Firm Sorting And Spatial Inequality*

Ilse Lindenlaub — *Yale University and NBER*
Ryungha Oh — *Yale University*
Michael Peters — *Yale University and NBER*

February 2021

Abstract

We study the importance of firm sorting for spatial inequality. If productive locations are able to attract the most productive firms, workers in unproductive locations are not only hurt through inferior location fundamentals but also lack access to technologically advanced producers. Thus, firm sorting across locations can act as an amplifier of spatial inequality. We develop a novel model of spatial firm sorting, in which heterogenous firms first choose a location and then hire workers in a frictional local labor market. Firms' location choices are guided by a fundamental trade-off between regional productivity and local labor market competition: Operating in productive locations increases output. However, sharing a labor market with other productive firms makes it hard to poach and retain workers. We show that sorting between firms and locations is positive if firm and location productivity are complements in production and labor market frictions are sufficiently large. We estimate our model using matched employer-employee data from Germany. In our application, we quantify the role of firm sorting for wage differences between East and West Germany. We find that approximately 30% of the East-West wage gap can be accounted for by firm sorting, i.e., by the fact that more productive firms settle in the West.

*Lindenlaub: ilse.lindenlaub@yale.edu. Oh: ryungha.oh@yale.edu. Peters: m.peters@yale.edu. We thank Juan Gambetta for outstanding research assistance.

1 Introduction

Economic outcomes within developed economies have been staggeringly unequal since the early 1980s. While most existing research focuses on inequality across people, a recent literature highlights the disparate economic fortunes across space. Like inequality at the individual-level, the spatial nature of growth has also changed. Traditionally, many economies have been characterized by spatial convergence. However, spatial inequality has been increasing in the past decades, with poor, rural locations falling further behind. Take Germany as an example. Urban areas are characterized by around 15% higher wages and value added; furthermore West Germany has almost 30% higher wages and 40% higher value added compared to the East. This phenomenon of a spatial divide has been attributed to the spatial sorting of more productive workers to cities or to productivity advantages of areas with increasingly larger population. However, much less is known about the role of firm sorting for inequality across space. Firms may be particularly important in this context since they are responsible for substantial part of workers' pay.[1]

In this project, we aim to fill this gap. We develop a theory of how heterogenous firms sort across space to better understand the implications for spatial inequality and to study place-based policies that aim to counter it. We use our framework to shed light on spatial disparities in wages and productivity in Germany, both between urban and rural areas, as well as between East and West.

Our point of departure is based on two observations: First, firms hire predominantly from the local labor market. In Germany, two thirds of hires come from the commuting zone a firm is located in, suggesting that the economy is segmented into many local labor markets. Second, a substantial share of firms' hires are workers that were already employed in other firms (more than 40%, and for large firms this number is much higher). This highlights the importance of firms' ability to poach workers from other firms, and therefore the importance of firms' competitiveness in the local labor market. Taken together, the ease of hiring workers from other firms at the local level should play a dominant role in firms' location choices, which is why it will play a central role feature of our theory.

In our model, firms on the one hand 'like' productive locations as they boost output. These are locations with good fundamentals, which we interpret broadly. They can stem from modern infrastructure, productive spillovers, existing input-output networks and workers'

[1] In two-way fixed effects regressions, firm effects typically explain around 20% of the variance of log-earnings Bonhomme et al. (2020).

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 reason is that this bulk of productive firms reduces the firm's own competitiveness in the local labor market, making it difficult to poach workers and having workers poached is a real risk.

As a result, a fundamental trade-off firms face when making their location decision is between location productivity and local competitiveness, with the latter depending on the endogenous composition of firms in the location. While location productivity increases firm output, fierce competition competition by other firms reduces a firm's own capability to hire and retain workers, curbing firm size.

To formalize this trade-off, we propose a spatial search model of the labor market, in which the sorting of heterogeneous firms to heterogenous locations is determined in equilibrium. Firms differ in productivity and first make their location choice and buy land where they decide to settle. Then, they hire workers in a frictional local labor market, which is characterized by wage posting and an endogenous firm ladder. In each local labor market, both unemployed and employed workers search for jobs, all aiming to climb the local firm (or job) ladder toward increasingly higher wages. Once matched, the worker-firm pair produces output that depends on the firm's productivity and the productivity of the location. Because of spatial sorting of firms, the local firm productivity distribution is endogenous and so is the firms' position in the local job ladder, which differs across locations: Sorting into a location with many productive firms implies a low local productivity rank for a given firm, compared to when it collocates with less productive firms. These ingredients capture the key trade-off we wish to analyze: Firms need to weigh their hiring potential against the location productivity when choosing their spot. To our knowledge, this is the first model that integrates on-the-job search with firms' location choices, highlighting the novel trade-off described above.

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 complements in production and if local labor market frictions are sufficiently large. Productive complementarities ensure that highly productive firms have a greater willingness to pay for the land in more productive places. In turn, sufficiently large labor market frictions ensure that the competition motive does not outweigh this productivity consideration. Settling in the same good location with many productive firms is costly because poaching and retaining workers is difficult. If frictions are large and job-to-job flows are small relative to the complementarities in production and the difference of location productivity, then competition is of limited importance and the technological motive dominates, leading to positive sorting.

Our theory also has important implications for equilibrium land prices. Land prices reflect firms' willingness to pay for different locations, and thus the net benefit—the productivity gain minus the costs from competition—from settling into a certain location. Our theory thus highlights that the land price gradient is intrinsically linked to the extent of firm sorting. Land prices of productive locations are higher relative to non-productive ones if location and firm productivity are sufficiently strong complements or if the frequency of job-to-job flows is relatively low. In these cases, firms all agree that productive locations are most desirable, so their land prices increase to achieve market clearing of between firms and land.

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

In theory, positive firm sorting across space, i.e., the pattern that more productive firms settle in more productive locations, affects spatial wage inequality in two ways. First, it tends to steepen the wage (job) ladders in more productive locations. Second, it leads to a stochastically better employment composition in more productive locations, by which more workers are employed in highly productive firms that pay higher wages. Both factors amplify the spatial wage premium of places that are more productive due to higher TFP.

To quantitatively assess the importance of spatial firm sorting for spatial inequality, we estimate our model and perform counterfactuals. We first extend our parsimonious model to make it amenable for quantitative work while preserving its key mechanism. We introduce labor mobility of unemployed workers with the requirement that—in equilibrium—they are indifferent between all locations. This way, population size and meeting rates become endogenous and vary across locations—a realistic feature that we want our model to capture. We then prove that our model is identified. We overcome a challenging identification problem, which is to disentangle location fundamentals from firm sorting: The *within*-location labor share allows us to identify the firm productivity distribution in each location (and thus firm sorting) while the *across*-location variation in value added pins down the location fundamentals (TFP). Our estimated model features positive sorting between firm productivity and location fundamental, even though we do not target this feature of the data.

Our model fits well the observed spatial inequality in wages between East and West Germany. To gauge the importance of spatial firm sorting for spatial inequality, we conduct a counterfactual, in which we match firms *randomly* to locations. This reveals that firm sorting can account for around 30% of the wage gap between East/West. The East is not only disadvantaged because of poor economic fundamentals. This weakness is amplified by the fact that low-productivity firms tend to cluster there. Finally, we highlight important interactions between firm sorting and the presence of search frictions and on-the-job search. If there was no on-the-job search (no job ladders), firm sorting would not affect inequality in our model.

Related literature Our project merges two strands of literature that have largely existed without any interaction: the literature on frictional labor markets and cross-sectional wage dispersion and the urban literature on spatial inequality.

On the one hand, we build on the literature on labor search and frictional wage dispersion (Burdett and Mortensen, 1998, Postel-Vinay and Robin, 2002). Two important findings are that, conditional on worker heterogeneity, firm heterogeneity accounts for a sizable share of the cross-sectional wage dispersion (15-30%) and so do search frictions and on the job search (10-40%).[2] Despite this evidence on the importance of firms and search for inequality, there has been no attempt to link *spatial* inequality to firm *sorting* into local labor markets that are characterized by search frictions and job ladders. Our paper aims to address this gap.

On the other hand, there is a large and growing literature on the sources of spatial wage inequality, with special focus on the urban wage premium in the U.S. (Glaeser and Maré, 2001, Duranton and Puga, 2004, Gould, 2007, Baum-Snow and Pavan, 2011, Moretti, 2011), in Spain (De La Roca and Puga, 2017), in France (Combes et al., 2008) and in Germany (Dauth et al., 2022). In turn, Heise and Porzio (2021) analyze the East-West German wage gap using a job ladder model, focussing on *worker* mobility and preference frictions as the main source of the spatial divide. In contrast to our work, what these papers leave out of the picture is the (endogenous) *spatial allocation of firms* as a driver of spatial inequality.[3]

Only a handful of papers have theoretically or quantitatively analyzed firms' location

[2]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 firm contribution of 18% while the contribution of the search channel is 10% in Denmark. Using a non-structural approach (two-way fixed effect regressions), firm effects typically explain around 20% of the variance of log-earnings or slightly less when correcting for limited mobility bias Bonhomme et al. (2020). For instance, Card et al. (2013) find that 21% of the wage variance can be accounted for by workplace heterogeneity in Germany and its importance for wage inequality has increased over time. Last, using a non-structural approach based on the Mean-Min (Mm) ratio to assess frictional wage dispersion leads to large estimates when accounting for on-the-job search, namely Mm ratios from 1.16 to 1.55 for the US (Hornstein et al. (2011)).

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

choices, all of which differ in focus and modeling choices from ours. Combes et al. (2012) disentangle firm selection from agglomeration economies in the productivity advantage of cities, by studying shifts and truncations of local productivity distributions. Behrens et al. (2014) develop a model of worker sorting, firm selection, and agglomeration economies to rationalize several stylized facts on the urban wage premium. Gaubert (2018) builds a model of spatial firm sorting, in which firms trade off agglomeration economies and labor costs, to analyze the efficiency impact of place-based policies. Bilal (2020) analyzes the effect of firms' location choices on spatial unemployment differences. With one exception (Bilal, 2020), these papers feature friction*less* labor markets. None of these papers analyzes spatial *firm sorting* in the presence of *on-the-job search* in local labor markets—two ingredients we will show interact in important ways to shape spatial inequality.[4]

Our paper is structured as follows. Section 2 lays out the model and Section 3 presents the main results. Section 4 contains model extensions. Section 5 contains motivating empirical evidence. Identification and estimation are in Section 6. Section 7 quantifies the role of spatial sorting for spatial inequality and the impact of counterfactuals. Section 8 concludes.

2 The Model

2.1 Environment

Time $t \in \mathbb{R}_+$ is continuous. The economy has a continuum of locations, and a continuum of firms and workers. Each firm chooses a single location to operate in and each location is a local labor market. Workers populate locations.

Locations differ in index ℓ and we use subscript ℓ for distributions that depend on location. Index ℓ determines the location's productivity $A(\ell)$, which is exogenous in our baseline model. We assume that $A(\ell)$ is strictly positive for all ℓ and continuously differentiable, and that locations are ordered such that high-ℓ is associated with high productivity, $\frac{\partial A(\ell)}{\partial \ell} > 0$. In each location, an exogenous (in the baseline model) amount of land is available, distributed with the continuously differentiable cdf R on $[\underline{\ell}, \overline{\ell}]$; $r > 0$ is the corresponding density.

Each location ℓ is populated by a unit mass of risk-neutral, spatially immobile and homogenous workers. Below, we extend our framework to allow for population mobility. Unemployed workers in location ℓ receive flow benefit $b(\ell)$ and employed workers receive a flow wage. Both search for jobs, as detailed below.

[4]In the model by Bilal (2020), workers only search when unemployed; there is no on-the-job search.

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

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

In each location, there is a labor market, in which workers and firms face search frictions and search is random. In the baseline model, we abstract from endogenous vacancy posting and labor mobility, so 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 rate of employed and unemployed workers *relative* to the destruction rate by $\varphi^E \equiv \lambda^E/\delta$ and $\varphi^U \equiv \lambda^U/\delta$. We provide extensions with labor mobility and with vacancy posting, in which we endogenize φ^E and φ^U through a local matching function.

In terms of wage setting, we assume that firms post wages with commitment as in Burdett and Mortensen (1998). Thus, firms do not engage in counter-offers even if their employees receive outside offers they could match. We denote the wage paid by firm y in location ℓ by $w(y, \ell)$. Hence, firm y in location ℓ receives flow profit $\pi(y, \ell) = z(A(\ell), y) - w(y, \ell)$ when employing a worker. We impose the following assumptions.

Assumption 1.

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

Assumption 1 is not particularly restrictive but simplifies some arguments. The assumptions of common support and zero profits for marginal firms imply that—despite the endogenous sorting of firms across locations—the resulting distributions of firm productivity have common support across locations, simplifying our analytical arguments. Part 2. will be guaranteed by ensuring that the output of the least productive firm equals the reservation wage, $w^R(\ell)$, i.e., $w^R(\ell) = z(A(\ell), \underline{y})$ for all ℓ. Below, we show how to ensure this property by appropriately choosing non-employment income $b(\ell)$ (a primitive) across locations.

We focus on the economy in steady state.

2.2 Equilibrium

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

Worker Decisions. Workers in each location ℓ face two decisions: first, whether or not to accept a job offer when unemployed, characterized by their reservation wage; and second, whether to accept a job offer when employed. We discuss them briefly since they are standard.

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

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

where F_ℓ is the endogenous wage offer distribution in location ℓ and $V^U(\ell)$ denotes the value of unemployment, solving the recursive equation

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

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

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

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

Note that we can always justify our Assumption 1.2 that $w^R(\ell) = z(A(\ell), \underline{y})$ based on the appropriate choice of primitive function $b(\ell)$, once F_ℓ is pinned down. A particularly tractable schedule arises in the case of $\lambda^U = \lambda^E$, where $b(\ell) = z(A(\ell), \underline{y})$ satisfies our assumption.

Firm Decisions. Firms face two decisions. First, they choose location ℓ to maximize expected discounted profits. They do so by taking location productivity, competition from other firms and land prices into account. Second, conditional on the location choice, firms post a wage to maximize expected profits from employment. We describe the two stages backwards.

Wage Posting. Given the wage offer distribution in location ℓ, F_ℓ, determined by competitors' wage posting strategies, the optimal wage posting rule of firm y in location ℓ maximizes the expected value from employing a worker:

$$\tilde{J}(y,\ell) \equiv \max_{w \geq w^R(\ell)} h(w,\ell)J(y,w,\ell) = \max_{w \geq w^R(\ell)} \underbrace{\frac{\lambda^F \delta}{\delta + \lambda^E(1-F_\ell(w))}}_{=h(w,\ell)} \underbrace{\frac{z(A(\ell),y)-w}{\rho+\delta+\lambda^E(1-F_\ell(w))}}_{=J(y,w,\ell)}, \tag{3}$$

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.[5]

Note that hiring rate, $h(w,\ell)$, positively depends on the posted wage, as a higher wage allows for more poaching. Crucially, it negatively depends on the local wage distribution F_ℓ (in the FOSD sense) because higher wages from competitors make it harder for the firm to hire. In turn, focussing on $J(y,w,\ell)$, a higher wage decreases flow profits but alleviates discounting since match duration becomes longer. A firm's match value $J(y,w,\ell)$ is also lower if F_ℓ is stochastically better since tougher labor market competition makes it difficult to retain workers. Note that the importance of the local wage distribution F_ℓ for expected

[5] The hiring rate of firm y in location ℓ is: $h(w,\ell) \equiv \lambda^F \left(\frac{\lambda^U u(\ell)}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))} + \frac{\lambda^E(1-u(\ell))}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))}E_\ell(w)\right)$, considering that a firm meets workers at rate λ^F from two pools: unemployment $u(\ell)$ (they will always accept the job), and employment $1-u(\ell)$ (they will accept if the new wage is higher than their current one). We denote the steady-state employment distribution by E_ℓ, where $E_\ell(w) = \delta\frac{F_\ell(w)}{\delta+\lambda^E(1-F_\ell(w))}$ (see (12) and (13)), so that $h(w,\ell)$ reduces to the expression in (3).

value $\tilde{J}(y,\ell)$ is large if the net contact rate $\varphi^E = \lambda^E/\delta$ is large.

There is a useful reformulation of problem (3) that highlights the firms' location choice trade-off that we will resolve below. To simplify exposition, we set $\rho \to 0$ for the remainder of the analysis. We then recognize that we can express $\tilde{J}(y,\ell))$, using *firm size* $l(w,\ell)$, given by:[6]

$$l(w,\ell) \equiv \frac{\lambda^F}{\delta\left(1+\varphi^E\left(1-F_\ell(w)\right)\right)^2}. \tag{4}$$

Firms that are higher ranked in the wage distribution of location ℓ (firms with higher $F_\ell(w)$) are larger, since they poach more and are being poached less, compared to lower ranked firms.

We can then frame the firm's wage posting problem in (3) as:

$$\max_{w \geq w^R(\ell)} l(w,\ell)(z(A(\ell),y)-w),$$

whereby the firm maximizes per worker profit times its employment size. This formulation highlights that a firms' location matters in two distinct ways. On the one hand, choosing 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, then competition is fierce (the wage offer distribution F_ℓ is stochastically better), and a given firm y's size becomes compressed. The competition channel is mitigated if φ^E is low, that is if labor market frictions are significant.

The firm's objective function (3) is supermodular in (w,y), implying that w is weakly increasing in y. In combination with a continuum of productivity levels $y \in [\underline{y},\overline{y}]$, this implies in our context that the solution to this problem is a wage function $w(\cdot,\ell)$ that is strictly increasing in y for each ℓ.[7] Therefore, the distribution of wage offers reflects the distribution of productivity across firms, $F_\ell(w(y,\ell)) = \Gamma_\ell(y)$, where Γ_ℓ is the *endogenous* productivity cdf of firms in location ℓ. It 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.[8]

Making this substitution, we solve the firm's problem to obtain the well-known wage

[6] The firm size can be interpreted as the hiring rate times the expected duration of a match once we take the limit $\rho \to 0$. Taking into account a consistency condition on λ^F (relating it to λ^E and λ^U), this is equivalent to the definition of firm size in Burdett and Mortensen (1998), which is the measure of workers employed at y divided by the measure of firms of productivity y. See Appendix A.1 for more details.

[7] If it is only weakly increasing, and an interval of firm types offers the same wage, then there exist a profitable deviation for a firm type in that interval to offer a slightly higher wage and increase the hiring probability by a discrete amount.

[8] And further, $E_\ell(w(y,\ell)) = G_\ell(y)$, where $G_\ell(y)$ is the probability of being employed in firms with productivity below y. In what follows, we will use G_ℓ instead of E_ℓ.

function under wage posting (Burdett and Mortensen (1998))

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

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

Location Choice. Given the wage function for each location ℓ, we can now specify the firm's location choice problem. We consider the expected value of firm p to settle in location ℓ:

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

taking into account that when this choice is made, the firm does not yet know its ex-post productivity draw y, and that it needs to pay land price $k(\ell)$ to settle in ℓ. Using $\tilde{J}(y,\ell)$ from (3) and wage function (5), $\overline{J}(p,\ell)$ can be expressed as:

$$\overline{J}(p,\ell) \quad = \quad \delta\lambda^F \int_{\underline{y}}^{\overline{y}} \int_{\underline{y}}^{y} \frac{\frac{\partial z(A(\ell),t)}{\partial y}}{\left[\delta + \lambda^E(1-\Gamma_\ell(t))\right]^2} dt d\Gamma(y|p) - k(\ell). \tag{6}$$

The expected value for firm p of settling in location ℓ is given by the expected value from employment—which depends on the endogenous firm distribution in location ℓ, Γ_ℓ—net of the price of land. The firm's location choice problem is then given by:

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

The solution to problem (7) describes firms' spatial choices and is at the center of our analysis. The FOC of this problem highlights the fundamental location choice trade-off faced by firms and can be expressed as follows (see Appendix A.3):

$$\delta\lambda^F \int_{\underline{y}}^{\overline{y}} \left(\frac{\partial \ln\left(\frac{\partial z(A(\ell),y)}{\partial y}\right)}{\partial \ell} + \frac{\partial \ln l(y,\ell)}{\partial \ell} \right) \frac{\partial z(A(\ell),y)}{\partial y} \frac{(1-\Gamma(y|p))}{\left[\delta + \lambda^E(1-\Gamma_\ell(y))\right]^2} dy = \frac{\partial k(\ell)}{\partial \ell} \tag{8}$$

where $\frac{\partial \ln l(y,\ell)}{\partial \ell}$ is the elasticity of firm size w.r.t. location ℓ (for fixed y) and $\frac{\partial \ln\left(\partial z(A(\ell),y)/\partial y\right)}{\partial \ell}$ is the elasticity of the firm's marginal product w.r.t. ℓ. Equation (8) determines the slope of the land price function $k(\ell)$. It highlights that equilibrium land prices reflect the 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 firm profits per employee (reflected in the positive elasticity of firm's marginal product w.r.t. ℓ). But if these are also the locations that attract many productive firms, then competition in high-ℓ locations is fierce, poaching and retaining workers is difficult, which reduces firm size. At the optimal location choice, this marginal (net) benefit of choosing ℓ equals its marginal cost, which is the increase in the price of land. If high-ℓ locations are overall more profitable, then $\partial k(\ell)/\partial\ell > 0$, reflecting that high-$\ell$ locations command higher land prices.

This FOC—along with land market clearing—pins down the equilibrium allocation of firms to locations, captured by Γ_ℓ:

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

where we define by $m(\ell, p)$ the joint matching density between (ℓ, p), with conditional densities $m_\ell(\ell|p)$ and $m_p(p|\ell)$.[9] They also pin down the land price schedule sustaining this allocation. Solving the differential equation (8), evaluated at the equilibrium assignment, yields

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

where $\overline{k}$ is a constant of integration. The land price in location ℓ is given by the cumulative marginal contributions of land to the match surplus (between firms and land) in all locations that are weakly less productive than ℓ. We anchor this land price function by choosing $\bar{k}$ such that the landowner whose land commands the lowest price in equilibrium obtains zero.[10]

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

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

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

Good Market Clearing. We assume that the good market clears in each location ℓ, so there is no trade across space, and workers, firms and land owners consume their entire income.

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

[10] In this competitive land market, firms maximizing expected profits and landowners maximizing 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.

Total income equals total consumption, which in turn equals total output in each location:

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

where we use firm size of y in ℓ, $l(y, \ell)$, defined in (4) (when replacing w by y, as discussed).

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

First, inflow into and outflow out of unemployment need to balance, pinning down the unemployment rate $u(\ell)$, where we embed the firms' optimal choice to only post wages that exceed the workers' reservation wage (so that unemployed workers accept all jobs):

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

giving the standard steady-state unemployment rate. In our baseline model, where we assume that transition rates do not vary across locations, $u(\ell) = u$ does not vary across ℓ either.

Second, inflow into and outflow out of employment in firms with productivity below y need to balance (for all y), where we take into account the (optimal) decision of employed workers to accept job offers of any firm that is more productive than the current one,

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

pinning down the cdf of employment in location ℓ, G_ℓ. 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, inducing workers to quit for better jobs when they find them (which happens at rate $\lambda^E(1 - \Gamma_\ell(y))$).

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

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

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: each firm of type* $y \in [\underline{y}, \overline{y}]$ *posts a wage to maximize* $\tilde{J}(y, \ell)$ *in* (3)*, pinning down* $w(\cdot, \ell)$*;*

3. *Optimal worker behavior: employed workers accept any job offer from more productive firms than the current one; unemployed workers accept any job with wage* $w(y,\ell) \geq w^R(\ell)$*, where* $w^R(\ell)$ *is pinned down by* (2)*;*

4. *Flow balance conditions* (12) *and* (13) *hold, pinning down* $u(\ell)$ *and* G_ℓ*;*

5. *Good market clearing: in each location* ℓ*,* (11) *holds.*

Requirement 1. is satisfied if firms choose locations optimally, land owners (or locations) choose firms optimally, and the land market clears. Optimal firm sorting means that firms maximize (7) so that a firm of type $p \in [\underline{p}, \overline{p}]$ chooses market ℓ, i.e., $m_\ell(\ell|p) > 0$, only if $\overline{J}(p,\ell) \geq \overline{J}(p,\ell')$ for all $\ell' \neq \ell$. Requirement 2. pins down the wage in each location ℓ as a function of ex-post firm productivity. Requirement 3. pins down the reservation wage. Requirement 4. on the flow-balance conditions pins down the steady-state employment distribution G_ℓ and unemployment rate $u(\ell)$. Requirement 5. on good market clearing ensures aggregate consistency in the absence of goods trade across space.

3 Equilibrium Analysis

3.1 Spatial Firm Sorting

We now analyze the patterns of firm sorting that occur in equilibrium. In particular, we provide conditions under which more productive firm types p (from an ex-ante perspective) sort into more productive locations ℓ. This is an allocation with positive assortative matching (PAM) and, as we show below, the empirically relevant case. For completeness, we also analyze the allocation with negative sorting (NAM) in Appendix B.2. We also show that sorting has distinct implications for the relationship between firms' local and global productivity rank.

Sufficient Conditions for Positive Sorting. We focus on pure assignments between (p,ℓ), i.e. on deterministic assignments under which any two firms of the same type are matched to the same location. Assignment $m_\ell(\ell|p)$ can then be expressed as a matching *function* $\mu : [\underline{\ell}, \overline{\ell}] \to [\underline{p}, \overline{p}]$. We now introduce the well-known concept of monotone sorting.

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

Under positive sorting, more productive firms sort into more productive locations. Moreover, $m_\ell(\ell|p)$ has positive mass only at a single point $p = \mu(\ell)$ (where firm p is assigned to location ℓ) and we can use (9) to express the endogenous distribution of firms in location ℓ as:[11]

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

To obtain *sufficient* conditions for positive sorting, recall that a firm p faces location choice (7), i.e. maximizing $\overline{J}(p,\ell)$ by choosing ℓ. Based on results from the literature on Monotone Comparative Statics (Milgrom and Shannon (1994)), any solution to this problem is increasing in p (and thus, μ is increasing as well) 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 strictly increasing. Note that strict supermodularity of $\overline{J}(p,\ell)$ in (p,ℓ) is sufficient for the strict single crossing property. This discussion leads to our first result on the sufficient conditions for positive sorting, summarized in Proposition 1.

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

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

Recall the firm's location choice problem $\max_\ell \overline{J}(p,\ell) = \max_\ell \int \tilde{J}(y,\ell)d\Gamma(y|p) - k(\ell)$. Since p shifts $\Gamma(y|p)$ in the FOSD sense, we have

$$\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell} > 0 \quad \text{if} \quad \frac{\partial^2 \tilde{J}(y,\ell)}{\partial y \partial \ell} > 0.$$

We show in the appendix that the sign ($\overset{\text{S}}{=}$) of the cross-partial derivative of $\tilde{J}$ w.r.t. (y,ℓ) is determined by the location choice trade-off between productivity gains and competition:

$$\frac{\partial^2 \tilde{J}(y,\ell)}{\partial y \partial \ell} \overset{\text{S}}{=} \underbrace{\frac{\partial \ln\left(\frac{\partial z(A(\ell),y)}{\partial y}\right)}{\partial \ell}}_{\text{Productivity Gains}} + \underbrace{\frac{\partial \ln l(y,\ell)}{\partial \ell}}_{\text{Competition}} \tag{14}$$

While the productivity gains from settling into high-ℓ locations are positive, the competition

[11] To see this, note:

$$\Gamma_\ell(y) = \int_{\underline{p}}^{\overline{p}} \Gamma(y|p)m_p(p|\ell)dp = \int_{\underline{\ell}}^{\overline{\ell}} \Gamma(y|\mu(\tilde{\ell}))\frac{m(\mu(\tilde{\ell}),\ell)}{r(\ell)}\frac{r(\tilde{\ell})}{q(\mu(\tilde{\ell}))}d\tilde{\ell} = \frac{1}{r(\ell)}\int_{\underline{\ell}}^{\overline{\ell}} \Gamma(y|\mu(\tilde{\ell}))m_\ell(\ell|\mu(\tilde{\ell}))r(\tilde{\ell})d\tilde{\ell} = \frac{1}{r(\ell)}\Gamma(y|\mu(\ell))r(\ell) = \Gamma(y|\mu(\ell))$$

effect is negative under positive sorting, as productive firms cluster in the best locations. Positive sorting then materializes if the productivity benefits outweigh the costs from competition that translate into lower expected firm size, so that sorting into better locations creates a net benefit.

We now seek to express (14) in terms of primitives and discipline its sign. In order to guarantee supermodularity of $\tilde{J}(y,\ell)$ in (y,ℓ) and thus of $\overline{J}(p,\ell)$ in (p,ℓ), we use (14) and the expression for firm size $l(y,\ell)$ in (4) to obtain the following sufficient condition for PAM:

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

In the case of positive sorting, the endogenous firm distribution in location ℓ is given by $\Gamma_\ell(y) = \Gamma(y|\mu(\ell)) = \Gamma(y|Q^{-1}(R(\ell)))$ and therefore $\frac{\partial \Gamma_\ell(y)}{\partial \ell} = \frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\frac{r(\ell)}{q(\mu(\ell))} < 0$, i.e., better locations have better firms in a FOSD sense, where we made use of the unique equilibrium matching function $\mu(\ell) = Q^{-1}(R(\ell))$ that satisfies land market clearing.[12]

Condition (15) states that in order for PAM to materialize, the marginal productivity gains of settling in a better location have to outweigh the costs of tougher competition.

The productivity gains on the L.H.S. are large if productivity differences across locations are large (the A-schedule is steep), and when complementarities of z in $(A(\ell),y)$ are strong. In turn, the R.H.S. captures the costs from sorting into a high-ℓ that stem from labor market competition. Under PAM, better locations have more productive firms, which reduces firm y's expected size. The extent to which this is the case depends on how the endogenous firm distribution varies across space (and on the impact of changes in Γ_ℓ on firm size); as well as on the degree of labor market frictions $\varphi^E = \lambda^E/\delta$. The cost from sorting into a high-ℓ region is low when λ^E, the rate at which employed workers meet firms, is small since then poaching and competition do not matter much. The cost is also low if δ is large so that match duration is mainly determined by workers separating into unemployment as opposed to quitting. In this case, hiring happens predominantly from unemployment and, again, poaching considerations

[12] Under positive sorting, the unique matching function that satisfies land market clearing between firms' ex-ante types and locations is such that $Q(\mu(\ell)) = R(\ell)$ and therefore $\mu(\ell) = Q^{-1}(R(\ell))$. To see that μ in this case satisfies our general market clearing condition (10), note that:

$$R(\ell) = \int_{\underline{\ell}}^{\ell}\int_{\underline{p}}^{\overline{p}} m_\ell(\tilde{\ell}|\tilde{p})q(\tilde{p})d\tilde{p}d\tilde{\ell} = \int_{\underline{\ell}}^{\ell}\int_{\underline{\ell}}^{\overline{\ell}} m_\ell(\tilde{\ell}|\mu(\hat{\ell}))q(\mu(\hat{\ell}))\mu'(\hat{\ell})d\hat{\ell}d\tilde{\ell} = \int_{\underline{\ell}}^{\ell} q(\mu(\tilde{\ell}))\mu'(\tilde{\ell})d\tilde{\ell} = Q(\mu(\ell)),$$

where we used a change of variable and the fact that under perfect sorting, the matching can be represented by a strictly increasing function μ, so that $m_\ell(\ell|p) = 1$ for $p = \mu(\ell)$ and zero otherwise ($m_\ell(\ell|p)$ is a Dirac measure in this case).

carry less weight. The ratio φ^E summarizes local labor market frictions and captures both of these considerations. A small φ^E weakens the competition channel so that it does not interfere with the motive for positive spatial sorting driven by location productivity.

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

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

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

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

The proof is in Appendix B.1, where we make the statements of 'sufficiently large ε^P' and 'sufficiently small φ^E' precise. Proposition 2 highlights in terms of primitives the key trade-off between productivity and competition that firms face when they make their location choice.

Note that, under the conditions from Proposition 2, the productivity gain from settling into high ℓ locations outweighs the cost from competition for *all* y-type firms. But—due to productive complementarities between (A, y)—the net benefit is especially high for those firms with high y, something that high ex-ante productivity p yields stochastically. This is the reason why highly productive firms are 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, $\partial\Gamma_\ell/\partial\ell \leq 0$.

Global vs Local Rank. We now show that, in the presence of spatial sorting, there is an important distinction between the local and the global productivity rank of firms. This result renders a unique and testable prediction of our theory that features local poaching by firms as a key ingredient. We define the difference between firm y's global (economy-wide) rank and its local rank as follows:

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

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 in others. We therefore average a firm y's local rank in location ℓ, $\Gamma_\ell(y)$, across locations using the matching density that indicates the distribution of y across space.[13]

Spatial sorting by firms has distinct implications for the shape of $D(\cdot)$. In particular, if sorting is monotone (either PAM or NAM), there is a concentration of highly productive firms in some locations and of little productive firms in others. Highly productive firms, who settle in locations with many other productive firms, face severe competition. As a result, their local rank is *low* relative to their global rank, yielding $D > 0$. The opposite is true for the least productive firms who are surrounded by other low-productivity peers in their locations, shielding them from competition by top firms. As a result, their local rank tends to be *high* compared to their global rank with $D < 0$. Finally, our assumption that location-productivity distributions have common support implies that $D(\underline{y}) = D(\overline{y}) = 0$ because the worst (best) firm economy-wide is also the worst (best) firm in any local labor market. In Proposition 6 (Appendix B.3), we formalize these statements and show that a common shape of D is the one depicted in Figure 1.

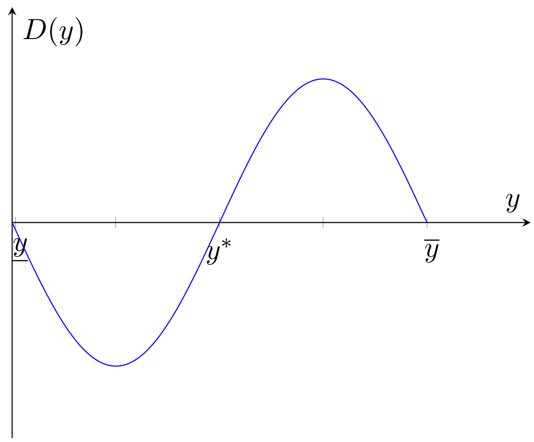
Note that this difference between global and local ranks is absent in the standard job-ladder model that consists of a single market. It is also absent in a multi-market model with no spatial sorting where local and global ranks coincide for all firm types.

[13] 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 distribution 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 density of $(\ell, y))$ and $n(y) := \int_{\underline{\ell}}^{\overline{\ell}} n(\ell, y)d\ell = \int_{\underline{\ell}}^{\overline{\ell}} \gamma(y|\mu(\ell))q(\mu(\ell))\mu'(\ell)d\ell$ is the corresponding marginal density of y. Note that for the second and third equality we are using the established property that matching is monotone here (the fact that we can express $p = \mu(\ell)$).

Figure 1: Spatial Firm Sorting and the Difference between Global and Local Productivity Ranks



3.2 Existence and Uniqueness

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

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

The proof is in Appendix B.4. We show existence of a fixed point in Γ_ℓ by construction. In turn, uniqueness arises because, under the conditions on primitives stated in Proposition 2, the impact of the endogenous firm composition on firms' value function leaves the *complementarity properties* of $\overline{J}$ unchanged.

3.3 Spatial Firm Sorting and Spatial Wage Inequality

In this section, we turn to spatial wage inequality. We aim to understand its drivers and particularly how firm sorting across locations affects it.

Consider the following measure of spatial wage inequality,

$$\frac{\mathbb{E}[w(y,\ell)|\ell]}{\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]} = \frac{w(\underline{y},\ell) + \int_{\underline{y}}^{\overline{y}} \frac{\partial w(y,\ell))}{\partial y}(1-G_\ell(y))dy}{w(\underline{y},\underline{\ell}) + \int_{\underline{y}}^{\overline{y}} \frac{\partial w(y,\underline{\ell})}{\partial y}(1-G_{\underline{\ell}}(y))dy}. \tag{16}$$

To illustrate the drivers of inequality, we consider how this measure varies as we increase ℓ, i.e., comparing the average wage of a location that is 'further away' from the lowest ranked one in terms of productivity:

$$\frac{\partial \frac{\mathbb{E}[w(y,\ell)|\ell]}{\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}]}}{\partial \ell} \overset{s}{=} \Bigg(\underbrace{\underline{y}\frac{\partial A(\ell)}{\partial \ell}}_{\text{intercept of job ladder}} + \int_{\underline{y}}^{\overline{y}} \underbrace{\frac{\partial^2 w(y,\ell)}{\partial y \partial \ell}}_{\text{steepness of job ladder}} (1-G_\ell(y)) + \underbrace{\frac{\partial w}{\partial y}\left(-\frac{\partial G_\ell(y)}{\partial \ell}\right)}_{\text{employment composition}} dy \Bigg). \tag{17}$$

There are two fundamental differences between the 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. In turn, these differences are important for understanding the three factors driving our measure of cross-location wage inequality: First, higher-ℓ locations have a higher intercept of the wage function and thus job ladder, due to higher location TFP. Second, there is larger spatial inequality if higher-ℓ locations have a steeper wage function (i.e., a *steeper job ladder*), which is predominantly impacted by their higher TFP along with complementarities between A and y in production; and moreover, as a result of positive firm sorting, by tougher competition among highly productive firms to poach workers. Third, higher-ℓ locations have a better firm composition (due to *spatial firm sorting*) so that more employment is clustered at the upper part of the wage schedule where wages are higher, compared to $\underline{\ell}$. We summarize:

Lemma 1. *Assume the conditions from Proposition 2 hold. Spatial wage inequality between locations* $\underline{\ell}$ *and* $\ell > \underline{\ell}$ *exists,* $\mathbb{E}[w(y,\ell)|\ell]/\mathbb{E}[w(y,\underline{\ell})|\underline{\ell}] > 1$*, because* ℓ *has (i) a higher wage function intercept,* $A(\ell)\underline{y}$*; (ii) a steeper wage function, provided that* $1-\Gamma(y|p)$ *is log-submodular in* (y,p)*; and (iii) a stochastically better firm, and thereby, employment distribution* G_ℓ*.*

In Appendix B.5, Remark 1, we give examples of densities γ that satisfy the condition in (ii). This decomposition of spatial wage inequality will guide our empirical and quantitative analysis below. A similar analysis for value added can be found in Appendix B.5, Remark 2.

3.4 Comparative Statics

We summarize how three important outcomes—firm wage, size and net poaching share — depend on the model's primitives, both within and across locations.[14] Details are in Appendix C.

We start with firm-level predictions *within* any given location ℓ (Appendix C.1). We are interested in how firm-level outcomes vary with local labor market *competition*—i.e. the local productivity rank of a firm $\Gamma_\ell(y)$—, and with local *labor market frictions*, φ^E. Together, competition and frictions determine the importance of on-the-job search for firm sorting. Indeed, the local productivity rank enters firm-level outcomes as an effective local rank, $\varphi^E\Gamma_\ell(y)$, so that the impact of competition is amplified if frictions are small (i.e., when φ^E is high).

Lemma 2 shows that a reduction in labor market frictions (i.e., an increase in φ^E), which makes the pool of on-the-job searchers more important for firms' hiring, increases the wage and relative firm size of any firm $y > \underline{y}$, and it decreases its net poaching share if φ^E is large enough.

[14] The net poaching share is the difference between poaching share and poached share, indicating how good a firm is at hiring and retaining workers.

Lemma 3 shows that in any given location, being higher up in the local job ladder (i.e., a higher *local rank* $\Gamma_\ell(y)$) induces the firm to post a higher wage to facilitate hiring workers from lower ranked firms, increasing poaching and firm size. Importantly, the global productivity rank has no impact on firm-level outcomes, once the local rank is accounted for.

Lemma 4 then acknowledges that the local rank $\Gamma_\ell(y)$ is itself endogenous and depends on matching function μ. We show how the firm productivity distribution Q and the land distribution R—through their impact on μ—affect wages, firm size, and net poaching.

Next we are interested in market-level outcomes—*average* wage, net poaching share and firm size—and how they vary across locations ℓ (Appendix C.2).

Lemma 5 shows that, given constant labor market frictions across space (i.e., $(\lambda^U, \lambda^E, \delta)$ do not depend on ℓ), the average net poaching share and average firm size are independent of ℓ, while the average wage is increasing in ℓ. The reasons for the increasing wage are twofold: Higher-ℓ locations have a higher wage intercept due to higher TFP (productivity channel); and also a steeper wage schedule due to a better firm composition (firm sorting channel). That is, job ladders in high-ℓ locations are steeper.

Finally, Lemma 6 shows how market-outcomes change if labor market frictions vary (while keeping A fixed). In line with our analysis on firm-level outcomes within locations (Lemma 2), locations with lower labor market frictions (i.e., higher φ^E) have a higher average wage and higher average firm size, but lower average net poaching share.

4 Extensions

In our baseline model, we focused on the necessary ingredients to analyze the trade-off firms face when sorting across space in the presence of local on-the-job search. Our main result on spatial sorting, however, extends to more general environments, highlighting that it is quite robust. We discuss the extension that allows for labor mobility—which we will use in our quantitative exercise—and summarize the results on endogenous productivity spillovers; endogenous land supply; and endogenous vacancy creation. The details are in Appendix D.

4.1 Labor Mobility

We now relax the assumption that labor is fully immobile and allow unemployed workers to settle in any location. This is an important extension, endogenizing the population size, $L(\ell)$, which is here allowed to vary across locations. As a result market tightness $\theta(\ell)$ (which is defined as the measure of vacancies over the measure of searchers) is endogenous and so are the firms' meeting rate, $\lambda^F(\ell)$, and workers' meeting rates, $(\lambda^U(\ell), \lambda^E(\ell))$, and the unemployment rate $u(\ell)$. These location-specific meeting rates create another channel that affects the costs from competition and thus sorting. We will show, however, that similar conditions as in the baseline model guarantee positive sorting of firms to locations.

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

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

where $V^U(\ell)$ is given by (1). If location ℓ' has a better wage distribution or higher non-employment income than location ℓ'' (causing a temporary imbalance $V^U(\ell') > V^U(\ell'')$), then workers will move into ℓ'. This puts downward pressure on market tightness and thus workers' meeting rates in ℓ' until the difference in the locations' attractiveness is arbitraged away.

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

We assume, as before, that the measure of vacancies in each ℓ satisfies $\mathcal{V}(\ell) = 1$ (by assumption, Γ_ℓ is associated with measure one of firms) and $\mathcal{U}(\ell) = L(\ell)(u(\ell) + \kappa(1 - u(\ell)))$ is the effective measure of searchers in ℓ (now endogenous). We define market tightness in location ℓ by $\theta(\ell) = \frac{\mathcal{V}(\ell)}{\mathcal{U}(\ell)}$. We then express the meeting rates as $\lambda^F(\ell) := p^F\big(\theta(\ell)\big), \lambda^U(\ell) := p^W\big(\theta(\ell)\big)$, and $\lambda^E(\ell) := \kappa p^W\big(\theta(\ell)\big), \kappa < 1$, with the standard properties of functions p^W (increasing and concave) and p^F (decreasing and convex), both continuously differentiable.

In order to guarantee supermodularity of $\overline{J}(p, \ell)$ and thus PAM of firm types and locations, we need to ensure that productivity gains from settling in higher-ℓ locations dominate the costs that stem from more competition and smaller firms size—just as in the baseline model (see (14)). However, there is one key difference, which is that meeting rates are now endogenous. Competition has therefore two components. It depends not only on the local firm composition (as before) but also on local congestion (new). How the firm size elasticity in (14) varies across space depends on this local congestion.

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

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 (they have not only higher non-employment income, but also a stochastically better wage distribution), and draw in more workers. Congestion forces therefore benefit firms in high-ℓ locations, $\partial\lambda^F/\partial\ell > 0$ and $\partial\lambda^E/\partial\ell < 0$, strengthening their desire to settle there. As a result, PAM materializes more easily than in the baseline model, in which congestion was absent (meeting rates were exogenous and the same across space). We now state our result on firm sorting under labor mobility formally. To do so, denote the minimum of the first term on the R.H.S. of (14) (over ℓ, y) again by ε^P.

Proposition 4. *If z is strictly supermodular, and the productivity gains from sorting into higher ℓ, ε^P, are sufficiently large, then there is positive sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(R(\ell))$.*

See Appendix D.1 for the proof. We use this extension in our quantitative analysis below.

4.2 Additional Extensions

Endogenous Productivity Spillovers. We can generalize our model by allowing for endogenous productivity spillovers. Instead of assuming that location productivity A increases in ℓ exogenously, we here assume that locations are ex-ante identical but ex-post—after firms make location choices—differ in productivity, depending on the endogenous composition of firms, i.e. $A(\ell) = \tilde{A}(\Gamma_\ell)$. We assume that $\tilde{A}$ is a positive function that negatively depends on Γ_ℓ. Thus, better (worse) firm distributions (in the FOSD sense) in high-ℓ locations lead to higher (lower) location productivity, $\frac{\partial A(\ell)}{\partial \ell} = \tilde{A}' \frac{\partial \Gamma_\ell}{\partial \ell} > (<)0$. This captures the idea that there are positive spillover effects from productive firms onto all firms in a location.

In Proposition 7 , Appendix D.2, we show that if the (endogenous) location productivity advantage, along with the impact on firms' marginal productivity, is large enough relative to the cost of more severe competition in these locations, then highly productive firms (with high-p) indeed settle into high-ℓ locations—similar to the baseline model. The interesting feature is that essentially the same conditions on primitives can sustain both positive or negative sorting in terms of (p, ℓ). The reason is that the impact of the *endogenous* firm composition on firms' value function $\overline{J}$ changes the *complementarity properties* of $\overline{J}$, even if we maintain supermodularity of z. Proposition 8 show the equilibrium exists (but it is not unique).

Endogenous Supply of Land. We can also endogenize the land distribution $R(\cdot)$, allowing for the possibility of changes in the matching of firms to locations in response to changes in the model environment, even if positive sorting is maintained. We model this endogeneity of land supply through pre-match investment by land developers.[15] In Proposition 9, Appendix D.3, we show that the key-tradeoff between productivity and competition remains in place in this extended model, again guiding firms' sorting choices. Proposition 10 shows existence.

Endogenous Vacancy Posting. Finally, we endogenize vacancy creation instead of assuming that each firm has one open vacancy at all times. This extension—like the one with labor mobility—endogenizes the firms' and workers' meeting rates, which can now vary across locations, creating a second channel that affects the costs from competition and thus sorting. In Proposition 11, Appendix D.4, we show that the by now familiar trade-off between productivity and competition determines firm sorting choices also in this case, only that the conditions on primitives are slightly modified.

5 Empirical Analysis

To quantify the importance of firm sorting for spatial inequality, we estimate our model using matched employer-employee data from Germany. In this section we describe the data and test some qualitative predictions of our theory. In Section 7 we highlights its quantitative implications.

5.1 Data Description

Out main dataset is the Linked-Employer-Employee-Data (LIAB) provided by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research. The LIAB data links information on establishments from the IAB Establishment Panel, a representative annual establishment survey, with information on all individuals employed at those establishments. We focus on establishments for the years 2009-2016. For these establishments we observe individual-level information for all employees that have worked at one of these establishments for at least one day. Because we only observe data at the level of the establishment, we use the terms establishments and firms interchangeably throughout the paper. This individual-level data is assembled from the official social security

[15] The approach is similar to the pre-match investment in Chade and Lindenlaub (2021) but without aggregate risk and in an otherwise different model.

records. We for example observe the workers' gender, education, full-time employment status, and gross daily wages. Moreover, we observe basic employer information the worker's entire employment history between 1975 and 2018. Whereas the original data is in a spell format, we transform it into a monthly panel.[16]

We augment the LIAB data with information from the Establishment History Panel 7518 (BHP). Using this data allows us to measure for each establishment in the LIAB - in addition to standard information like total employment, average wages, sales or costs of inputs - total inflows and outflows of workers. We also include establishment-level fixed effects, provided by the FDZ for all establishments in the LIAB-BHP. These are estimated in Card et al. (2013) for the period 2010-2017 using the methodology developed in Abowd et al. (1999) (henceforth, AKM). For our final sample we focus on the period 2010-2017. We also drop establishments with less than 10 employees and whose mean real daily wage across the sample period is lower than ten Euros.[17].

Local Labor Markets We base our analysis on 204 travel-to-work areas (*Arbeitsmarktregionen*), which are similar to the US commuting zones. In Table 7 in Section E.1 in the Appendix, we report some basic statistics across commuting zones. On average, daily wages per full-time employee are 120EUR, average value added per full-time employee is 95,000EUR.[18] Furthermore, firms have 21 full time employees on average, the biggest 10% of employers in a local market account for 65% of employment and the average locality has a population of about 400,000. We also report the quantiles of these statistics to describe their spatial variation. Comparing the 10% and the 90% quantile of the respective local labor market distribution reveals that there are large differences across local labor markets. The 90-10 difference of daily wages is about 60EUR, i.e., approximately 50% of the average daily wage. Similarly, the distribution of average value added across locations is quite dispersed. In terms of the local firm size distribution, average size ranges from 15 to 27 employees and the employment share of the top 10% firms is between 50% and 60% across locations. Finally, locations differ substantially in their population size, ranging between ninety thousand and one million inhabitants.

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

[17]To compute real wages we deflate nominal wages using the German CPI (Table 61111-0001 in the GENESIS database of the Federal Statistical Office)

[18]We measure value added at the firm-level as the difference between sales and input costs as reported in the LIAB data. We deflate input costs using an index of wholesale prices (Table 61281-0001 in GENESIS database of the Federal Statistical Office). See also Bruns (2019).

We measure the parameters describing the frictional labor market, i.e. λ^E, λ^U, and δ, directly from our data. In the model, λ^E is a job offer arrival rate of an employed worker. We proxy this rate by the realized transition rate in the data. For the contact rate λ^U and the rate of job destruction δ, since unemployed workers accept all offers and separations to unemployment are exogenous, the arrival rates are equal to the realized rates. We measure these flows at the monthly frequency and then take the average to arrive at one number per local labor market.

5.2 Local Labor Markets: On-the-Job Search, Sorting and Spatial Inequality

Our theory stresses the role of on-the-job search (OJS) for the location choice of firms and for spatial inequality. In this section we empirically document that OJS routinely takes place within local labor markets, that firms are indeed spatially sorted and that spatial inequality is substantial.

On-the-Job Search In our context, OJS is an important source of firms' hiring. 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.[19] Given our focus on *local* labor markets, we also compute firms' share of local employment-to-employment (EE) inflows and local unemployment-to-employment (UE) flows, which we define as the share of all EE (or UE) inflows that stem from the same local labor market.

In Table 1, we summarize these measures at the firm-level (Panel A) and at the level of the local labor market (Panel B). There are two main take-aways. First, OJS is salient in the data. On average, about one third of all worker inflows, measured at the monthly level[20], are hires from other firms. Moreover, the variation across firms is large. At the top of the distribution (P90), almost 50% of inflows stem from other firms. At the bottom (P10), only 16% of firms' hires are from other firms. Second, a significant share of hiring—both from unemployment and from employment—stems from firms' local labor markets. In terms of EE flows, two out of three hires from other firms are done locally; and in terms of the UE flows, this is true for well over half of the hires from unemployment.

[19] 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.

[20] Since we restrict our sample only to the panel cases for Table 1, we measure these flows at the monthly level.

Table 1: OJS and Local Labor Markets

	Mean	S.D.	P10	P25	P50	P75	P90
Firm level (N = 4,690)							
Poaching	0.31	0.12	0.16	0.23	0.31	0.38	0.44
Share of local EE	0.62	0.20	0.34	0.50	0.64	0.78	0.85
Share of local UE	0.57	0.23	0.32	0.42	0.56	0.76	0.89
Commuting zone level (N = 204)							
Poaching	0.30	0.05	0.25	0.27	0.30	0.33	0.36
Share of local EE	0.61	0.10	0.46	0.56	0.64	0.67	0.72
Share of local UE	0.61	0.11	0.46	0.54	0.63	0.66	0.76

Notes: Data source: LIAB dataset, 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 weight firms by total employment. The commuting zone level statistics are weighted by number of firms in that location.

Finally, when we look at these statistics across labor markets, a similar pattern emerges whereby for the average commuting zone, 60% of all hires (both from employment and unemployment) are from the local labor market. Interestingly, the cross-sectional variation across markets is smaller than the variation across firms.

In Section E.4 in the Appendix, we provide further evidence that firms' competitiveness in their local labor market is an important determinant of their hiring, size and wage policy. We first use the variation across firms within local labor markets and show that the local productivity rank (as opposed to the nation-wise productivity rank) is the main determinant of firms' poaching shares, their size and their wage policies. In particular, we show that firms that are higher up in local job ladder have higher poaching shares, are larger and pay higher wages. We then use variation across local labor markets and relate the local size and wage distributions to the contact rates λ^E and λ^U and the destruction rate δ. As predicted by the theory we for example show that both the average and the dispersion of firm size and wages are increasing in the efficiency of EE hiring, i.e, in φ^E, and are independent in the efficiency of UE hiring, φ^U.

Spatial Sorting We now turn to the evidence on spatial sorting by firms. Consider first Table 2, in which we perform a variance-covariance decomposition of the distribution of firm productivity y across space. Specifically, we decompose the cross-sectional variation of y into the part that is explained by the 204 local labor markets in our sample and the residual. Table 2 shows that around 13% of the firm-level variation in productivity is due to variation across space and thus due to spatial sorting of firms. Recall that our measure of y already controls for differences in location productivity across space, so that the remaining spatial variation of y is due to the pure sorting effect. The fact that there is sorting while there is also ample overlap in the productivity distribution across commuting zones supports our model assumption of common support of firm productivity distributions across space (Assumption 1).

Table 2: Variance Decomposition of Firm Productivity

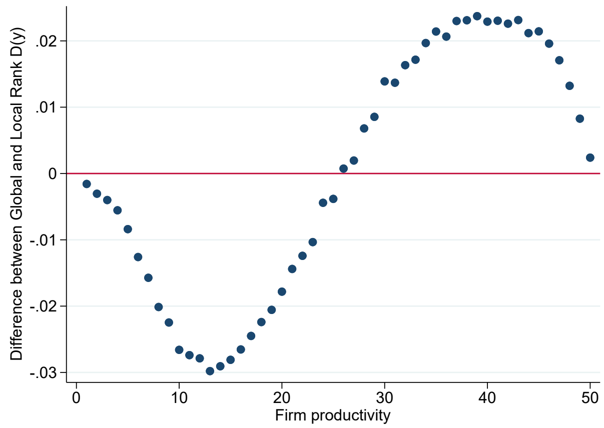
	Variance	Share
Total	0.0346	100
Local Labor Market	0.0044	12.84
Residual	0.0302	87.16

Notes: Data source: LIAB-BHP. Regression of y (residualized firm AKM effects) on commuting zone dummies. Weighted by mean number of full-time employees.

In Corollary 6 and Figure 1, we showed that our model suggests an alternative test of spatial sorting in that sorting implies an "S-shaped" relationship between firm productivity y and the difference between their global and local rank $D(y)$. In Figure 2, we plot this relationship in the data. The resemblance to our theoretical result displayed in Figure 1 is striking. On the horizontal axis, we order firms by their global rank in Germany (which moves one-to-one with firms' y) and categorize them into 100 equally sized bins (percentiles of the global productivity distribution in Germany). On the vertical axis, we display the average of the difference between the global and the local rank for each productivity bin. Hence, a negative number means that the global rank is below the local rank.

Figure 2 highlights two features of our theory. First, the least productive firms in Germany are at the bottom of the productivity ladder in all locations. Similarly, the most productive firms in Germany are at the top of the productivity ladder in all locations. Hence, $D(y)$ is close to zero at the two extremes. In our model, this is an implication of our assumption of common support of the distributions $\Gamma(y|p)$ for all p. The data seems to support this assumption. Second, as explained in Corollary 6, the remaining "S-shape" is an implication of firm sorting. Globally unproductive firms sort into locations where they are relatively more productive because their competitors are—like themselves—on average unproductive. Hence, for them 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 collocate with other productive firms, i.e.

Figure 2: Global versus Local Productivity Ranks



Notes: We rank firms by their residualized AKM fixed effects and put them in 50 bins of equal size. For each bin, we measure its local rank and global rank and calculate the average difference between the two. Data source: LIAB-BHP.

within their local labor market they are relatively unproductive and therefore $D(y) > 0$.

While Figure 2 suggests that there is spatial sorting of firms, it does not give insights into which form of sorting—positive or negative, in the way defined above—takes place. In Section E.5 in the Appendix, we provide evidence that is suggestive of sorting being positive. In particular, we ask whether local productivity (in our theory $A(\ell)$) or firms hypothetical local rank in location ℓ ($\Gamma_\ell(y)$) is a better predictor for firm y to actually settle in ℓ. We find that $A(\ell)$ has more predictive power for more productive firms, while the local rank $\Gamma_\ell(y)$ is more successful to empirically predict the location choice of less productive firms. These results indicate that productive and unproductive firms resolve the trade-off between local productivity and competition differently and that sorting is driven by a positive complementarity between location productivity and firm productivity.

Spatial Inequality A key object of interest of our analysis is the extent to which firm sorting amplifies spatial wage inequality. Table 3 shows that the spatial heterogeneity in average wages is substantial. In particular, a set of region FE at the level of the local labor market explains about 20% of the entire dispersion of log wages.

Table 3: Variance decomposition, log wages

	Total	Region FE	Residual	Cov
Variance	0.3341	0.0686	0.2654	0.0000
Share	100.00	20.53	79.44	00.00

Notes: The table reports a decomposition of log wages at the individual level into the part across local labor markets ("Region FE") and the residual.

Our theory, especially equation (17), highlights that this spatial variation reflects differences in the local job ladder. In particular, positive spatial sorting by firms implies that productive locations attract better firms, which leads to both a steeper job ladder and an employment distribution, which is concentrated among the upper end of the increasing wage schedule. Both of these forces tend to increase spatial inequality.

In Figure 3 below we provide direct evidence for this mechanism. In the left panel, we show that the returns to an E-E flow are higher in better locations. More precisely, we run a regression of the form

$$\ln w_{i\ell t} = \delta_\ell + \sum_{\ell=1}^{R} \gamma_\ell 1\left[EE_i = 1\right] + x'_{i\ell t}\zeta + u_{i\ell t} \tag{18}$$

, where $w_{i\ell t}$ denotes the wage of individual i in region ℓ at time t, δ_ℓ and γ_ℓ are region fixed effects and $1\left[EE_i = 1\right]$ is a dummy variable whether individual i experienced an E-E flow within the last month. The vector $x'_{i\ell t}$ absorbs a variety of individual characteristics like age and education. Our main coefficients of interest are γ_ℓ, which capture the regional heterogeneity in the returns to job-to-job flows. In the left panel of Figure 3 we depict a scatter plot of γ_ℓ against an the regional index ℓ, where localities are sorted - in accordance to our theory - by $A(\ell)$. This ranking stems from our structural estimation, which we explain in detail below. The figure shows first that $\gamma_\ell > 0$ for (essentially) all locations. Hence, job-flows are on average associated with higher wages. More importantly, γ_ℓ is higher in high ℓ locations, that is more productive locations have indeed a steeper job ladder as implied by our theory. Qualitatively, these differences are meaningful: a single job-to-job move in a productive location increases wages by about 3% more than in an unproductive location. Given that annual wages of stayers increase by about 3%, this means that a job-to-job flow at the top of the spatial productivity distribution is worth about one year of wage growth.

In the right panel of Figure 3 we focus directly on the differences between average wages and wage of recently employed individuals across space. More productive locations offer higher wages both because of their superior location productivity and their better firm composition. To highlight the role of the latter channel, we compare the average wage *relative* to the wage paid to workers that just entered the labor force out of unemployment. By scaling average wages with the average wage of the recently unemployed, we control for differences in the wage intercept, which directly reflects location productivity A_ℓ, and focus squarely on the employment composition. Figure 3 shows this statistic is increasing in the location index ℓ, as implied by a steeper job ladder in more productive locations.

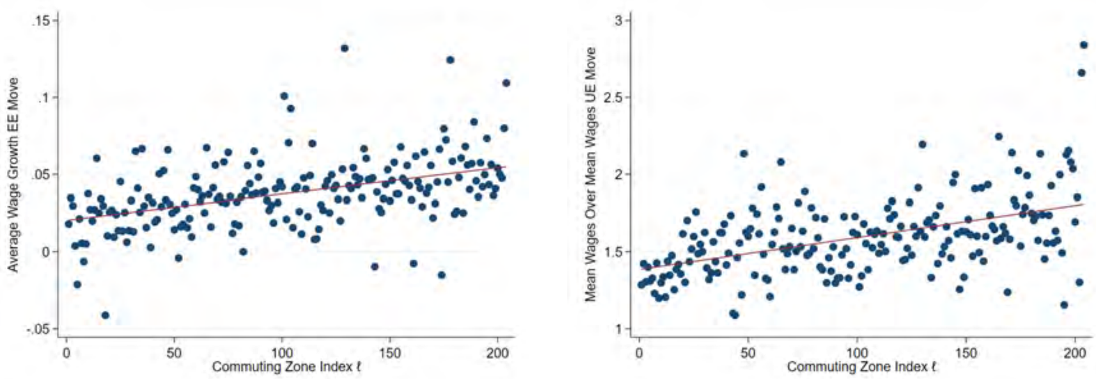


Figure 3: Spatial Inequality and the Local Job Ladder

6 Estimation

To evaluate the quantitative importance of firm sorting for spatial wage inequality, we estimate our model. We first outline the parameterization. We then show that our model is identified and discuss the estimation strategy. Finally, we present the estimation results and the fit between model and data.

6.1 Quantitative Model

Our quantitative model is based on the extension with labor mobility of the unemployed (Section 4.1). This feature is important since, even though we observe a high degree of local hiring, local labor markets are not perfectly segmented in the data. Moreover, labor mobility causes the (now endogenous) population size as well as meeting rates of workers and firms to vary across locations—in line with the evidence. To be able to fit the observed meeting rates across space, while still ensuring that unemployed workers are indifferent between all locations, we introduce a function B that depends on ℓ (to be estimated), where we interpret $B(\ell)$ as a location preference, which could e.g. reflect local amenities. We postulate that unemployed workers' flow payoff is $b(\ell)B(\ell)$, while employed workers' flow payoff is $w(y,\ell)B(\ell)$.

We choose a semi-parametric estimation approach below. We assume that the meeting rates are based on a Cobb-Douglas matching function $M(\ell) = \mathcal{A}\mathcal{V}(\ell)^{\alpha}\mathcal{U}(\ell)^{1-\alpha}$, where $\mathcal{U}(\ell) = L(\ell)(u(\ell)+\kappa(1-u(\ell))$ is the measure of effective job searchers, κ is the relative matching efficiency of employed workers, and $\lambda^E(\ell) = \kappa\lambda^U(\ell)$; and where $\mathcal{A}$ is the overall matching efficiency and $\mathcal{V}(\ell) = 1$ is the measure of vacant jobs in each location. We set $\alpha = 0.5$. In turn, we assume that the production function is multiplicative, $z(A(\ell), y) = A(\ell)y$;[21] and the distribution from which firms draw their ex-post productivity is Pareto, $\Gamma(y \mid p) = 1 - y^{-\frac{1}{p}}$, where we normalized the location parameter, $\underline{y} = 1$, and the scale parameter is $1/p$. As a result, firms with higher ex-ante productivity p draw their ex-post productivity from a stochastically better distribution with fatter tail. In Appendix F.1, we provide a detailed discussion, justifying the pareto assumption based on log-linear tails of the observed firm productivity distributions (using AKM fixed effects). Further, we need to normalize land distribution R (for identification reasons, see below), and assume it is uniform. We do not assume any functional forms for (Q, A, B) but will identify and estimate them non-parametrically.

6.2 Identification

We need to identify the ranking of locations $[\underline{\ell}, \overline{\ell}]$, functions ($Q$, A, B), which indicate the firms' ex-ante productivity distribution, the location TFP and the location amenities, and the parameters of the matching function $(\kappa, \mathcal{A})$. We will show that these objects are identified.

First, the ranking of locations, which lets us assign an ℓ to each place, is identified from any observable statistic that increases in location productivity, e.g., by a ranking of the average wage across locations (see Lemma 5, Appendix C.2.1).

Second, the ex-ante firm productivity distribution, Q, is identified from variation of the labor share across ℓ. We can show that, under the pareto assumption, the labor share in location ℓ is $LS(\ell) := 1-\mu(\ell)$, which—under PAM—is decreasing in ℓ. Thus, high-ℓ locations with more productive firms (higher $p = \mu(\ell)$) have *lower* labor share. The intuition is that locations with more productive firms on average also have more productivity dispersion, reducing the competition among firms at the top and thus their labor share.[22] Given the matching function μ, we can identify Q, using $\mu(\ell) = Q^{-1}(R(\ell))$ (since R is normalized).

[21]We could specify $g(A(\ell))$ for some function g (instead of just $A(\ell)$), but we cannot separately identify g and A.
[22]For a similar result in a one-market economy, see Gouin-Bonenfant (2020), who assumes Pareto of value added, z, not y.

Third, given the firm allocationμ, we can identify the location productivity schedule A from the variation in average value added across space:

$$A(\ell) = (1 - \mu(\ell))\mathbb{E}[z(A(\ell), y)], \tag{19}$$

where we make use of the pareto assumption. That is, the variation in value added across locations that is *not* accounted for by firm sorting μ must be driven by location TFP A.

Fourth, the labor market matching function parameters are identified as follows. The relative matching efficiency of employed workers is identified from their relative job finding rate. In turn, the overall matching efficiency is identified from a combination of the job finding rate of unemployed workers and the average firm size in any location:

$$\mathcal{A} = \left(\frac{\delta + \kappa\lambda^U(\ell)}{\delta + \lambda^U(\ell)} (\lambda^U(\ell))^{\frac{1}{2}} \left(1 + \frac{\delta}{\lambda^U(\ell)}\right) \bar{l}(\ell) \right)^2, \tag{20}$$

where δ is the job destruction rate, identified from the observed job separation rate.

Fifth, we identify the amenity schedule B, using the indifference condition that unemployed workers' value of search is equalized across space:

$$\rho V^U = B(\ell)A(\ell)\left(1 + 2(\lambda^E(\ell))^2 \int_1^\infty y^{-\frac{1}{\mu(\ell)}} \frac{1}{\mu(\ell)} y^{-\frac{1}{\mu(\ell)}-1} \int_1^y \frac{1}{[\delta + \lambda^E(\ell)t^{-\frac{1}{\mu(\ell)}}]^2} dtdy\right), \tag{21}$$

when imposing the normalization $\rho V^U = 1$. We now summarize this discussion:

Proposition 5 (Identification)**.** *Under the assumed functional forms (summarized in Assumption 5, Appendix F.2), the model is identified.*

We provide more details on the derivations of this section in the proof, see Appendix F.2.

6.3 Estimation Strategy and Results

Our identification argument provides us with a concrete estimation protocol that we will closely follow. As with identification, we will proceed in five steps.

First, we specify our location unit as *Commuting Zone* (CZ), of which there are 204 in Germany. We assign each CZ an index ℓ, based on the first principal component of several variables that—according to the model—are increasing in location productivity (average location wage, average location value added, gross sales). In addition, we include the location's mean log worker fixed effect from the AKM regression (provided by the FDZ) into the principal component analysis.[23] This procedure gives us a (productivity) ranking of locations. Examples of some of the highest ranked commuting zones are Munich and Frankfurt while among the lowest ranked we have Goerlitz (East) and Kyffhaeuserkreis (rural East).

Second, we obtain $\mu(\ell)$ from the observed labor share in ℓ. Because our model is highly stylized (e.g., it does not feature any noise in the firm-location matching process), we smooth out the noise in the data, which is possibly due to measurement error. Specifically, we approximate the data series that we target in this estimation by flexible polynomials as a function of ℓ before feeding them into the model. Since the (approximated) labor share is decreasing in ℓ (Figure 4, top left panel), we obtain an increasing matching function μ (top right panel), indicating that high-ℓ locations are characterized by stochastically better firm distributions.

Third, we estimate location TFP based on the average (approximated) value added across locations using (19). Since, value added is strongly increasing in ℓ (Figure 4, middle left panel), we obtain an increasing A-schedule, even after controlling for firm sorting through μ (Figure 4, middle right panel).

Fourth, we obtain the overall matching efficiency from the median observed matching rate and median location's firm size, using (20) (see Table 13, Appendix F.3). In turn, we obtain the relative matching efficiency of the employed workers from the data as $\kappa = 0.65 \cdot \frac{\lambda^E}{\lambda^U}$, where we use the Germany-wide job finding rates for λ^E and λ^U and where we adjust this fraction further to take job-to-job moves into account only if they are associated with a higher wage, see Table 13.[24] Moreover, to avoid using noisy CZ-specific matching rates from the data directly, we 'estimate' the (endogenous) meeting rates across locations by targeting how the average firm size varies across space (Figure 4, bottom left panel). And we verify the fit of these estimated λ's with the empirical job finding rates ex-post in the spirit of an over-identification test. More specifically, we back out $\lambda^U(\ell)$ in each ℓ based on the observed average firm size $\bar{l}(\ell)$ (using our model equations, see Appendix F.2 for the derivation):[25]

$$\left(1 + \frac{\delta}{\lambda^U(\ell)}\right) \bar{l}(\ell) = \mathcal{A}^{\frac{1}{\alpha}} \frac{\delta + \lambda^U(\ell)}{\delta + \kappa\lambda^U(\ell)} \left(\frac{1}{\lambda^U(\ell)}\right)^{\frac{1}{\alpha}}$$

We then derive $\lambda^E = \kappa\lambda^U$ and $\lambda^F(\ell) = \lambda^U(\ell)/\theta(\ell)$. We obtain a slightly decreasing λ^U, in line with the data, implying lower meeting rates for workers and higher meeting rates for firms in more productive locations. See Figure 4, bottom right panel.

[23] This way, we interpret location productivity broadly, also allowing for worker productivity to be factored in.

[24] Sorkin (2019) finds that around 65% of EE moves are associated with wage gains.

[25] The average firm-size by CZ is less noisy than the CZ-specific λ^U, which is based on a much smaller sample.

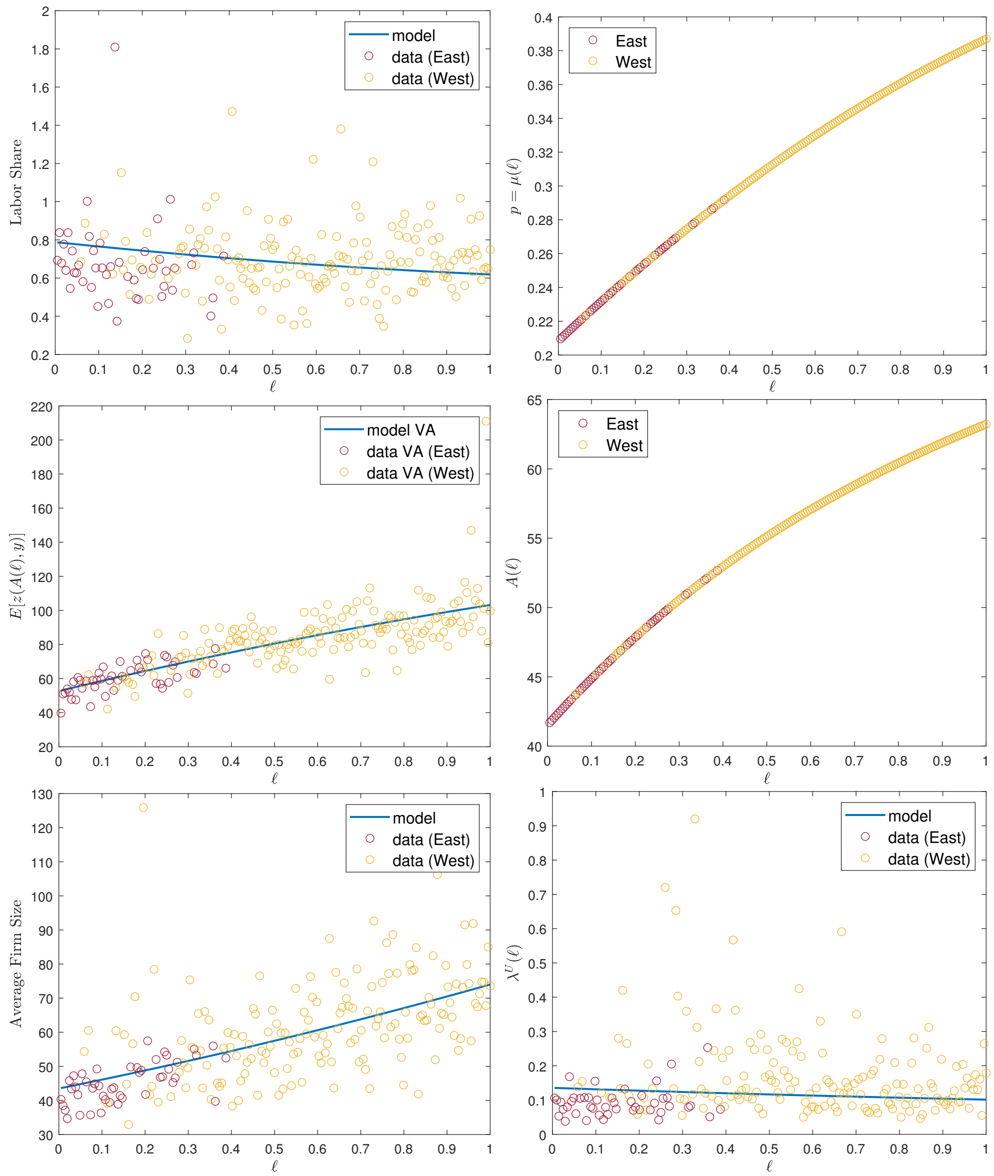


Figure 4: Fit and Estimation Results

Fifth, with $(\mu(\ell), A(\ell), \lambda^E(\ell))$ for each ℓ in hand, we use (21) to back out the amenity schedule B that ensures that unemployed workers are indifferent between all locations. Figure 11 (Appendix F.3) shows that amenities are slightly decreasing in the location productivity index. If not, high-ℓ locations would attract all workers due to higher TFP *and* better firms.

Note that the model equations we used for estimation were based on the premise of positive sorting of firms across space. Given the estimation output, we verify that the value of firm p of settling in ℓ, $\overline{J}(\ell, p)$, is indeed supermodular in (p, ℓ), verifying the premise. Importantly, at no point of the estimation did we target PAM.

6.4 Model Validation

We clearly fit the targeted (approximated) data series of labor share, value added and firm size exactly, which is due to our estimation approach (Figure 4, left column). Importantly, for model validation, we also fit several non-targeted features of the data quite well. In Figure 5, we show the model's performance regarding the average wage, left panel, whose slope we are slightly overestimating; average population size (middle panel);[26] and land price, which is convex in ℓ both in data and model (right panel).

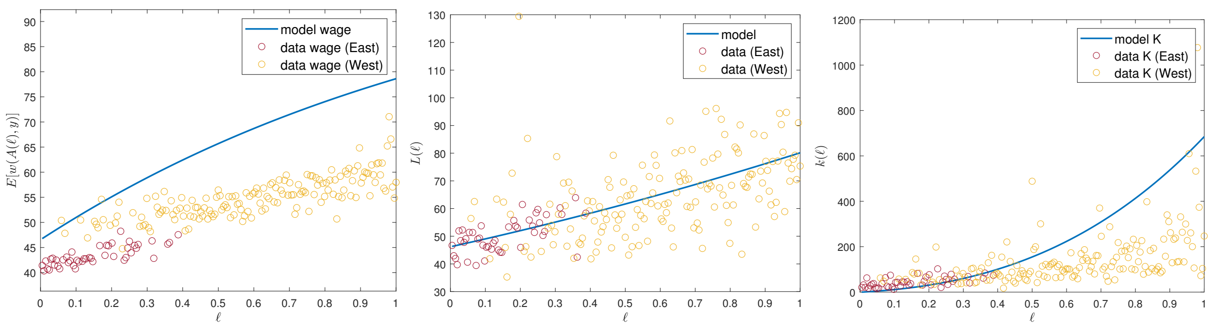


Figure 5: Fit: Non-Targeted Moments

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

7 The Drivers of Spatial Inequality

This section uses the estimated model to decompose and understand the drivers of spatial inequality, where we focus on firm sorting, on-the-job search, and spatial frictions. In the text, we concentrate on the East-West divide in Germany. Appendix F.5 analyzes the urban-rural gaps as well as inequality across commuting zones.

7.1 Spatial Inequality Through the Lens of Our Model

We first show that the estimated model captures the observed spatial inequality between East and West Germany quite well. In the data, daily nominal wages in the West exceed those in the East by 42% (Table 14), where we weigh these firm-level statistics by employment. Comparing data and model, we find that our model slightly understates spatial inequality in wages but still fits the non-targeted West-East gap with 34% relatively well.

[26] We normalize population size in each CZ by the number of firms located there.

Table 4: West-East Wage Inequality

	Data	Model
West	73.64	85.75
East	51.84	63.82
West/East	1.42	1.34

We now use our model to shed light on how much of West-East inequality is due to differences in location fundamentals versus firm sorting. To this end, we first conduct the following wage variance decomposition that lets us separate the between-region from the within-region variation:

$$\mathrm{Var}(\log w(A(\ell), y)) = \mathrm{Var}(\mathbb{E}[\log w(A(\ell), y) \mid \widehat{\ell}]) + \mathbb{E}[\mathrm{Var}(\log w(A(\ell), y) \mid \widehat{\ell})]$$

where $\widehat{\ell} \in \{East, West\}$. Using the model, we can further decompose the first term—across-location component—into a pure location component (local TFP), firm component (differences in mean firm productivity across $\widehat{\ell}$) and a sorting component (sorting between $\widehat{\ell}$—and thus A—and p, where higher p ex-ante implies stochastically better y ex-post):[27]

$$\begin{aligned}
\mathrm{Var}(\mathbb{E}[\log w(A(\ell), y) \mid \widehat{\ell}]) &= \mathrm{Var}(\mathbb{E}[\log A(\ell) \mid \widehat{\ell}] + \mathbb{E}[\log \tilde{w}(\ell, y) \mid \widehat{\ell}]) \\
&= \mathrm{Var}(\mathbb{E}[\log A(\ell) \mid \widehat{\ell}]) + \mathrm{Var}(\mathbb{E}[\log \tilde{w}(\ell, y) \mid \widehat{\ell}]) + \\
&\quad 2\,\mathrm{Cov}(\mathbb{E}[\log A(\ell) \mid \widehat{\ell}], \mathbb{E}[\log \tilde{w}(\ell, y) \mid \widehat{\ell}])
\end{aligned}$$

Table 5 shows the results. Around 15% of the overall cross-sectional inequality in wages can be accounted for by the East-West wage gap. Further decomposing this across-component into how much is due to location heterogeneity A (first component) versus firm sorting (second plus third component), we find that the contribution of firm sorting to spatial inequality is quantitatively important. Around a half of the variation in wages across space is accounted for by the fact that better firms settle in more productive (West) locations.

We produce a comparable decomposition for spatial inequality in *value added* with similar results, see Appendix F.4.1.

[27] We defined $\tilde{w}$ as the component of the wage that is independent of $A(\ell)$:

$$\tilde{w}(\ell, y) := \frac{w(A(\ell), \ell)}{A(\ell)} = y - \left[1 + \varphi^E(\ell)\,(1 - \Gamma_\ell(y))\right]^2 \int_{\underline{y}}^{y} \frac{1}{\left[1 + \varphi^E(\ell)\,(1 - \Gamma_\ell(t))\right]^2} dt.$$

Table 5: Wage Variance Decomposition, East-West

	Variance	Across	$A(\ell)$	$\tilde{w}(\ell)$	2Cov
$\log w$	0.081	14.75	51.54	7.96	40.51

What is the mechanism for why firm sorting and differences in location fundamentals lead to higher wages in the West? Our theoretical decomposition of the spatial wage premium (17) sheds light on this when applied to the West-East wage gap, where $\underline{\ell}$ pertains to the East, and $\ell > \underline{\ell}$ pertains to the West. First, a higher TFP in high-ℓ (West) locations leads to a larger intercept of the wage function and makes it steeper, compared to low-ℓ (East) locations—both aspects fuel spatial inequality. Second, due to spatial firm sorting, higher-ℓ (West) locations have a better firm composition so that more employment is clustered at the upper part of the wage schedule where wages are higher, compared to $\underline{\ell}$ (East), which also increases the spatial wage gap.

7.2 Counterfactuals

We now consider several counterfactual exercises to isolate the role of firm sorting, on-the-job search and spatial frictions in spatial inequality. In each of them, we unpack the mechanism behind the resulting inequality changes by analyzing how the different components of spatial wage inequality (17)—location differences in wage intercept, slope, and employment composition—are affected. When conducting the counterfactuals we maintain Assumption 1. As a result, the overall firm distribution stays the same as in the baseline model.[28]

The Role of Firm Sorting. We first assess the role of firm sorting in spatial inequality by shutting down sorting entirely. This random matching benchmark lets us quantify how much of the West-East gaps is due to the fact that better firms tend to settle into West Germany. See Appendix F.6 for the technical details on the implementation of this exercise.

Table 6, column 2, indicates that spatial firm sorting can account for approximately 32% of the East-West wage gap. Our conclusion is that the East is not only disadvantaged because of poor economic fundamentals (low A). This drawback is amplified by the fact that low-productivity firms tend to cluster in those locations.

[28] By not taking into account changes in firm selection in these counterfactuals, we neglect a potentially interesting margin of adjustment. On the other hand, this approach allows us to analyze all three counterfactuals in a coherent manner since at least one of them—shutting down firms' hiring frictions—would be intractable when relaxing Assumption 1 (details are available upon request). Moreover, under this approach we obtain a conservative estimate of inequality reduction in each of the counterfactuals. We ensure Assumption 1 in the counterfactuals, i.e. $w^R(\ell) = A(\ell)\underline{y}$ for all ℓ, by adjusting $b(\ell)$.

Table 6: West-East Wage Inequality—Counterfactual Models

	Model	No Sorting	No OJS	No Worker Frictions	No Firm Frictions
Wage West/East	1.34	1.23	1.24	1.05	1.02

Zooming into (17), the decomposition of the spatial wage premium, helps us understand why the lack of firm sorting curbs inequality. First, random matching of firms to locations reduces job ladder differences across space. While the job ladder was steeper in a typical West compared to East location at baseline (compare yellow solid with red solid wage schedules in Figure 6, left), this difference vanishes when firm sorting is absent (compare yellow and red dashed schedules). Moreover, while in the baseline model the employment distribution in West stochastically dominates that in East (compare yellow solid cdf with red solid cdf, right panel) with the effect that considerably more employment is clustered at the upper part of the wage schedule, this difference is almost entirely undone in this counterfactual. The reason is that in the baseline model, the FOSD of the Western employment distribution is mostly driven by firm sorting and thus a better firm mix. Both effects—changes in relative steepness of the job ladder and in the employment compositions—mitigate spatial wage inequality.

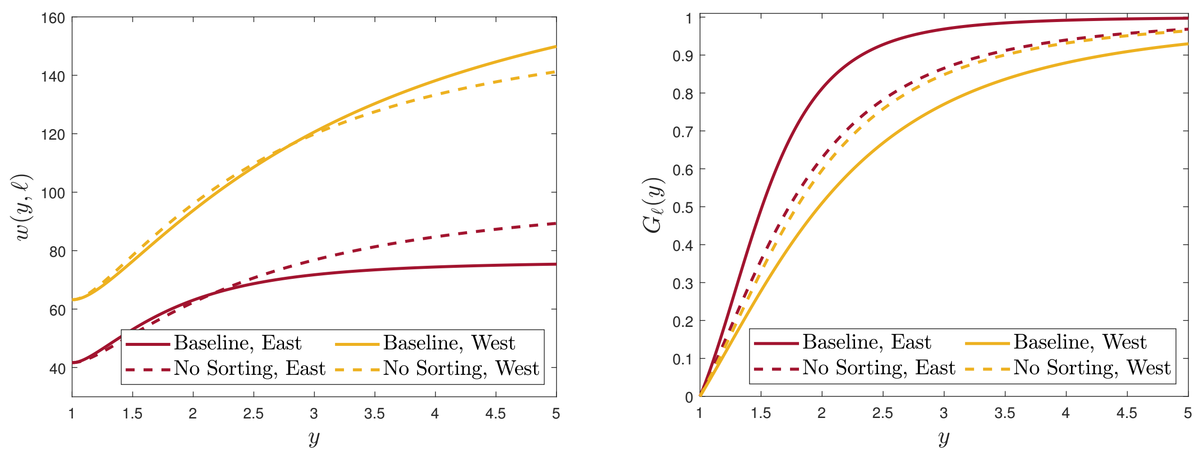


Figure 6: No Sorting: Wages and Employment Distribution

The Role of On-The-Job Search. To illustrate that on-the-job search has a quantitatively important role in spatial wage inequality, we consider a counterfactual where the EE transition rate is very small ($\kappa \to 0$, in practice 10% of baseline), while not vanishing entirely to avoid the 'spatial version' of the Diamond paradox.[29] Technical details are in Appendix F.6.

[29] We have a discontinuity at $\kappa = 0$: The value of search becomes $\rho V^U(\ell) = A(\ell)B(\ell)$ and cannot be equalized across ℓ. Under our calibration, $A(\ell)B(\ell)$ is decreasing, so all workers would sort into $\underline{\ell}$ and get the same monopsony wage in that

This 'No OJS' case decreases the West-East wage gap by 29% (compare columns 3 to 1 in Table 6). This is again due to both, changes in job ladders and in employment distributions across space. Job ladders are flattening everywhere (compare solid baseline functions to dashed counterfactual ones in Figure 7, left), reducing the economic attractiveness of West over East and also spatial inequality. As a result, workers' location choices are mostly guided by preferences $B(\ell)$, leading to a large inflow into low-ℓ (East) locations. This drives up firms' meeting rates in East, flipping the modularity properties of $\bar{J}(p, \ell)$ to submodular. Negative sorting of firms across space is the consequence, undoing the stochastic dominance of the employment distribution of West (right panel) and further curbing the West-East wage gap.

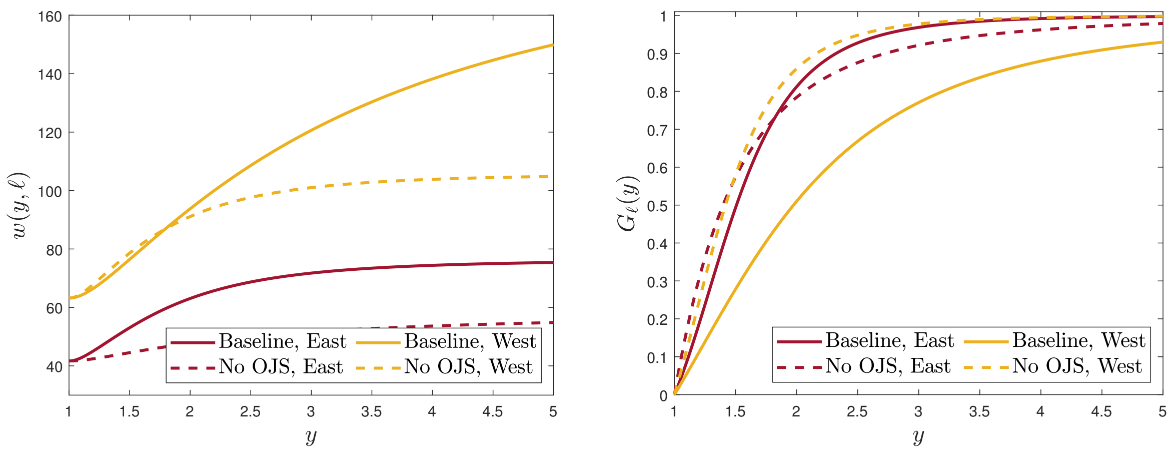


Figure 7: No (Small) On-The-Job Search: Wages and Employment Distribution

We stress that there is an important interaction between firm sorting and on-the-job search. With small OJS intensity, firm sorting becomes negative and actually *reduces* spatial inequality. The firm mix is now better in the East, improving the employment composition there. Thus, more workers in the East are at the upper part of wage ladder (which is much flatter than in the baseline), mitigating spatial inequality. In turn, in the limit with no OJS, all job ladders collapse and firm sorting has no effect on spatial inequality.

Finally, it is interesting to note that as we decrease κ in a more continuous way, firm sorting patterns (and therefore inequality) changes in a smooth way from PAM to NAM. That is, for intermediate κ sorting is a mix of PAM and NAM. There exists an $\ell^*(\kappa) \in (\underline{\ell}, \overline{\ell})$, above which there is positive sorting and below which there is negative sorting, see the left panel of Figure 8 (going from PAM in green, to a mix of PAM and NAM in orange, to NAM in purple). As κ decreases, ℓ^* increases, and thus for small enough levels of OJS sorting is

location. The equilibrium degenerates. (Spatial) inequality collapses to zero.

NAM, as described above.

The reason why PAM gradually becomes NAM is driven by worker reallocation across space, which determines firm contact rates λ^F. For intermediate levels of OJS, the decreasing market size (and thus decreasing $\lambda^F(\cdot)$) dominates the productivity forces (increasing $A(\cdot)$) for low-ℓ regions—leading to submodular $\overline{J}$ and thus NAM; in turn, market size and productivity forces reinforce each other for high-ℓ regions—preserving supermodular $\bar{J}$ and therefore PAM. Population density in ℓ depends on preferences and average wages (determined by firm composition). Stronger preferences attract more workers to $\underline{\ell}$ among lower-ℓ locations ($\ell < \ell^*$) while higher average wages attract more people to $\overline{\ell}$ among higher ranked locations ($\ell > \ell^*$), see (orange function in the middle and right panels). Lastly, we have a discontinuity of firm sorting, population size and wage at ℓ^*, where the modularity properties of $\bar{J}$ flip. Workers in ℓ slightly larger than ℓ^* can more easily find a job (smaller population) but receive lower wages on average (worse firm types), compared to those in ℓ slightly smaller than ℓ^*.

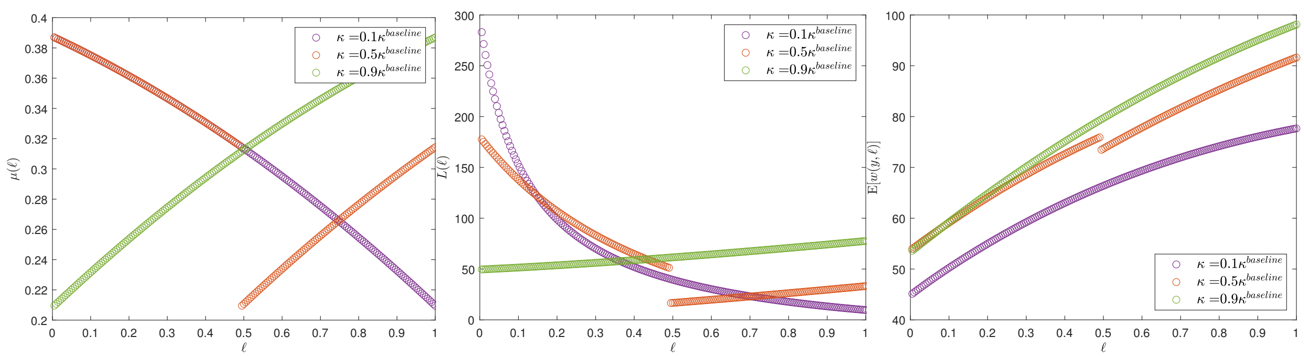


Figure 8: $\kappa = 10, 50, 90\%$ of the Baseline: Sorting, Population Size, Average Wage

The Role of Spatial Frictions. In the final counterfactual, we elicit the role of spatial frictions in spatial inequality. We sequentially shut down the model's two major frictions, pertaining to both workers and firms: first, the mobility friction that induces *workers* in the baseline model to settle in non-productive locations with poor firm mix, which we eliminate by equalizing location preferences $B(\ell) = B$; second, the hiring friction that *firms* faced in the baseline model, when forced to hire on-the-job searchers from their local labor market only. We shut down the firms' friction by integrating the labor market, so that the economy consists of a single job ladder with all firms hiring from everywhere. Note that neither of these counterfactuals changes the incentives of firms to sort across space, i.e., PAM is still in place. Technical details on the implementation of these counterfactuals are in Appendix F.6.

Shutting down the *mobility friction of workers* reduces spatial inequality substantially. Comparing columns 4 and 1 of Table 6 indicates that if non-productive locations lose their amenity advantage, so that bulk of their workers flow into the most productive regions inhabited by productive firms, the West-East wage gap declines by 85%. Zooming into the decomposition of the spatial wage premium (17) sheds light on the forces behind these shifts.

First, the wage schedule—and thus job ladder—in the West becomes flatter compared to the East, relative to the baseline model (Figure 9, left), reducing the West-East divide.

However, the main trigger of reduced inequality is the increase in population size in productive locations (high ℓ, which coincides with West Germany). It leads to severe congestion in these labor markets $\partial\lambda^E/\partial\ell << 0$ and prevents the job ladder in high ℓ locations from working effectively. This force counteracts the spatial sorting of firms and deteriorates the employment distribution G_ℓ in West compared to East relative to the baseline model, thereby pushing toward less spatial inequality.[30] This is illustrated in Figure 9, right, which shows that in this counterfactual G_ℓ of a typical West location no longer stochastically dominates the corresponding employment distribution of a typical East location (i.e., it is no longer true that $\partial G_\ell/\partial\ell < 0$).

Next, we additionally shut down the *spatial hiring friction of firms*. The counterfactual economy features a single integrated labor market as opposed to the many segmented local labor markets from the baseline model. Firms are effectively characterized by the productivity index $z = A(\ell)y$ and workers climb the global z-job ladder, facing no geographic constraints as to which firms can recruit them. Comparing column 5 with column 4 of Table 6 indicates that eliminating both the friction on the worker *and* the firm side further reduces spatial inequality, almost entirely eliminating the wage gap between the East and the West.

Our wage premium decomposition again provides guidance into the driving forces behind these shifts. Although the job ladder steepness does not change much (Figure 9, left), the intercept in West declines. With a single economy-wide job ladder, each location's wage intercept is determined by that global ladder, leading to intercepts that are increasing more slowly across ℓ compared to baseline with a dampening effect on inequality. The main driver, however, are shifts in employment composition in high-ℓ versus low ℓ-locations. Positive firm

[30] To see these countervailing forces clearly, note that based on the employment distribution G_ℓ in (13):

$$\frac{\partial G_\ell(y)}{\partial \ell} \overset{\text{s}}{=} \underbrace{\frac{\partial \Gamma_\ell}{\partial \ell}}_{\text{Firm Sorting }(-)} (\delta^2 + \delta\lambda^E(\ell)) \underbrace{-\frac{\partial \lambda^E(\ell)}{\partial \ell}}_{\text{Congestion }(+)} \Gamma_\ell(y)(1-\Gamma_\ell(y))\delta.$$

sorting across space still pushes toward stochastically better employment distributions G_ℓ in high-ℓ (West) locations (fueling spatial inequality); but this factor is counteracted by important composition shifts that improve the employment distributions in low-ℓ (East) locations (dampening inequality). Here these composition shifts are not due to congestion forces as above (λ^E is equalized across ℓ in this global labor market) but due to the differential positioning of local firms in the global job ladder.[31] In high-ℓ (West) locations, firms at the bottom of the local job ladder were disproportionally hurt under labor market segmentation (they faced severe competition). But—due to a high location fundamental A that increases their z in the global job ladder—they are globally competitive under labor market integration. They gain employment relative to more productive firms in their locations, so weight is shifting towards less productive firms, deteriorating G_ℓ. The opposite is true for low-ℓ locations (East). Due to positive firm sorting, the worst firms in those locations were shielded from tough competition under segmentation. Under labor market integration, however, they lose employment, shifting weight toward more productive firms in their locations, improving G_ℓ. Indeed, Figure 9 (right, dashed functions) indicates that G_ℓ in a typical East CZ now stochastically dominates the employment distribution in a typical West CZ.

Inequality versus Aggregate Output. While our main focus is on understanding spatial wage inequality, we now briefly discuss how aggregate output is affected by each counterfactual. First, positive firm sorting increases total output due to the complementarity between local TFP and firm productivity in z. Under random matching, the output loss is 11%. Second, without OJS, total output decreases even more (about 26%) since (i) firms now sort negatively in space, forgoing the productive complementarities, (ii) job ladders essentially collapse, preventing the efficient allocation of workers to productive firms; and (iii) workers cluster in less productive regions. Finally, and contrary to the previous two counterfactu-

[31] Note that in this counterfactual with global labor market, G_ℓ is no longer given by (13) but by the following:

$$G_\ell(y) = \frac{\int_{\underline{y}}^{y} \frac{\tilde{g}(A(\ell)y')}{\tilde{\gamma}(A(\ell)y')}\gamma_\ell(y')dy'}{\int_{\underline{y}}^{\overline{y}} \frac{\tilde{g}(A(\ell)y')}{\tilde{\gamma}(A(\ell)y')}\gamma_\ell(y')dy'}$$

where $\tilde{\gamma}$ and $\tilde{g}$ are the densities corresponding to cdf's:

$$\tilde{\Gamma}(t) = \mathbb{P}(z < t) = \int_{\underline{\ell}}^{\overline{\ell}} \Gamma\left(\frac{t}{A(\ell')}\middle|\mu(\ell')\right) dR(\ell')$$
$$\tilde{G}(z) = \delta \frac{\tilde{\Gamma}(z)}{\delta + \lambda^E\left(1 - \tilde{\Gamma}(z)\right)}.$$

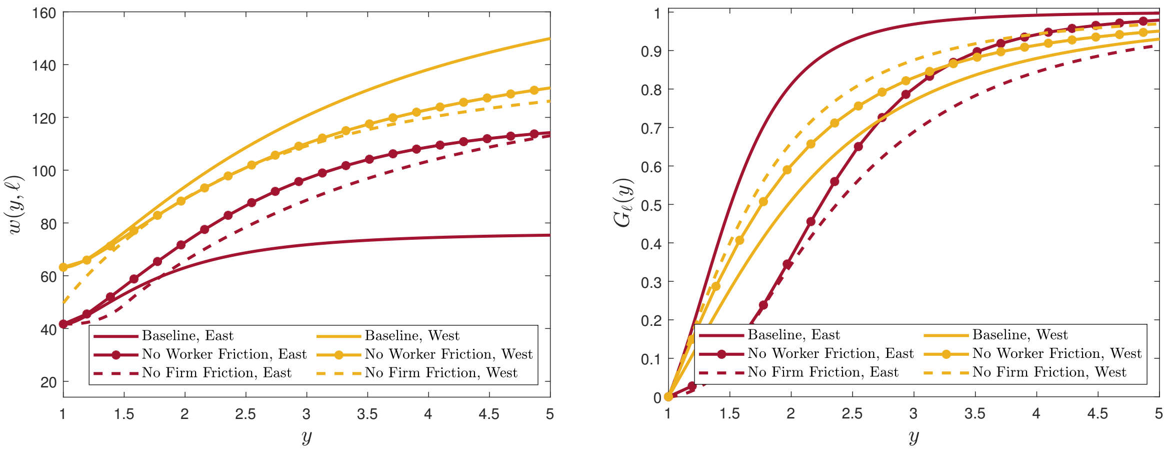


Figure 9: No Spatial Frictions: Wages and Employment Distribution

als, removing spatial frictions increases total output (about 8% and 11% respectively) by attracting more workers to productive locations with productive firms, either by removing the amenity advantage of non-productive regions or by integrating the German labor market.

8 Conclusion

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

We study firms' location decision in the context of a model with on-the-job search and spatially segmented labor markets. Our theory highlights that firms face a fundamental trade-off when deciding which labor market to enter in. 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 as they lose employees quickly and have a hard time poaching workers from other firms. The degree of firm sorting in equilibrium thus depends on the balance of these forces.

We characterize this trade-off analytically and provide sufficient conditions for positive sorting, i.e. an allocation, in which more productive firms settle in more productive locations. We show that positive sorting emerges as the unique equilibrium outcome if firm and location productivity are complements and labor market frictions are sufficiently large. We also derive the equilibrium schedule of land prices and show that the land price gradient reflects both productivity difference and local labor market competition.

Using matched employer-employee data from Germany, we estimate our model to quantify the role of firm sorting in spatial inequality. In our application, we assess the importance of firm sorting for wage differences between East and West Germany. We find that approximately 30% of the East-West wage gap can be accounted for by firm sorting, i.e., by the fact that more productive firms settle in the West.

References

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

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

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

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

Bilal, A.: 2020, 'The Geography of Unemployment'.

Bonhomme, S., K. Holzheu, T. Lamadon, E. Manresa, M. Mogstad, and B. Setzler: 2020, 'How Much Should we Trust Estimates of Firm Effects and Worker Sorting?'. Working Paper 27368, National Bureau of Economic Research.

Bossler, M., G. Geis, and J. Stegmaier: 2018, 'Comparing survey data with an official administrative population: Assessing sample-selectivity in the IAB Establishment Panel'. *Quality and quantity* **52**(2), 899–920.

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'. **39**(2), 257–273.

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

Chade, H. and I. Lindenlaub: 2021, 'Risky Matching'. *The Review of Economic Studies*. rdab033.

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*. jvac004.

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

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

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

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

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

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

Heise, S. and T. Porzio: 2021, 'The Aggregate and Distributional Effects of Spatial Frictions'. Working Paper.

Hornstein, A., P. Krusell, and G. L. Violante: 2011, 'Frictional Wage Dispersion in Search Models: A Quantitative Assessment'. *American Economic Review* **101**(7), 2873–98.

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.

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

Appendix

A Derivations

A.1 Firm Size

As explained in footnote 6, we can re-interpret the firm size as being the product between hiring probability and the expected duration of a match, which coincides with expression (4):

$$\begin{aligned} l(y,\ell) &= \lambda^F \left(\frac{\lambda^U \delta}{\lambda^U \delta + \lambda^E \lambda^U} + \frac{\lambda^E \lambda^U}{\lambda^U \delta + \lambda^E \lambda^U} \frac{\delta}{\delta + \lambda^E (1 - \Gamma_\ell(y))} \Gamma_\ell(y) \right) \frac{1}{\rho + \delta + \lambda^E (1 - \Gamma_\ell(y))} \\ &= \lambda^F \frac{\delta}{\delta + \lambda^E (1 - \Gamma_\ell(y))} \frac{1}{\rho + \delta + \lambda^E (1 - \Gamma_\ell(y))} \\ &\to \lambda^F \frac{\delta}{[\delta + \lambda^E (1 - \Gamma_\ell(y))]^2} \quad \text{if } \rho \to 0 \end{aligned}$$

As the measure of matching from firm side and worker side should coincide for the consistency, the following equation between meeting rates should hold: $\lambda^F = \lambda^U u + \lambda^E (1 - u) = \lambda^U \frac{\delta}{\delta + \lambda^U} + \lambda^E \frac{\lambda^U}{\delta + \lambda^U} = \frac{\delta + \lambda^E}{\delta + \lambda^U} \lambda^U$. Plugging this into our definition, the firm size becomes $l(y,\ell) = \frac{\lambda^U (\delta + \lambda^E)}{\delta + \lambda^U} \frac{\delta}{[\delta + \lambda^E (1 - \Gamma_\ell(y))]^2}$ which is equivalent to the definition of Burdett and Mortensen (1998), who define it as:

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

A.2 Wage posting

Using our definition of firm size $l(w(y,\ell)) = \frac{\lambda^F \delta}{\delta + \lambda^E (1 - \Gamma_\ell(y))} \frac{1}{\rho + \delta + \lambda^E (1 - \Gamma_\ell(y))}$, we can rewrite a firm's profits $\tilde{J}(y,\ell) := l(w(y,\ell))(z(A(\ell), y) - w(y,\ell))$. The envelope condition is:

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

And so:

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

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

$$\begin{aligned} w(y,\ell) &= z(A(\ell),y) \\ &\quad - \left[\delta + \lambda^E(1-\Gamma_\ell(y))\right]\cdot\left[\rho+\delta+\lambda^E(1-\Gamma_\ell(y))\right]\left\{\int_{\underline{y}}^{y} \frac{\frac{\partial z(A(\ell),t)}{\partial y}}{[\delta+\lambda^E(1-\Gamma_\ell(t)]\cdot[\rho+\delta+\lambda^E(1-\Gamma_\ell(t))]}dt\right\} \\ &\quad - \left[\delta + \lambda^E(1-\Gamma_\ell(y))\right]\cdot\left[\rho+\delta+\lambda^E(1-\Gamma_\ell(y))\right]\frac{\tilde{J}(\underline{y},\ell)}{\delta} \end{aligned} \tag{22}$$

Using this result, $\tilde{J}$ becomes:

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

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

Imposing the zero profit assumption of the lowest productive firm type in each location, $\tilde{J}(\underline{y},\ell) = 0$, and additionally imposing $\rho = 0$, we obtain wage function (5) from (22).

A.3 Location Choice: Firm's FOC

The FOC of problem (7) is given by:

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

$$= \frac{\partial k(\ell)}{\partial \ell}$$

Using firm size $l(y,\ell)$ from (4), and rearranging, this becomes (8).

B Baseline Model: Proofs and Additional Results

B.1 Proof of Proposition 2

Apply integration by parts to (6) to obtain:

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

Assessing under which conditions $\overline{J}(p,\ell)$ is supermodular (or submodular) in (p,ℓ), which is a sufficient condition for the (reverse) single crossing property of $\overline{J}(p,\ell)$ in (p,ℓ), we differentiate w.r.t. (p,ℓ):

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

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

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

note that under positive sorting, there is a unique way of matching up firms' ex-ante types with locations, $Q(\mu(\ell)) = R(\ell)$, and based on our assumption of strictly positive densities r, q, the assignment is one-to-one: μ is a strictly increasing function, where the firm type p assigned to location ℓ is given by $\mu(\ell) = Q^{-1}(R(\ell))$. Pure positive sorting also implies that the endogenous firm distribution in location ℓ is given by $\Gamma_\ell(y) = \Gamma(y|Q^{-1}(R(\ell)))$ where we used $\mu(\ell) = Q^{-1}(R(\ell))$. Then, $\frac{\partial \Gamma_\ell(y)}{\partial \ell} = \frac{\partial \Gamma}{\partial p} \frac{r(\ell)}{q(\mu(\ell))}$. So in order to guarantee supermodularity of $\overline{J}(p,\ell)$ in (p,ℓ), we need to ensure that:

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

which is a condition in terms of primitives only, but several objects depend on ℓ and y. In order to specify bounds that hold uniformly in (ℓ, y), let

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

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

In turn, t^P exists (and it is also strictly positive and bounded) since we assumed 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 $[\underline{\ell}, \overline{\ell}]$ with strictly positive densities (q, r).

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

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

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

B.2 Propositions 2 for the Case of NAM and Proof

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 emerged if $\overline{J}(p,\ell)$ was strictly supermodular in (p,ℓ) (see Proposition 1), sorting is negative if $\overline{J}(p,\ell)$ is strictly submodular.

Proposition 1′. *Sorting is negative in equilibrium if $\overline{J}(p,\ell)$ is strictly submodular in (p,ℓ).*

Furthermore, the sufficient conditions for NAM in terms of primitives can be summarized as follows:

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

Proof. To derive sufficient conditions for negative sorting to occur in equilibrium, we can proceed similar as in Section B.1 of this Appendix. In order for $\frac{\partial^2 \overline{J}_\ell(p)}{\partial p \partial \ell}$ to be (strictly) negative, it suffices that the integrand is negative for all $y \in [\underline{y}, \overline{y}]$ and strictly so for some set of y. In turn, for this it suffices that

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

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 R.H.S. is negative, implying that the L.H.S. of the inequality needs to be *sufficiently negative*. Following similar steps as for PAM, the sufficient condition for NAM in terms of primitives reads:

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

Again, we define the uniform bounds (just swapping min and max due to the sign changes):

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

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

A sufficient condition for $\overline{J}$ to be submodular in (ℓ, p) is therefore:

$$\varepsilon^N < 2\varphi^E t^N.$$

So, equilibrium sorting is NAM, either if ε^N is sufficiently small, or if φ^E is sufficiently small (making the R.H.S. approach zero). □

Remark. Note that under the sufficient conditions for negative sorting in Proposition 2′, $\partial k/\partial \ell < 0$, as locations with higher ℓ are less attractive to firms. We show in the proof of Proposition 3 below how the constant of integration in the solution for land price function k can be chosen to ensure that all land prices are positive, so that the individual rationality constraints for land owners are satisfied despite a negative marginal land price.

B.3 Proposition on Global vs Local Rank and Proof

In this section we state the formal results on $D(y)$, which we described in the text. We maintain the following assumption.

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

We can then show the following results.

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

1. *If there is no spatial sorting of firms, $\Gamma_{\ell'} = \Gamma_{\ell''}$ for all $\ell' \neq \ell''$, then $D(y) = 0$ for all $y \in [\underline{y}, \overline{y}]$.*
2. *If there is sorting of firms across space, $\Gamma_{\ell'} \neq \Gamma_{\ell''}$ 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. Part 1. follows from the premise of no sorting, in which case:

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

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

Part 2., second statement, follows from examining the slope of D at $y = \{\underline{y}, \overline{y}\}$. Differentiate D w.r.t. y to obtain:

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

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

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

where Var_r is our notation for the variance of an object, taking the land distribution r into account. Both expressions are strictly negative if $\mathrm{Var}_r[\gamma(\underline{y}|\mu(\ell))] > 0$ and $\mathrm{Var}_r[\gamma(\overline{y}|\mu(\ell))] > 0$, which is the case under our assumptions that $\gamma(\underline{y}|p)$ and $\gamma(\overline{y}|p)$ vary in p.

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^* such that $D(y^*) = 0$ and at that point D crosses zero from below. If D has several interior zeros, then the first and the last one share this 'crossing-from-below' property, proving the claim. □

Example with Unique Interior Zero. We parameterize our model as follows.

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

where $\ell \in [a, b]$. It follows that under PAM, $\mu(\ell) = \ell$ and $\Gamma_\ell(y|\mu(\ell)) = y^{\mu(\ell)} = y^\ell$. If $a = 1$ and $b = 2$, then we can solve for the zeros of D in closed form, and one can show that the unique interior zero is $y^* = 0.4998$ in this case.

B.4 Proof of Proposition 3

We want to show that a fixed point in Γ_ℓ exists and we will do so by construction. For concreteness, suppose the conditions of Proposition 2, Part 1. hold.

In a first step, consider an assignment $\mu(\ell) = Q^{-1}(R(\ell))$, yielding a unique firm distribution across locations $\Gamma_\ell = \Gamma(y|\mu(\ell))$ and a unique wage function w in (5).

In a second step, we will show that the pair $(k(\ell), \mu(\cdot))$ is the unique (up to a constant of integration) Walrasian equilibrium of the land market, where $\mu(\cdot)$ as in the first step and

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

By construction, $\mu(\cdot)$ 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

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

To reduce notation, we introduce

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

and so firm p solves

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

and from the first-order condition (FOC) we obtain $\frac{\partial \mathcal{J}(p,\ell)}{\partial \ell} = \frac{\partial \mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell}$ and hence $p = \mu(\ell)$. To show that this 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

$$\mathcal{J}(p,\ell) - \mathcal{J}(p,\ell') \geq \int_{\ell'}^{\ell} \frac{\partial \mathcal{J}(\mu(\hat{\ell}),\hat{\ell})}{\partial \ell} d\hat{\ell}. \tag{23}$$

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

To see that this equilibrium is unique, we first note that under our assumptions, Theorem 10.28 in Villani (2009) implies that there exists a unique optimal assignment μ, which is deterministic. Second, regarding $k(\ell)$, note that it must be strictly increasing to ensure that if $\ell > \hat{\ell}$ then ℓ is chosen by a strictly higher p than $\hat{\ell}$, 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 unique (up to an additive constant $\bar{k}$) and given by the one above then follows from Theorem 10.28 and Remark 10.30 in Villani (2009).

In turn, if the conditions of Part 2. of Proposition 2 hold (negative sorting), then the proof is similar. In the existence argument, just replace supermodularity with submodularity of $\mathcal{J}$ and note that for any $\ell > \ell'$, $\mu(\ell) < \mu(\ell')$. In the uniqueness argument, Theorem 10.28 in Villani (2009) again implies that there exists a unique optimal assignment μ, which is deterministic. Regarding price $k(\ell)$, note that it must be strictly *decreasing* to ensure that if $\ell > \hat{\ell}$ then ℓ is chosen by a strictly *lower* p than $\hat{\ell}$, 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. In particular, it must be that

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

so that the land price is positive for each location. □

B.5 Proof of Lemma 1

We need to show that under the conditions of the Lemma, (17) is positive, due to all three components being positive.

(i) follows directly from the fact that A is strictly increasing in ℓ and the wage intercept being increasing in A.

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

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

The first line is positive under the sufficient conditions for PAM, which we assume to hold. In turn, the second line is positive if

$$\Big(- \frac{\partial \Gamma_\ell(y)}{\partial \ell} \Big) \frac{\partial \Gamma_\ell(y)}{\partial y} + \frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}(1 - \Gamma_\ell(y)) \geq 0$$

or if,

$$\frac{\partial \Gamma_\ell(y)}{\partial \ell} \frac{\partial \Gamma_\ell(y)}{\partial y} + \frac{\partial^2 \Gamma_\ell(y)}{\partial y \partial \ell}(1 - \Gamma_\ell(y)) \geq 0$$

where we replaced 'minus' with 'plus' in the first term. And this is true if

$$\frac{\frac{\partial^2 \Gamma_\ell}{\partial y \partial \ell}(1 - \Gamma_\ell)}{\frac{\partial \Gamma_\ell}{\partial y}\Big(- \frac{\partial \Gamma_\ell}{\partial \ell} \Big)} \geq 1$$

or, equivalently, if $1 - \Gamma_\ell$ is log-submodular. A sufficient condition for this is that $1 - \Gamma$ is log-submodular, as claimed.

(iii) follows from (13) and the fact that with positive firm sorting, $\partial \Gamma_\ell / \partial \ell \leq 0$. □

Remark 1. Several classes of densities γ give rise to log-submodular $1 - \Gamma$. Examples include the multinomial density, the multivariate hypergeometric density and the multinormal density with certain covariance structure (see Karlin and Rinott, 1980, (1.3) for more examples). What these densities have in common is that they are *strongly* log-submodular, and as a result their log-submodularity is preserved by integration when the integrand is the composition of a (strongly) log-submodular density and a PF_2 function (Polya frequency functions). Note that the indicator function of a convex set is log-concave, and therefore PF_2. Thus:

$$1 - \Gamma(y|p) = \int_{\underline{y}}^{\overline{y}} \mathbf{1}_{[y, \overline{y}]}(s) \gamma(s|p) ds$$

is log-submodular if γ is strongly sub-modular (see Karlin and Rinott, 1980, Definition 1.2). Note that Karlin and Rinott call strongly submodular functions MRR_2 (multivariate reverse rule of order 2).

Remark 2. Similar to the driving forces of spatial wage inequality, we can analyze the driving forces behind spatial inequality in *value added*. To do so, we consider the following measure of spatial inequality in value added:

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

To shed light on the drivers of spatial value-added inequality, consider:

$$\frac{\partial \frac{\mathbb{E}[A(\ell)y|\ell]}{\mathbb{E}[A(\underline{\ell})y|\underline{\ell}]}}{\partial \ell} =$$

$$\frac{1}{\mathbb{E}[A(\underline{\ell})y|\underline{\ell}]} \times \Bigg(\underbrace{\underline{y}A'(\ell)}_{\text{intercept of VA-schedule}} + \underbrace{\left(\int_{\underline{y}}^{\overline{y}}(1 - G_\ell(y)dy)\right)A'(\ell)}_{\text{steepness of VA-schedule}} + \underbrace{A(\ell)\int_{\underline{y}}^{\overline{y}}\left(-\frac{\partial G_\ell(y)}{\partial \ell}\right)dy}_{\text{employment composition}}\Bigg). \quad (24)$$

As with wages, here there are three factors driving cross-location value added inequality: (i) Higher-ℓ locations have a higher intercept and also (ii) a steeper slope of the value added schedule compared to $\underline{\ell}$, both due to higher location TFP. (iii) Higher-ℓ locations have a better firm composition (due to *spatial firm sorting*) so that more employment is concentrated at the upper part of the value added schedule, compared to $\underline{\ell}$.

C Baseline Model: Comparative Statics

Throughout this exercise, we focus on an economy that satisfies the sufficient conditions for positive sorting of firms across space, i.e. PAM in (p, ℓ).

C.1 Firm-Level Predictions

We first focus on the firm-level implications of our theory for a given location ℓ. We focus on three firm-level outcomes.

1. The *wage function* was derived under firms' optimal wage posting decisions and given by:

$$w(y, \ell) = z(A(\ell), y) - \left[1 + \varphi^E(1 - \Gamma_\ell(y))\right]^2 \left\{ \int_{\underline{y}}^{y} \frac{\frac{\partial z(A(\ell);t)}{\partial y}}{\left[1 + \varphi^E(1 - \Gamma_\ell(t))\right]^2} dt \right\} \quad (25)$$

2. The poaching share $h^E(y, \ell)$ of a firm y in location ℓ is its share of hires from employment. Similarly, the poached share $s^E(y, \ell)$ is its share of worker separations to another job. The poaching share is strictly increasing and the poached share strictly decreasing in firm's local rank. As a summary measure of both statistics, we focus on the *net poaching*

share of a firm y in location ℓ, which is the difference between poaching share and poached share:[32]

$$n^E(y,\ell) := h^E(y,\ell) - s^E(y,\ell) = \frac{\varphi^E}{\varphi^E+1}\Gamma_\ell(y) - \frac{\varphi^E(1-\Gamma_\ell(y))}{1+\varphi^E(1-\Gamma_\ell(y))} \tag{26}$$

3. The *firm size* of y relative to the smallest firm in location ℓ is derived from equation (4):

$$\frac{l(y,\ell)}{l(\underline{y},\ell)} = \left(\frac{1+\varphi^E}{1+\varphi^E(1-\Gamma_\ell(y))}\right)^2. \tag{27}$$

C.1.1 Impact of Local Labor Market Frictions

We use φ^E as our summary statistic for labor market frictions. An increase in φ^E means lower search frictions, either through a higher job arrival rate for employed workers or lower job destruction rate, or both.

Lemma 2. *In any location ℓ, and for any firm type $y \in (\underline{y}, \overline{y})$, a reduction in labor market frictions (i.e., an increase in φ^E) increases the wage and relative firm size, and it decreases the net poaching share if φ^E is large enough.*

Proof. To obtain the effects of φ^E on the net poaching share (26) and relative firm size (27), multiply and divide by δ (or δ^2), so that they become functions of φ^E. Then, differentiate each outcome w.r.t. φ^E:

For firm size, the sign of the comparative static is straightforward.

For the net poaching share, differentiate $n^E(y,\ell)$ with respect to φ^E,

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

This is a quadratic polynomial of φ^E with inverse U-shape for all $\Gamma_\ell(y) \in [0,1)$. Thus, $n^E(y,\ell)$

[32] The poaching share $h^E(y,\ell)$ is derived taking into account the firm's hiring rate $h(y,\ell) = \frac{\lambda^F\delta}{\delta+\lambda^E(1-\Gamma_\ell(y))}$:

$$h^E(y,\ell) := \frac{\lambda^F \frac{\lambda^E(1-u(\ell))}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))} E_\ell(w(y))}{h(y,\ell)} = \frac{\lambda^E}{\lambda^E+\delta}\Gamma_\ell(y)$$

The poached share $s^E(y,\ell) = (\lambda^E(1-\Gamma_\ell(y)))/(\delta+\lambda^E(1-\Gamma_\ell(y)))$ reflects that firms' workers get poached only if the offer is from a firm with higher rank. Both shares range from 0 to $\frac{\lambda^E}{\delta+\lambda^E} = \frac{\varphi^E}{1+\varphi^E}$, with this upper bound increasing in φ^E.

is decreasing in φ^E if φ^E is greater than the larger root, given by

$$\varphi^E \geq \sqrt{\frac{\Gamma_\ell(y)^3}{(1-\Gamma_\ell(y))(1-\Gamma_\ell(y)+\Gamma_\ell(y)^2)^2}} + \frac{-1+\Gamma_\ell(y)}{1-\Gamma_\ell(y)+\Gamma_\ell(y)^2}.$$

Finally, regarding the wage function (25), we can express it as:

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

Differentiate this expression w.r.t. φ^E to obtain:

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

which has the sign of

$$\begin{aligned} & -(1-\Gamma_\ell(y))\left(1+\varphi^E(1-\Gamma_\ell(t))\right) + (1-\Gamma_\ell(t))\left(1+\varphi^E(1-\Gamma_\ell(y))\right) \\ \Leftrightarrow \quad & -(1-\Gamma_\ell(y)) + (1-\Gamma_\ell(t)) \geq 0 \end{aligned}$$

since $t \in [\underline{y}, y]$. Hence, the wage increases in φ^E. □

Lower labor market frictions can be driven by longer match duration (through lower δ), whereby the pool of employed workers becomes more important for firm hiring, incentivizing firms to post higher wages. As a result, poaching share, poached share and relative firm size increase. In turn, lower frictions can also stem from a higher worker contact rate (higher λ^E), implying that workers are more likely to be poached. Again, wages increase to shield firms from competition, thereby raising the poaching share, poached share and relative firm size. Finally, the *net* poaching share depends positively on the poaching and negatively on the poached share. For a firm whose local productivity rank is low, the net poaching share is mostly determined by the poached share (it can hardly hire from other firms, independently of how severe the frictions are). As a result, its poaching share does not increase much if labor market frictions lessen while its poached share increases significantly. Thus, the net poaching share decreases for a low-productivity firm. In turn, for a firm whose local productivity rank is high, the net poaching share is predominantly determined by its poaching share. Thus, as labor market frictions decline, its net poaching share can increase. But it decreases given that φ^E is large enough.

C.1.2 Impact of Local Productivity Rank

Lemma 3. *In any location ℓ, a firm's wage, net poaching share and relative firm size are all increasing in its local productivity rank $\Gamma_\ell(y)$.*

Proof. We apply a change of variable to re-express the wage to depend on the local productivity rank (instead of productivity y), which we denote in short by $\mathcal{R} := \Gamma_\ell(y)$. Consider a firm in ℓ with rank $\mathcal{R} \in [0,1]$. Its wage is given by:

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

$$w(\mathcal{R},\ell) = z(A(\ell),\Gamma_\ell^{-1}(\mathcal{R})) - [\delta + \lambda^E(1-\mathcal{R})]^2 \int_0^{\mathcal{R}} \frac{\frac{\partial z(A(\ell),\Gamma_\ell^{-1}(x))}{\partial y}}{[\delta + \lambda^E(1-x)]^2} \frac{1}{\gamma_\ell(\Gamma_\ell^{-1}(x))} dx \tag{28}$$

where $\gamma_\ell(\Gamma_\ell^{-1}(x))$ is the pdf at xth quantile. Differentiate w.r.t. $\mathcal{R}$:

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

which is positive. In turn, the effect of the local productivity rank on the other firm level outcomes follows from differentiating (26), (4) and (27) w.r.t. $\Gamma_\ell(y)$. □

C.1.3 Impact of Productivity Distribution and Land Distribution

Despite our focus on an allocation with positive firm sorting, the model primitives—specifically, the firm productivity distribution Q and the land distribution R—impact firm sorting through the matching function μ, thereby affecting a firm's local rank $\Gamma_\ell(y) = \Gamma(y|\mu(\ell))$. We now investigate how improvements in the two distributions (in a FOSD sense) affect our firm-level outcomes by altering local productivity ranks.

Lemma 4. *Consider a firm with productivity $y \in (\underline{y}, \overline{y})$ in location ℓ. A FOSD shift in Q decreases the firm's wage, net poaching share and relative firm size. In turn, a FOSD in R increases the firm's wage, net poaching share and relative firm size.*

Proof. To prove this result, we need to show how FOSD changes in (Q, R) affect μ and thus $\Gamma_\ell(y) = \Gamma(y|\mu(\ell))$.

First, consider a FOSD shift in Q. To this end, parameterize Q by t, $Q(p|t)$ with $Q_t \leq 0$ (strict for some set of positive measure of p). Then $\mu_t = -\frac{Q_t}{q(\mu(\ell))} \geq 0$ and $\partial\Gamma_\ell/\partial t = \frac{\partial\Gamma}{\partial p}\mu_t \leq 0$. The result then follows from combining this with the results from Lemma 3.

Second, consider a FOSD shift in R. To this end, parameterize R by t, $R(\ell|t)$ with $R_t \leq 0$ (strict for some set of positive measure of ℓ). Then $\mu_t = \frac{R_t}{q(\mu(\ell))} \leq 0$ and $\partial\Gamma_\ell/\partial t = \frac{\partial\Gamma}{\partial p}\mu_t \geq 0$. The result then follows from combining this with the results from Lemma 3. $\square$

The reason for why a FOSD shift in Q and R move our firm-level outcomes in different ways is that they shift the matching function μ in opposite directions. For instance, a FOSD shift in R (making the land distribution stochastically better so that there is more land supply in highly productive locations) shifts the matching function μ down for all ℓ. As a result, every firm with given ex-ante productivity p gets matched to a better location. As the supply of good locations becomes relatively more abundant, there is less competition for these locations among firms and overall less competition of firms within each local labor market. This *increases* the local rank of a firm of any given y, and thus its wage, net poaching share and (relative) firm size. The opposite logic applies when the firm productivity distribution Q improves, since competition in each local labor market tightens in this case.

C.1.4 Impact of Local Productivity

For completeness, we examine how firm-level outcomes depend on location productivity. Conditional on the firms' local productivity rank (i.e. conditional on a fixed matching function μ satisfying PAM), local productivity $A(\ell)$ has a positive impact on the wage (this follows directly from (25)) but has no impact on the net poaching share or relative firm size. The reason is that the net poaching share and firm size are determined by the *relative* productivity of firms in a location, while $A(\ell)$ shifts the productivity of all firms in location ℓ in the same way, leaving the productivity ranking unaltered.

C.2 Market-Level Predictions

We are interested market-level outcomes:

1. Average wage:

$$\overline{w}(\ell) := \int_{\underline{y}}^{\overline{y}} w(y,\ell) d\Gamma_\ell(y) = w(\underline{y},\ell) + \int_{\underline{y}}^{\overline{y}} \frac{\partial w(y,\ell)}{\partial y}(1-\Gamma_\ell(y))dy$$

2. Average Net Poaching Share:

$$\overline{n}^E(\ell) := \int_{\underline{y}}^{\overline{y}} n^E(y,\ell) d\Gamma_\ell(y) = \frac{1}{2}\frac{\varphi^E}{1+\varphi^E} - \left(1 - \frac{1}{\varphi^E}\log\left(1+\varphi^E\right)\right)$$

3. Average Relative Firm Size:

$$\bar{l}(\ell) := \int \frac{l(y,\ell)}{l(\underline{y},\ell)} d\Gamma_\ell(y) = 1 + \varphi^E,$$

where we used integration by parts.

4. Wage Dispersion:

$$\frac{w(\overline{y},\ell)}{w(\underline{y},\ell)} = \frac{z(A(\ell),\overline{y}) - \int_{\underline{y}}^{\overline{y}} \frac{\partial z(A(\ell);t)}{\partial y}\frac{l(t,\ell)}{l(\overline{y},\ell)}dt}{z(A(\ell),\underline{y})}. \tag{29}$$

5. Firm size dispersion:

$$\frac{l(\overline{y},\ell)}{l(\underline{y},\ell)} = \left(1+\varphi^E\right)^2 \tag{30}$$

Note that the *net poaching rate dispersion* in location ℓ is simply constant, $\frac{n^E(\overline{y},\ell)}{n^E(\underline{y},\ell)} = -1$.

C.2.1 Outcomes across Locations (Constant Labor Market Frictions).

We start analyzing how outcomes vary across locations ℓ, varying location productivity A but keeping labor market frictions fixed.

Lemma 5. *Assume that $(\lambda^U, \lambda^E, \delta)$ are constant across locations. Then, the average net poaching share and average firm size are independent of ℓ, and so is the firm size dispersion. Moreover, the average wage is increasing in ℓ, while the effect on wage dispersion is ambiguous.*

Proof. We can re-express the average wage as:

$$\begin{aligned}
\overline{w}(\ell) &= w(\underline{y},\ell) + \int_{\underline{y}}^{\overline{y}} \frac{\partial w(y,\ell)}{\partial y}(1-\Gamma_\ell(y))dy \\
&= w(\underline{y},\ell) + \int_{\underline{y}}^{\overline{y}} 2\varphi^E\left(1+\varphi^E(1-\Gamma_\ell(y))\right)\gamma_\ell(y)\int_{\underline{y}}^{y}\frac{\frac{\partial z(A(\ell),t)}{\partial y}}{(1+\varphi^E(1-\Gamma_\ell(t)))^2}dt(1-\Gamma_\ell(y))dy \\
&= w(\underline{y},\ell) + \int_0^1 2\varphi^E\left(1+\varphi^E(1-\mathcal{R})\right)(1-\mathcal{R})\left(\int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})}\frac{\frac{\partial z(A(\ell),t)}{\partial y}}{(1+\varphi^E(1-\Gamma_\ell(t)))^2}dt\right)d\mathcal{R}
\end{aligned}$$

where we plug in $\frac{\partial w}{\partial y}$ in the second line, and apply the change-of-variables $\mathcal{R} = \Gamma_\ell(y) \in [0,1]$ ($d\mathcal{R} = \gamma_\ell(y)dy$) in the third line. Note that now only the inner (not the outer) integral in the third line depends on ℓ. It is increasing in ℓ since (i) $\frac{\frac{\partial z(t,A(\ell))}{\partial y}}{(1+\varphi^E(1-\Gamma_\ell(t)))^2}$ is increasing in ℓ (follows from our sufficient conditions for PAM); (ii) $\Gamma_\ell^{-1}(\mathcal{R})$ is increasing in ℓ (follows from $\frac{\partial\Gamma_\ell(y)}{\partial\ell} < 0$ under PAM). Thus, under our sufficient condition for PAM, the average wage is increasing in ℓ.

Nex, consider the average firm size (see footnote 6 for the derivation of $l(y,\ell)$):

$$\bar{l}(\ell) = \int_{\underline{y}}^{\overline{y}}(1-u)\frac{g_\ell(y)}{\gamma_\ell(y)}\gamma_\ell(y)dy$$

where $l(y,\ell) = (1-u)\frac{g_\ell(y)}{\gamma_\ell(y)}$. Therefore,

$$\bar{l}(\ell) = (1-u)\int_{\underline{y}}^{\overline{y}} g_\ell(y)dy = 1-u,$$

confirming that average firm size is independent of ℓ.

The remaining claims follow directly from the formulae of the average net poaching share and firm size dispersion above. □

When assuming labor market frictions are constant across locations, the average net poaching share does not vary across space since the net poaching share of each firm with the same local productivity rank is invariant in ℓ. Similarly, in this case, average firm size is the same across locations as total employment (determined by labor market frictions) and the total measure of firms is the same across markets.

Finally, even for constant labor market frictions, the average wage increases in ℓ if (i) $w(\underline{y},\ell) = z(A(\ell),\underline{y})$ is increasing in ℓ so that higher ℓ locations are more productive (productivity channel), (ii) firm composition improves with higher ℓ (firm sorting channel), which—after a change of variable—renders the wage function supermodular in firm's local rank $\mathcal{R}$ and ℓ, so that the wage gradient in $\mathcal{R}$ becomes steeper in higher-ℓ locations. That is, job ladders in high-ℓ locations are steeper. While (i) is true by assumption, (ii) is guaranteed in an equilibrium with positive sorting.

C.3 Outcomes across Locations (Varying Labor Market Frictions)

We now investigate how these market-outcomes change if labor market frictions vary while keeping A fixed.

Lemma 6. *A reduction in labor market frictions (i.e., an increase in φ^E) increases the average wage and average firm size, and it also increases wage dispersion and firm size dispersion. In turn, it lowers the average net poaching share.*

Proof. The claim on the average wage and the average relative firm size follow in a straightforward way from differentiating these statistics w.r.t. φ^E.

In turn, for the claim on the average net poaching share, $\overline{n}^E(\ell)$, differentiate $\overline{n}^E(\ell)$ with respect to φ^E:

$$\frac{\partial \overline{n}^E(\ell)}{\partial \varphi^E} = \frac{1}{2}\frac{1}{(1+\varphi^E)^2} - \frac{1}{(\varphi^E)^2}\log(1+\varphi^E) + \frac{1}{\varphi^E}\frac{1}{1+\varphi^E}.$$

We now show that this expression is negative for $\varphi^E > 0$. First, let's define $f(x) := \frac{1}{2}x^2 + x + \log(1-x)$ where $f(x)$ is $\frac{\partial \overline{n}^E(\ell)}{\partial \varphi^E}$ multiplied with $(\varphi^E)^2$ and where we use $x = \frac{\varphi^E}{1+\varphi^E} \in (0,1)$. Thus, it suffices to show that $f(x) < 0$ for $x \in (0,1)$. Observe that $f(0) = 0$. Next, $f'(x) = x + 1 - \frac{1}{1-x} = \frac{-x^2}{1-x} < 0$.

The claim on firm-size dispersion follows directly from differentiating (30) w.r.t. φ^E.

Finally, for the claim on wage dispersion, note that (29) depends on φ^E only through $\frac{l(t,\ell)}{l(\overline{y},\ell)}$, where

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

and thus wage dispersion increases in φ^E. □

In line with our analysis on firm-level outcomes within locations above, locations with lower labor market frictions (i.e., higher φ^E) have a higher *average* wage and higher *average* firm size.

Also, firm-size dispersion increases as local labor market frictions become smaller. With low levels of frictions, more productive firms can reap the full benefit from their poaching potential, hire many on-the-job searchers and grow large relative to low-productivity firms.

Firm size dispersion in market ℓ is tightly linked to wage dispersion in that market; see (29). Wage dispersion increases in the firm size dispersion between the 'average' firm (where, to compute this average, the weight of each firm is given by its marginal productivity) and the most productive firm in a location. Thus, it increases as labor market frictions lessen.[33]

Finally, the average net poaching share decreases as labor market frictions become smaller. From Lemma 3, as frictions lessen, the net poaching share decreases for low productivity firms, but for highly productive firms, the net poaching share tends to increase. Lemma 6 shows that if we take the *average* of net poaching shares across firm types, then this average decreases in φ^E. Also, note that the net poaching share *dispersion* is independent of local labor market frictions.

D Model Extensions: Propositions and Proofs

D.1 Labor Mobility

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

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

where we will denote more compactly:

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

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

$$\frac{\partial^2 \bar{J}(p,\ell)}{\partial \ell \partial p} = \frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p} \lambda^F(\ell) + \frac{\partial \hat{J}(p,\ell)}{\partial p} \frac{\partial \lambda^F(\ell)}{\partial \ell}. \tag{31}$$

[33] If we assumed a multiplicative production function, the relationship between wage and firm size dispersion becomes even clearer. What matters for wage dispersion then is the dispersion between the (unweighted) average firm size and the size of the top firm in the market. Wage dispersion is increasing in that statistic. The above expression reads:

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

Apply integration by parts to $\hat{J}(p,\ell)$:

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

and then obtain its derivatives:

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

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

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

We want to find conditions under which (31) is positive (i.e., $\overline{J}(p,\ell)$ is supermodular in (p,ℓ)). A sufficient condition is that the integrands of these expressions are positive for all $y \in [\underline{y},\overline{y}]$ and strictly so for some set of y. We can capture the equation with one single integral and use $-\frac{\partial \Gamma(y|p)}{\partial p} > 0$ to obtain that a sufficient condition for PAM is:

$$\frac{\frac{\partial^2 z(A(\ell),y)}{\partial y \partial A(\ell)}\frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(A(\ell),y)}{\partial y}} > \frac{2\left(\lambda^E(\ell)\left(-\frac{\partial \Gamma_\ell}{\partial \ell}\right) + \frac{\partial \lambda^E}{\partial \ell}(1-\Gamma_\ell(y))\right)}{\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))} - \frac{\frac{\partial \lambda^F(\ell)}{\partial \ell}}{\lambda^F(\ell)}. \tag{32}$$

Define ε^P as the baseline model which is the minimum of the L.H.S. It is strictly positive under our assumptions and the premise. Under labor mobility, the R.H.S. depends on the endogenous market tightness $\theta(\ell)$. Thus, the sufficient conditions for PAM from the baseline model are not readily applicable. Instead, we argue that the R.H.S. is bounded away from infinity. Thus, (32) holds for a large enough ε^P.

To see this, first note that θ is *decreasing* in ℓ. Or, equivalently, λ^E is decreasing in ℓ and λ^F is increasing in ℓ. This follows from the fact that the value of unemployment is increasing in ℓ for fixed λ^E. Recall the value of unemployment:

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

As the wage is increasing in firm's local productivity and thus local rank, we can express it as a function of the firm rank in the local productivity distribution $\mathcal{R}$ instead of the firm's productivity y: $t = w(\mathcal{R}, \ell)$. Using $F_\ell(t) = \mathcal{R}$, a change of variable gives:

$$\rho V^U = z(A(\ell), \underline{y}) + \lambda^E(\ell)\Bigg[\int_0^1 \frac{1 - \mathcal{R}}{\delta + \lambda^E(\ell)(1 - \mathcal{R})} \frac{\partial w(\mathcal{R}, \ell)}{\partial \mathcal{R}} d\mathcal{R}\Bigg] \tag{33}$$

Fixing, for the moment, $\lambda^E(\ell) = \lambda^E$,

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

Using equation 28,

$$\frac{\partial^2 w(\mathcal{R}, \ell)}{\partial \ell \partial \mathcal{R}} = 2\frac{\lambda^E}{\delta}\left(1 + \frac{\lambda^E}{\delta}(1 - \mathcal{R})\right)\frac{\partial}{\partial \ell}\int_{\underline{y}}^{\Gamma_\ell^{-1}(\mathcal{R})} \frac{\frac{\partial z(A(\ell), h)}{\partial y}}{(1 + \frac{\lambda^E}{\delta}(1 - \Gamma_\ell(h)))^2} dh. \tag{34}$$

Observe that $\Gamma_\ell^{-1}(\mathcal{R})$ is increasing in ℓ since $\frac{\partial}{\partial \ell}\Gamma_\ell < 0$. In addition, under the sufficient condition for PAM (which we maintain throughout), $\frac{\frac{\partial z(A(\ell),y)}{\partial y}}{(1+\frac{\lambda^E}{\delta}(1-\Gamma_\ell(y)))^2}$ is also increasing in ℓ. Thus, the wage function is supermodular in $(\mathcal{R}, \ell)$. This ensures that $\frac{\partial \rho V^U}{\partial \ell}\big|_{\lambda^E(\ell)=\lambda^E} > 0$.

As a result and because V^U is increasing in λ^E, in order for the equilibrium indifference condition to hold (value of unemployment, V^U, is equalized across ℓ), it must be that λ^E is decreasing in ℓ. And so θ is decreasing in ℓ and λ^F is increasing in ℓ, rendering the second term on the R.H.S. in (32) negative.

It then suffices that the first (positive) term on the R.H.S. of (32) is bounded. Note that $\theta(\ell) \in [\underline{\theta}, \overline{\theta}]$ for all ℓ. This is due to the fact that $\theta(\cdot)$ is a continuous function over a compact support. To see this, note that $\lambda^E(\cdot)$ (and thus $\theta(\cdot)$) is implicitly defined by (33), where, in equilibrium, V^U is a number that no longer depends on ℓ. All functions in this expression $(z, \partial w/\partial \mathcal{R})$ are continuous in ℓ (see (34)), and thus λ^E inherits this property and so does θ. The first term on the R.H.S. of (32) is thus within the bounds:

$$\left[\frac{2\kappa\mathcal{A}\left(-\frac{\partial \Gamma}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)}{\frac{\delta}{\underline{\theta}^\alpha}}, \frac{2\kappa\mathcal{A}\left(-\frac{\partial \Gamma}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)}{\frac{\delta}{\overline{\theta}^\alpha}}\right]$$

Define:

$$\tilde{t}^P \equiv \frac{2\kappa\mathcal{A}}{\frac{\delta}{\overline{\theta}^\alpha}}\left(max_{y,\ell}\left(-\frac{\partial\Gamma}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)\right),$$

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).

As a result, PAM obtains if:

$$\varepsilon^P > \tilde{t}^P.$$

□

D.2 Endogenous Spillovers

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

$$A(\ell) = \int_{\underline{y}}^{\overline{y}} (1 - \Gamma_\ell(y))dy, \tag{35}$$

since for $\underline{y} = 0$ this is equivalent to $A(\ell) = \int_{\underline{y}}^{\overline{y}} y d\Gamma_\ell(y)$ so that productivity spillovers take the form of the *average* firm productivity in a location. For concreteness, we will assume spillover function (35) in this section.

Assumption 3. *Productivity in location ℓ is endogenous and given by* (35).

Note that ex-ante, before any sorting takes place, the 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 where 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 as it is related to the productivity of firms settling there.

Proposition 1 on the necessary and sufficient conditions for sorting applies as is. Proposition 2 applies with minor modifications. For $\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell} > 0$ and hence for positive sorting, we can again unpack (14) from the baseline model and obtain the analogue of (15) as:

$$\frac{\frac{\partial^2 z(A(\ell),y)}{\partial y \partial A(\ell)}}{\frac{\partial z(A(\ell),y)}{\partial y}} \int_{\underline{y}}^{\overline{y}} -\frac{\partial \Gamma_\ell(y)}{\partial \ell} dy \geq \frac{2\lambda^E}{\delta + \lambda^E(1 - \Gamma(y|Q^{-1}(R(\ell)))}\left(-\frac{\partial \Gamma(y|Q^{-1}(R(\ell)))}{\partial p}\right)\frac{r(\ell)}{q(Q^{-1}(R(\ell)))} \tag{36}$$

with the difference of *endogenous* location productivity differences $\frac{\partial A(\ell)}{\partial \ell} = \int_{\underline{y}}^{\overline{y}} -\frac{\partial \Gamma_\ell(y)}{\partial \ell} dy > 0$. Under positive sorting in (p, ℓ), productivity A increases in ℓ as firm distributions improve in those locations. 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 in these locations, then highly productive firms (with high-p) indeed settle into high-ℓ locations — similar to the baseline model.

In turn, negative sorting in (p, ℓ) can be supported when (14) holds with the opposite sign, rendering $\overline{J}$ submodular. Under NAM, location productivity A decreases in ℓ as the firm distribution deteriorates in high-ℓ locations, $\frac{\partial A(\ell)}{\partial \ell} = \int_{\underline{y}}^{\overline{y}} -\frac{\partial \Gamma_\ell(y)}{\partial \ell} dy < 0$. If this location productivity disadvantage in high-ℓ locations is large enough relative to the benefit of less severe poaching competition, then high-p firms indeed settle into *low*-ℓ locations.

We now state the sorting result under endogenous spillovers formally, where we redefine the minimum (maximum) productivity gains (losses) from sorting into high-ℓ locations as:

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

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

where $A(\ell) = \int_{\underline{y}}^{\overline{y}} 1 - \Gamma(y | Q^{-1}(R(\ell)) dy$ in the first line, and $A(\ell) = \int_{\underline{y}}^{\overline{y}} 1 - \Gamma(y | Q^{-1}(1 - R(\ell)) dy$ in the second. Note that under our assumptions $\varepsilon^P > 0$ and $\varepsilon^N < 0$.

Proposition 7. *Suppose Assumption 3 holds.*

1. *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 is positive sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(R(\ell))$.*
2. *If z is strictly supermodular, and either the productivity losses from sorting into higher ℓ, ε^N, are sufficiently large (in absolute terms), or the competition forces φ^E are sufficiently small, then there is negative sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(1 - R(\ell))$.*

The proof of this result resembles the proof of Proposition 2, with the proposed re-definition of ε^P and ε^N, and is therefore omitted.

Next, we provide the existence result for this extension.

Proposition 8. *Suppose Assumption 3 holds along with the conditions from Proposition 7 (either Part 1. or 2.). Then, an equilibrium exists but it is not unique.*

Proof. The existence proof follows the same steps as in the baseline model (proof of Proposition 3). First, focus on the equilibrium with PAM. In Step 1, the assumed $\mu(\ell) = Q^{-1}R(\ell)$ — by uniquely pinning down Γ_ℓ — now also uniquely pins down the spillover function $A(\ell) = \int(1-\Gamma_\ell(y))dy$. We then seek to construct a fixed point in Γ_ℓ as in the baseline model.

Second, focus on the equilibrium with NAM. In Step 1, the assumed $\mu(\ell) = Q^{-1}(1-R(\ell))$ — by uniquely pinning down Γ_ℓ — now also uniquely pins down the spillover function $A(\ell) = \int(1-\Gamma_\ell(y))dy$. We then seek to construct a fixed point in Γ_ℓ as in the case with PAM. Note the particularity that under the conditions from the proposition, the marginal land price $\partial k/\partial \ell = \frac{\partial \mathcal{J}(\mu(\ell),\ell)}{\partial \ell} < 0$ is negative, so the land price is decreasing in ℓ. By choosing the appropriate constant of integration $\bar{k}$ when solving the firm's FOC to problem (7) (which is an ordinary differential equation in k), we can still ensure positive land prices for all ℓ, and thus satisfy individual rationality of land owners who only participate in the market if $k(\ell) \geq 0$ for all ℓ. In particular, we choose $\bar{k}$ such that $k(\bar{\ell}) = 0$ (where, under NAM, $\bar{\ell}$ is the location with the worst firm distribution). All $\ell < \bar{\ell}$ earn a strictly positive land price, and the highest price is paid for $\underline{\ell}$, $k(\underline{\ell}) = \bar{k}$. Moreover, following the same steps as in the case of PAM, we can verify that under this land price schedule $p = \mu(\ell) = Q^{-1}(1 - R(\ell))$ is the globally optimal matching for firms. □

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

D.3 Endogenous Land Supply

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

Environment. We maintain from the baseline model that locations can be ranked by productivity and are indexed by $\ell \in [\underline{\ell}, \bar{\ell}]$. Contrary to the baseline model, we now bring to life the *land developers*, who are initially heterogenous 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 a location). In turn, the firms' payoffs are as in the baseline model.

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 the same equilibrium requirements hold plus land developers invest optimally. 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 agents that a developer with ability ψ anticipates, with 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 the 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 unique 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

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

Sorting. 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:

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

where they take the economy-wide investment threshold ψ^* and thus land supply R as given. Using the same definitions for ε^P and ε^N as in the baseline model, we can show.

We now prove our main result in this extension.

Proposition 9.

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

2. *If z is strictly submodular, and either the productivity gains from sorting into higher ℓ, ε^N, are sufficiently small, or the competition forces φ^E are sufficiently small, then there is negative sorting of firms p to locations ℓ with $p = \mu(\ell) = Q^{-1}(1 - R(\ell, \psi^*))$.*

Proof. We here focus on the case of positive sorting (Part 1.). Cross-differentiating $\overline{J}(p, \ell; \psi^*)$

w.r.t. (p, ℓ) yields again:

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

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:

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

which again can be guaranteed if the productivity advantage from sorting into a high-type location outweighs the disadvantage stemming from fiercer competition.

A sufficient condition for this inequality to hold is:

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

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 for positive sorting also applies here:

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

A suitably high ε^P or low φ^E again guarantees positive sorting—as in the baseline model. The proof for negative sorting (Part 2.) is analogous. □

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. This, in turn, affects land distribution R and thus the matching function μ. So despite consistently focusing on the case of pure positive sorting in (ℓ, p), one can now analyze how investment and thereby the 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 location, and thus spatial sorting and spatial inequality.

Existence. Equilibrium existence thus obtains by solving for that threshold ψ^* such that a developer with ability $\psi = \psi^*$ is indifferent between investing and not investing given the Walrasian equilibrium in the land market consistent with that threshold. This then uniquely pins down $\Gamma_\ell(y) = \Gamma(y|\mu(\ell, \psi^*))$ under positive sorting and thereby the other equilibrium objects, as detailed in the proof of existence in the baseline model.

In particular, ψ^* solves $U_1(\psi^*) - U_0(\psi^*) = c(\psi^*)$, where we have made explicit that U_0 and U_1 are functions of ψ^*. Integrating $U_1 - U_0$ by parts yields

$$\int \frac{\partial k(\ell, \psi^*))}{\partial \ell} \Delta R(\ell) d\ell = c(\psi^*), \tag{37}$$

where $\Delta R \equiv R_0 - R_1 \geq 0$. In sum, given any ψ^* that solves (37), we obtain the land distribution $R(\cdot, \psi^*)$, which then yields a Walrasian equilibrium in the matching stage given by $(k(\cdot, \psi^*), \mu(\cdot, \psi^*))$. In turn, this pins down $U_1(\psi^*) - U_0(\psi^*)$ in the investment stage, which is higher than the cost $c(\psi)$ for $\psi \geq \psi^*$ and lower otherwise. This rationalizes $R(\cdot, \psi^*)$, completing the equilibrium construction.

Proposition 10. *Assume the conditions from Proposition 9 hold (either Part 1. or 2.). An equilibrium exists and exhibits $\psi^* \in (0, 1)$.*

Proof. We are seeking a fixed point in ψ^*. To see that it exists and exhibits $\psi^* \in (0, 1)$, note that $U_1 - U_0$ is strictly positive and continuous for all $\psi^* \in [0, 1]$, and c diverges to infinity when ψ^* goes to zero and it is zero at $\psi^* = 1$. Hence, there is at least one solution to $U_1(\psi^*) - U_0(\psi^*) = c(\psi^*)$ where $U_1 - U_0$ crosses c *from below* (i.e. a stable equilibrium in the sense that small perturbations cause adjustments that bring the economy back to equilibrium). For any given ψ^*, the Walrasian equilibrium of the land market can be constructed as in the baseline model (see Appendix B.4). $\square$

D.4 Endogenous Vacancy Posting

We assume that, when firms p choose locations ℓ, they decide 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 the total meetings between workers and firms in location ℓ are given by

$$\mathcal{M}(\ell) = \mathcal{A}\mathcal{V}(\ell)^{\alpha}\mathcal{U}(\ell)^{1-\alpha} \tag{38}$$

where $\mathcal{V}(\ell)$ is the number 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 weighted number of workers who search, both unemployed and employed ones. We denote the relative meeting rate of an employed worker compared to an unemployed worker by κ. We define market tightness in location ℓ by $\theta(\ell) = \frac{\mathcal{V}(\ell)}{\mathcal{U}(\ell)}$. Then, we can express the meeting rates as $\lambda^F(\ell) = \mathcal{A}\theta(\ell)^{\alpha-1}, \lambda^U(\ell) = \mathcal{A}\theta(\ell)^{\alpha}$, and $\lambda^E(\ell) = \kappa\mathcal{A}\theta(\ell)^{\alpha}$, with the standard properties. We impose the following assumptions on the matching function and vacancy posting costs.

Assumption 4.

1. *Total meetings in location ℓ are given by (38) with $1 > \alpha\,(2\kappa - 1)$.*
2. *Vacancy posting cost c is C^2 with $c' > 0, c'' > 0$, and $c'(0) = 0$.*

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

$$\mathcal{V}(\ell) = \int v(p,\ell)m_p(p|\ell)dp.$$

The effective measure of workers searching for a job in location ℓ is the weighted sum of unemployed and employed workers, where the weight is given by relative meeting rate of employed workers κ:

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

Plugging both $\mathcal{V}(\ell)$ and $\mathcal{U}(\ell)$ into the definition of market tightness, $\theta(\ell) = \frac{\mathcal{V}(\ell)}{\mathcal{U}(\ell)}$ and simplifying the expression, we obtain

$$\theta(\ell)\frac{\delta+\kappa\mathcal{A}\theta(\ell)^{\alpha}}{\delta+\mathcal{A}\theta(\ell)^{\alpha}}=v(\mu(\ell),\ell) \tag{39}$$

and note that θ is strictly increasing in the vacancy posting rate of firms v in any given market ℓ under Assumption 4.1. Equation (39) implicitly determines the equilibrium local market tightness $\theta(\ell)$ as function of the vacancy posting rate $v(p,\ell)$.

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

$$\begin{aligned}\overline{J}(p,\ell) &= \max_{v}\{\lambda^F(\ell)v\hat{J}(p,\ell)-c(v)\}-k(\ell)\\ \text{where } \hat{J}(p,\ell) &= \delta\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{y}\frac{\frac{\partial z(A(\ell),t)}{\partial y}}{[\delta+\lambda^E(\ell)(1-\Gamma_\ell(t))]^2}dtd\Gamma(y|p)\end{aligned}$$

We now state our main result of this extension.

Proposition 11.

1. *If z is strictly supermodular, and the competition forces $1/\delta$ are sufficiently small, then there is positive sorting of firms p to locations ℓ with $p=\mu(\ell)=Q^{-1}(R(\ell))$.*
2. *If z is strictly submodular, and the competition forces $1/\delta$ are sufficiently small, then there is negative sorting of firms p to locations ℓ with $p=\mu(\ell)=Q^{-1}(1-R(\ell))$.*

Proof. Before discussing Part 1. and 2. separately, we begin with some observations that will be useful for both.

The first-order condition with respect to the vacancy posting rate is given by

$$\lambda^F(\ell)\hat{J}(p,\ell)=c'(v(p,\ell)), \tag{40}$$

allowing us to solve for the vacancy posting rate $v(p,\ell)$ of firm p in location ℓ. We can then compute the expected value $\overline{J}(p,\ell)$ and its derivatives as:

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

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 (40) with respect to ℓ.

Apply integration by parts to $\hat{J}(p,\ell)$:

$$\hat{J}(p,\ell) = \delta \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(A(\ell),y)}{\partial y}}{[\delta + \lambda^E(\ell)(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|p))dy \tag{41}$$

and then obtain its derivatives:

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

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

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

Part 1. First note that $\hat{J}(p,\ell)$ is increasing in p, $\frac{\partial}{\partial p}\hat{J}(p,\ell) > 0$. Then, $\overline{J}(p,\ell)$ is supermodular in (p,ℓ) if

$$\frac{\partial^2 \hat{J}(p,\ell)}{\partial \ell \partial p}\lambda^F(\ell) + \frac{\partial \hat{J}(p,\ell)}{\partial p}\frac{\partial \lambda^F(\ell)}{\partial \ell} > 0, \ \frac{\partial \lambda^F(\ell)\hat{J}(p,\ell)}{\partial \ell} > 0.$$

A sufficient condition is that the integrands of these expressions are positive for all $y \in [\underline{y}, \overline{y}]$ and strictly so for some set of y. Note that the first equation is the derivative of the second one with respect to p. The difference of integrands of the two is that the first one has $-\frac{\partial \Gamma(y|p)}{\partial p}$ instead of $1-\Gamma(y|p)$ in addition to the common term. Using $-\frac{\partial \Gamma(y|p)}{\partial p} > 0$ and $1-\Gamma(y|p) \geq 0$, both conditions are satisfied if

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

We can re-express this inequality by replacing the meeting rates by functions of market tightness:

$$\frac{\frac{\partial^2 z(A(\ell),y)}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(A(\ell),y)}{\partial y}} > \frac{2\kappa \mathcal{A} \theta_\ell^\alpha \left(-\frac{\partial \Gamma_\ell}{\partial \ell} + \frac{\alpha}{\theta_\ell}\frac{\partial \theta_\ell}{\partial \ell}(1-\Gamma_\ell(y))\right)}{\delta + \kappa \mathcal{A} \theta_\ell^\alpha (1-\Gamma_\ell(y))} + \frac{1-\alpha}{\theta_\ell}\frac{\partial \theta_\ell}{\partial \ell}$$

Plugging in $\Gamma_\ell(y) = \Gamma(y|Q^{-1}(R(\ell)))$ and $\frac{\partial \Gamma_\ell(y)}{\partial \ell} = \frac{\partial \Gamma}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}$ (both of which hold under PAM), this inequality becomes

$$\frac{\frac{\partial^2 z(A(\ell),y)}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{\frac{\partial z(A(\ell),y)}{\partial y}} >$$
$$\frac{2\kappa \mathcal{A}\theta(\ell)^\alpha \left(-\frac{\partial \Gamma}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)}{\delta + \kappa \mathcal{A}\theta(\ell)^\alpha(1-\Gamma(y|Q^{-1}(R(\ell))))} + \frac{\frac{\partial \theta(\ell)}{\partial \ell}}{\theta(\ell)}\left(\frac{2\kappa \mathcal{A}\theta(\ell)^\alpha \alpha(1-\Gamma(y|Q^{-1}(R(\ell))))}{\delta + \kappa \mathcal{A}\theta(\ell)^\alpha(1-\Gamma(y|Q^{-1}(R(\ell))))} + 1 - \alpha\right), \quad (42)$$

which is our sufficient condition for PAM under endogenous vacancy posting.

Define ε^P as the baseline model which is the minimum of the L.H.S. It is strictly positive under our assumptions and the premise. Moreover, define on the R.H.S.

$$t^P \equiv \max_{y,\ell} \frac{2\kappa \mathcal{A}\theta(\ell)^\alpha \left(-\frac{\partial \Gamma}{\partial p}\frac{r(\ell)}{q(Q^{-1}(R(\ell)))}\right)}{\delta + \kappa \mathcal{A}\theta(\ell)^\alpha(1-\Gamma(y|Q^{-1}(R(\ell))))} + \frac{\frac{\partial \theta(\ell)}{\partial \ell}}{\theta(\ell)}\left(\frac{2\kappa \mathcal{A}\theta(\ell)^\alpha \alpha(1-\Gamma(y|Q^{-1}(R(\ell))))}{\delta + \kappa \mathcal{A}\theta(\ell)^\alpha(1-\Gamma(y|Q^{-1}(R(\ell))))} + 1 - \alpha\right).$$

Under endogenous vacancy posting, t^P depends on $\theta(\ell) = v(Q^{-1}(R(\ell)), \ell)$, which is an endogenous object. Thus, the sufficient conditions for PAM from the baseline model are not readily applicable. Instead, we argue that if δ is large, then t^P converges to zero and thus (42) holds for a strictly positive ε^P.

First, if $\delta \to \infty$, then $\hat{J}(p,\ell) \to 0$, which follows from the definition of $\widehat{J}$, equation (41). To check how θ behaves as $\delta \to \infty$, plug in $\lambda^F(\ell) = \mathcal{A}\theta(\ell)^{\alpha-1}$ into equation (40),

$$\hat{J}(\mu(\ell),\ell) = c'(v(\theta(\ell)))\left(\mathcal{A}\theta(\ell)^{\alpha-1}\right)^{-1}, \quad (43)$$

where $v(\theta(\ell))$ is implicitly defined by (39). Since $c'(v(\theta(\ell)))$ and $\theta(\ell)^{1-\alpha}$ are both strictly increasing in θ and zero at $\theta(\ell) = 0$, and since we have just shown that $\hat{J}(p,\ell) \to 0$ as δ

becomes large, we conclude that $\theta(\ell)$ converges to zero if $\delta \to \infty$. As a result, the first fraction in t^P converges to zero as $\delta \to \infty$, and so does $\frac{2\kappa\mathcal{A}\theta(\ell)^\alpha \alpha(1-\Gamma(y|Q^{-1}(R(\ell))))}{\delta+\kappa\mathcal{A}\theta(\ell)^\alpha(1-\Gamma(y|Q^{-1}(R(\ell))))}$ in the second term.

Next, differentiating equation (43) with respect to ℓ, we obtain after some rearrangement:

$$\frac{\frac{\partial\theta(\ell)}{\partial\ell}}{\theta(\ell)} = \frac{\frac{\partial\mu(\ell)}{\partial\ell}\frac{\partial\hat{J}(\mu(\ell),\ell)}{\partial p}}{c''(v(\theta(\ell)))v'(\theta(\ell))\frac{\theta(\ell)}{v(\theta(\ell))}\frac{v(\theta(\ell))}{c'(v(\theta(\ell)))} + 1 - \alpha}.$$

The denominator on the right hand side is bounded from below by $1-\alpha$. Also, $\frac{\partial\hat{J}(\mu(\ell),\ell)}{\partial p}$ converges to zero if δ goes to infinity since:

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

Furthermore, our assumptions guarantee that $\frac{\partial\mu(\ell)}{\partial\ell} = \frac{r(\ell)}{q(\mu(\ell))}$ is bounded since the pdfs q and r are continuous and defined over the compact sets $[\underline{p},\overline{p}]$ and $[\underline{\ell},\overline{\ell}]$, respectively. Thus, $\frac{\partial\theta(\ell)}{\partial\ell}\frac{1}{\theta(\ell)}$ converges to zero if δ becomes large. Therefore, we can find sufficiently large and finite δ that guarantees that $\varepsilon^P > t^P$ and therefore PAM.

Part 2. In turn, $\overline{J}(p,\ell)$ is submodular in p and ℓ, thereby guaranteeing NAM, if

$$\frac{\partial^2\hat{J}(p,\ell)}{\partial\ell\partial p}\lambda^F(\ell) + \frac{\partial\hat{J}(p,\ell)}{\partial p}\frac{\partial\lambda^F(\ell)}{\partial\ell} < 0, \quad \frac{\partial\lambda^F(\ell)\hat{J}(p,\ell)}{\partial\ell} < 0.$$

It suffices that the integrands involved in these expressions are negative for all $y \in [\underline{y},\overline{y}]$ and strictly so for some set of y. Following the same logic as in Part 1, these conditions are satisfied if

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

Define ε^N as in the baseline as the maximum of the L.H.S. Note that under a submodular production function, ε^N exists and negative. Define t^N as the minimum of the R.H.S. Then, following similar steps as in Part 1, we can show that t^N converges to zero if δ increases. Thus, we can find a sufficiently large and finite δ that yields $\varepsilon^N < t^N$, and therefore NAM. $\square$

For intuition, consider the case of PAM. Competition forces are 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 (new channel). This increases market tightness in good locations, further disadvantaging firms. Hence, competition in good locations is amplified by endogenous vacancy posting. To compensate for stronger competition arising from both 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 obtains.

E Empirical Exercises

E.1 Descriptive Statistics

Table 7: Spatial Heterogeneity: Descriptive Statistics

	Mean	S.D.	P10	P25	P50	P75	P90
Wages	119.34	22.25	89.66	102.37	118.22	133.87	153.79
Value Added	94.62	27.61	63.67	74.76	89.31	113.25	136.50
Average firm size	61.59	12.35	45.23	52.46	59.79	73.29	74.77
Share emp. top 10%	0.53	0.05	0.46	0.49	0.54	0.56	0.58
Pop. density	247.16	328.56	74.85	105.25	151.55	245.61	422.18
Population	405,822	593,441	89,239	127,522	193,713	419,795	1,026,787

Notes: Data source: LIAB-BHP for wages and size, EP for value added. Number of observations is 204. The table reports moments of the distributions of various outcomes across commuting zones (*Arbeitsmarktregionen*). Except for population and population density, moments are weighted using the number of firms in each commuting zone. Wages refer to daily wages, and are reported in 2015 Euros. Value added is in thousands of 2015 Euros. Finally, for wages and value added, we calculate the firm-weighted averages within each commuting zone, where the weights are the number of full-time employees of a firm.

E.2 Measuring Firm Productivity y.

Our theory highlights the role of innate firm-productivity y in determining firms' local productivity rank, $\Gamma_\ell(y)$, and their global productivity rank, $\int_\ell \Gamma_\ell(y)r(\ell)d\ell$. Measuring y empirically, however, is complicated by the fact that firms—according our theory—are sorted spatially and that their output z and wages w depend on not only their own productivity y but also on the location productivity, $A(\ell)$. Therefore, we cannot readily rank firms by either of these observable (wage or value added) statistics.

In order to obtain a clean measure of y, we implement the following procedure. Our starting point are the firm fixed effects from a standard two-way fixed effects (AKM) wage regression and we denote this estimate for firm f in location ℓ by $y_{f,\ell}^{AKM}$. In doing so, we control for the effect of worker sorting across firms or locations, which is not present in our theory. But we still have to isolate in this measure the effects of y from those of $A(\ell)$.

To control for the location productivity $A(\ell)$, note that our theory implies that if the production function takes the multiplicative form $z(y, A(\ell)) = A(\ell)y$, the highest output level in location ℓ is achieved by the top firm $\overline{y}$:

$$\ln z(\overline{y}, A(\ell)) = \ln A(\ell) + \ln \overline{y},$$

where, because of our assumption of common support of $\Gamma(y|p)$, $\overline{y}$ is common across locations

ℓ. Hence, the regional variation of $\ln z(\overline{y}, A(\ell))$ only reflects variation of $A(\ell)$ and not the effect of sorting. We therefore measure the maximum output in location ℓ by the maximum value added, $\overline{va}_\ell$, and use it to residualize the estimated firm fixed effects. Formally, we run the regression

$$y_{f,\ell}^{AKM} = \alpha + \beta \ln \overline{va}_\ell + u_{f,\ell}, \tag{44}$$

and take the estimated residuals $\hat{u}_{f,\ell}$ as our measure of y. In practice, to minimize measurement error, we proxy $\overline{va}_\ell$ with the 90% quantile of the local value added distribution. In Section E.3 in the Appendix, we discuss this approach in more detail. In particular, we show that—for a particular parametrization of our model—residualization (44) is dictated by our theory and yields a clean measure of y, purged from the location productivity $A(\ell)$.

Note that this procedure does not affect firms' *local* rank $\Gamma_\ell(y)$, which is based on firms' relative productivity differences *within* locations. Hence, the measurement choice in (44) is immaterial for most of our empirical results that only rely on this local rank. It is only when we look at cross-location differences and exploit firms' *global* rank that (44) matters. In our quantitative analysis in Section 7, we will be able to identify y separately from $A(\ell)$ in the context of our model.

E.3 Residualization of AKM effects to measure y

AKM estimation decomposes wages into two components: workers and firms. We abstract from worker ability differences in the baseline model and focus on the firm side. Under assumptions for AKM estimation being valid, AKM firm fixed effects reflect the wage function $w(y, \ell)$ in our model:

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

Under the spatial sorting of firms, wage is determined by many factors including the location's productivity $A(\ell)$, the local firm distribution Γ_ℓ, and a firm's productivity y. To attain a cleaner measure of y, we residualize AKM firm fixed effects with the top firm's value added which is a proxy of $z(A(\ell), \overline{y})$.

Example Assume output function is multiplicative, i.e. $z(A(\ell), y) = A(\ell)y$, which is our measure of value added, and $y \sim U[0, 2p]$. Then, $y = 2pq$ where q is the local productivity rank. Rearranging the wage as a function of the local productivity rank q, we obtain

$$\log w(q,\ell) = \log A(\ell) + \log \mu(\ell) + 2\log q + \log 2 + \log \frac{\varphi^E}{1+\varphi^E}$$

If the true data generating process is as assumed, then subtracting the top firm's value added from the AKM firm fixed estimates gives the following residualized firm FE:

$$\log w^r(q,\ell) = \log \mu(\ell) + 2\log q + \text{const.} \tag{45}$$

Since $y = 2\mu(\ell)q$ (see above), the residualized firm FE becomes $\log w^r = \log y + \log q + \text{const.}$, and is therefore a good measure of $\log y$ (with an uncorrelated noise term $\log q$).

In practice, however, as AKM firm fixed effects and the value added come from two separate sources, it is difficult to directly subtract the top firm's value added from AKM firm fixed effects. Moreover, this derivation relies on specific functional forms. Thus, we emphasize that the estimated residuals are a *proxy* for y. A potential complication can arise in regression (45) if the omitted variable bias in this regression is severe: Due to a positive correlation between $\log A(\ell)$ (in the residual) and $\log \mu(\ell)$ (regressor) under PAM, we may overestimate the coefficient in the residualization regression. Still, note that our measure of $\log y$ is conservative since we may attribute some of the firms' sorting effect (stemming from $\mu(\ell)$) to the effect of locations' productivity $A(\ell)$. Thus, even if there is a bias, it would underestimate the impacts of firms' spatial sorting.

E.4 Local Labor Markets: Qualitative Predictions of our Theory

In this section we provide qualitative evidence for a variety of implications of our theory. We first look at the variation across labor markets, focussing on differences in labor market frictions. We then look at the firm-level variation within labor markets.

Variation across labor markets Table 8 provides an alternative way of understanding the importance of local labor markets, where we relate the local distribution of firm size, firm wages and the net poaching shares to our measures of local labor market frictions. In Lemma 6 we showed that both the average and the dispersion of firm size and wages are increasing in the efficiency of EE hiring, i.e, in φ^E, and are independent in the efficiency of UE hiring, φ^U. We also showed that the dispersion in net poaching shares is independent of local labor

Table 8: OJS and Local Labor Market Outcomes

Panel A: Levels						
	Size		Wage		NPS	
φ^E	2.782*	6.852***	7.013***	14.913***	0.001	0.004
	(1.081)	(1.625)	(1.529)	(1.923)	(0.002)	(0.002)
φ^U		-5.087**		-9.873***		-0.003
		(1.847)		(2.243)		(0.003)
Panel B: Dispersion						
	Size		Wage		NPS	
φ^E	1.849*	3.885***	0.146***	0.341***	0.012	-0.034
	(0.758)	(1.143)	(0.039)	(0.058)	(0.017)	(0.025)
φ^U		-2.545		-0.244***		0.058*
		(1.404)		(0.055)		(0.024)

Notes: Data source: LIAB-BHP for size, wage, and NPS; LIAB for φ^E and φ^U. Regressions at the commuting zone level ($N = 204$). In Panel A , the dependent variables are the average local firm size (relative the smallest firm), the average local wage and the average local net poaching share. In Panel B, the dependent variables are the local dispersion of the respective variables, which we measure as the ratio between the mean of the top decile and the mean of the bottom decile of the corresponding distribution. The variable φ^E is defined as $\varphi^E = \lambda^E/\delta$, where λ^E is the EE rate, and δ is the separation rate. Similarly, $\varphi^U = \lambda^U/\delta$ for λ^U the UE rate. We normalize both φ^E and φ^U before running the regression. See Section 5.1 for details on measurement.

market frictions while the average level is decreasing in φ^E.

Table 8 shows that these predications are qualitatively borne out in our data. In Panel A we focus on the average level of the respective variable in a commuting zone; in Panel B, we focus on the local dispersion of the respective variable. The first four columns show that the local frequency of EE contacts relative to the rate of job destruction, φ^E, is an important determinant of both the average level and the dispersion of firm size and wages in the commuting zone. By contrast, the efficiency of search out of the unemployment pool, captured by φ^U, is less important.

Regional differences in the net poaching share are only weakly correlated with local measures of labor market frictions. For the case of dispersion, this is exactly what our theory implies. For the case of the average level, our theory implies that the coefficient should be negative while the effect in the data ranges from zero to slightly negative. However, as we pointed out when discussing our comparative statics, two effects are at play, with the negative effect being the dominant one.

Variation across firms within labor markets Our theory makes tight predictions on the relationship between firm-level outcomes and the firms' competitiveness in its local labor market, as captured by the firms' *local rank* $\Gamma_\ell(y)$. As we showed in Lemma 3, wages, poaching shares and size are all increasing in $\Gamma_\ell(y)$. We now assess these prediction empirically by considering a specification of the form

$$\ln \mathcal{D}_{f,\ell} = \alpha + \beta\Gamma_\ell(y) + X'_{f,\ell}\gamma + \epsilon_{f,\ell}, \tag{46}$$

where $\mathcal{D}_{f,\ell}$ denotes the dependent variable (either $n^E(y_f,\ell)$ or $l(y_f,\ell)$ or $\ln w(y_f,\ell)$) and $X'_{f,\ell}$ is a vector of additional controls.

We first focus on firms' net poaching share, $n^E(y_f,\ell)$, and their size, $l(y_f,\ell)$. Our theory implies that these statistics should only be determined by firms' local rank $\Gamma_\ell(y)$, not by their global one. In Panel A of Table 9, we show the estimation results of (46) for the firm-level net poaching share. In column 1 we report the simple correlation between the net poaching share and the local productivity rank. Consistent with our theory, there is a strong positive correlation: the higher firms are in the local productivity ladder, the more successful they are in poaching from other firms (and avoiding to be poached) in the local labor market. In columns 2 and 3, we control for firm-level characteristics and a labor market fixed effect. This reduces the coefficient but leaves the positive correlation intact. In column 4, we predict the net poaching share by the firm's *global* rank (instead of its local rank). Again, there is a positive relationship owing to the positive correlation between the global and the local rank. In the last column, we include both the local and the global rank. Our theory implies that the local rank should be a sufficient statistic for firms' net poaching share. This is exactly what the data shows: Holding the local productivity rank constant, the global productivity rank is no longer significant.

Table 9: The local rank and firm outcomes

	M1	M2	M3	M4	M5
Panel A: LHS = Net poaching share					
Local Rank	0.24***	0.17***	0.17***		0.18***
	(0.01)	(0.01)	(0.01)		(0.01)
Global Rank				0.15***	-0.01
				(0.01)	(0.01)
Panel B: LHS = Log full-time employees					
Local Rank	0.96***	0.87***	0.87***		0.79***
	(0.02)	(0.01)	(0.01)		(0.06)
Global Rank				0.81***	0.09
				(0.02)	(0.07)
Panel C: LHS = Log real wage of full-time employees					
Local Rank	1.13***	0.87***	0.89***		0.17**
	(0.02)	(0.02)	(0.02)		(0.07)
Global Rank				0.95***	0.77***
				(0.01)	(0.06)
Controls	N	Y	Y	Y	Y
Location FE	N	N	Y	N	N

Notes: Data source: LIAB-BHP. Observation is at the firm level. Standard errors clustered by commuting zone are in parentheses. Panel A: Regression weighted by mean (across years) number of full-time employees. Controls are P90 commuting zone log real value added, and 3-digit industry fixed effects. Number of observations is 251,404. Panel B: Controls are log mean full-time employees and P90 log mean real value added, both at commuting zone level, maximum age of firm in the data, and 3-digit industry fixed effects. Number of observations is 262,998. Panel C: Regression weighted by mean (across years) number of full-time employees. Controls are P90 commuting zone log real value added, log mean wage at the commuting zone level, commuting zone unemployment rate, 3-digit industry fixed effects, and mean share of full-time and marginal employees. Number of observations is 262,666.

Panel B of Table 9 focuses on the determinants of firm size. The structure is identical to Panel A. The first three columns again show a strong positive relationship between firm size and the firms' local rank. In particular, as in the case of the net poaching share, the relationship between firm size with the local and the global rank is similar (columns 3-4). Column 5 then shows that the local, not the global, rank is the main predictor of firm size.

Finally, in Panel C, we study firm-level wages. Again, we find a strong positive relationship with the local productivity rank, which also holds true within local labor markets (see column 3). The last column shows that the distinction between the local and global productivity rank is less clear cut. In fact, our theory suggests why this is the case: In contrast to firm size and poaching shares, wages depend directly on location productivity $A(\ell)$ conditional on firms' local rank. And because of the aforementioned challenges in separating $A(\ell)$ from y, the global rank may still reflect some part of local productivity $A(\ell)$, rendering it significant.

E.5 Postive Firm Sorting

To test this prediction empirically, we conduct the following exercise. We assign firms to five bins based on their productivity y. For each bin or group g, we calculate the share of firms of type g in location ℓ, $s_{g,\ell} = N_{g,\ell}/\sum_g N_{g,\ell}$, where $N_{g,\ell}$ denotes the number of firms of type g in location ℓ. We then run a regression of the form

$$s_{g,\ell} = \alpha + \sum_{g=1}^{G} \beta_g^A A(\ell) + \gamma A(\ell) + X_\ell' \rho + u_{g,\ell}, \tag{47}$$

and report the coefficients β_g^A. These coefficients capture the effect of location productivity $A(\ell)$ on the relative probability of firms of type g to settle in location ℓ. Similarly, to test the role of local competition, we run the same specification as in (47), except that we use firms' hypothetical local rank in lieu of location productivity $A(\ell)$.[34] The corresponding coefficient $\beta_g^{\mathcal{R}}$ captures the extent to which firm sorting depends on the local rank a given firm would have in location ℓ if it settled there.

In Table 10, we report the results of both specifications. We report three specifications with different local controls. All of them paint a consistent picture. For local productivity $A(\ell)$ (see the columns denoted by it), the coefficients β_g^A tend to be increasing in g, that is location productivity is a stronger predictor for highly productive firms (Group 5) to settle in the respective location compared to little productive ones (Group 1). This pattern is consistent with PAM being driven by a complementarity between location productivity and firm productivity. For the local rank (see columns denoted by it), the opposite pattern emerges: The fact that $\beta_g^{\mathcal{R}}$ is decreasing in g is consistent with the idea that low type firms are more responsive to the threat of competition and hence settle in locations where their local rank is relatively higher. These results indicate that productive and unproductive firms resolve the trade-off between local productivity and competition differently, lending support to positive sorting of firms across space—in line with our theory.

[34]For each type g, we calculate its hypothetical local rank in ℓ as the share of firms with a worse type, i.e. $\mathcal{R}_{g,\ell} = \sum_{j\leq g} N_{j,\ell}/\sum_j N_{j,\ell}$.

Table 10: Location Patterns by Firm Type

	M1		M2		M3	
	Local rank	A_ℓ	Local rank	A_ℓ	Local rank	A_ℓ
Group 1 (low)	1.27***	0.47	1.11***	0.69**	1.12***	0.55**
	(0.08)	(0.28)	(0.09)	(0.22)	(0.08)	(0.21)
Group 2	0.85***	0.47*	0.64**	0.73***	0.69***	0.61**
	(0.13)	(0.21)	(0.15)	(0.21)	(0.13)	(0.20)
Group 3	0.15	0.87***	-0.17	1.04***	-0.05	0.93***
	(0.24)	(0.19)	(0.27)	(0.23)	(0.25)	(0.22)
Group 4	-1.33**	0.99***	-2.02***	1.10***	-1.72**	1.00***
	(0.49)	(0.22)	(0.55)	(0.26)	(0.52)	(0.25)
Group 5 (high)	0.00	0.74**	0.00	0.87**	0.00	0.77**
	(.)	(0.26)	(.)	(0.29)	(.)	(0.29)
A_ℓ	N	N	Y	N	Y	N
Local rank	N	N	N	Y	N	Y
Unemployment rate	N	N	Y	Y	Y	Y
Skill level	N	N	N	N	Y	Y
N	1,015	1,015	1,015	1,015	1,015	1,015

Notes: Data source: LIAB-BHP. Each column is a separate specification. Regression is at the commuting zone-firm group level, where the lowest group (Group 1) refers to the firms with lowest y and, equivalently, the highest group refers to the firms with the highest y. Column label 'Local rank' means that the coefficients refer to the interaction between group dummies and local ranks; column label 'A_ℓ' means that the coefficients refer to the interaction between group dummies and A_ℓ. A_ℓ is defined as the logarithm of the mean value added in that commuting zone. The control 'Skill level' refers to the mean (across firms) share of college graduates in the commuting zone. $N = 1,015$ for all regressions, equal to 203 commuting zones (we omit one due to data limitations) times 5 firm groups. Standard errors clustered by district are in parentheses. The coefficient for Group 5 interacted with local rank cannot be identified since there is no variation in the data: firms in this group are always ranked top.

E.6 BHP vs LIAB-BHP: Representativeness

In this part we argue that LIAB-BHP is very similar to BHP, so that, despite LIAB-BHP not being sampled in a random way, it is alike to a random sample. To begin with, Table 11 compares summary statistics for both LIAB-BHP and BHP. For poaching and poached shares, as well as wages, the moments are remarkably similar[35], both at the firm level and at the commuting zone level. Take the firm-level wage as an example. Not only it is true that the mean differs by 0.64EUR - equivalent to 0.5% of the wage in LIAB-BHP - but also the standard deviation and median are essentially identical. The industry composition of both datasets is also remarkably similar.

It is important to note, however, that firm size is larger, on average, in the LIAB-BHP. This can be seen on Table 12. As Bossler et al. (2018) explains, this is due to the fact that

[35]This is after performing the sample selection described in the main text. If instead we use all firms in each dataset, mean employment in LIAB-BHP is 3 times as large, and average wage is 5% higher.

the EP sampling design neglects newly founded (i.e. smaller) establishments, but it does not affect any economic indicator significantly, when comparing the data against representative administrative data.

Table 11: Summary statistics for BHP and LIAB-BHP

		Mean	SD	Median	CV
Panel A: Firm level					
LIAB-BHP	Poaching	0.56	0.17	0.55	0.30
	Poached	0.49	0.17	0.49	0.35
	Wage	123.15	53.82	115.31	0.44
	FTE	61.59	241.38	25.00	3.92
BHP	Poaching	0.57	0.16	0.55	0.28
	Poached	0.50	0.16	0.49	0.32
	Wage	122.51	52.38	114.66	0.43
	FTE	50.64	220.11	21.00	4.35
Panel B: Commuting zone level					
LIAB-BHP	Poaching	0.56	0.02	0.56	0.04
	Poached	0.49	0.02	0.49	0.04
	Wage	119.81	21.26	118.12	0.18
	FTE	61.59	12.35	59.79	0.20
BHP	Poaching	0.57	0.02	0.57	0.04
	Poached	0.50	0.03	0.49	0.06
	Wage	119.93	20.10	1147.62	0.17
	FTE	50.64	8.97	50.70	0.18

Notes: Panel A: Number of observations is 263,753 for LIAB-BHP, 177,049 for BHP. Poaching, poached, and wages' statistics weighted by number of full-time employees. Panel B: Number of observations is 204 in both cases.

Table 12: Size distribution for BHP and LIAB-BHP

N employees	LIAB-BHP	BHP
10	0.03	0.02
15	0.26	0.31
20	0.40	0.48
25	0.50	0.58
30	0.58	0.65
35	0.63	0.70
40	0.67	0.74
45	0.71	0.77
50	0.74	0.80
70	0.82	0.86
90	0.86	0.90
150	0.93	0.95
300	0.97	0.98
500	0.99	0.99

Notes: Table shows the empirical cumulative distribution function for number of employees, for both LIAB-BHP and BHP. For example, 0.03 in the first row means that 3% of firms have ten or less employees on LIAB-BHP; that same moment is 2% for BHP.

F Estimation

F.1 The Pareto Assumption

The Distribution of Firm-Size in Germany In our quantitative model we assume that the ex-post productivity distribution is a (bounded) Pareto distribution. In the section we show that empirically the firm size distribution in Germany indeed has a pareto tail.

Suppose y is pareto distributed with

$$P(y \leq q) = 1 - \left(\frac{y_0}{q}\right)^{\zeta}$$

Then

$$P(z \leq k) \quad = \quad P(Ay \leq k) = P\left(y \leq \frac{k}{A}\right) 1 - \left(\frac{y_0 A}{k}\right)^{\zeta}.$$

Hence, z is distributed pareto with tail ζ and lower bound $y_0 A$. Now define $S(k)$ to be the counter-cumulative distribution function $S(k) \equiv 1 - P(z \leq k) = P(z > k) = \left(\frac{y_0 A}{k}\right)^{\zeta}$. Then

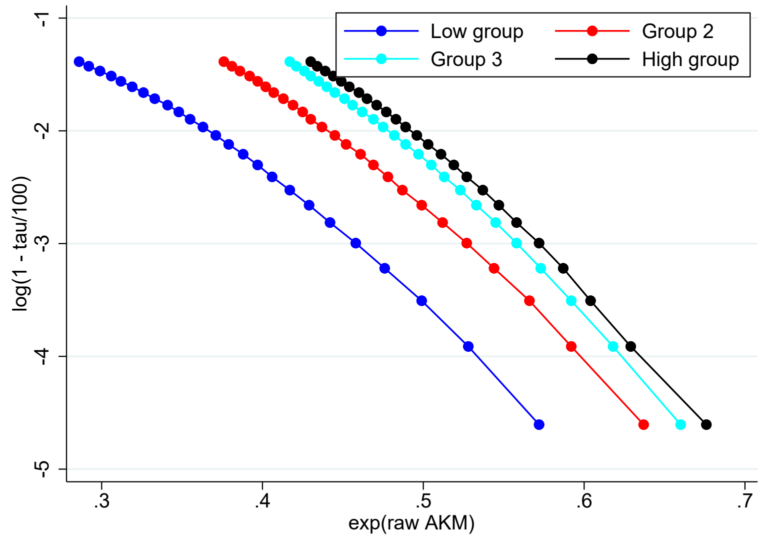
$$\ln S(k) = \zeta \ln (y_0 A) - \zeta \ln k, \tag{48}$$

that is the relationship between $\ln S(k)$ and $\ln k$ should be linear.

To test the linearity restriction contained in equation (48), let x_τ denote the τth percentile of the estimated AKM firm effects for $\tau = 1, ...99$. Equation (48) implies that the relationship between $\ln\left(1 - \frac{\tau}{100}\right)$ and $\ln(x_\tau)$ should be linear. To allow for differences across locations, we consider this relationship for four groups of commuting zones as ordered by their location productivity A_ℓ. We construct these groups based on the quantiles of the A_ℓ distribution. that is each bin accounts for 25% of locations in our sample.[36] The results are contained in Figure 10.

[36] Strictly speaking, in our theory, each group contains a continuum of locations, each with their region-specific tail index $p = \mu(\ell)$. The average distribution of productivity when pooled across locations within a bin, is not exactly pareto. However, it has a pareto tail, which is given by the lowest tail of the productivity distributions within the group. Formally, let $\mathcal{G} \subset [\underline{\ell}, \overline{\ell}]$ be a subset of locations. The productivity distribution of all firms in $\ell \in \mathcal{G}$ then has a pareto tail $\zeta(\mathcal{G})$ given by $\zeta(\mathcal{G}) = \min_{\ell \in \mathcal{G}} \{p(\ell)^{-1}\}$.

Figure 10: Testing the Pareto Assumption



Notes: The figure shows the relation between $\ln S(k)$ and $\ln k$ in the four different groups of commuting zones as defined by their productivity A_ℓ.

Figure 10 contains two results. First, the relationship is indeed approximately linear as implied by (48). Secondly, the "rightward" shift across space is vividly apparent. Note that (48) implies that the location of the regression is a function of A_ℓ.

F.2 Identification

We first summarize our functional form assumptions:

Assumption 5. *We assume the following functional forms and normalizations:*

1. *The matching function is given by* $M(\ell) = \mathcal{A}V(\ell)^{1/2}U(\ell)^{1/2}$.
2. *The ex-post firm productivity distribution is given by* $\Gamma(y \mid p) = 1 - y^{-\frac{1}{p}}$.
3. *Land distribution* R *is normalized,* $R \sim U[\underline{\ell}, \overline{\ell}]$.
4. *Normalize* $\rho V^U = 1$.

Proof of Proposition 5. We need to identify the ranking of locations $[\underline{\ell}, \overline{\ell}]$, functions (Q, A, B), and the parameters of the matching function $(\kappa, \mathcal{A})$.

First, we discussed how to rank locations, assigning $\ell \in [\underline{\ell}, \overline{\ell}]$ to each location, in the text.

Second, Q is identified from $p = \mu(\ell) = Q^{-1}(R(\ell))$, and $\mu(\ell)$ can be obtained from a location's labor share as $LS(\ell) = 1 - \mu(\ell)$. To see this note that aggregate output in location

ℓ equal to aggregate wages plus aggregate profits and land prices in equilibrium:

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

where we use the integration by part for the second line. So the labor share is given by:

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

At the same time, aggregate output can be expressed as follows, using that Γ is pareto and the firm size expression from (4):

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

Plugging aggregate output back into $LS(\ell)$ above, we obtain $LS(\ell) = 1 - \mu(\ell)$.

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

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

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

$$
\begin{aligned}
\lambda^E(\ell) &= M(V(\ell), U(\ell)) \frac{\kappa(1-u(\ell))}{u(\ell)+\kappa(1-u(\ell))} \frac{1}{(1-u(\ell))L(\ell)} \\
&= \mathcal{A} V(\ell)^{\alpha} (u(\ell)+\kappa(1-u(\ell)))^{1-\alpha} L(\ell)^{1-\alpha} \frac{\lambda^E(\ell)}{\delta+\lambda^E(\ell)} \frac{\delta+\lambda^U(\ell)}{\lambda^U(\ell)} \frac{1}{L(\ell)} \\
&= \mathcal{A} \left(\frac{\delta+\lambda^E(\ell)}{\delta+\lambda^U(\ell)}\right)^{1-\alpha} \frac{\lambda^E(\ell)}{\delta+\lambda^E(\ell)} \frac{\delta+\lambda^U(\ell)}{\lambda^U(\ell)} L(\ell)^{-\alpha} \\
\Rightarrow \quad L(\ell) &= \mathcal{A}^{\frac{1}{\alpha}} \frac{\delta+\lambda^U(\ell)}{\delta+\kappa\lambda^U(\ell)} \left(\frac{1}{\lambda^U(\ell)}\right)^{\frac{1}{\alpha}} \qquad (49)
\end{aligned}
$$

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

$$L(\ell) = \left(1 + \frac{\delta}{\lambda^U(\ell)}\right)\bar{l}(\ell) \tag{50}$$

Equalizing (49) and (50), and solving for $\mathcal{A}$, while imposing $\alpha = .5$, gives the equation in the text.

Fifth, to derive the value of unemployment, our starting point is:

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

which is the same as in the baseline model, except that $b(\ell)$ is augmented and now $B(\ell)b(\ell)$. Next, note that the reservation wage is implicitly given by:

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

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

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

Plug $b(\ell)$ back into V^U above, and use a change of variable (to re-express F using Γ_ℓ) and make use of the pareto assumption of Γ to obtain:

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

which allows us to back out $B(\ell)$ for each ℓ, given the normalization $\rho V^U = 1$. $\square$

F.3 Estimation Results

Table 13: Calibrated Parameters

Parameter	Value	Calibration
δ	0.0084	monthly EU transition rate (LIAB)
κ	0.3162	monthly UE and EE transition rate (LIAB)
$\mathcal{A}$	0.5547	median λ^U

Notes: The table reports the structural parameters for the quantitative exercise.

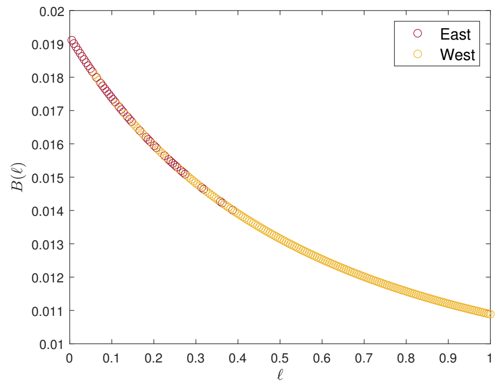


Figure 11: Estimation Result: Amenity Schedule

F.4 East-West Comparison: Additional Results

F.4.1 Value Added

West-East Inequality in Value Added.

Table 14: West-East Inequality

	Value Added	
	Data	Model
West	112.65	129.16
East	76.38	85.22
West/East	1.47	1.52

Variance Decomposition of Value Added.

The decomposition of the across-component of the spatial value added variance is given by:

$$\begin{aligned}\mathrm{Var}(\mathbb{E}[\log z(A(\ell), y) \mid \widehat{\ell}]) &= \mathrm{Var}(\mathbb{E}[\log A(\ell) \mid \widehat{\ell}] + \mathbb{E}[\log y \mid \widehat{\ell}]) \\ &= \underbrace{\mathrm{Var}(\mathbb{E}[\log A(\ell) \mid \widehat{\ell}])}_{\text{location}} + \underbrace{\mathrm{Var}(\mathbb{E}[\log y \mid \widehat{\ell}])}_{\text{firm}} + \underbrace{2\,\mathrm{Cov}(\mathbb{E}[\log A(\ell) \mid \widehat{\ell}], \mathbb{E}[\log y \mid \widehat{\ell}])}_{\text{PAM}}\end{aligned}$$

Table 15: Value Added Variance Decomposition, East-West

	Variance	Across	Location	Firm	PAM
$\log z$	0.233	8.29	32.11	18.78	49.11

Counterfactuals.

Table 16: West-East Value Added Inequality: Counterfactual Models

	Model	No Sorting	No Worker Frictions	No Firm Frictions	No OJS
Value Added West/East	1.52	1.24	1.18	1.03	1.15

F.5 Other Dimensions of Spatial Inequality

Rural-Urban Inequality.

Table 17: Urban-Rural Inequality

	Value Added		Wage	
	Data	Model	Data	Model
Urban	114.33	139.83	74.65	90.41
Rural	94.97	113.40	63.02	77.90
Urban/Rural	1.20	1.23	1.18	1.16

Table 18: Urban-Rural: Variance Decomposition

	Variance	Across	$A(\ell)$	$\tilde{w}(\ell)$	2Cov
$\log z$	0.233	3.20	30.78	19.82	49.40
$\log w$	0.081	5.55	50.74	8.28	40.98

Table 19: Urban-Rural Inequality: Counterfactual Models

	Model	No Sorting	No Worker Frictions	No Firm Frictions	No OJS
Value Added Urban/Rural	1.23	1.14	1.04	1.02	1.12
Wage Urban/Rural	1.16	1.14	1.01	1.01	1.21

Inequality across Commuting Zones.

Table 20: Commuting Zones: Variance Decomposition

	Variance	Across	$A(\ell)$	$\tilde{w}(\ell)$	2Cov
$\log z$	0.233	17.49	31.51	19.29	49.20
$\log w$	0.081	30.75	51.17	8.11	40.71

F.6 Counterfactuals: Technical Details

F.6.1 The Role of Firm Sorting

We assume that the unemployment benefit $b(\ell)$ is adjusted so that the reservation in each ℓ remains the same, i.e. $w^R(\ell) = A(\ell)\underline{y}$. In particular,

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

where F_ℓ and (λ^U, λ^E) are all determined in this counterfactual without spatial firm sorting. The wage function in each ℓ is strictly increasing in y as before, and so $F_\ell(w(y,\ell)) = \Gamma_\ell(y) = \Gamma(y)$, where the last equality follows since the ex-post productivity distribution is the same across regions under random matching.

As unemployed workers are freely mobile across regions, we calculate $\lambda^E(\ell)$ for each ℓ to equalize the value of search, given $B(\ell)$ and $A(\ell)$:

$$\rho V^U = B(\ell)A(\ell)\left[1 + 2(\lambda^E(\ell))^2 \int_1^\infty (1-\Gamma(y))\gamma(y) \int_1^y \frac{1}{[\delta + \lambda^E(\ell)(1-\Gamma(t))]^2} dtdy\right],$$

where Γ is the productivity distribution of firms in Germany (not ℓ-specific) as we impose random matching of firms to locations. Note that compared to baseline, we need to determine a new value of search, ρV^U, to calculate $\lambda^E(\ell)$. We choose the ρV^U that guarantees the same total population size as in the baseline economy. In practice, we solve a fixed point in ρV^U so that it satisfies both welfare equalization of workers as well as the population constraint.

Once we determine λ^E, we can compute λ^U.

F.6.2 The Role of On-The-Job Search

We decrease $\kappa = \frac{\lambda^E}{\lambda^U}$ to a small positive number. For this counterfactual exercise, the modularity properties of $\bar{J}$ may change, so we need to additionally solve for the sorting decision of firms. In turn, the population size in each location (and thus worker and firm meeting rates) also depends on the firm composition in each ℓ, which affects workers' wages. Thus, we need to solve for a fixed point in the firm allocation μ (and thus, Γ_ℓ).

For a given allocation of firms μ, we find the meeting rate λ^U as a function of $(\ell; \kappa, \mu(\ell), \rho V^U(\kappa))$ that equalizes the value of search for unemployed workers across space:

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

The details underlying this step are the same as in the "No-Sorting" counterfactual above. Note that the value is calculated assuming $w^R(\ell) = A(\ell)\underline{y}$ as before and this assumption is supported by adjusting $b(\ell)$. In particular, $b(\ell)$ is defined as in "No-Sorting" but with $F_\ell, (\lambda^U, \lambda^E)$ under "No-OJS". In turn, we use a given allocation $\mu(\ell)$ for computing F_ℓ. With $\lambda^U(\ell)$ for each ℓ in hand, we compute $\lambda^F(\ell) = \mathcal{A}^{\frac{1}{\alpha}}(\lambda^U(\ell))^{1-\frac{1}{\alpha}}$, as well as the value of a firm type p from sorting into ℓ (excluding the land price):

$$\delta\lambda^F(\ell)A(\ell)\int_{\underline{y}}^{\overline{y}}\int_{\underline{y}}^{y}\frac{1}{[\delta+\kappa\lambda^U(\ell)(1-\Gamma(t|\mu(\ell)))]^2}dtd\Gamma(y|p).$$

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

F.6.3 The Role of Spatial Frictions

Eliminate Worker Mobility Friction. We remove the preference heterogeneity across regions and set B to the average baseline $B(\ell)$. As in the previous counterfactuals, we first solve for $\lambda^U(\ell)$ for each ℓ so that the value of search for unemployed workers is equalized across space:

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

We again adjust $b(\ell)$, with F_ℓ which is the same as in baseline PAM equilibrium and (λ^U,λ^E) solved above. With λ^U in hand, we solve for $(\lambda^E(\ell),\lambda^F(\ell),L(\ell))$, and thus the new equilibrium.

Eliminate Firm Hiring Friction. If we assume the labor market is integrated so that the economy has a single job ladder, the model becomes similar to the basic wage posting model with the productivity distribution $\tilde{\Gamma}(t)=\int_{\underline{\ell}}^{\overline{\ell}}\Gamma\left(\frac{t}{A(\ell')}\middle|\mu(\ell')\right)dR(\ell')$. Employed workers accept a job offer if the new wage is higher than the current one and the wage function is increasing in z. The employment distribution becomes $\tilde{G}(z)=\delta\frac{\tilde{\Gamma}(z)}{\delta+\lambda^E\left(1-\tilde{\Gamma}(z)\right)}$.

The value of search does not depend on a specific location (i.e., is independent of ℓ) since there exists neither location preferences (B is equalized across ℓ) or differences in job opportunities (global job ladder). Thus, by construction, the value of search is the same everywhere. We therefore find the economy-wide meeting rates λ^E,λ^U so that the total population size from the baseline model is respected, $\bar{L}=\mathcal{A}^{\frac{1}{\alpha}}\frac{\delta+\lambda^U}{\delta+\kappa\lambda^U}(\lambda^U)^{-\frac{1}{\alpha}}$. Similar to the other counterfactual exercises, we adjust the unemployment flow benefit $b(\ell)=b$ (note that it is independent of ℓ) so that $w^R(\ell)=w^R=A(\underline{\ell})\underline{y}$. As the labor market is integrated, the reservation is also determined economy-wide, not region-specific.