Firm Sorting and Spatial Inequality:

Online Appendix

by Ilse Lindenlaub Ryungha Oh Michael Peters

- FOR ONLINE PUBLICATION -

Throughout this Online Appendix, we indicate figures, tables, and equations within this appendix by O.# ('O' for Online). In turn, figures, tables, and equations from the main paper are denoted by just 1,2,.... Figures, tables, and equations from the main appendix are denoted by A.#.

OA-1 Additional Theoretical Results and Proofs: Baseline Model

OA-1.1 The Case of Negative Sorting

OA-1.1.1 Proposition 1 for the Case of NAM

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 emerges if $\overline{J}(p,\ell)$ is strictly supermodular in (p,ℓ), sorting is negative if $\overline{J}(p,\ell)$ is strictly submodular.

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

Proposition O1 (Negative Spatial Sorting of Firms)**.** *If z is strictly submodular, and either the productivity gains from sorting into higher ℓ are sufficiently small, or the competition forces are sufficiently small (sufficiently small φ^E), then there is negative sorting of firms p to locations ℓ with $p=\mu(\ell)=Q^{-1}(1-R(\ell))$.*

Proof. To derive sufficient conditions for negative sorting in equilibrium, we can proceed similarly as in Appendix B.1. In order for $\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell}$ 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 positive measure of y. In turn, for this it is sufficient that

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

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

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

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

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

A sufficient condition for $\overline{J}$ to be submodular in (ℓ, p) is therefore $\varepsilon^N < 2\varphi^E t^N$. So, equilibrium sorting is negative either if ε^N is sufficiently small or if φ^E is sufficiently small. □

Remark. Note that under the sufficient conditions for negative sorting in Proposition O1, $\partial k/\partial \ell < 0$, as locations with higher ℓ are less attractive to firms. We show in the proof of Proposition 2 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. As a result, the individual rationality constraints for land owners are satisfied despite the decreasing land price schedule.

OA-1.1.2 Proposition 2 for the Case of NAM

If the conditions of Proposition O1 hold (negative sorting), then the proof of Proposition 2 is similar to the one under PAM in the text. 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 land 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, i.e.,

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

so that the land price is positive for each location.

OA-2 Theoretical Results and Proofs: Model Extensions

In this section, we discuss two extensions of our baseline model. In Section OA-2.1, we endogenize location productivity $A(\ell)$ by allowing for spillovers across firms. In Section OA-2.2, we allow firms to decide how many vacancies to post, which endogenizes the local job finding/filling rate. We derive sufficient conditions for *positive* sorting of firms across space; the case of negative sorting is similar and omitted for brevity.

OA-2.1 Endogenous Spillovers

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

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

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

Assumption O1. *Productivity in location ℓ is endogenous and given by* (OA-1).

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

Proposition 1 applies with minor modifications. For $\frac{\partial^2 \overline{J}(p,\ell)}{\partial p \partial \ell} > 0$ and hence for positive sorting to obtain, we can again unpack (13) from the baseline model and obtain the following sufficient condition:

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

The main change relative to the baseline model is that differences in location productivity are endogenous. In particular, under positive sorting in (p, ℓ), productivity A increases in ℓ because these locations have access to better firms: $\frac{\partial A(\ell)}{\partial \ell} = \int_{\underline{y}}^{\overline{y}} -\frac{\partial \Gamma_\ell(y)}{\partial \ell} dy > 0$. If this

location productivity advantage, along with the impact on firms' marginal productivity, is large enough relative to the cost of more severe poaching competition, highly productive firms (those with high-p) indeed settle into high-ℓ locations—similar to the baseline model.

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

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

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

Proposition O2. *Suppose that Assumption O1 holds. If z is strictly supermodular, and either the productivity gains from sorting into higher ℓ, ε^P, are sufficiently large, or the competition forces φ^E are sufficiently small, then there is positive sorting of firms p to locations ℓ.*

The proof of this result resembles the proof of Proposition 1 and is therefore omitted.

Next, we provide the existence result for this extension.

Proposition O3. *Suppose Assumption O1 holds along with the conditions from Proposition O2. 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 2). First, focus on the equilibrium with PAM. The postulated $\mu(\ell) = Q^{-1}(R(\ell))$—by uniquely pinning down Γ_ℓ—uniquely pins down the spillover function $A(\ell) = \int (1 - \Gamma_\ell(y))dy$. We then construct a fixed point in Γ_ℓ as in the baseline model. □

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

OA-2.2 Endogenous Vacancy Posting

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

With endogenous vacancy posting, meeting rates $\lambda^F(\ell)$ and $\lambda^E(\ell)$ depend on ℓ. We assume

that total meetings between workers and firms in location ℓ are given by

$$\mathcal{M}(\mathcal{V}(\ell),\mathcal{U}(\ell)) = \mathcal{A}\mathcal{V}(\ell)^{\alpha}\mathcal{U}(\ell)^{1-\alpha}, \tag{OA-3}$$

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

Assumption O2.

1. *Total meetings in location ℓ are given by (OA-3) with $0 < \alpha < 1$.*
2. *Vacancy posting cost c is C^2 with $c' > 0, c'' > 0$, $c'(0) = 0$, and $\lim_{v\to 0} \frac{vc''(v)}{c'(v)} := \underline{c} > 0$.*

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

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

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

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

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

$$\theta(\ell)\frac{\delta + \kappa\mathcal{A}\theta(\ell)^{\alpha}}{\delta + \mathcal{A}\theta(\ell)^{\alpha}} = \mathcal{V}(\ell). \tag{OA-4}$$

Equation (OA-4) implicitly determines the equilibrium local labor market tightness, $\theta(\ell)$, as a function of the measure of vacancies, $\mathcal{V}(\ell)$, in any given market ℓ. Note that, under Assumption O2.1., θ is strictly increasing in $\mathcal{V}(\ell)$. To see this, differentiate (OA-4) with respect to θ:

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

which is positive under Assumption O2.1. ($\alpha < 1$) and achieves its minimum (equal to δ^2) at $\theta = 0$.

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

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

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

Proposition O4. *If z is strictly supermodular, the productivity gains from sorting into higher ℓ are sufficiently large, and the competition forces are sufficiently small (i.e., $1/\delta$ is sufficiently small), then there is positive sorting of firms p to locations ℓ.*

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

$$\lambda^F(\ell)\hat{J}(p,\ell) = c'(v(p,\ell)). \tag{OA-5}$$

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

$$\begin{aligned}
\bar{J}(p,\ell) &= \lambda^F(\ell)v(p,\ell)\hat{J}(p,\ell) - c(v(p,\ell)) - k(\ell)\\
\frac{\partial\bar{J}(p,\ell)}{\partial p} &= \frac{\partial\hat{J}(p,\ell)}{\partial p}\lambda^F(\ell)v(p,\ell)\\
\frac{\partial^2\bar{J}(p,\ell)}{\partial\ell\partial p} &= \frac{\partial^2\hat{J}(p,\ell)}{\partial\ell\partial p}\lambda^F(\ell)v(p,\ell) + \frac{1}{c''(v(p,\ell))}\frac{\partial\lambda^F(\ell)\hat{J}(p,\ell)}{\partial\ell}\frac{\partial\hat{J}(p,\ell)}{\partial p}\lambda^F(\ell) + \frac{\partial\hat{J}(p,\ell)}{\partial p}\frac{\partial\lambda^F(\ell)}{\partial\ell}v(p,\ell).
\end{aligned} \tag{OA-6}$$

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 (OA-5) with respect to ℓ. We will characterize conditions under which (OA-6) is positive so that PAM between (p,ℓ) arises. To that end, we will specify conditions under which, at $p=\mu(\ell)$, $\frac{\partial^2\hat{J}(p,\ell)}{\partial\ell\partial p} > 0$ in the first term (Step 1) and the remaining two terms are also positive (Step 2). In each step, the conditions we specify will involve taking the limit $\frac{1}{\delta}\to 0$.

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

$$\hat{J}(p,\ell) = \frac{1}{\delta}\int_{\underline{y}}^{\overline{y}}\frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1+\frac{\lambda^E(\ell)}{\delta}(1-\Gamma_\ell(y))]^2}(1-\Gamma(y|p))dy. \tag{OA-7}$$

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

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

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

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

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

The main difference compared with the baseline model is the endogeneity of market tightness (and, thus, of $\lambda^E(\ell)$ and $\lambda^F(\ell)$). Therefore, it is important to understand the behavior of local market tightness. Note that, under PAM, $\mathcal{V}(\ell)=v(\mu(\ell),\ell)$, which follows from the definition of $\mathcal{V}(\ell)$. Based on (OA-4), we denote $v(\theta(\ell)):=v(\mu(\ell),\ell)$, where v is an increasing function of θ. Moreover, we have $\lim_{\theta\to0}v(\theta)=0$, $\lim_{\theta\to0}\frac{\partial v(\theta)}{\partial\theta}=1$, and $\lim_{\theta\to0}\frac{v(\theta)}{\theta}=\lim_{\theta\to0}\mathcal{U}(\ell)=1$. Using $\lambda^F(\ell)=\mathcal{A}\theta(\ell)^{\alpha-1}$ and FOC (OA-5), we have

$$\hat{J}(\mu(\ell),\ell)=c'(v(\theta(\ell)))(\mathcal{A}\theta(\ell)^{\alpha-1})^{-1}. \tag{OA-8}$$

If $\delta\to\infty$, then $\hat{J}(p,\ell)\to0$, which follows from the definition of $\widehat{J}$; see (OA-7). Since $c'(v(\theta))$ and $(\theta^{\alpha-1})^{-1}$ are both strictly increasing in θ and zero at $\theta=0$, we conclude that $\lim_{\delta\to\infty}\theta(\ell)=0$ and $\lim_{\delta\to\infty}v(\theta(\ell))=0$.

Differentiating (OA-8) with respect to ℓ we obtain (after some rearrangement) the elasticity of market tightness under PAM.

Plugging in the expressions for $\frac{\partial}{\partial\ell}\hat{J}(p,\ell)$, $\frac{\partial}{\partial p}\hat{J}(p,\ell)$, and $\hat{J}(p,\ell)$ from above gives

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

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

$$\left. + \quad \frac{\int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell}}{[1 + \frac{\lambda^E(\ell)}{\delta}(1 - \Gamma_\ell(y))]^2} (1 - \Gamma(y|\mu(\ell)))\, dy + \frac{\partial \mu(\ell)}{\partial \ell} \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1 + \frac{\lambda^E(\ell)}{\delta}(1 - \Gamma_\ell(y))]^2} \left(- \frac{\partial}{\partial p} \Gamma(y|\mu(\ell)) \right) dy}{\delta \mathcal{A}^{-1} \theta(\ell)^{2-\alpha} \frac{\partial v(\theta(\ell))}{\partial \theta} c''(v(\theta(\ell))) + (1-\alpha) \int_{\underline{y}}^{\overline{y}} \frac{\frac{\partial z(y,A(\ell))}{\partial y}}{[1 + \frac{\lambda^E(\ell)}{\delta}(1 - \Gamma_\ell(y))]^2} (1 - \Gamma(y|\mu(\ell)))\, dy} \right). \tag{OA-9}$$

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

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

Using (OA-8), this is $\delta \hat{J}(\mu(\ell), \ell) \frac{c''(v(\theta(\ell)))\theta(\ell)}{c'(v(\theta(\ell)))}$, where $\frac{c''(v(\theta(\ell)))\theta(\ell)}{c'(v(\theta(\ell)))}$ converges to $\underline{c}$ under Assumption O2.2. Thus, we can characterize the limit of the elasticity of market tightness:

$$\lim_{\delta \to \infty} \frac{\frac{\partial \theta(\ell)}{\partial \ell}}{\theta(\ell)} = \frac{\int_{\underline{y}}^{\overline{y}} \frac{\partial^2 z(y,A(\ell))}{\partial y \partial A(\ell)} \frac{\partial A(\ell)}{\partial \ell} (1 - \Gamma(y|\mu(\ell)))dy + \frac{\partial \mu(\ell)}{\partial \ell} \int_{\underline{y}}^{\overline{y}} \frac{\partial z(y,A(\ell))}{\partial y} \left(- \frac{\partial}{\partial p} \Gamma(y|\mu(\ell)) \right) dy}{[1 - \alpha + \underline{c}] \int_{\underline{y}}^{\overline{y}} \frac{\partial z(y,A(\ell))}{\partial y} (1 - \Gamma(y|\mu(\ell)))\, dy} \tag{OA-10}$$

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

We now return to our task of signing (OA-6).

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

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

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

This holds as $\delta \to \infty$ since the RHS vanishes (recall that we showed in (OA-10) that the elasticity of market tightness is bounded from above as $\delta \to \infty$) while the LHS remains strictly positive.

Step 2. Next, we show that the sum of the last two terms in (OA-6) is positive when

$p = \mu(\ell)$. After some algebra, they become

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

Evaluating this equation at $p = \mu(\ell)$ and using (OA-8), we obtain

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

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

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

where we used Assumption O2.2. Observing that both $\delta\hat{J}(\mu(\ell),\ell)$ and $\delta\frac{\partial J(p,\ell)}{\partial \ell}$ converge to some positive numbers as we consider the limit $\delta \to \infty$, the above inequality becomes in this limit

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

where we substituted in (OA-10). This holds if

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

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

From Steps 1 and 2, we conclude that $\frac{\partial^2 \bar{J}(p,\ell)}{\partial \ell \partial p}$ is positive along $p = \mu(\ell)$, which shows that PAM is indeed an equilibrium. □

The economic intuition for Proposition O4 is as follows. Labor market competition is strong in high-ℓ locations not only because there are better firms than in low-ℓ regions (due

to positive sorting—as in the baseline model), but also because more productive firms tend to post more vacancies. This new channel increases market tightness in good locations and hence further discourages firms from settling there. Hence, competition in productive locations is amplified by endogenous vacancy posting. To compensate for this stronger competition that arises from both firm composition and congestion, we require the productivity gains from settling into high-ℓ locations to be large enough or, stated differently, competition to be sufficiently muted (through low $1/\delta$), so that PAM can be sustained in equilibrium.

OA-3 Spatial Firm Sorting: Global vs. Local Rank

In this section we devise an additional test for the presence of firm sorting in the data. We show that firm sorting has distinct implications for the relationship between the local and the global (economy-wide) productivity rank of firms.

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

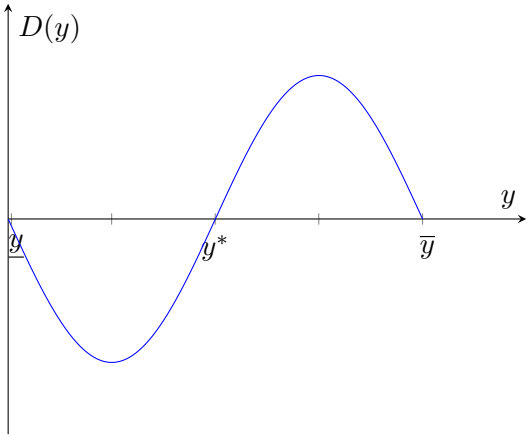
$$D(y) := \underbrace{\int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) r(\ell) d\ell}_{\text{Global Rank}} - \underbrace{\int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) \frac{\gamma(y|\mu(\ell))r(\ell)}{\int_{\underline{\ell}}^{\overline{\ell}} \gamma(y|\mu(\hat{\ell}))r(\hat{\ell})d\hat{\ell}}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

The global rank reflects the firm's position in the economy-wide productivity ranking. By contrast, the local rank reflects the firm's position in the productivity ranking of its location. It takes into account that firms of a given type y can be found in all locations but, because of sorting, they are more prevalent in some locations than others. We therefore average the local rank of firm type y, $\Gamma_\ell(y)$, across locations using the density that describes the distribution of y across space (see the proof of Proposition O5 for the detailed derivation of the local rank).

Spatial sorting by firms has specific implications for the shape of D. If sorting is monotone, there is a concentration of highly productive firms in some locations and of much less productive firms in others. As a consequence, the local rank of highly productive firms is *low* relative to their global rank, which yields $D > 0$. The opposite is true for the least productive firms who are surrounded by other low-productivity peers in their locations. As a result, their local rank tends to be *high* compared with their global rank, with $D < 0$. Finally, $D(\underline{y}) = D(\overline{y}) = 0$ because the worst (best) firm economy-wide is also the worst (best) firm in any local labor market. Note that the difference between global and local ranks is absent (i.e., $D(y) = 0$ for all y) if there is no firm sorting. Figure OA-1 depicts D for a parametric

example with spatial sorting.[39]

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



We now show that the shape depicted in Figure OA-1 is a robust feature of spatial firm sorting. To do so, we maintain the following regularity assumption.

Assumption O3. *Both $\gamma(\underline{y}|p)$ and $\gamma(\overline{y}|p)$ are not constant in p.*

We can then show the following results.

Proposition O5 (Firm Sorting and the Difference between Global and Local Productivity Ranks)**.** *Suppose Assumption O3 holds.*

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

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

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

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

Our definition of local rank reflects the *average* local rank of any given firm y: $\int_{\underline{\ell}}^{\overline{\ell}} \Gamma_\ell(y) n_\ell(\ell|y) d\ell$,

[39] Suppose that $R(\ell) = \frac{\ell - a}{b - a}$, $Q(p) = \frac{p - a}{b - a}$, and $\Gamma(y|p) = y^p$ for $b > a > 0$, $p \in [a, b]$ and $\ell \in [a, b]$. Thus, under PAM, $\mu(\ell) = \ell$ and $\Gamma_\ell(y|\mu(\ell)) = y^{\mu(\ell)} = y^\ell$. If $a = 1$ and $b = 2$, we can solve for the zeros of D in closed form, giving the unique interior zero at $y^* = 0.5$; see Figure OA-1. Note that this example does not satisfy Assumption O3 for $\gamma(\underline{y}|p)$, which however is only sufficient (not necessary) for the result.

where $n_\ell(\ell|y)$ is defined as the (endogenous) location density conditional on y,

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

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

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

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

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

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

Differentiate D wrt y to obtain

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

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

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

where Var_r is our notation for the variance of a random variable, taking land distribution r into account. Both expressions are *strictly negative* if $\text{Var}_r[\gamma(\underline{y}|\mu(\ell))] > 0$ and $\text{Var}_r[\gamma(\overline{y}|\mu(\ell))] > 0$, which is the case under Assumption O3.

Since D starts at zero and first decreases, it is strictly negative for small $y > \underline{y}$; and since it ends at zero in a decreasing manner, it must be that for high $y < \overline{y}$ it is strictly

positive. Hence, there must be at least one $y^* \in (\underline{y}, \overline{y})$ such that $D(y^*) = 0$ and at that point D crosses zero from below. If this interior crossing is unique, then $y^* = y^{**}$. In turn, if D has several interior zeros, then the first one, y^*, and the last one, $y^{**} > y^*$, share this 'crossing-from-below' property, proving the claim. □

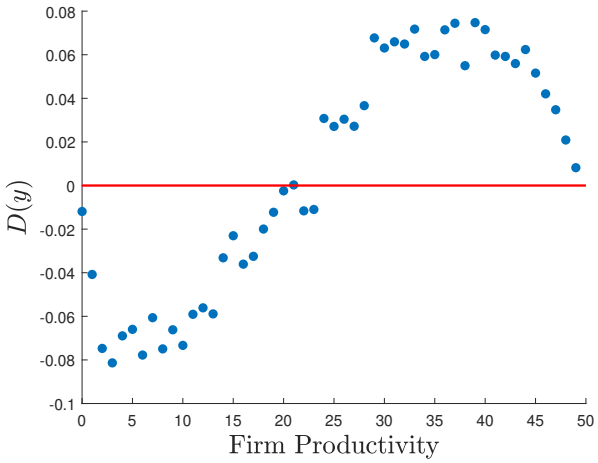
Evidence. To detect spatial sorting empirically, Proposition O5 and Figure OA-1 suggest a simple test: If there is monotone spatial sorting, there is an S-shaped relationship between the difference in firms' global and local ranks, $D(y)$, and productivity y. In contrast to our other "tests" of firm sorting in this appendix, this one does *not* rely on any parametric restriction on the local firm productivity distributions (i.e., we can dispense with the Pareto assumption).

Implementing this test in practice, requires a measure of firm productivity y. Measuring y empirically is complicated by the fact that firms—according our theory—are sorted spatially and, thus, their output per worker z depends not only on their productivity y but also on location productivity $A(\ell)$. To purge firm output per worker z from local productivity $A(\ell)$, we exploit the fact that under the assumption of common support, output per worker of the least productive firm in location ℓ is given by $\underline{y}A(\ell)$ and hence should only reflect $A(\ell)$. In practice, we therefore measure y as sales per worker divided by the 1% quantile of the sales per worker distribution in location ℓ.[40]

In Figure OA-2, we plot the relationship between $D(y)$ and y in the data. On the horizontal axis, we order firms by their global productivity rank and categorize them into 50 equally sized bins (based on percentiles of the global productivity distribution). On the vertical axis, we display the average of the difference between global and local ranks for each productivity bin. As in Figure OA-1, there is clear S-shape. Globally unproductive firms sort into locations with a high concentration of unproductive competitors. Hence, their global rank is below their average local rank, i.e., $D(y) < 0$. In turn, for globally productive firms, the opposite pattern arises: They co-locate with other productive firms—i.e., within their local labor market they are relatively unproductive compared to their economy-wide productivity—and therefore $D(y) > 0$. Recall that if there is no spatial firm sorting, we would observe that D is a horizontal line and zero everywhere.

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

Figure OA-2: Difference between Global and Local Productivity Rank



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

OA-4 Application: The West-East and Urban Wage Premium

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

In Table OA-1, we report wages for East-West and urban-rural labor markets in both data and model. The wage gap between West and East Germany is 28%. The urban-rural wage gap is slightly smaller but still substantial: On average, wages are almost 20% higher in urban labor markets. Our model matches these dimensions of spatial inequality quite well.

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

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

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

Table OA-2 reports the results of our no-sorting counterfactual applied to West-East and urban-rural inequality. Firm sorting again plays a quantitatively important role. Without it, the West-East wage gap in our model would be 19% rather than 22%. Similarly, the urban wage premium would decrease from 16% to 13%.

Table OA-2: West-East and Urban-Rural Inequality: No-Sorting Counterfactual

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

OA-5 Quantitative Results: Additional Robustness

OA-5.1 Relation of Labor Share and Firm Productivity: Log-Normal

We now demonstrate that the property of the negative relationship between local labor share and local firm productivity from our quantitative model does not hinge on the Pareto assumption on $\Gamma(y|p)$. To do so, we provide simulations for a common alternative functional form of firm productivity. Specifically, we assume that $\Gamma(y|p)$ follows a log-normal distribution, whereby $\log(y)$ follows a normal distribution with parameters $(p, 0.5^2)$ and, in equilibrium, $(\mu(\ell), 0.5^2)$. That is, firms with higher ex ante type (higher μ) have higher mean productivity ex post. In line with the discussion in the paper, we left-truncate this productivity distribution such that its density is sufficiently decreasing for all μ (we truncate all local productivity distributions at the median productivity of the worst location in order to maintain a common $\underline{y}$). Figure OA-3 shows that, as in the Pareto case, there is a negative relationship between local labor shares and firm ex ante productivity under the log-normal assumption. As a result, we back out an *increasing* matching function $\mu(\cdot)$ (indicating PAM between firms and locations) from the decreasing local labor share function $LS(\cdot)$ observed in the data.

To gain intuition for the negative relationship between local labor share and local firm productivity, Figure OA-4 plots for two locations—the top and bottom CZ—firm-level labor shares $Ls(\cdot, \ell)$ (left panel) and the cdf of (weighted) employment (right panel), informing the two terms $\frac{\partial Ls(y,\ell)}{\partial y}\left(-\frac{\tilde{G}_\ell(y)}{\partial \ell}\right)$ in the integrand of equation (A.10): In each location, firm-level labor shares decrease in productivity y; moreover, the weighted employment distribution of the top location is stochastically better (under PAM). Thus, a decreasing empirical local labor share function $LS(\cdot)$ calls—under decreasing firm-level labor shares $Ls(\cdot, \ell)$—for PAM so that in high ℓ more employment is concentrated in top firms with low labor shares.

Figure OA-3: Identifying Firm Sorting from Local Labor Shares: Truncated Log-Normal $\Gamma(y|p)$

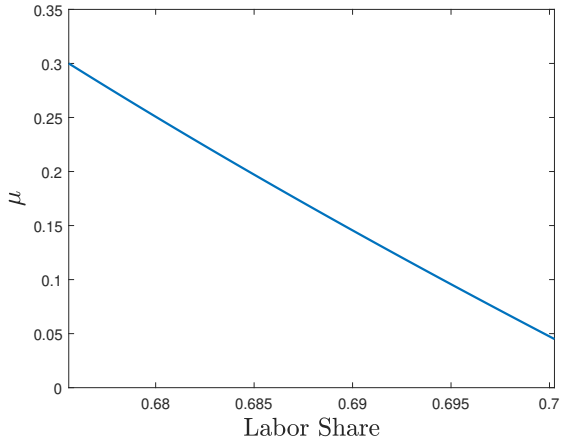
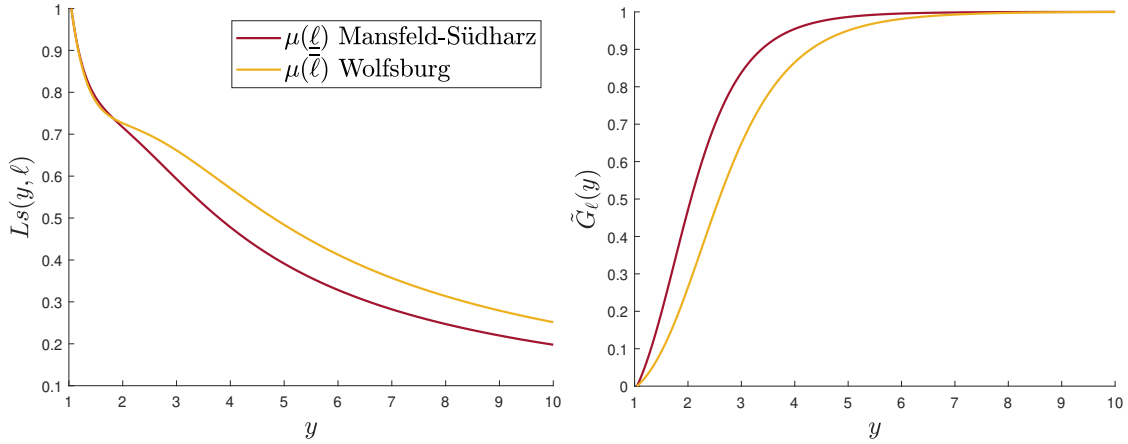


Figure OA-4: Firm-Level Labor Shares and Employment Distribution: Truncated Log-Normal $\Gamma(y|p)$



OA-5.2 Estimation Conditional on Industry

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

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

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

OA-5.3 Estimation on an Alternative Data Source (FDZ)

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

OA-5.4 Estimation of a Model with Imperfect Worker Mobility

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

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

$$\frac{L(\ell)}{\overline{L}} = \frac{(V^U(\ell))^\nu}{\sum_k (V^U(k))^\nu}. \tag{OA-11}$$

Choice probabilities (OA-11) replace the search value equalization condition from our baseline model. Dispersion parameter ν represents the elasticity of the local population with respect to the local value of search. In the limit, when ν goes to infinity, our model reduces to the baseline case, in which migration frictions are absent.

Estimation. We pin down parameter ν by computing the elasticity of the local population size with respect to local average wages, which is as a function of ν, leveraging relevant elasticity estimates available in the literature. Using the value of unemployed workers in

equation (20), the elasticity of the local population size with respect to local average wages can be approximated as (ignoring a general equilibrium constant)

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

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

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

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

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

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

Counterfactual. To compute the counterfactual without firm sorting, i.e., $\Gamma_\ell(y) = \Gamma(y)$, we find the schedules $\{L(\ell), \lambda^U(\ell), d(\ell)\}$ that satisfy (OA-11), which replaces the condition that equalizes the value of search across locations in the baseline model. The other equilibrium conditions remain the same as in our baseline model. Table 3 summarizes our main counterfactual based on this model extension.

OA-5.5 Estimation of a Model with Endogenous Firm Selection

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

Setup. We assume that firms and workers can observe productivity y only when they meet. Since the value of employment increases with wages, and wages increase with productivity,

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

By combining $A(\ell)\underline{y}(\ell) = w^R(\ell)$ with reservation wage equation (A.18) and wage equation (5), we obtain the following equation that implicitly defines cutoff $\underline{y}(\ell)$:

$$A(\ell)\underline{y}(\ell) = \widehat{b} + (1-\kappa)\varphi^U(\ell) \int_{\underline{y}(\ell)}^{\overline{y}} (1-\Gamma_\ell(t)) 2\varphi^E(\ell)\gamma_\ell(y) \int_{\underline{y}(\ell)}^{\overline{y}} \frac{A(\ell)}{(1+\varphi^E(\ell)(1-\Gamma_\ell(t)))^2} dt dy. \tag{OA-12}$$

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

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

$$\text{jfr}(\ell) = \lambda^U(\ell)(1-\Gamma_\ell(\underline{y}(\ell))) \tag{OA-13}$$

$$u(\ell) = \frac{\text{jfr}(\ell)}{\text{jfr}(\ell)+\delta(\ell)}. \tag{OA-14}$$

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

$$\lambda^U(\ell) = \mathcal{A}(u(\ell)+\kappa(1-u(\ell)))^{-1/2} \left(\frac{\bar{l}(\ell)}{1-u(\ell)}\right)^{-1/2}, \tag{OA-15}$$

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

$$\mathbb{E}[z(y,\ell)|\ell] = A(\ell)\frac{1}{1-\Gamma_\ell(\underline{y}(\ell))} \int_{\underline{y}(\ell)}^{\overline{y}} y g_\ell(y) dy, \tag{OA-16}$$

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

Identification. When the cutoff, $\underline{y}(\ell)$, is greater than $\underline{y}$ (which is the case we focus on), we can identify the firm type $p = \mu(\ell)$ that settled in location ℓ from its labor share in the same way as in the baseline model. For the remaining parameters, first note that we can

express $\lambda^U(\ell)$ as a function of $\underline{y}(\ell)$ by combining (OA-13), (OA-14), and (OA-15)

$$\lambda^U(\ell) = \mathcal{A}\left(\frac{\delta(\ell) + \kappa\lambda^U(\ell)(1-\Gamma_\ell(\underline{y}(\ell)))}{\lambda^U(\ell)(1-\Gamma_\ell(\underline{y}(\ell)))}\bar{l}(\ell)\right)^{-1/2}. \tag{OA-17}$$

We then jointly identify $(\underline{y}(\ell), \lambda^U(\ell), A(\ell))$ for each ℓ using three equations, (OA-12), (OA-16) and (OA-17), along with separation rate $\delta(\ell)$, average firm size $\bar{l}(\ell)$, mean value added $\mathbb{E}[z(y,\ell)|\ell]$, the overall matching efficiency $\mathcal{A}$, and the relative matching efficiency κ from the baseline estimation. And we recover job finding rate from (OA-13).

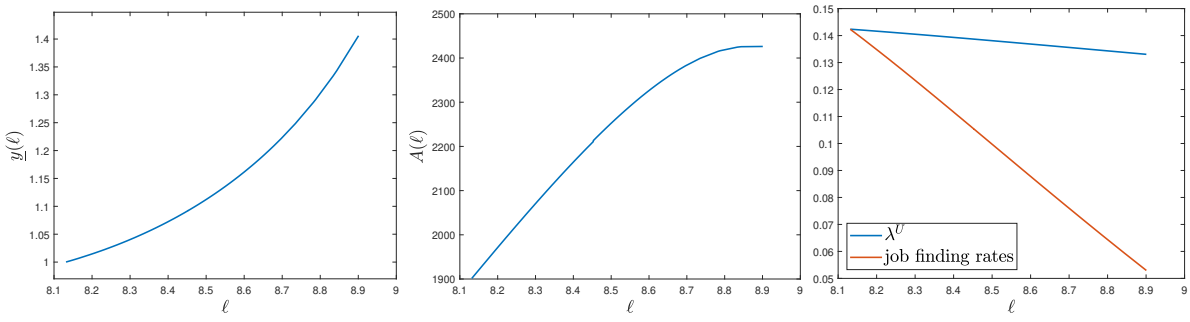
Because unemployed workers have positive income $\widehat{b}$, we assume that the government no longer provides subsidies, and all workers spend a fraction ω of their income on housing. With this assumption, we can estimate local amenities $B(\ell)$ and housing supply $h(\ell)$ using two equations: (i)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

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

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

Results. We report the estimation results of the main model objects graphically, focusing on the changes compared to the baseline model. The left panel of Figure OA-5 shows the key new object, productivity cutoff $\underline{y}(\ell)$. It is increasing in ℓ, which indicates that workers in prosperous places are more selective, also leading to a declining job finding rate in ℓ (see orange line, right panel). In turn, estimated local TFP is still increasing in ℓ (middle panel).

Figure OA-5: Estimation Results



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