this case the optimal assignment will indeed be positive.

Finally, there is a unique way of matching up two one-dimensional, continuously differentiable distributions (Q,R) in a positively assortative way that preserves the measures, given by $Q(\mu(\ell)) = R(\ell) \;\Leftrightarrow\; \mu(\ell) = Q^{-1}(R(\ell))$. We conclude that under the premise, *any* equilibrium satisfies positive sorting with $\mu'(\ell) = r(\ell)/q(\mu(\ell)) > 0$. □

Remark 1. In Step 2 of the proof of Proposition 1, the only deviation from positive sorting we consider is that of *strictly* negative sorting for a set of agents. Note that since we restrict attention to pure assignments, we do not need to consider the case of weakly negative sorting.

Remark 2. Clearly, a weaker sufficient condition for (A.2) to be positive would be based on the entire integral, which is different from our approach in the proof of Proposition 1 in which we aim for (more demanding) sufficient conditions that the integrand be positive *for all y*. Working with the integrand proved tractable, while working with the integral did not: First, none of the standard integral inequalities, which could have helped signing (A.2), apply; second, even when making functional form assumptions, the integral does not admit an explicit solution.

We give the detailed interpretation of the condition for positive sorting on the integrand (A.3) in the text, after stating condition (13). We note that the RHS of (A.3) is equivalent to (the negative of) the firm size elasticity in ℓ in (13) (see also (14)).

A.2 Proof of Proposition 2

We first 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., any equilibrium features PAM of firms to locations.

We therefore consider the 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 a 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 SA.1.4 for the derivation.

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

$$\mathcal{J}(p,\ell)-\mathcal{J}(p,\ell')\geq\int_{\ell'}^{\ell}\frac{\partial\mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial\ell}d\hat{\ell}. \tag{A.4}$$

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.4) 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 globally 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 by Proposition 1, there exists a unique optimal assignment μ, which is deterministic (and satisfies PAM). Second, the uniqueness of $k(\ell)$ (up to a constant of integration) then follows from Remarks 10.29 and 10.30 in Villani (2009). □

A.3 Proof of Proposition 3

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

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

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:

$$\int_{\underline{y}}^{y}\frac{\partial Z(t,\ell)}{\partial\ell}dt\geq-\frac{\left(\frac{\lambda^E}{\delta}\right)^2\frac{\partial\Gamma_\ell(y)}{\partial y}\left(-\frac{\partial\Gamma_\ell(y)}{\partial\ell}\right)+\frac{\lambda^E}{\delta}\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)\frac{\partial^2\Gamma_\ell(y)}{\partial y\partial\ell}}{\frac{\lambda^E}{\delta}\gamma_\ell(y)\left(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y))\right)}\int_{\underline{y}}^{y}Z(t,\ell)dt. \tag{A.5}$$

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

Second, consider $y > \underline{y}$. We obtain the following sufficient condition for (A.5):

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

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.5). The condition ensuring this is given by

$$\min_{y,\ell} \left(\frac{\int_{\underline{y}}^{y} \frac{\partial Z(t,\ell)}{\partial \ell} dt}{\int_{\underline{y}}^{y} Z(t,\ell)dt} \right) \geq \max_{y,\ell} \left(-\frac{\frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}}{\gamma_\ell(y)} \right). \tag{A.6}$$

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.6) is sufficient for w to be supermodular in (y,ℓ), which we can further unpack as:

$$\min_{\ell,y} \frac{\int_{\underline{y}}^{y} \frac{\partial^2 z(t,A(\ell))}{\partial \ell \partial y} dt}{\int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y} dt} \geq \left(1 + \frac{\lambda^E}{\delta}\right)^2 \left(\max_{y,\ell} \left(-\frac{\frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}}{\gamma_\ell(y)} \right) + 2\frac{\lambda^E}{\delta} \max_{y,\ell} \left(\frac{\int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y} \left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell}\right) dt}{\int_{\underline{y}}^{y} \frac{\partial z(t,A(\ell))}{\partial y} dt} \right) \right). \tag{A.7}$$

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:

$$\min_{\ell,y} \frac{\frac{\partial A(\ell)}{\partial \ell}}{A(\ell)} \geq \left(1 + \frac{\lambda^E}{\delta}\right)^2 \left(\max_{y,\ell} \left(-\frac{\frac{\partial^2 \Gamma(y|Q^{-1}(R(\ell)))}{\partial y \partial p} \mu'(\ell)}{\gamma(y|Q^{-1}(R(\ell)))} \right) + 2\frac{\lambda^E}{\delta} \max_{y,\ell} \left(\frac{\int_{\underline{y}}^{y} \left(-\frac{\partial \Gamma_\ell(t)}{\partial \ell}\right) dt}{y - \underline{y}} \right) \right). \tag{A.8}$$

Note that the first term on the RHS has a well-defined maximum since we are maximizing a continuous function over a compact set and 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.8) holds, i.e., if the elasticity of $A(\ell)$ wrt ℓ on the LHS of (A.8) is sufficiently large. Finally note that when (A.8) 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.8) since the first term on the RHS of (A.8) 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). □

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 $\mathcal{R}$:

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

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

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

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.

A.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

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

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

A.5 Proof of Proposition 5

<u>Preliminaries.</u> 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

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

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

$$\frac{\partial LS(\ell)}{\partial \ell} = \frac{\partial Ls(\underline{y},\ell)}{\partial \ell} + \int_{\underline{y}}^{\overline{y}} \frac{\partial^2 Ls(y,\ell)}{\partial y \partial \ell}(1-\tilde{G}_\ell(y)) + \frac{\partial Ls(y,\ell)}{\partial y}\left(-\frac{\partial \tilde{G}_\ell(y)}{\partial \ell}\right) dy < 0, \tag{A.9}$$

which uses integration by parts and 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.9) 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.9) is negative. Thus, under PAM and large enough labor market frictions, i.e., if φ^E is small enough, $LS(\cdot)$ is decreasing in ℓ. We now provide the details.

Step 1. We focus on the second term in (A.9) 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:

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

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

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

Hence, under the multiplicative production function $z = yA$, we obtain[42]

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

Step 2. As for the third term in (A.9), 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:

$$\frac{\partial Ls(y,\ell)}{\partial y} = (1 - Ls(y,\ell))\frac{2\varphi^E\gamma_\ell(y)}{1+\varphi^E(1-\Gamma_\ell(y))} - Ls(y,\ell)\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{z(y,A(\ell))}. \tag{A.10}$$

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

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.10) to be negative it is therefore sufficient that

$$(1+\varphi^E)^2 \min_{\ell,y}\left(\frac{z(\underline{y},A(\ell)}{z(y,A(\ell))}\right) > 2\varphi^E \max_{\ell,y}\left(\gamma_\ell(y)\frac{z(y,A(\ell))}{\frac{\partial z(y,A(\ell))}{\partial y}}\right) + 2\varphi^E + (\varphi^E)^2 \qquad \text{(A.11)}$$

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.11) 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)

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

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

$$\frac{\partial^2 \log \gamma(y|p)}{\partial y \partial p} \geq 2\varphi^E \max_{y,p}\left(-\frac{\partial^2 \Gamma(y|p)}{\partial p \partial y}\right) + 2(\varphi^E)^2 \max_{y,p}\left(\frac{\partial \Gamma(y|p)}{\partial y}\left(-\frac{\partial \Gamma(y|p)}{\partial p}\right)\right), \tag{A.12}$$

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 (i.e., that $\gamma(y|p)$ is log-supermodular in (y, p)). For (A.12) 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.12) 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.9) is negative for sufficiently small φ^E.

Thus, the integrand of (A.9) 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.9) is negative. □

A.6 Proof of Corollary 1

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

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

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 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 firms' 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. We assume that the exogenous functions $B(\cdot)$ and $\delta(\cdot)$ do not vary with ℓ (i.e., $B(\ell) = B = 1$ and $\delta(\ell) = \delta$). This is to tie our hands since there certainly exist some assumptions on how these exogenous objects depend on ℓ that would deliver the result of positive sorting but those conditions would have little to do with the specific mechanism of our model. Moreover, for illustration, we assume that local housing supply is given by $h(\ell) = d(\ell)^\xi$, where ξ is the housing supply elasticity.

Proposition A1. *Suppose $B(\ell) = B = 1$ and $\delta(\ell) = \delta$ and local housing supply is given by $h(\ell) = d(\ell)^\xi$. 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 exists an equilibrium with positive sorting in (p,ℓ).*

Proof. See Appendix SA.3.1. □

C Data and Sample Restrictions

C.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 CZ. If applicable, we adjust the variables to the monthly level, 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. 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.

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

Data Description. We use worker-/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.[43] Surveyed establishments (the 'panel cases') are followed between 2009-2016 and we observe individual-level information for *all* their employees. This individual-level data, which includes workers' gender, education, full-time employment status, gross daily wages and work district, is assembled from official social security records. 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.[44]

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

[43] Because we only observe data at the level of the establishment, we use 'establishments' and 'firms' interchangeably.
[44] If a new spell starts in the middle of a month, we assign the month to the longest spell within the 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 $t-2$), 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 scenario (ii), we consider it likely that the new job was already lined up when the worker left the previous one. Data Source: Linked Employer-Employee Data (LIAB).

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

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

Labor Market Transition Rates. In our regression analysis, we construct measures of 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.[45] 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).

C.3 Variables from Other Data Sources

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

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

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

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

C.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 SA.4.1, 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 CZs into East or West Germany based on whether the districts they consist of belong to Eastern or Western states. Many CZs contain more than one district; however, there are no commuting zones containing districts from both East and West Germany.

Urban-Rural. We categorize CZs 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

[45] 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.

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 Empirical Analysis

D.1 Spatial Wage Inequality: Real versus Nominal

In our main analysis, we use—consistent with our theory—nominal wages to measure inequality both within and across locations. In Table A.1 we report average wage premia after deflating wages with local CPIs. Doing so decreases spatial inequality by about one third.

Table A.1: 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.

D.2 Decomposition of Life-Time Earnings

In this exercise, we study the impact of heterogeneous job ladders across space on spatial inequality in lifetime earnings. We compare lifetime 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 lifetime income inequality across space keeping job ladders *the same* across regions.

Table A.2: Decomposition of Lifetime Earnings Inequality in Top versus Bottom 25% of Local Labor Markets

	Top 25%	Bottom 25%
Wage Growth (total)	0.595	0.362
Wage Growth of Stayers	0.540	0.364
Wage Growth of Unemployed	0.407	0.313
Wage Growth of EE Movers	0.712	0.381
Starting Wage (Monthly, €)	3670.44	2365.23
Wage after 15 Years (Monthly, €)	5854.35	3221.44
After-15-years Spatial Wage Inequality (data)	1.817	
After-15-years Spatial Wage Inequality (counterfactual, same EE wage growth)	1.664	
After-15-years Spatial Wage Inequality (counterfactual, same Stayer wage growth)	1.720	
Contribution of Job Ladder Differences to Spatial Wage Inequality	0.187	
Contribution of Stayer Wage Growth Differences to Spatial Wage Inequality	0.119	

Notes: Data source: LIAB. Top and bottom 25% of local labor markets (CZs) are categorized based on GDP per capita. The last two rows report the percentage differences between the actual inequality in lifetime income (row 7) and counterfactual inequality in lifetime income (row 8 and 9), that is, 19% $\sim (81.7 - 66.4)/81.7$, and 12% $\sim (81.7 - 72.0)/81.7$.

The results are in Table A.2. If poor and rich regions had the same job ladder, spatial inequality in lifetime income would be around 19% lower than under the heterogeneous job ladders that we observe in the data, i.e., the rich region would be characterized by only 66.4% higher wages than the poor one, instead of the observed 81.7% (∼5854.35/3221.44-1).

E Identification

We prove identification of our model under the following assumption:

Assumption A1. *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:

$$\begin{aligned}
\lambda^E(\ell) &= \frac{M(\mathcal{V}(\ell),\mathcal{U}(\ell))}{(u(\ell)+\kappa(1-u(\ell)))L(\ell)}\kappa \\
&= M(\mathcal{V}(\ell),\mathcal{U}(\ell))\frac{\kappa(1-u(\ell))}{u(\ell)+\kappa(1-u(\ell))}\frac{1}{(1-u(\ell))L(\ell)} \\
&= \mathcal{A}\mathcal{V}(\ell)^{\frac{1}{2}}\big(u(\ell)+\kappa(1-u(\ell))L(\ell)\big)^{\frac{1}{2}}\frac{\lambda^E(\ell)}{\delta(\ell)+\lambda^E(\ell)}\frac{\delta(\ell)+\lambda^U(\ell)}{\lambda^U(\ell)}\frac{1}{L(\ell)} \\
\Rightarrow \quad L(\ell) &= \mathcal{A}^2\frac{\delta(\ell)+\lambda^U(\ell)}{\delta(\ell)+\kappa\lambda^U(\ell)}\left(\frac{1}{\lambda^U(\ell)}\right)^2. \qquad \text{(A.13)}
\end{aligned}$$

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

$$L(\ell) = \left(1+\frac{\delta(\ell)}{\lambda^U(\ell)}\right)\bar{l}(\ell). \qquad \text{(A.14)}$$

Equalizing (A.13) and (A.14), 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:

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

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,

$$\tau d(\ell)h(\ell) = w^U(\ell)u(\ell)L(\ell), \tag{A.15}$$

housing market clearing,

$$h(\ell) = \omega\frac{w^U(\ell)}{d(\ell)}u(\ell)L(\ell) + \omega\frac{\mathbb{E}[w(y,\ell)|\ell]}{d(\ell)}(1-u(\ell))L(\ell), \tag{A.16}$$

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

$$\tau = \frac{1}{\omega}\frac{\mathcal{T}}{\frac{\sum_{\hat{\ell}}(1-u(\hat{\ell}))L(\hat{\ell})}{\sum_{\hat{\ell}}u(\hat{\ell})L(\hat{\ell})} + \mathcal{T}}. \tag{A.17}$$

We computed (A.17) 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.15) and the second one uses a combination of (A.15) and housing market clearing (A.16), 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.17).

Equation (A.17) pins down τ given the observed housing expenditure share and replacement rate $(\omega, \mathcal{T})$. For each location ℓ and given τ, equations (A.16)–(A.15) 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:

$$\rho V^U(\ell) = d(\ell)^{-\omega} B(\ell) w^U(\ell) + \tilde{b}(\ell) + d(\ell)^{-\omega} B(\ell) \lambda^U(\ell) \left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right], \tag{A.18}$$

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 $w^R(\ell) = z(\underline{y}, A(\ell))$:

$$\tilde{b}(\ell) = d(\ell)^{-\omega} B(\ell) \left(z(\underline{y}, A(\ell)) - w^U(\ell) - (\lambda^U(\ell) - \lambda^E(\ell)) \left[\int_{w^R(\ell)}^{\overline{w}} \frac{1 - F_\ell(t)}{\delta(\ell) + \lambda^E(\ell)(1 - F_\ell(t))} dt \right] \right). \tag{A.19}$$

Plug $\tilde{b}(\ell)$ back into V^U in (A.18), 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). □

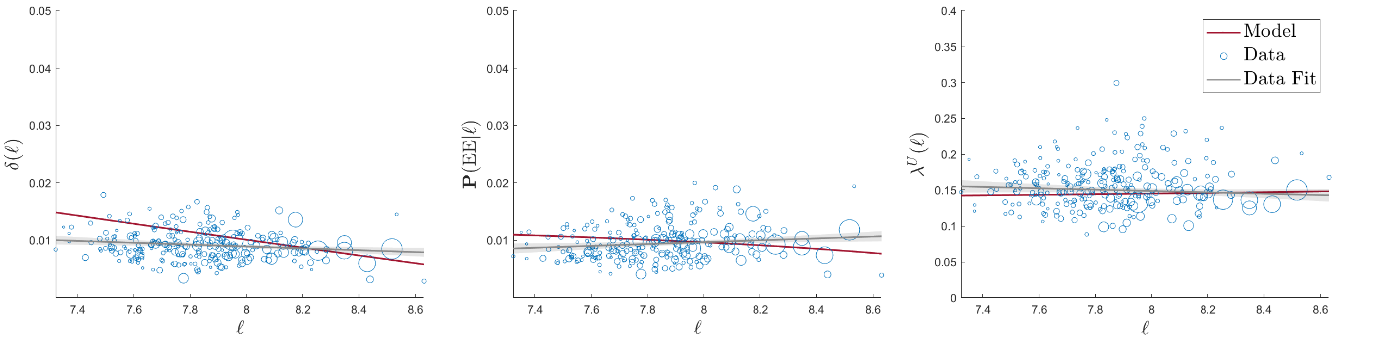
F Estimation Results

Table A.3: 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)

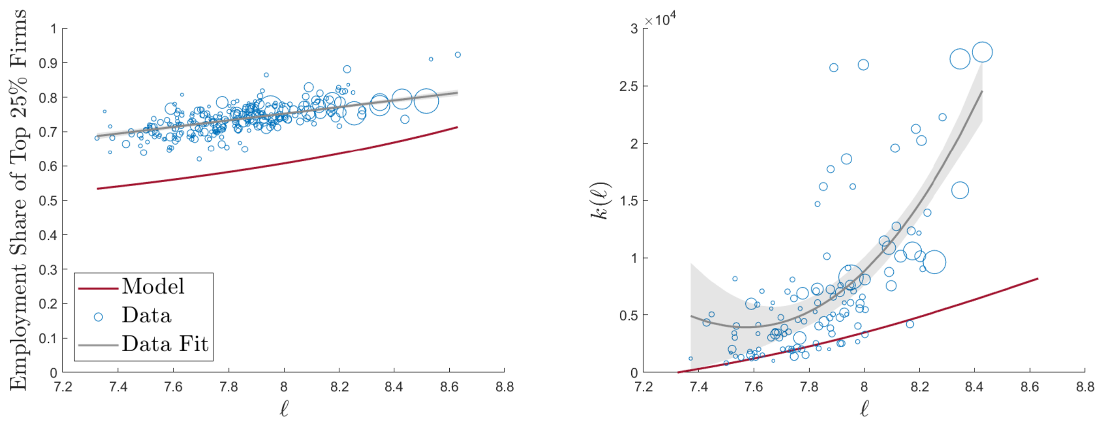
Notes: For details on data sources and variable definitions, see Appendix C. To obtain the relative matching efficiency of employed workers, κ, we use Germany-wide job-finding rates λ^E based on monthly EE transition rates, counting a transition only for those workers who move to a new firm with higher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

Figure A.1: Over-Identification: Transition Rates δ, $\mathbb{P}(EE|\ell)$ and λ^U in Data and Model



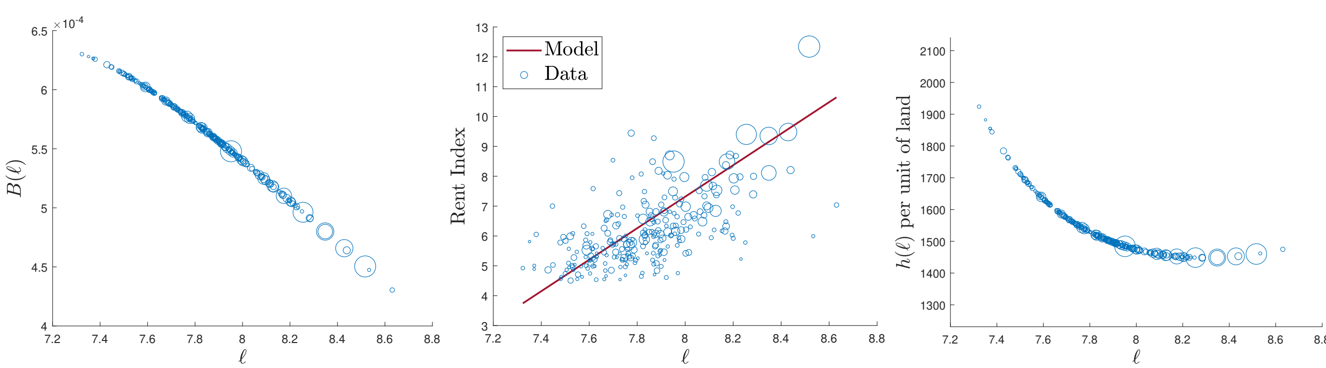
Notes: Data Sources: LIAB. For details on how job destruction rate δ (left panel) and UE transition rate λ^U (right panel) are constructed, see Appendix C.2. In the middle panel, we display the probability of an EE transition, $\mathbb{P}(EE|\ell)$. In the model, workers change jobs if they receive an offer (which happens at rate $\lambda^E(\ell)$) and if its wage exceeds their current one. To compute the data counterpart, we thus multiply the observed EE probability by 0.597, which represents the fraction of EE moves associated with a wage gain. 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.2: 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 C.2 and C.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.3: Additional Parameter Estimates: Location Preference Schedule (left); Housing Supply (right) Obtained from Residential Rents (middle)



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

Table A.4: 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 C.1 and C.3 for the definition of the independent variables.

Firm Sorting and Spatial Inequality: Supplementary Appendix

by Ilse Lindenlaub Ryungha Oh Michael Peters

- FOR ONLINE PUBLICATION -

Throughout this Supplementary Material, we indicate figures, tables, and equations within this appendix by SA.#. In turn, figures, tables, and equations from the main paper are denoted by just 1,2,.... Figures, tables, and equations from the main appendix are denoted by A.#.

SA.1 Baseline Model: Derivations

SA.1.1 Alternative Formulation of Wage-Posting Problem

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

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

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.[46] Using firm size expression (SA.2) (Appendix SA.1.2), we obtain (4).

SA.1.2 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):

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

[46] 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 (SA.1).

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

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

SA.1.3 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,

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

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

Then:

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

Plugging (SA.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 $\rho=0$ (as stated in footnote 5), we obtain wage function (5) from (SA.4).

SA.1.4 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:

$$k(\ell) = \bar{k} + \int_{\underline{\ell}}^{\ell} \int_{\underline{y}}^{\overline{y}} \frac{\partial^2 \tilde{J}(y,\hat{\ell})}{\partial y \partial \ell} \left(1-\Gamma(y|\mu(\hat{\ell}))\right) dy d\hat{\ell},$$

where $\bar{k}$ 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.

SA.1.5 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.

SA.2 Additional Theoretical Results and Proofs: Baseline Model

SA.2.1 Proposition 1 for the Case of Negative Sorting

We complement the case of positive sorting from the text with the analysis of negative sorting. In contrast to the case of positive sorting, which is optimal if $\overline{J}(p,\ell)$ is strictly supermodular in (p,ℓ), optimal sorting is negative if $\overline{J}(p,\ell)$ is strictly submodular. The sufficient conditions for NAM in terms of primitives can be summarized as follows:

Proposition SA1 (Negative Spatial Sorting of Firms)**.** *If z is strictly submodular, and either the productivity gains from sorting into higher ℓ are sufficiently small, or the competition forces are sufficiently small (sufficiently small φ^E), then any equilibrium features negative sorting in (p,ℓ).*

Proof. The proof follows the steps of the one of Proposition 1, which is why we are brief. To derive sufficient conditions under which negative sorting is optimal (Step 1), i.e., under which $\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell}$ is (strictly) negative, it suffices that the integrand of this cross-partial is negative for all $y \in [\underline{y}, \overline{y}]$ and strictly so for some set of positive measure of y. In turn, for this it is sufficient that

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

Note that here, in contrast to the case of PAM, workers anticipate negative sorting $\frac{\partial \Gamma_\ell(y)}{\partial \ell} > 0$ and so the RHS is negative, implying that the LHS of the inequality needs to be *sufficiently negative.* Following similar steps as for PAM, the sufficient condition for NAM in terms of primitives reads:

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

We again define uniform bounds (just swapping min and max due to the sign changes)

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

A sufficient condition for $\overline{J}$ to be submodular in (ℓ,p) is therefore $\varepsilon^N < 2\varphi^E t^N$.

In Step 2, we follow the same approach as for Proposition 1 to show that any optimal assignment satisfies NAM under the premise. Toward a contradiction, suppose that there is PAM for at least one pair of firms and locations. It is then straightforward to show that under this conjecture, $\overline{J}$ is strictly *sub*modular when evaluated at this pair. Thus, there exists a blocking pair to PAM, rendering this assignment non-optimal.

In Step 3, as for Proposition 1, it follows that any equilibrium features negative sorting, where we combine the insights from Steps 1 and 2 with the properties of the distributions (R, Q). □

Remark 1. Note that under the sufficient conditions for negative sorting in Proposition SA1, $\partial k/\partial \ell < 0$, as locations with higher ℓ are less attractive to firms. This ensures that ℓ is chosen by a strictly *lower* p than $\hat{\ell}$ when $\ell > \hat{\ell}$, so $k(\ell)$ is almost-everywhere differentiable. Also, $k(\ell)$ is continuous since any jumps would lead to some profitable deviation by some ℓ near the jump. Indeed, the land price is again unique up to an additive constant $\bar{k}$, where in this case $\bar{k}$ needs to be high enough to ensure that the individual rationality condition for all landowners holds, $k(\ell) \geq 0$ for all ℓ, i.e.,

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

Remark 2. Under the conditions of Proposition SA1 existence of equilibrium follows from the same steps as in Proposition 2, just replace supermodularity with submodularity of $\mathcal{J}$ and note that for any $\ell > \ell'$, $\mu(\ell) < \mu(\ell')$. Uniqueness also follows from the same arguments as in Proposition 2.

SA.3 Theoretical Results and Proofs: Model Extensions

In this appendix, we discuss several extensions of our baseline model. In Section SA.3.1 we provide the proof of Proposition A1, which deals with the case of labor mobility and housing. In Section SA.3.2, we endogenize location productivity $A(\ell)$ by allowing for spillovers across firms. In Section SA.3.3, we allow firms to decide how many vacancies to post, which endogenizes the local job finding/filling rate. In section SA.3.4, we consider endogenous land supply. In all these extensions, we derive sufficient conditions for *positive* sorting of firms across space; the case of negative sorting is similar and omitted for brevity.

SA.3.1 Labor Mobility and Residential Housing: Proof of Proposition A1

We here present the proof of Proposition A1, stated in Appendix B.

Proof. We will provide sufficient conditions for positive sorting to be an equilibrium.

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

$$\frac{\partial^2 \bar{J}(p,\ell)}{\partial \ell \partial p} = \frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p} \lambda^F(\ell) + \frac{\partial \hat{J}(p,\ell)}{\partial p} \frac{\partial \lambda^F(\ell)}{\partial \ell}. \tag{SA.5}$$

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

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

and then compute its derivatives:

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

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

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

Plugging these derivatives into (SA.5), we can write (SA.5) as a single integral. Then, a sufficient condition for (SA.5) 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 obtain the following sufficient condition for PAM:

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\left(\lambda^E(\ell)\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right) + \frac{\partial \lambda^E(\ell)}{\partial \ell}(1-\Gamma_\ell(y))\right)}{\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))} - \frac{\frac{\partial \lambda^F(\ell)}{\partial \ell}}{\lambda^F(\ell)}. \tag{SA.6}$$

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, (SA.6) 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.13) in Appendix E)

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

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. Setting $t = w(\mathcal{R}, \ell)$ and using $F_\ell(t) = \mathcal{R}$, a change of variables yields:

$$\rho V^U(\ell) = \tilde{d}(\ell)^{-\frac{\omega}{1+\xi}} \left(z(\underline{y}, A(\ell)) + \lambda^E(\ell) \left[\int_0^1 \frac{1-\mathcal{R}}{\delta + \lambda^E(\ell)(1-\mathcal{R})} \frac{\partial w(\mathcal{R}, \ell)}{\partial \mathcal{R}} d\mathcal{R} \right] \right). \tag{SA.7}$$

We now differentiate value (SA.7) wrt ℓ for a fixed $\lambda^U(\ell) = \lambda^U$ (and thus fixed $\lambda^E = \kappa\lambda^U$):

$$\begin{aligned} \frac{\partial \rho V^U}{\partial \ell}\bigg|_{\lambda^U(\ell)=\lambda^U} &= \tilde{d}(\ell)^{-\frac{\omega}{1+\xi}} \times \left(\frac{\partial z}{\partial A} \frac{\partial A(\ell)}{\partial \ell} + \lambda^E \int_0^1 \frac{1-\mathcal{R}}{\delta + \lambda^E(1-\mathcal{R})} \frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \ell \partial \mathcal{R}} d\mathcal{R} \right) \\ &\quad - \frac{\omega}{1+\xi} \tilde{d}^{-\frac{\omega}{1+\xi}-1} \frac{\partial \tilde{d}(\ell)}{\partial \ell} \times \left(z(\underline{y}, A(\ell)) + \lambda^E \left[\int_0^1 \frac{1-\mathcal{R}}{\delta + \lambda^E(1-\mathcal{R})} \frac{\partial w(\mathcal{R}, \ell)}{\partial \mathcal{R}} d\mathcal{R} \right] \right). \end{aligned} \tag{SA.8}$$

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 (SA.8) 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 $(\mathcal{R}, \ell)$:

$$\begin{aligned} w(\mathcal{R}, \ell) &= z(\Gamma_\ell^{-1}(\mathcal{R}), A(\ell)) - [\delta + \lambda^E(1-\mathcal{R})]^2 \int_0^{\mathcal{R}} \frac{\frac{\partial z(\Gamma_\ell^{-1}(x), A(\ell))}{\partial y}}{[\delta + \lambda^E(1-x)]^2} \frac{1}{\gamma_\ell(\Gamma_\ell^{-1}(x))} dx \\ \frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \mathcal{R} \partial \ell} &= 2\frac{\lambda^E}{\delta} \left(1 + \frac{\lambda^E}{\delta}(1-\mathcal{R}) \right) \frac{\partial}{\partial \ell} \int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})} \frac{\frac{\partial z(t, A(\ell))}{\partial y}}{(1 + \frac{\lambda^E}{\delta}(1-\Gamma_\ell(t)))^2} dt. \end{aligned} \tag{SA.9}$$

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 (SA.9), $\frac{\partial z(y, A(\ell))}{\partial y} / (1 + \frac{\lambda^E}{\delta}(1 - \Gamma_\ell(y)))^2$, 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 (SA.8) is positive.

In turn, to see that the second line of (SA.8) is negative note that

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

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, taking the same limit of the first line shows that it remains positive:

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

Thus, by continuity of V^U in ξ, there exists a finite $\widehat{\xi}$ such that for $\xi>\widehat{\xi}$, the positive effect stemming from the first line of (SA.8) dominates the negative effect stemming from the second line, which renders (SA.8) positive. Thus, 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 $\tilde{d}(\ell)^{-\omega/(1+\xi)}$ in (SA.7). 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 (SA.6) negative.

For (SA.6) to hold, it then suffices that the first (and the only positive) term on the RHS—given by $2\lambda^E(\ell)\big(-\frac{\partial\Gamma_\ell}{\partial\ell}\big)/(\delta+\lambda^E(\ell)(1-\Gamma_\ell(y)))$—is bounded and "dominated" by the LHS. Note that $\lambda^E(\cdot)$ is implicitly defined by (SA.7), where, in equilibrium, V^U is a number that no longer depends on ℓ. If there is PAM, $\mu'(\ell)>0$, all functions in (SA.7), i.e. $(\tilde{d},z,\partial w/\partial\mathcal{R})$, are continuous in ℓ on $\ell\in[\underline{\ell},\overline{\ell}]$ (see (SA.9)), and thus $\lambda^E(\cdot)$ inherits this property. It follows that $\lambda^E(\cdot)$ is bounded and, as above, we denote its upper bound by $\overline{\lambda}^E=\max_\ell\lambda^E(\ell)$.

We now show that this implies that the first term on the RHS of (SA.6) is bounded from above. Recall that $-\frac{\partial\Gamma_\ell(y)}{\partial\ell}=-\frac{\partial\Gamma(y|\mu(\ell))}{\partial\ell}=-\frac{\partial\Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}$ under (the conjecture of) positive sorting and define

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

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, (SA.6) holds if

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

which holds if

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\lambda^E(\ell)}{\delta}\left(-\frac{\partial \Gamma_\ell(y)}{\partial \ell}\right) \text{ for all } (y,\ell),$$

which holds if $\varepsilon^P > 2\frac{1}{\delta}\tilde{t}^P$. Thus, positive sorting is optimal for firms if ε^P is large enough or if $1/\delta$ is small enough. These conditions ensure that (i) inequality (SA.6) holds; and thereby that (ii) $\Gamma_\ell^{-1}(\mathcal{R})$ is differentiable and increasing in ℓ and the integrand of (SA.9) is increasing in ℓ, all of which we had postulated above.

That an equilibrium with PAM exists then follows from the steps in the first part of Proposition 2, i.e., from the construction of a fixed point in Γ_ℓ (where Γ_ℓ satisfies positive sorting as shown above), when appropriately adjusting $\overline{J}(p,\ell)$ and $k(\ell)$ to this setting with labor mobility. □

SA.3.2 Endogenous Spillovers

While we can pursue the analysis with a general spillover function, a natural specification is

$$A(\ell) = \int_{\underline{y}}^{\overline{y}} (1-\Gamma_\ell(y))dy, \tag{SA.10}$$

since, for $\underline{y} = 0$, this is equivalent to $A(\ell) = \int_{\underline{y}}^{\overline{y}} y d\Gamma_\ell(y)$ and productivity spillovers take the form of the *average* firm productivity in a location. For concreteness, we will assume:

Assumption SA1. *Productivity in location ℓ is endogenous and given by* (SA.10).

Note that ex ante, before any sorting takes place, location index ℓ carries no information about productivity as all locations are identical in this dimension. Thus, the ordering of ℓ is arbitrary, but land distribution R over any given ordering $[\underline{\ell}, \overline{\ell}]$ still indicates (ex ante) heterogeneity of locations, whereby some of them are in scarce supply compared to others. Ex post, after firms sort into locations, the index ℓ will also indicate heterogeneity in location productivity, determined by the productivity of firms that settle there.

For $\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell} > 0$ to obtain when agents conjecture positive sorting (and hence for positive sorting to be optimal), we can again unpack (13) from the baseline model and obtain the following sufficient condition:

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)}}{\frac{\partial z(y,A(\ell))}{\partial y}} \int_{\underline{y}}^{\overline{y}} -\frac{\partial \Gamma_\ell(y)}{\partial \ell} dy$$
$$\geq \quad \frac{2\lambda^E}{\delta + \lambda^E(1-\Gamma(y|Q^{-1}(R(\ell)))} \left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\right) \frac{r(\ell)}{q(Q^{-1}(R(\ell)))}. \qquad \text{(SA.11)}$$

The main change relative to the baseline model is that differences in location productivity are endogenous. In particular, under positive sorting in (p, ℓ), productivity A increases in ℓ because these locations have access to better firms: $\frac{\partial A(\ell)}{\partial \ell} = \int_{\underline{y}}^{\overline{y}} -\frac{\partial \Gamma_\ell(y)}{\partial \ell} dy > 0$. If this location productivity advantage, along with the impact on firms' marginal productivity, is large enough relative to the cost of more severe poaching competition, highly productive firms (those with high-p) indeed settle into high-ℓ locations—similar to the baseline model.

We now state the sorting result under endogenous spillovers formally. To this end, we re-define the minimum productivity gains from sorting into high-ℓ locations as

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

where $A(\ell) = \int_{\underline{y}}^{\overline{y}} 1 - \Gamma(y|Q^{-1}(R(\ell))dy$. Note that under our assumptions $\varepsilon^P > 0$.

Proposition SA2. *Suppose that Assumption SA1 holds. If z is strictly supermodular, and either the productivity gains from sorting into higher ℓ, ε^P, are sufficiently large, or the competition forces φ^E are sufficiently small, then there exists an equilibrium with positive sorting in (p, ℓ).*

The proof of this result resembles Step 1 in the proof of Proposition 1 and the first part (existence) of Proposition 2, where we note that any optimal Γ_ℓ uniquely pins down the spillovers $A(\ell) = \int(1 - \Gamma_\ell(y))dy$. To avoid repetition, we omit the details.

Remark. Note that while an equilibrium with positive sorting exists, it will no longer be unique as far as the firm-location allocation is concerned. This is common under endogenous spillovers, since the coordination of agents affects the equilibrium. Both positive or negative sorting in (p, ℓ) can be self-sustained under identical primitives.

SA.3.3 Endogenous Vacancy Posting

To allow for vacancy posting, we assume that, when a firm of type p chooses a location ℓ, it also decides how many vacancies, $v(p, \ell)$, to post subject to a vacancy posting cost $c(v)$. Thus, firms decide about vacancies before drawing ex post productivity y.

With endogenous vacancy posting, meeting rates $\lambda^F(\ell)$ and $\lambda^E(\ell)$ depend on ℓ. We assume that total meetings between workers and firms in location ℓ are given by

$$\mathcal{M}(\mathcal{V}(\ell),\mathcal{U}(\ell)) = \mathcal{A}\mathcal{V}(\ell)^{\alpha}\mathcal{U}(\ell)^{1-\alpha}, \tag{SA.12}$$

where $\mathcal{V}(\ell)$ is the measure of vacancies in ℓ, $\mathcal{A}$ is matching efficiency, and α is the elasticity of matches with respect to vacancies. In turn, $\mathcal{U}(\ell)$ is the measure of job searchers in ℓ. As before, we define market tightness in location ℓ by $\theta(\ell) = \frac{\mathcal{V}(\ell)}{\mathcal{U}(\ell)}$. Then, the meeting rates are given by $\lambda^F(\ell) = \mathcal{A}\theta(\ell)^{\alpha-1}, \lambda^U(\ell) = \mathcal{A}\theta(\ell)^{\alpha}$, and $\lambda^E(\ell) = \kappa\mathcal{A}\theta(\ell)^{\alpha}$. We impose the following assumptions on matching function and vacancy costs.

Assumption SA2.

1. *Total meetings in location ℓ are given by (SA.12) with* $0 < \alpha < 1$.
2. *Vacancy posting cost c is C^2 with $c' > 0, c'' > 0$, $c'(0) = 0$, and* $\lim_{v\to 0} \frac{vc''(v)}{c'(v)} := \underline{c} > 0$.

The total measure of vacancies, $\mathcal{V}(\ell)$, is determined by the vacancy posting decision of firms in ℓ:

$$\mathcal{V}(\ell) = \int_{\underline{p}}^{\overline{p}} v(p,\ell)m_p(p|\ell)dp.$$

The effective measure of workers searching for a job in location ℓ is

$$\mathcal{U}(\ell) = u(\ell) + \kappa(1-u(\ell)) = \frac{\delta}{\delta+\lambda^U(\ell)} + \kappa\frac{\lambda^U(\ell)}{\delta+\lambda^U(\ell)}.$$

Plugging both $\mathcal{V}(\ell)$ and $\mathcal{U}(\ell)$ into $\theta(\ell) = \frac{\mathcal{V}(\ell)}{\mathcal{U}(\ell)}$ and simplifying yields

$$\theta(\ell)\frac{\delta+\kappa\mathcal{A}\theta(\ell)^{\alpha}}{\delta+\mathcal{A}\theta(\ell)^{\alpha}} = \mathcal{V}(\ell). \tag{SA.13}$$

Equation (SA.13) implicitly determines the equilibrium local labor market tightness, $\theta(\ell)$, as a function of the measure of vacancies, $\mathcal{V}(\ell)$, in any given market ℓ. Note that, under Assumption SA2.1., θ is strictly increasing in $\mathcal{V}(\ell)$. To see this, differentiate (SA.13) with respect to θ:

$$\frac{\partial\mathcal{V}(\ell)}{\partial\theta} \overset{\text{s}}{=} \kappa\mathcal{A}^2(\theta^{\alpha})^2 + (\kappa+1+\alpha(\kappa-1))\mathcal{A}\delta\theta^{\alpha} + \delta^2,$$

which is positive when $\alpha < 1$ and achieves its minimum (equal to δ^2) at $\theta = 0$.

The expected value of firm p of settling in location ℓ is now given by:

$$\overline{J}(p,\ell) = \max_{v\geq 0}\{\lambda^F(\ell)v\hat{J}(p,\ell) - c(v)\} - k(\ell)$$
$$\text{with } \hat{J}(p,\ell) = \delta \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{\frac{\partial z(t,A(\ell))}{\partial y}}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(t))]^2} dt d\Gamma(y|p).$$

We now state our main result of this extension of our model.

Proposition SA3. *If z is strictly supermodular, the productivity gains from sorting into higher ℓ are sufficiently large, and the competition forces are sufficiently small (i.e., $1/\delta$ is sufficiently small), then there exists an equilibrium with positive sorting in (p,ℓ).*

Proof. Conjecture that positive sorting between firms and locations is optimal, as in the baseline model. The firm's first-order condition with respect to the vacancy posting rate is given by

$$\lambda^F(\ell)\hat{J}(p,\ell) = c'(v(p,\ell)). \tag{SA.14}$$

This equation implicitly solves for the optimal vacancy posting rate of firm p in location ℓ, $v(p,\ell)$. We can then compute expected value $\overline{J}(p,\ell)$ and its derivatives as:

$$\begin{aligned}
\bar{J}(p,\ell) &= \lambda^F(\ell)v(p,\ell)\hat{J}(p,\ell) - c(v(p,\ell)) - k(\ell) \\
\frac{\partial \bar{J}(p,\ell)}{\partial p} &= \frac{\partial \hat{J}(p,\ell)}{\partial p}\lambda^F(\ell)v(p,\ell) \\
\frac{\partial^2 \bar{J}(p,\ell)}{\partial \ell \partial p} &= \frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p}\lambda^F(\ell)v(p,\ell) + \frac{1}{c''(v(p,\ell))}\frac{\partial \lambda^F(\ell)\hat{J}(p,\ell)}{\partial \ell}\frac{\partial \hat{J}(p,\ell)}{\partial p}\lambda^F(\ell) + \frac{\partial \hat{J}(p,\ell)}{\partial p}\frac{\partial \lambda^F(\ell)}{\partial \ell}v(p,\ell).
\end{aligned} \tag{SA.15}$$

The second line uses the envelope theorem. In the third line, we use $\frac{\partial v(p,\ell)}{\partial \ell} = \frac{1}{c''(v(p,\ell))}\frac{\partial \lambda^F(\ell)\hat{J}(p,\ell)}{\partial \ell}$, obtained by differentiating (SA.14) with respect to ℓ. We will characterize conditions under which (SA.15) is positive so that PAM between (p,ℓ) arises. To that end, we will specify conditions under which, at $p = \mu(\ell)$, $\frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p} > 0$ in the first term (Step 1) and the remaining two terms are also positive (Step 2). In each step, the conditions we specify will invoke the limit $\frac{1}{\delta} \to 0$.

First, we derive a few useful equations. Applying integration by parts to $\hat{J}(p,\ell)$ yields

$$\hat{J}(p,\ell) = \frac{1}{\delta}\int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1 + \frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|p))dy. \tag{SA.16}$$

We can then compute the derivatives of $\hat{J}(p,\ell)$ as

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

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

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

Second, we make some observations for the case $\delta \to \infty$ (capturing small competition forces, $\frac{1}{\delta}$).

The main difference compared with the baseline model is the endogeneity of market tightness (and, thus, of $\lambda^E(\ell)$ and $\lambda^F(\ell)$). Therefore, it is important to understand the behavior of local market tightness. Note that, under PAM, $\mathcal{V}(\ell) = v(\mu(\ell),\ell)$, which follows from the definition of $\mathcal{V}(\ell)$. Based on (SA.13), we denote $v(\theta(\ell)) := v(\mu(\ell),\ell)$, where v is an increasing function of θ. Moreover, we have $\lim_{\theta\to 0} v(\theta) = 0$, $\lim_{\theta\to 0}\frac{\partial v(\theta)}{\partial \theta} = 1$, and $\lim_{\theta\to 0}\frac{v(\theta)}{\theta} = \lim_{\theta\to 0}\mathcal{U}(\ell) = 1$. Using $\lambda^F(\ell) = \mathcal{A}\theta(\ell)^{\alpha-1}$ and FOC (SA.14), we have

$$\hat{J}(\mu(\ell),\ell) = c'(v(\theta(\ell)))(\mathcal{A}\theta(\ell)^{\alpha-1})^{-1}. \tag{SA.17}$$

If $\delta\to\infty$, then $\hat{J}(p,\ell)\to 0$, which follows from the definition of $\widehat{J}$; see (SA.16). Since $c'(v(\theta))$ and $(\theta^{\alpha-1})^{-1}$ are both strictly increasing in θ and zero at $\theta = 0$, we conclude that $\lim_{\delta\to\infty}\theta(\ell) = 0$ and $\lim_{\delta\to\infty} v(\theta(\ell)) = 0$.

Differentiating (SA.17) with respect to ℓ we obtain (after some algebra) the elasticity of market tightness under PAM. Plugging in the expressions for $\frac{\partial}{\partial \ell}\hat{J}(p,\ell)$, $\frac{\partial}{\partial p}\hat{J}(p,\ell)$, and $\hat{J}(p,\ell)$ from above gives

$$\begin{aligned}\frac{\frac{\partial \theta(\ell)}{\partial \ell}}{\theta(\ell)} = &\left(1+\frac{2\alpha\frac{\lambda^E(\ell)}{\delta}\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}(1-\Gamma_\ell(y))}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^3}(1-\Gamma(y|\mu(\ell)))\,dy}{\delta\mathcal{A}^{-1}\theta(\ell)^{2-\alpha}\frac{\partial v(\theta(\ell))}{\partial \theta}c''(v(\theta(\ell)))+(1-\alpha)\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|\mu(\ell)))\,dy}\right)^{-1}\\
\times &\left(-\frac{\int_{\underline{y}}^{\overline{y}}\left(\frac{2\lambda^E(\ell)\frac{\partial z(y,A(\ell))}{\partial y}\left(-\frac{\partial\Gamma_\ell}{\partial \ell}\right)}{\delta[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^3}\right)(1-\Gamma(y|\mu(\ell)))\,dy}{\delta\mathcal{A}^{-1}\theta(\ell)^{2-\alpha}\frac{\partial v(\theta(\ell))}{\partial \theta}c''(v(\theta(\ell)))+(1-\alpha)\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|\mu(\ell)))\,dy}\right.\\
&\left.+\frac{\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|\mu(\ell)))\,dy+\frac{\partial \mu(\ell)}{\partial \ell}\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}\left(-\frac{\partial}{\partial p}\Gamma(y|\mu(\ell))\right)dy}{\delta\mathcal{A}^{-1}\theta(\ell)^{2-\alpha}\frac{\partial v(\theta(\ell))}{\partial \theta}c''(v(\theta(\ell)))+(1-\alpha)\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|\mu(\ell)))\,dy}\right).\end{aligned} \tag{SA.18}$$

As $\delta \to \infty$, the first line converges to 1 and the second line vanishes. Focus on the third line. In the denominator, the first term is

$$\delta \mathcal{A}^{-1}\theta(\ell)^{2-\alpha}\frac{\partial v(\theta(\ell))}{\partial \theta}c''(v(\theta(\ell))) = \delta c'(\theta(\ell))\mathcal{A}^{-1}\theta(\ell)^{1-\alpha}\frac{c''(v(\theta(\ell)))\theta(\ell)}{c'(v(\theta(\ell)))}.$$

Using (SA.17), this is $\delta \hat{J}(\mu(\ell),\ell)\frac{c''(v(\theta(\ell)))\theta(\ell)}{c'(v(\theta(\ell)))}$, where $\frac{c''(v(\theta(\ell)))\theta(\ell)}{c'(v(\theta(\ell)))}$ converges to $\underline{c}$ under Assumption SA2.2. Thus, we can characterize the limit of the elasticity of market tightness:

$$\lim_{\delta\to\infty}\frac{\frac{\partial\theta(\ell)}{\partial\ell}}{\theta(\ell)} = \frac{\int_{\underline{y}}^{\overline{y}}\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}(1-\Gamma(y|\mu(\ell)))dy + \frac{\partial\mu(\ell)}{\partial\ell}\int_{\underline{y}}^{\overline{y}}\frac{\partial z(y,A(\ell))}{\partial y}\left(-\frac{\partial}{\partial p}\Gamma(y|\mu(\ell))\right)dy}{[1-\alpha+\underline{c}]\int_{\underline{y}}^{\overline{y}}\frac{\partial z(y,A(\ell))}{\partial y}\left(1-\Gamma(y|\mu(\ell))\right)dy} \tag{SA.19}$$

which, for all ℓ, is bounded from above by a positive and finite constant.

We now return to our task of signing (SA.15).

Step 1. We first show that $\frac{\partial^2\hat{J}(p,\ell)}{\partial p\partial\ell} > 0$ along the assignment $p=\mu(\ell)$. It is sufficient to ensure that the following inequality holds if $\delta\to\infty$:

$$\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial\ell}\left(-\frac{\partial\Gamma(y|\mu(\ell))}{\partial p}\right)}{[1+\frac{\kappa\mathcal{A}\theta(\ell)^\alpha}{\delta}(1-\Gamma(y|\mu(\ell)))]^2}dy$$
$$> 2\frac{\kappa\mathcal{A}\theta(\ell)^\alpha}{\delta}\int_{\underline{y}}^{\overline{y}}\frac{\partial z(y,A(\ell))}{\partial y}\left(-\frac{\partial\Gamma(y|\mu(\ell))}{\partial p}\right)\left(\frac{\partial\mu(\ell)}{\partial\ell}\left(-\frac{\partial\Gamma(y|\mu(\ell))}{\partial p}\right)+\alpha\frac{\frac{\partial\theta(\ell)}{\partial\ell}}{\theta(\ell)}(1-\Gamma(y|\mu(\ell)))\right)dy.$$

This holds as $\delta\to\infty$ since the RHS vanishes (recall that we showed in (SA.19) that the elasticity of market tightness is bounded from above as $\delta\to\infty$) while the LHS remains strictly positive.

Step 2. Next, we show that the sum of the last two terms in (SA.15) is positive when $p=\mu(\ell)$. After some algebra, this sum becomes

$$\frac{1}{c''(v(p,\ell))}\frac{\partial\lambda^F(\ell)\hat{J}(p,\ell)}{\partial\ell}\frac{\partial\hat{J}(p,\ell)}{\partial p}\lambda^F(\ell) + \frac{\partial\hat{J}(p,\ell)}{\partial p}\frac{\partial\lambda^F(\ell)}{\partial\ell}v(p,\ell)$$
$$= \frac{\partial\hat{J}(p,\ell)}{\partial p}\frac{(\lambda^F(\ell))^2}{c''(v(p,\ell))}\left((\alpha-1)\frac{\frac{\partial\theta(\ell)}{\partial\ell}}{\theta(\ell)}\left(\hat{J}(p,\ell)+\frac{c''(v(p,\ell))v(p,\ell)}{\lambda^F(\ell)}\right)+\frac{\partial\hat{J}(p,\ell)}{\partial\ell}\right).$$

Evaluating this equation at $p=\mu(\ell)$ and using (SA.17), we obtain

$$\frac{\partial\hat{J}(\mu(\ell),\ell)}{\partial p}\frac{(\lambda^F(\ell))^2}{c''(v(\theta(\ell)))}\left((\alpha-1)\frac{\frac{\partial\theta(\ell)}{\partial\ell}}{\theta(\ell)}\left(1+\frac{c''(v(\theta(\ell)))v(\theta(\ell))}{c'(v(\theta(\ell)))}\right)\hat{J}(\mu(\ell),\ell)+\frac{\partial\hat{J}(\mu(\ell),\ell)}{\partial\ell}\right).$$

A sufficient condition for this to be positive if $\delta\to\infty$ is that the term in parentheses is positive, i.e.,

$$\delta\frac{\partial \hat{J}(\mu(\ell),\ell)}{\partial \ell} > (1-\alpha)\frac{\frac{\partial\theta(\ell)}{\partial\ell}}{\theta(\ell)}(1+\underline{c})\delta\hat{J}(\mu(\ell),\ell),$$

where we used Assumption SA2.2. Observing that both $\delta\hat{J}(\mu(\ell),\ell)$ and $\delta\frac{\partial J(p,\ell)}{\partial\ell}$ converge to some positive numbers as we consider $\delta\to\infty$, the above inequality becomes in this limit

$$\left(1-\frac{(1-\alpha)(1+\underline{c})}{1-\alpha+\underline{c}}\right)\int\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial\ell}(1-\Gamma(y|\mu(\ell)))dy$$
$$> \frac{(1-\alpha)(1+\underline{c})}{1-\alpha+\underline{c}}\frac{\partial\mu(\ell)}{\partial\ell}\int_{\underline{y}}^{\overline{y}}\frac{\partial z(y,A(\ell))}{\partial y}\left(-\frac{\partial}{\partial p}\Gamma(y|\mu(\ell))\right)dy,$$

where we substituted in (SA.19). This holds if

$$\int\frac{\partial^2 z(y,A(\ell))}{\partial y\partial A(\ell)}\frac{\partial A(\ell)}{\partial\ell}(1-\Gamma(y|\mu(\ell)))dy > \frac{(1-\alpha)(1-\underline{c})}{\alpha\underline{c}}\frac{\partial\mu(\ell)}{\partial\ell}\int_{\underline{y}}^{\overline{y}}\frac{\partial z(y,A(\ell))}{\partial y}\left(-\frac{\partial}{\partial p}\Gamma(y|\mu(\ell))\right)dy,$$

which is a condition on primitives (recall that $\mu(\ell)=Q^{-1}(R(\ell))$). As in our baseline model, we can define the maximum of the RHS over ℓ as t^V and the minimum of the LHS over ℓ as ε^V. Then, the inequality holds if $\varepsilon^V > t^V$, i.e., if complementarities of z in (y,ℓ) are large enough.

From Steps 1 and 2, we conclude that $\frac{\partial^2\bar{J}(p,\ell)}{\partial\ell\partial p}$ is positive along $p=\mu(\ell)$, which shows that positive sorting is indeed optimal under the premise.

That an equilibrium with PAM exists then follows from the steps in the first part of Proposition 2, i.e., from the construction of a fixed point in Γ_ℓ (where Γ_ℓ satisfies positive sorting as shown above), when appropriately adjusting $\overline{J}(p,\ell)$ and $k(\ell)$ to this setting with vacancy posting. □

Remark. The economic intuition for Proposition SA3 is as follows. Labor market competition is strong in high-ℓ locations not only because there are better firms than in low-ℓ regions (due to positive sorting—as in the baseline model), but also because more productive firms tend to post more vacancies. This new channel increases market tightness in good locations and hence further discourages firms from settling there. Hence, competition in productive locations is amplified by endogenous vacancy posting. To compensate for this stronger competition that arises from both firm composition and congestion, we require the productivity gains from settling into high-ℓ locations to be large enough or, stated differently, competition to be sufficiently muted (through low $1/\delta$), so that PAM can be sustained in equilibrium.

SA.3.4 Endogenous Land Supply

We first describe the environment and equilibrium. We then prove the sorting result.

We maintain from the baseline model that locations can be ranked by productivity and are indexed by $\ell \in [\underline{\ell}, \overline{\ell}]$. Contrary to the baseline model, we now bring to life the *land developers*, who are initially heterogeneous in their ability to do this job $\psi \sim U[0,1]$ (where we assume the uniform distribution for convenience). Land developers are risk neutral.

Before entering the land market, developers face a binary investment choice with stochastic returns: If they invest, they draw the land they need to develop from a stochastically better distribution R_1, compared to when they do not invest (in which case they draw from R_0). Investment is costly, and this cost negatively depends on the land developer's ability ψ. The investment cost is given by a function c, with $c(\psi) \geq 0$ for all ψ, and where c is strictly decreasing and differentiable on $[0,1]$. Further, $c(1) = 0$ and $\lim_{\psi \to 0} c(\psi) = +\infty$.

After developers' draw their location characteristic ℓ and develop the land, firms again match pairwise with locations in a competitive market. If a land developer with ability ψ invests, then his expected payoff is $\int k(\ell) dR_1(\ell) - c(\psi)$; in turn, it is $\int k(\ell) dR_0(\ell)$ if he does not invest (where k is again the land price associated with ℓ). In turn, the firms' payoffs are as in the baseline model.

We now describe the equilibrium. Let $a : [0,1] \to \{0,1\}$ be a measurable investment function, where $a(\psi) = 0$ if a developer with ability ψ does not invest, and $a(\psi) = 1$ if he does. For a given a, the distribution of land ℓ is $R(\cdot, a)$, a mixture of R_1 and R_0 with weights given by the measure of developers who invest and do not invest (see below).

An *equilibrium* consists of an investment function a plus the equilibrium objects from the baseline model $(w, k, m, \Gamma_\ell, G_\ell, u, w^R)$ such that land developers invest optimally in addition to the usual equilibrium requirements. That is, for all ψ, $a(\psi) = 1$ if and only if the net benefit from investing is higher than from not investing, $U_1 - c(\psi) \geq U_0$, where

$$U_i = \int k(\ell) dR_i(\ell), \quad i = 0, 1;$$

is the expected utility from investment choice $i = 0, 1$, taking investment risk into account. We construct an equilibrium as follows. Consider the investment stage. For any investment choices of other developers and for the corresponding land price function in the matching stage, the developer invests if and only if $U_1 - c(\psi) \geq U_0$. Since the land price function k strictly increases in ℓ in the positive sorting equilibrium that we aim to construct, and since R_1 strictly FOSD R_0, we have that $U_1 - U_0 > 0$. Thus, in any equilibrium we have

$$a(\psi) = \begin{cases} 1 & \text{if } \psi \geq \psi^* \\ 0 & \text{if } \psi < \psi^*, \end{cases}$$

where ability threshold $\psi^* \in (0,1)$ characterizes a, and where wlog we have set $a(\psi^*) = 1$. Thus, given the binary nature of the investment decision, in any equilibrium a is characterized by an ability threshold above which developers optimally decide to invest.

For any given investment function a (summarized by threshold ψ^*), we obtain the endogenous land distribution (recall that we assumed that ψ is uniformly distributed):

$$R(\ell, \psi^*) = (1-\psi^*)R_1(\ell) + \psi^* R_0(\ell),$$

and the Walrasian equilibrium of the land market is $(\mu(\cdot, \psi^*), k(\cdot, \psi^*))$, where $\mu(\ell, \psi^*) = Q^{-1}(R(\ell, \psi^*))$ under positive sorting and

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

The conditions for sorting remain similar to those in the baseline model. To see this, note that the firm's location choice problem (when anticipating PAM) is:

$$\max_{\ell} \overline{J}(p, \ell; \psi^*) \quad = \quad \delta\lambda^F \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{\frac{\partial z(t, A(\ell))}{\partial y}}{\left[\delta + \lambda(1-\Gamma(t|\mu(\ell, \psi^*)))\right]^2} dt d\Gamma(y|p) - k(\ell, \psi^*),$$

where it takes the economy-wide investment threshold ψ^* and thus land supply R as given. Using the same definition for ε^P as in the baseline model, we now prove the main result of this extension.

Proposition SA4. *If z is strictly supermodular, and either the productivity gains from sorting into higher ℓ, ε^P, are sufficiently large, or the competition forces φ^E are sufficiently small, then there exists an equilibrium with positive sorting in (p, ℓ).*

Proof. Cross-differentiating $\overline{J}(p, \ell; \psi^*)$ w.r.t. (p, ℓ) yields again:

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

only that the matching function now depends on ψ^*. 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. So, it suffices that for all (e, ℓ, ψ^*)

$$\frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\lambda^E}{\delta + \lambda^E(1 - \Gamma(y|\mu(\ell, \psi^*)))} \left(-\frac{\partial \Gamma}{\partial p} \frac{\partial \mu(\ell, \psi^*)}{\partial \ell} \right).$$

A sufficient condition for this inequality to hold is:

$$\min_{\ell,y} \frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(y,A(\ell))}{\partial y}} > \frac{2\lambda^E}{\delta} \max_{\ell,y,\psi^*} \left(-\frac{\partial \Gamma(y|Q^{-1}(\psi^* R_0(\ell) + (1-\psi^*)R_1(\ell)))}{\partial p} \frac{\psi^* r_0(\ell) + (1-\psi^*) r_1(\ell)}{q(Q^{-1}(\psi^* R_0(\ell) + (1-\psi^*)R_1(\ell)))} \right).$$

We define ε^P as in the baseline model. Note that it exists based on the same arguments as before. Moreover, let

$$t^P := \max_{\ell,y,\psi^*} \left(-\frac{\partial \Gamma(y|Q^{-1}(\psi^* R_0(\ell) + (1-\psi^*)R_1(\ell)))}{\partial p} \frac{\psi^* r_0(\ell) + (1-\psi^*) r_1(\ell)}{q(Q^{-1}(\psi^* R_0(\ell) + (1-\psi^*)R_1(\ell)))} \right) > 0,$$

which is positive and finite since the function we are maximizing is continuous in (ℓ, y, ψ^*), where (ℓ, y, ψ^*) are all defined over compact sets (recall that $\psi^* \in [0,1]$). Hence, the familiar sufficient condition renders $\overline{J}$ supermodular in this context: $\varepsilon^P > 2\varphi^E t^P$. A sufficiently high ε^P or low φ^E makes positive sorting optimal—as in the baseline model.

That an equilibrium with PAM exists then follows from the steps in the first part of Proposition 2, i.e., from the construction of a fixed point in Γ_ℓ (where Γ_ℓ satisfies positive sorting as shown above), when appropriately adjusting $\overline{J}(p, \ell)$ and $k(\ell)$ to this setting with endogenous land. □

Remark. Note that despite the stylized setting, this extension captures the important feature that the benefits of land investment—and therefore land supply—are guided by land price $k(\cdot)$, which in turn reflects the demand for land with different characteristics. For instance, if k is strongly increasing in ℓ, reflecting that high-quality land is relatively scarce, this encourages more developers to invest and so the land supply in high-ℓ locations increases, which affects land distribution R. So, despite consistently focusing on the case of pure positive sorting in (ℓ, p), one could use this extension to analyze how land supply R changes with a subsidy to invest (captured by a shift or curvature change of the investment cost function) or with varying land demand (captured by changes in Q) or productivity (A). Changes in R will then affect the matching between firms and locations, and thus spatial sorting and inequality.

SA.4 Additional Predictions and Evidence on Firm Sorting

We now provide additional results, which can be used to detect firm sorting in the data and complements our analysis on local labor shares. We first show that spatial firm sorting increases productivity dispersion in high-ℓ locations (Proposition SA5 and Corollary SA1). 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 SA6). We provide empirical support for both.

SA.4.1 Spatial Firm Sorting: Local Productivity Dispersion

Theory. We first 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. This result applies to productivity distributions $\Gamma(y|p)$ in which stochastic dominance wrt p is more pronounced for higher y.

Proposition SA5 (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:

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

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

In Corollary SA1, we show that the Pareto assumption satisfies the distributional requirement of Proposition SA5 and renders positive sorting not only sufficient but also *necessary* for the result.

Corollary SA1 (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 SA5, we saw that the quantile ratio of productivity, $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$, is increasing in ℓ if

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

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

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

And therefore

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

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

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

□

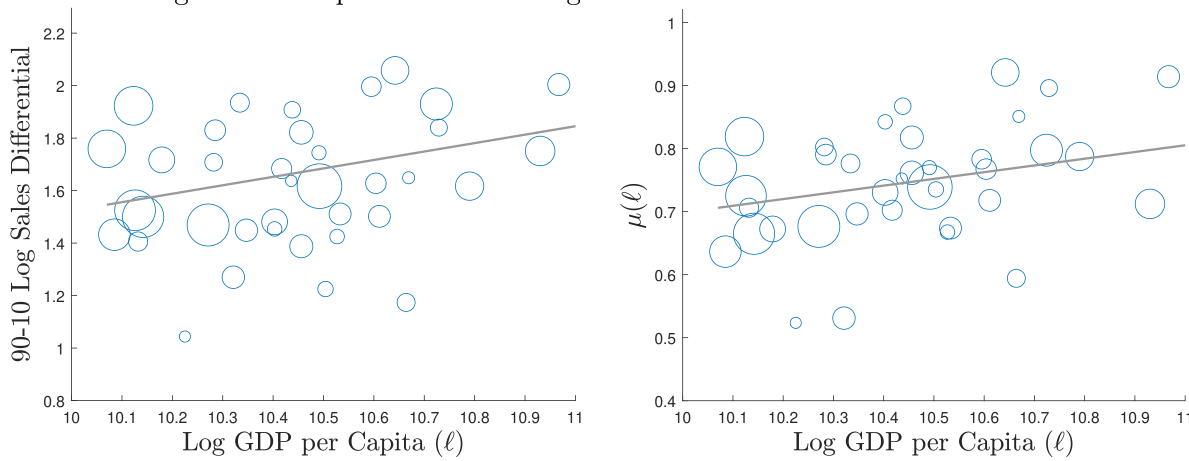
Firm-Level Evidence. We provide additional evidence on positive firm sorting using firm-level productivity indicators. We consider this analysis as only supplementary to our evidence based on local labor shares since the firm-level productivity data is based on the Establishment Panel, which has a relatively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

Within-Location Dispersion of Productivity and Sales. Corollary SA1 (Appendix SA.4.1) 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.[47] Figure SA.1 (left) plots the difference of the 90% and 10% quantile of the distribution of log sales per worker. Based on Corollary SA1, the positive relationship between sales dispersion and ℓ indicates positive firm sorting across space.

[47] 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 SA.1: 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 48). In the right panel, we plot $-1/\beta_\ell$ from regression (SA.20), 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 SA1. 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

$$\log(1 - \mathbb{P}[z(A(\ell), y) \leq k]) = \alpha_\ell + \beta_\ell \log(k) + \epsilon, \tag{SA.20}$$

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.[48]

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

[48]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 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 that provides the number of firms with fewer than 10, 10-50, 50-250, more than 250 employees at the district-year level.

estimated $\mu(\ell)$ against (log) GDP per capita, as shown in the right panel of Figure SA.1, is consistent with positive firm sorting across space.

SA.4.2 Spatial Firm Sorting: Global vs. Local Rank

We now devise an additional test for the presence of firm sorting in the data. We show that firm sorting has distinct implications for the relationship between the local and the global (economy-wide) productivity rank of firms. In contrast to our "tests" of firm sorting discussed above, this section does *not* rely on any parametric restrictions on local firm productivity distributions.

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

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

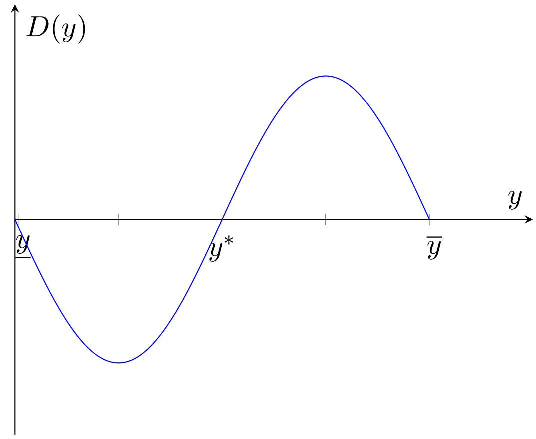
The global rank reflects the firm's position in the economy-wide productivity ranking. By contrast, the local rank reflects the firm's position in the productivity ranking of its location. It takes into account that firms of a given type y can be found in all locations but, because of sorting, they are more prevalent in some locations than others. We therefore average the local rank of firm type y, $\Gamma_\ell(y)$, across locations using the density that describes the distribution of y across space (see the proof of Proposition SA6 for the detailed derivation of the local rank).

Spatial sorting by firms has specific implications for the shape of D. If sorting is monotone, there is a concentration of highly productive firms in some locations and of much less productive firms in others. Thus, the local rank of highly productive firms is *low* relative to their global rank, which yields $D > 0$. The opposite is true for the least productive firms who are surrounded by other low-productivity peers in their locations. As a result, their local rank tends to be *high* compared with their global rank, with $D < 0$. Finally, $D(\underline{y}) = D(\overline{y}) = 0$ because the worst (best) firm economy-wide is also the worst (best) firm in any local labor market. Note that the difference between global and local ranks is absent (i.e., $D(y) = 0$ for all y) if there is no firm sorting. Figure SA.2 depicts D for a parametric example with spatial sorting.[49]

We now show that the shape depicted in Figure SA.2 is a robust feature of spatial firm sorting. To do so, we maintain the following regularity assumption.

[49] Suppose that $R(\ell) = \frac{\ell-a}{b-a}$, $Q(p) = \frac{p-a}{b-a}$, and $\Gamma(y|p) = y^p$ for $b > a > 0$, $p \in [a, b]$ and $\ell \in [a, b]$. Thus, under PAM, $\mu(\ell) = \ell$ and $\Gamma_\ell(y|\mu(\ell)) = y^{\mu(\ell)} = y^\ell$. If $a = 1$ and $b = 2$, we can solve for the zeros of D in closed form, giving the unique interior zero at $y^* = 0.5$; see Figure SA.2. Note that this example does not satisfy Assumption SA3 for $\gamma(\underline{y}|p)$, which however is only sufficient (not necessary) for the result.

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



Assumption SA3. *Both $\gamma(\underline{y}|p)$ and $\gamma(\overline{y}|p)$ are not constant in p.*

We can then show the following results.

Proposition SA6 (Firm Sorting and the Difference between Global and Local Productivity Ranks)**.** *Suppose Assumption SA3 holds.*

i. If there is no spatial firm sorting, $\Gamma_{\ell'} = \Gamma_{\ell''}$ for all $\ell' \neq \ell''$, then $D(y) = 0$ for all $y \in [\underline{y}, \overline{y}]$.

ii. If there is spatial firm sorting, $\Gamma_{\ell'} \neq \Gamma_{\ell''}$ for almost all $\ell' \neq \ell''$, then $D(y) = 0$ for $y = \{\underline{y}, \overline{y}\}$; in turn, there exists a firm type $y^* \in (\underline{y}, \overline{y})$ such that for all $y < y^*$, $D(y) < 0$, and a type $y^{**} \in (\underline{y}, \overline{y})$ with $y^{**} \geq y^*$ such that for all $y > y^{**}$, $D(y) > 0$.

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

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

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

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

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

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

case

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

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

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

Differentiate D wrt y to obtain

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

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

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

where Var_r is our notation for the variance of a random variable, taking land distribution r into account. Both expressions are *strictly negative* if $\text{Var}_r[\gamma(\underline{y}|\mu(\ell))] > 0$ and $\text{Var}_r[\gamma(\overline{y}|\mu(\ell))] > 0$, which is the case under Assumption SA3.

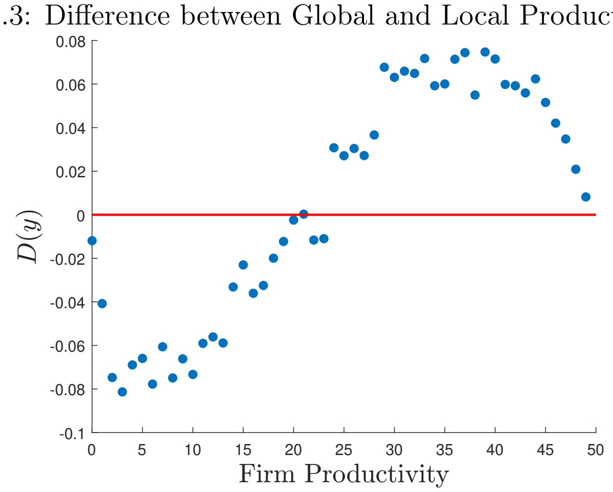
Since D starts at zero and first decreases, it is strictly negative for small $y > \underline{y}$; and since it ends at zero in a decreasing manner, it must be that for high $y < \overline{y}$ it is strictly positive. Hence, there must be at least one $y^* \in (\underline{y}, \overline{y})$ such that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

Evidence. To detect spatial sorting empirically, Proposition SA6 and Figure SA.2 suggest a simple test: If there is monotone spatial sorting, there is an S-shaped relationship between the difference in firms' global and local ranks, $D(y)$, and productivity y. In contrast to our other "tests" of firm sorting in this appendix, this one does *not* rely on any parametric restriction on the local firm productivity distributions (i.e., we can dispense with the Pareto assumption).

Implementing this test in practice, requires a measure of firm productivity y. Measuring y empirically is complicated by the fact that firms—according our theory—are sorted spatially and, thus, their output per worker z depends not only on their productivity y but also on location productivity $A(\ell)$. To purge firm output per worker z from local productivity $A(\ell)$, we exploit the fact that under the assumption of common support, output per worker of the least productive firm in location ℓ is given by $\underline{y}A(\ell)$ and hence should only reflect $A(\ell)$. In practice, we therefore measure y as sales per worker divided by the 1% quantile of the sales per worker distribution in location ℓ.[50]

In Figure SA.3, we plot the relationship between $D(y)$ and y in the data. On the horizontal axis, we order firms by their global productivity rank and categorize them into 50 equally sized bins (based on percentiles of the global productivity distribution). On the vertical axis, we display the average of the difference between global and local ranks for each productivity bin. As in Figure SA.2, there is clear S-shape. Globally unproductive firms sort into locations with a high concentration of unproductive competitors. Hence, their global rank is below their average local rank, i.e., $D(y) < 0$. In turn, for globally productive firms, the opposite pattern arises: They co-locate with other productive firms—i.e., within their local labor market they are relatively unproductive compared to their economy-wide productivity—and therefore $D(y) > 0$. Recall that if there is no spatial firm sorting, we would observe that D is a horizontal line and zero everywhere.

Figure SA.3: Difference between Global and Local Productivity Rank



Notes: Data source: Establishment Panel. We rank firms by their residualized sales per worker and group them in 50 bins of equal size. For each bin, we measure firms' rank in the local sales distribution (local rank) and in the global sales distribution (global rank) and plot the average difference between global and local rank, denoted by $D(y)$.

[50] We used the 1% quantile instead of the observed minimum sales to mitigate the effect of outliers.

SA.5 Characteristics of Local Labor Markets

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

Table SA.1: 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 C.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.

Table SA.2: 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 C.1 for more details on how the displayed variables are defined.

SA.6 Counterfactuals and Policy Exercise: Technical Details

SA.6.1 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.19). 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 $\lambda^U(\ell) = \lambda^E(\ell)/\kappa$.

SA.6.2 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.

SA.6.3 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 $\tilde{\Gamma}(z) = \int \Gamma\Big(\frac{z}{A(\ell)}\Big|\mu(\ell)\Big)dR(\ell)$. 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 $\tilde{G}(z) = \delta\frac{\tilde{\Gamma}(z)}{\delta+\lambda^E\big(1-\tilde{\Gamma}(z)\big)}$; and, in terms of the local employment distribution, G_ℓ is no longer given by (12) but by $G_\ell(y) = \big(\int_{\underline{y}}^{y}\frac{\tilde{g}(A(\ell)y')}{\tilde{\gamma}(A(\ell)y')}\gamma_\ell(y')dy'\big)/\big(\int_{\underline{y}}^{\overline{y}}\frac{\tilde{g}(A(\ell)y')}{\tilde{\gamma}(A(\ell)y')}\gamma_\ell(y')dy'\big)$. We keep the estimated schedules $(A(\cdot), B(\cdot), h(\cdot))$ from the baseline model in place.

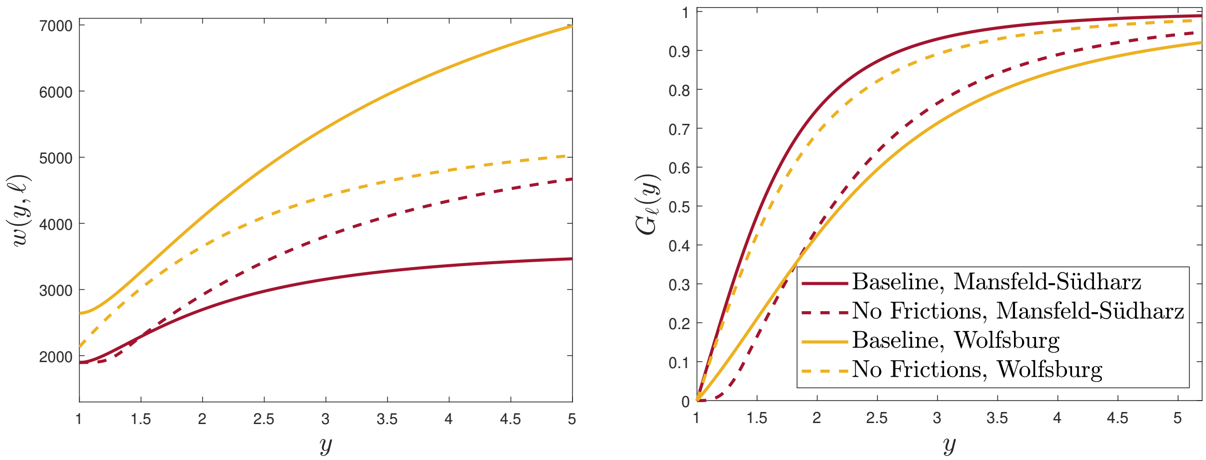
Note that $(\delta, \lambda^U, \lambda^E, \tilde{b})$ are now all constant across ℓ: There are no differences in the job-separation rate, as we compute the economy-wide δ from the average location-specific separation rates, $\delta = \int \delta(\ell)\frac{L(\ell)}{\bar{L}}dR(\ell)$ (where $\delta(\ell)$ and $L(\ell)$ are taken from the baseline model); further, all workers—irrespective of their residence location—have the same chances to find jobs, so there are economy-wide meeting rates for employed workers, λ^E, and unemployed workers, λ^U. We determine these rates using the total population size $\overline{L}$ from the baseline model and $\bar{L} = \mathcal{A}^{\frac{1}{\alpha}}\frac{\delta+\lambda^U}{\delta+\kappa\lambda^U}\left(\lambda^U\right)^{-\frac{1}{\alpha}}$, which we derive from average firm size $\bar{l}(\ell) = (1-u)\overline{L}$ and (24). Similar to the previous counterfactual exercises, we adjust the unemployment flow benefit $\tilde{b}(\ell) = \tilde{b}$ (only that here it is independent of ℓ) so that $w^R(\ell) = w^R = A(\underline{\ell})\underline{y}$, i.e., in an integrated labor market the reservation wage is also determined 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 $B(\ell)$ (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, local population size $L(\ell)$ adjusts so that housing market clearing holds.

Figure SA.4 displays the components of decomposition (16) for the baseline model (solid) and the counterfactual without spatial hiring frictions (dashed).

Figure SA.4: No-Spatial-Frictions Counterfactual: Wages and Employment Distribution



SA.6.4 Firm Sorting and Place-Based Policies

When reducing place-based subsidies (through a reduction in local TFP for some locations), the modularity properties of $\bar{J}$ may change, so we need to re-solve for the sorting decision of firms in this counterfactual equilibrium. The population size in each location (and thus worker and firm meeting rates) depends on the local firm composition, but at the same time impacts firms' sorting choices. We therefore need to solve for a fixed point in the firm allocation.

Given the counterfactual local TFP schedule, $\tilde{A}(\cdot)$, postulate an allocation of firms to locations $m(\ell,p)$ that is measure-preserving. Given $m(\ell,p)$, we first obtain Γ_ℓ from (8), and then find meeting rate λ^U and housing price d (both as a function of $(\ell;\kappa,\Gamma_\ell,\rho V^U)$) so that—given the counterfactual local TFP $\tilde{A}(\cdot)$ and the estimated schedules $(B(\cdot),h(\cdot))$ from the baseline model—the value of search is equalized across space, and local housing market clearing holds:

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

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

where the value of search is again calculated assuming $w^R(\ell) = \tilde{A}(\ell)\underline{y}$, supported by adjusting $\tilde{b}(\ell)$;[51] and where we set the new value of search, ρV^U, to achieve consistency with the total population size from the baseline economy, $\bar{L} = \int L(\ell)dR(\ell)$. Based on unemployed workers' welfare equalization, we obtain $\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)$ and therefore $\lambda^E(\ell;\kappa,\Gamma_\ell,\rho V^U) = \kappa\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)$. With λ^U for each ℓ in hand, we can also compute $\lambda^F(\ell;\kappa,\Gamma_\ell,\rho V^U) = \mathcal{A}^{\frac{1}{\alpha}}(\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U))^{1-\frac{1}{\alpha}}$, as well as the *match value* of a firm type p and location ℓ,

[51] In particular, $\tilde{b}(\ell)$ is defined as in (A.19) but using $\tilde{A}(\ell)$, F_ℓ (which we can compute based on the postulated Γ_ℓ) and $(\lambda^U(\ell),\lambda^E(\ell))$ obtained above.

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

To find the optimal allocation $\hat{m}(\ell,p)$, we maximize the sum of this value across all (p,ℓ)-pairs, subject to land market clearing, using a linear program. If $m=\hat{m}$, we have found the equilibrium. If not, we use $\hat{m}$ as a new starting point and repeat the same steps, until convergence.

SA.7 Application: The West-East and Urban Wage Premium

In our main analysis we focus on spatial wage differences between prosperous and disadvantaged local labor markets, as measured by their GDP per capita. In this section, we turn to two additional dimensions of spatial inequality: differences in economic prosperity between (i) East and West Germany and (ii) rural and urban commuting zones.

In Table SA.3, we report wages for East-West and urban-rural labor markets in both data and model. The wage gap between West and East Germany is 28%. The urban-rural wage gap is slightly smaller but still substantial: On average, wages are almost 20% higher in urban labor markets. Our model matches these dimensions of spatial inequality quite well.

Table SA.3: West-East and Urban-Rural Inequality: Monthly Wages (in €)

	Data	Model
Monthly Wage, West	3491.13	3480.72
Monthly Wage, East	2731.63	2845.38
West-East Gap	28%	22%
Monthly Wage, Urban	3510.01	3508.87
Monthly Wage, Rural	2984.37	3028.00
Urban-Rural Gap	18%	16%

Notes: Data source: German Federal Statistical Office. Wages are monthly and in 2015 Euros. We sum the total wagebill and the number of employees for urban and rural (and also for West and for East) commuting zones, and take their ratio. For definitions of 'Urban' and 'Rural', see Appendix C.4.

Table SA.4 reports the results of our no-sorting counterfactual applied to West-East and urban-rural inequality. Firm sorting again plays a quantitatively important role. Without it, the West-East wage gap in our model would be 19% rather than 22%. Similarly, the urban wage premium would decrease from 16% to 13%. 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 (due to enhanced competition for workers) and rendering a "better" employment composition in these places.

Table SA.4: West-East and Urban-Rural Inequality: No-Sorting Counterfactual

	Model	No Sorting
West-East Wage gap	22%	19%
Urban-Rural Wage Gap	16%	13%

SA.8 Quantitative Results: Additional Robustness

SA.8.1 Relation of Labor Share and Firm Productivity: Log-Normal

We now demonstrate that the property of the negative relationship between local labor share and local firm productivity from our quantitative model does not hinge on the Pareto assumption on $\Gamma(y|p)$. To do so, we provide simulations for a common alternative functional form of firm productivity. Specifically, we assume that $\Gamma(y|p)$ follows a log-normal distribution, whereby $\log(y)$ follows a normal distribution with parameters $(p, 0.5^2)$ and, in equilibrium, $(\mu(\ell), 0.5^2)$. That is, firms with higher ex ante type (higher μ) have higher mean productivity ex post. In line with the discussion in the paper, we left-truncate this productivity distribution such that its density is sufficiently decreasing for all μ (we truncate all local productivity distributions at the median productivity of the worst location in order to maintain a common $\underline{y}$). Figure SA.5 shows that, as in the Pareto case, there is a negative relationship between local labor shares and firm ex ante productivity under the log-normal assumption. As a result, we back out an *increasing* matching function $\mu(\cdot)$ (indicating PAM between firms and locations) from the decreasing local labor share function $LS(\cdot)$ observed in the data.

To gain intuition for the negative relationship between local labor share and local firm productivity, Figure SA.6 plots for two locations—the top and bottom CZ—firm-level labor shares $Ls(\cdot, \ell)$ (left panel) and the cdf of (weighted) employment (right panel), informing the two terms $\frac{\partial Ls(y,\ell)}{\partial y}\left(-\frac{\tilde{G}_\ell(y)}{\partial \ell}\right)$ in the integrand of equation (A.9): In each location, firm-level labor shares decrease in productivity y; moreover, the weighted employment distribution of the top location is stochastically better (under PAM). Thus, a decreasing empirical local labor share function $LS(\cdot)$ calls—under decreasing firm-level labor shares $Ls(\cdot, \ell)$—for PAM so that in high ℓ more employment is concentrated in top firms with low labor shares.

Figure SA.5: Identifying Firm Sorting from Local Labor Shares: Truncated Log-Normal $\Gamma(y|p)$

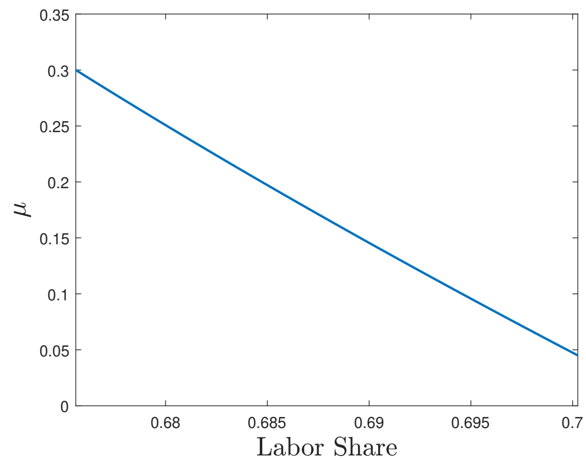
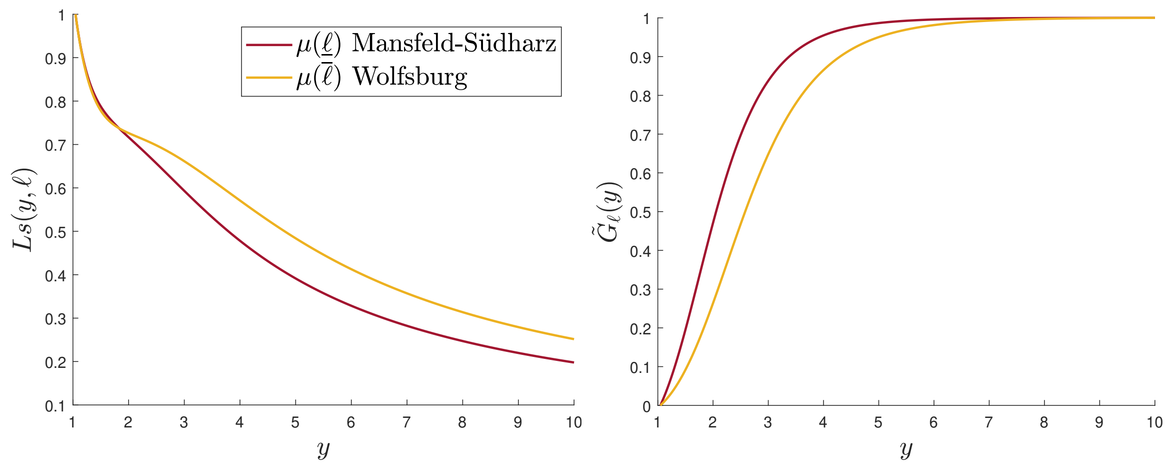


Figure SA.6: Firm-Level Labor Shares and Employment Distribution: Truncated Log-Normal $\Gamma(y|p)$



SA.8.2 Estimation Conditional on Industry

In Table 1, we show that local labor shares are decreasing in log GDP per capita even when we control for the local industrial composition, suggesting that firms sort positively across space also *within* industries. However, to assess the quantitative impact of within-industry firm sorting, we need to use our model. We proceed in two ways.

We first analyze whether our main results on the quantitative importance of firm sorting are robust to controlling for regional differences in industry composition. To do so, we first residualize local labor shares with respect to industrial employment shares. Given this residualized labor share schedule, we re-estimate our model and perform the No-Sorting counterfactual, in which we allocate firms randomly across space. Table 3 shows firm sorting has an even larger effect on spatial inequality than in the baseline model that did not control for industries.

An alternative way to show that our results are not driven by local differences in the industrial composition is to focus on a single industry. We focus on the manufacturing sector, which—given that it produces tradable goods—we believe is a sector that our model can fit better than others. To calibrate our model, we use local labor shares, average value added per employee and average firm size from the manufacturing sector. For local unemployment rates, due to data limitations, we use the average unemployment rate at the CZ-level as in our baseline analysis. Table 3 summarizes our main counterfactual using this alternative estimation.

SA.8.3 Estimation on an Alternative Data Source (FDZ)

In our main analysis we rely on regional-level data on labor shares, value added and firm size from the German Federal Statistical Office. We now show that we arrive at similar conclusions when exclusively using firm- and worker-level data from the FDZ. More specifically, we obtain firm-level value added from the Establishment Panel of the FDZ and construct the local labor shares based on this variable. In addition, rather than ranking locations by their GDP per capita, we rank them by their average value added per full-time employee. Table 3 summarizes our main counterfactual using these alternative data.

SA.8.4 Estimation of a Model with Imperfect Worker Mobility

We consider an extension of our model that accounts for imperfect spatial mobility of workers.

Setup. We assume that workers receive preference shocks $\epsilon(\ell)$, which follow an i.i.d. Frechet distribution with shape parameter ν, and their value when choosing a region ℓ is $V^U(\ell)\epsilon(\ell)$. The population distribution across space can then be expressed in closed form as:

$$\frac{L(\ell)}{\bar{L}} = \frac{(V^U(\ell))^\nu}{\sum_k (V^U(k))^\nu}. \tag{SA.21}$$

Choice probabilities (SA.21) replace the search value equalization condition from our baseline model. Dispersion parameter ν represents the elasticity of the local population with respect to the local value of search. In the limit, when ν goes to infinity, our model reduces to the baseline case, in which migration frictions are absent.

Estimation. We pin down parameter ν by computing the elasticity of the local population size with respect to local average wages, which is as a function of ν, leveraging relevant elasticity estimates available in the literature. Using the value of unemployed workers in equation (20), the elasticity of the local population size with respect to local average wages can be approximated as (ignoring a general equilibrium constant)

$$\frac{\partial \ln L(\ell)}{\partial \ln \mathbb{E}[w|\ell]} = \nu \frac{\partial}{\partial \ln \mathbb{E}[w|\ell]} \ln \left(A(\ell)\underline{y} + \lambda^E(\ell) \int_{w^R(\ell)} \frac{1 - F_\ell(t)}{\rho + \delta + \lambda^E(\ell)(1 - F_\ell(t))} dt \right) - \nu\omega \frac{\partial \ln d(\ell)}{\partial \ln \mathbb{E}[w|\ell]}.$$

Except for the scale parameter ν, we can compute the first term in the above expression by running a regression of local population size on local average wages, using our estimates from the baseline model; and we compute the second term using the housing market clearing condition. We obtain:

$$\frac{\partial \ln L(\ell)}{\partial \ln \mathbb{E}[w|\ell]} = \nu(1.01 - \omega) = 0.738\nu.$$

In the literature, the elasticity of migration flows with respect to income commonly ranges between 2 and 4 (e.g., Allen and Donaldson, 2020). We are interested in long-term effects, so we take a value at the higher end. Since we are considering long-run migration, we assume an elasticity of 4, which implies $\nu = 5.42$. Finally, we pin down amenity schedule $B(\cdot)$ by matching the spatial population distribution in the data. We do so by taking the ratio of the population size of two regions and plugging in the value of search (20), which yields the following equation:

$$B(\ell)d(\ell)^{-\omega} = \left(A(\ell)\underline{y} + \lambda^E(\ell) \int_{w^R(\ell)} \frac{1 - F_\ell(t)}{\rho + \delta + \lambda^E(\ell)(1 - F_\ell(t))} dt \right)^{-1} \rho V^U(\underline{\ell}) \left(\frac{L(\ell)}{L(\underline{\ell})} \right)^{\frac{1}{\nu}}.$$

Note that the assumption of imperfect worker mobility does not affect the estimation of any other parameters.

Counterfactual. To compute the counterfactual without firm sorting, i.e., $\Gamma_\ell(y) = \Gamma(y)$, we find the schedules $\{L(\ell), \lambda^U(\ell), d(\ell)\}$ that satisfy (SA.21), which replaces the condition that equalizes the value of search across locations in the baseline model. The other equilibrium conditions remain the same as in our baseline model. Table 3 summarizes our main counterfactual based on this model extension.

SA.8.5 Estimation of a Model with Endogenous Firm Selection

In this extension, we allow the types of firms that are active in each local labor market to be endogenously determined.

Setup. We assume that firms and workers can observe productivity y only when they meet. Since the value of employment increases with wages, and wages increase with productivity, workers accept job offers only when productivity exceeds a region-specific cutoff $\underline{y}(\ell) \in [\underline{y}, \overline{y})$. Note that a firm of productivity $\underline{y}(\ell)$ makes zero profit, and $A(\ell)\underline{y}(\ell) = w^R(\ell)$. We further assume that the unemployment benefit equals $\widehat{b} > 0$, which we choose such that $\underline{y}(\underline{\ell}) = \underline{y}$.

By combining $A(\ell)\underline{y}(\ell) = w^R(\ell)$ with reservation wage equation (SA.9) and wage equation (5), we obtain the following equation that implicitly defines cutoff $\underline{y}(\ell)$:

$$A(\ell)\underline{y}(\ell) = \widehat{b} + (1-\kappa)\varphi^U(\ell) \int_{\underline{y}(\ell)}^{\overline{y}} (1-\Gamma_\ell(t)) 2\varphi^E(\ell)\gamma_\ell(y) \int_{\underline{y}(\ell)}^{\overline{y}} \frac{A(\ell)}{(1+\varphi^E(\ell)(1-\Gamma_\ell(t)))^2} dtdy. \quad \text{(SA.22)}$$

If there does not exist a cutoff that satisfies this equation, workers accept all jobs, i.e., $\underline{y}(\ell) = \underline{y}$, and all firms earn (weakly) positive profits.

In turn, if there exists an endogenous cutoff, we need to distinguish job arrival rate $\lambda^U(\ell)$ from job finding rate jfr(ℓ) in each ℓ. The local job finding rate of unemployed workers and the local unemployment rate are given by:

$$\text{jfr}(\ell) = \lambda^U(\ell)(1-\Gamma_\ell(\underline{y}(\ell))) \quad \text{(SA.23)}$$

$$u(\ell) = \frac{\text{jfr}(\ell)}{\text{jfr}(\ell)+\delta(\ell)}. \quad \text{(SA.24)}$$

As in our baseline model, assuming the total measure of vacancies in each region equals one, the population size can be expressed as a function of the unemployment rate and the average firm size, $L(\ell) = \frac{1}{1-u(\ell)}\bar{l}(\ell)$. This determines the local job arrival rate as

$$\lambda^U(\ell) = \mathcal{A}(u(\ell) + \kappa(1-u(\ell)))^{-1/2} \left(\frac{\bar{l}(\ell)}{1-u(\ell)}\right)^{-1/2}, \quad \text{(SA.25)}$$

where we use $\mathcal{A}$ from the baseline estimation. Average local value added is then

$$\mathbb{E}[z(y,\ell)|\ell] = A(\ell)\frac{1}{1-\Gamma_\ell(\underline{y}(\ell))} \int_{\underline{y}(\ell)}^{\overline{y}} y g_\ell(y) dy, \quad \text{(SA.26)}$$

where we take into account that, in each labor market, only a subset of firms is active.

Identification. When the cutoff, $\underline{y}(\ell)$, is greater than $\underline{y}$ (which is the case we focus on), we can identify the firm type $p = \mu(\ell)$ that settled in location ℓ from its labor share in the same way as in the baseline model. For the remaining parameters, first note that we can express $\lambda^U(\ell)$ as a function of $\underline{y}(\ell)$ by combining (SA.23), (SA.24), and (SA.25)

$$\lambda^U(\ell) = \mathcal{A}\left(\frac{\delta(\ell) + \kappa\lambda^U(\ell)(1-\Gamma_\ell(\underline{y}(\ell)))}{\lambda^U(\ell)(1-\Gamma_\ell(\underline{y}(\ell)))}\bar{l}(\ell)\right)^{-1/2}. \quad \text{(SA.27)}$$

We then jointly identify $(\underline{y}(\ell), \lambda^U(\ell), A(\ell))$ for each ℓ using three equations, (SA.22), (SA.26) and (SA.27), along with separation rate $\delta(\ell)$, average firm size $\bar{l}(\ell)$, mean value added $\mathbb{E}[z(y,\ell)|\ell]$, the overall matching efficiency $\mathcal{A}$, and the relative matching efficiency κ from the baseline estimation. And we recover job finding rate from (SA.23).

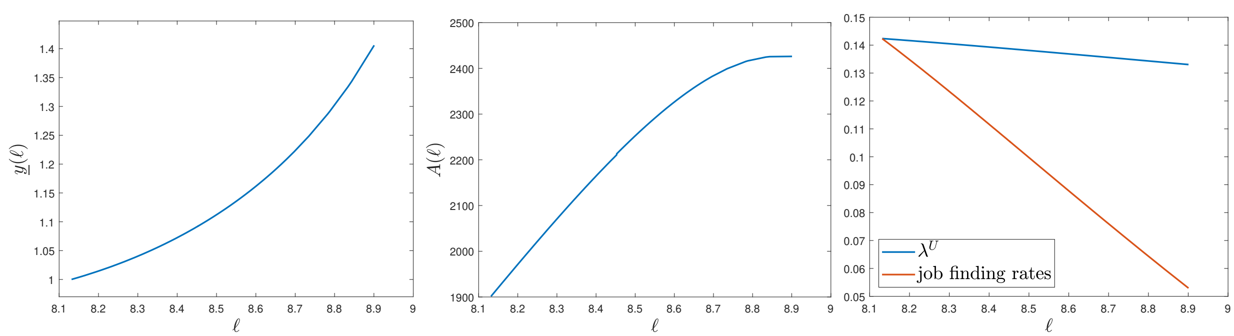
Because unemployed workers have positive income $\widehat{b}$, we assume that the government no longer provides subsidies, and all workers spend a fraction ω of their income on housing. With this assumption, we can estimate local amenities $B(\ell)$ and housing supply $h(\ell)$ using two equations: (i) an equation that is obtained by slightly modifying the value of search of unemployed workers in (20) to endogenize firm selection, and (ii) housing market clearing. They are respectively given by:

$$B(\ell)d(\ell)^{-\omega} = \left(A(\ell)\underline{y}(\ell) + \varphi^E(\ell) \int_{\underline{y}(\ell)}^{\overline{y}} (1-\Gamma_\ell(t)) 2\varphi^E(\ell)\gamma_\ell(y) \int_{\underline{y}(\ell)}^{\overline{y}} \frac{A(\ell)}{(1+\varphi^E(\ell)(1-\Gamma_\ell(t)))^2} dtdy \right)^{-1}$$

$$d(\ell)h(\ell) = \omega(u(\ell)\widehat{b} + (1-u(\ell))\mathbb{E}[w|\ell])L(\ell).$$

Results. We report the estimation results of the main model objects graphically, focusing on the changes compared to the baseline model. The left panel of Figure SA.7 shows the key new object, productivity cutoff $\underline{y}(\ell)$. It is increasing in ℓ, which indicates that workers in prosperous places are more selective, also leading to a declining job finding rate in ℓ (see orange line, right panel). In turn, estimated local TFP is still increasing in ℓ (middle panel).

Figure SA.7: Estimation Results



Counterfactual. For the counterfactual exercise without firm sorting, we set $\Gamma_\ell(y) = \Gamma(y)$ and determine the population distribution, while allowing the endogenous cutoff $\underline{y}(\ell)$ to be consistently determined in each local labor market. We solve for the population in each ℓ subject to the total population constraint, housing market clearing, and welfare (i.e., search value) equalization. Table 3 summarizes our main counterfactual based on this model extension.