Spatial Firm Sorting and Local Monopsony Power: Supplementary Appendix

by Ilse Lindenlaub Ryungha Oh Michael Peters

- FOR ONLINE PUBLICATION -

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

SA.1 Baseline Model: Derivations

SA.1.1 Alternative Formulation of Wage-Posting Problem

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

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

where $h(w,\ell)$ is the hiring rate of a firm posting w in location ℓ, and $J(y,w,\ell)$ is firm y's discounted flow profit when posting w in that location.[54] Using firm size expression (SA.2) (Appendix SA.1.2), we obtain (4).

SA.1.2 Firm Size

Firm size can be derived in two ways. First, we interpret the model's firm size as the product of the hiring rate and the expected duration of a match, which coincides with expression (3):

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

[54] The hiring rate of firm y in location ℓ is $h(w,\ell) \equiv \lambda^F \left(\frac{\lambda^U u(\ell)}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))} + \frac{\lambda^E(1-u(\ell))}{\lambda^U u(\ell)+\lambda^E(1-u(\ell))} E_\ell(w) \right)$, considering that a firm meets workers at rate λ^F from two pools: unemployment $u(\ell)$ (they will always accept the job), and employment $1-u(\ell)$ (they will accept if the new wage is higher than their current one). We denote the steady-state employment distribution by E_ℓ, where $E_\ell(w) = \delta \frac{F_\ell(w)}{\delta+\lambda^E(1-F_\ell(w))}$ (see (11) and (12)), so that $h(w,\ell)$ reduces to the expression in (SA.1).

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

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

SA.1.3 Wage Posting

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

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

And so,

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

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

Then:

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

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

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

where $\tilde{J}(\underline{y};\ell) = l(w(\underline{y},\ell))(z(\underline{y},A(\ell)) - w^R(\ell))$.

Imposing Assumption 1.2 (zero profits of the least productive firm type in each location, $\tilde{J}(\underline{y},\ell)=0$) as well as $\rho = 0$ (as stated in footnote 8), we obtain wage function (5) from (SA.4).

SA.1.4 Land Price Schedule

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

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

The FOC reads

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

Solving this differential equation, when evaluated at the equilibrium assignment, yields land price schedule k.

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

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

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

SA.1.5 Land Market Clearing

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

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

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

SA.2 Additional Theoretical Results and Proofs: Baseline Model

SA.2.1 Proposition 1 for the Case of Negative Sorting

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SA.3 Theoretical Results and Proofs: Model Extensions

In this appendix, we discuss several extensions of our baseline model. In Section SA.3.1, we provide the proof of Proposition A2, which deals with the case of labor mobility and housing. In Section SA.3.2, we endogenize location productivity $A(\ell)$ by allowing for spillovers across firms. In Section SA.3.3, we allow firms to decide how many vacancies to post, which endogenizes the local job finding/filling rate. In section SA.3.4, we consider endogenous land supply. In all these extensions, we derive sufficient conditions for *positive* sorting of firms across space; the case of negative sorting is similar and omitted for brevity.

SA.3.1 Labor Mobility and Residential Housing: Proof of Proposition A2

We here present the proof of Proposition A2, stated in Appendix B.

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

The expected value for firm p from settling in location ℓ is given by

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

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

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

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

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

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

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

and then compute its derivatives:

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

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

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

Plugging these derivatives into (SA.5), we can write (SA.5) as a single integral. Then, a sufficient condition for (SA.5) to be positive (i.e., a sufficient condition for $\overline{J}(p,\ell)$ to be supermodular in (p,ℓ)) is that this integrand is positive for all $y \in [\underline{y}, \overline{y}]$ and strictly so for a set of y of positive measure. Using $-\frac{\partial \Gamma(y|p)}{\partial p} \geq 0$, we obtain the following sufficient condition for PAM:

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

Define ε^P as the minimum of the LHS (as in the baseline model). It is strictly positive under our assumptions and the premise. Under labor mobility, the RHS depends on endogenous market tightness $\theta(\ell)$ through meeting rates $(\lambda^F(\ell), \lambda^E(\ell))$. Thus, the sufficient conditions for PAM from the baseline model are not readily applicable. Instead, we argue that the RHS is bounded. Thus, (SA.6) holds for a large enough ε^P, made precise below. We proceed in 3 steps.

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

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

Using the government budget constraint,

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

the housing market clearing condition,

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

the local population size (for a derivation, see (A.19) in Appendix E)

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

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

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

Denote by

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

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

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

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

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

We will show that the first line is positive while the second line is negative. However, for large enough ξ, the second line becomes sufficiently small, rendering the overall expression positive.

To see that the first line of (SA.8) is positive under the premise, denote firm y's local productivity rank by $\mathcal{R} = \Gamma_\ell(y)$. We apply a change of variables to wage function (5) (with $\Gamma_\ell(t) = x,\ \gamma_\ell(t)dt = dx$) and take the cross-partial derivative wrt $(\mathcal{R}, \ell)$:

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

Suppose that $\Gamma_\ell^{-1}(\mathcal{R})$ is increasing in ℓ (which is true if $\frac{\partial}{\partial \ell}\Gamma_\ell \leq 0$). In addition, suppose that, for any given λ^E such that $\underline{\lambda}^E \leq \lambda^E \leq \overline{\lambda}^E$ with $\underline{\lambda}^E = \min_\ell \lambda^E(\ell)$ and $\overline{\lambda}^E = \max_\ell \lambda^E(\ell)$, the integrand of (SA.9), $\frac{\partial z(y, A(\ell))}{\partial y} / (1 + \frac{\lambda^E}{\delta}(1-\Gamma_\ell(y)))^2$, is also increasing in ℓ. Both of these statements are true under the sufficient conditions for PAM that we provide below, so that the wage function is supermodular in $(\mathcal{R}, \ell)$. This ensures that the first line of (SA.8) is positive.

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

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

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

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

Importantly, taking the same limit of the first line of (SA.8) shows that it remains positive:

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

Thus, by continuity of V^U in ξ, there exists a finite $\widehat{\xi}$ such that for $\xi > \widehat{\xi}$, the positive effect stemming from the first line of (SA.8) dominates the negative effect stemming from the second line, which renders (SA.8) positive. Thus, the value of unemployment is increasing in ℓ for a *fixed* λ^U (and thus a fixed $\lambda^E = \kappa\lambda^U$) if housing supply elasticity ξ is sufficiently large.

Step 2. A similar argument shows that for large enough ξ, V^U is increasing in λ^U since its positive effect on $\left(\frac{\partial z}{\partial A}\frac{\partial A(\ell)}{\partial \ell}+\lambda^E\int_0^1\frac{1-\mathcal{R}}{\delta+\lambda^E(1-\mathcal{R})}\frac{\partial^2 w(\mathcal{R},\ell)}{\partial\ell\partial\mathcal{R}}d\mathcal{R}\right)$ dominates its (ambiguous) effect on $\tilde{d}(\ell)^{-\omega/(1+\xi)}$ in (SA.7). Denote the level of the housing supply elasticity for which this is (weakly) true by $\tilde{\xi}$, and so for $\xi > \tilde{\xi}$, V^U is increasing in λ^U. Going forward we assume that $\xi > \max\{\widehat{\xi}, \tilde{\xi}\}$, consistent with our premise that the housing supply elasticity is "large enough".

Step 3. This discussion implies that for the equilibrium indifference condition of searching workers to hold (i.e., the value of unemployment, V^U, is equalized across ℓ), it must be that λ^U (and thus λ^E) is decreasing in ℓ, and so θ is decreasing in ℓ while λ^F is increasing in ℓ. This renders the second and third term on the RHS in (SA.6) negative.

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

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

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

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

Then, (SA.6) holds if

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

which holds if

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

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

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

□

SA.3.2 Endogenous Spillovers

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SA.3.3 Endogenous Vacancy Posting

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

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

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

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

Assumption SA2.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

SA.3.4 Endogenous Land Supply

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

only that the matching function now depends on ψ^*. In order for this expression to be (strictly) positive, it suffices that the integrand is positive for all $y \in [\underline{y}, \overline{y}]$ and strictly so for some set of y of positive measure. So, it suffices that for all (e, ℓ, ψ^*)

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

A sufficient condition for this inequality to hold is:

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

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

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

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

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

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

SA.4 Additional Predictions and Evidence on Firm Sorting

We now provide additional results, which can be used to detect firm sorting in the data and complements our analysis on local labor shares. We first show that spatial firm sorting increases productivity dispersion in high-ℓ locations (Proposition SA5 and Corollary SA1). Second, we show that firm sorting has testable implications for the relationship between the local and the global (economy-wide) productivity rank of firms (Proposition SA6). We provide empirical support for both.

SA.4.1 Spatial Firm Sorting: Local Productivity Dispersion

Theory. We first show that positive firm sorting also implies that high-ℓ locations have more productivity dispersion, captured by the quantile ratio $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$ (where $t', t'' \in (0,1)$ and $t'' > t'$), which is increasing in ℓ. This result applies to productivity distributions $\Gamma(y|p)$ in which stochastic dominance wrt p is more pronounced for higher y.

Proposition SA5 (Firm Sorting & Local Productivity Dispersion)**.** *If there is positive firm sorting across space, then the quantile ratio of local productivity, $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$, is increasing in ℓ, provided that the elasticity of $(-\Gamma_p/\Gamma_y)$ with respect to y exceeds 1.*

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

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

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

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

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

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

This holds if:

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

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

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

Corollary SA1 (Firm Sorting & Local Productivity Dispersion: Pareto Case)**.** *If and only if there is positive firm sorting across space, then both the quantile ratio of local productivity, $\Gamma_\ell^{-1}(t'')/\Gamma_\ell^{-1}(t')$, and the quantile difference of the log value added distribution, $\Pi_\ell^{-1}(t'') - \Pi_\ell^{-1}(t')$ are increasing in ℓ (where we denote by $\Pi_\ell(z)$ the cdf of log value added $\log(z)$).*

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

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

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

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

And therefore

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

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

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

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

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

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

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

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

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

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

It follows that

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

□

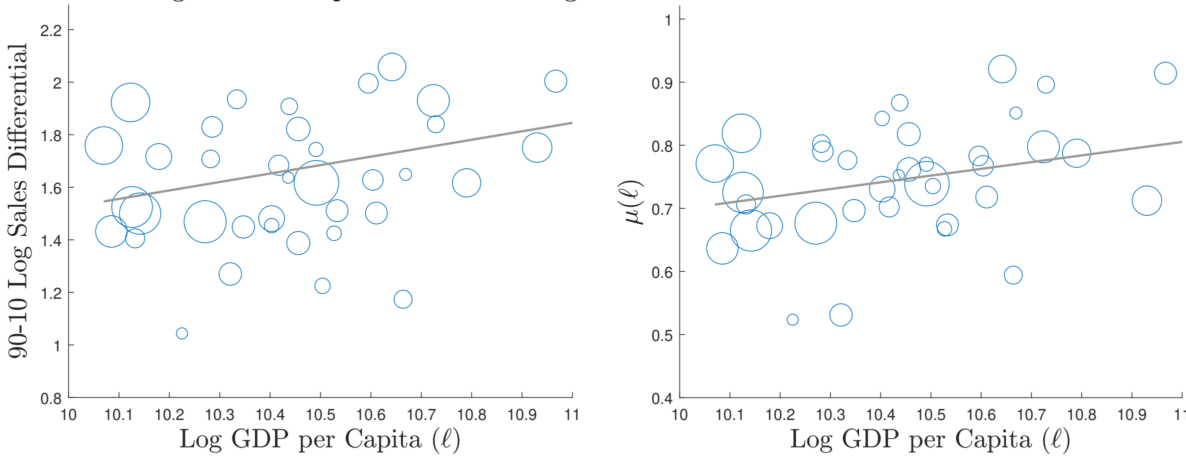
Firm-Level Evidence. We provide additional evidence on positive firm sorting using firm-level productivity indicators. We consider this analysis as only supplementary to our evidence based on local labor shares since the firm-level productivity data is based on the Establishment Panel, which has a relatively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

Within-Location Dispersion of Productivity and Sales. Corollary SA1 (Appendix SA.4.1) suggests a test of positive firm sorting based on how the local dispersion of (log) output per worker varies across space. If and only if sorting is positive, then high-ℓ locations are characterized by more dispersion in output per worker.

When assessing this prediction we measure output per worker at the firm level by sales per worker.[55] Figure SA.1 (left) plots the difference of the 90% and 10% quantile of the distribution of log sales per worker. Based on Corollary SA1, the positive relationship between sales dispersion and ℓ indicates positive firm sorting across space.

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

Figure SA.1: Spatial Firm Sorting: Evidence from Establishment Panel



Notes: Data source: Establishment Panel. The left panel shows a scatter plot between the log difference of the 90th and 10th quantile of firm sales per worker against local log GDPpc. We compute the 90th and 10th percentiles using frequency weights, where we weigh each firm observation by the number of firms in the same size class (see footnote 56). In the right panel, we plot $-1/\beta_\ell$ from regression (SA.20), where $z(y, A(\ell))$ in the dependent variable is measured as sales per worker. For each location ℓ, quantile k is taken from the local firm productivity distribution, where we use the same frequency weights as those in the left panel. Coefficient, β_ℓ, is weighted by the number of firms in each NUTS2 region. The size of the markers indicates the size of the region (number of firms in each NUTS2 region).

Pareto Tails of Firm Productivity. When assessing firm sorting based on the spatial variation in local labor shares or in the dispersion of sales per worker, we implicitly assume that firm productivity y in each ℓ follows a Pareto distribution with shape parameter $1/p$ (which in equilibrium becomes $1/\mu(\ell)$), see Corollaries 1 and SA1. Positive sorting of firms across locations means that μ is increasing and thus richer locations have a thicker Pareto tail of the local productivity distribution. To assess this prediction, we proxy firm productivity by sales per worker and estimate the Pareto shape parameter at the NUTS2 regional level by implementing the following regression at the local level

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

where $k = z^{(1)}, z^{(2)}, \cdots, z^{(n_\ell - 1)}, z^{(n_\ell)}$, and $(z^{(j)}, n_\ell)$ are the j-th order statistics and the number of firms in region ℓ, respectively.[56]

Under the assumptions of multiplicative technology and Pareto productivity distributions as well as the validity of our productivity proxy, regression coefficient β_ℓ captures the Pareto shape parameter $\frac{1}{\mu(\ell)}$. The R^2 of these regional Pareto regressions varies between 0.7 and 0.9, which suggests that the Pareto assumption is reasonable. Furthermore, the positive slope of the

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

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

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

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

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

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

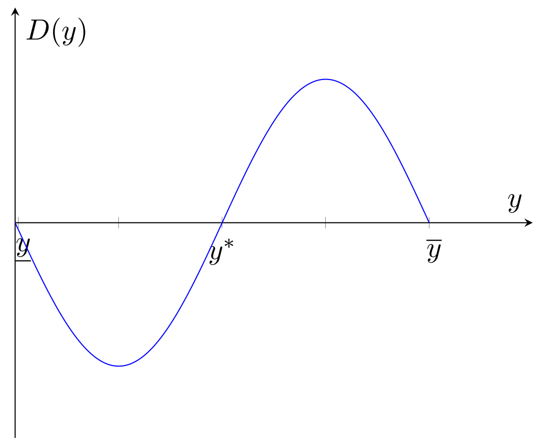
The global rank reflects the firm's position in the economy-wide productivity ranking. By contrast, the local rank reflects the firm's position in the productivity ranking of its location. It takes into account that firms of a given type y can be found in all locations but, because of sorting, they are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

Spatial sorting by firms has specific implications for the shape of D. If sorting is monotone, there is a concentration of highly productive firms in some locations and of much less productive firms in others. Thus, the local rank of highly productive firms is *low* relative to their global rank, which yields $D > 0$. The opposite is true for the least productive firms who are surrounded by other low-productivity peers in their locations. As a result, their local rank tends to be *high* compared with their global rank, with $D < 0$. Finally, $D(\underline{y}) = D(\overline{y}) = 0$ because the worst (best) firm economy-wide is also the worst (best) firm in any local labor market. Note that the difference between global and local ranks is absent (i.e., $D(y) = 0$ for all y) if there is no firm sorting. Figure SA.2 depicts D for a parametric example with spatial sorting.[57]

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

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

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



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

We can then show the following results.

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

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

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

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

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

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

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

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

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

case

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

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

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

Differentiate D wrt y to obtain

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

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

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

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

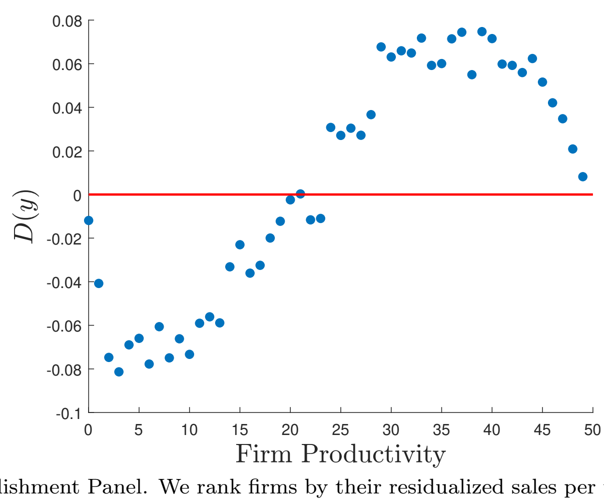
Since D starts at zero and first decreases, it is strictly negative for small $y > \underline{y}$; and since it ends at zero in a decreasing manner, it must be that for high $y < \overline{y}$ it is strictly positive. Hence, there must be at least one $y^* \in (\underline{y}, \overline{y})$ such that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

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

Implementing this test in practice, requires a measure of firm productivity y. Measuring y empirically is complicated by the fact that firms—according our theory—are sorted spatially and, thus, their output per worker z depends not only on their productivity y but also on location productivity $A(\ell)$. To purge firm output per worker z from local productivity $A(\ell)$, we exploit the fact that under the assumption of common support, output per worker of the least productive firm in location ℓ is given by $\underline{y}A(\ell)$ and hence should only reflect $A(\ell)$. In practice, we therefore measure y as sales per worker divided by the 1% quantile of the sales per worker distribution in location ℓ.[58]

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

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



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

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

SA.5 Characteristics of Local Labor Markets

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

Table SA.1: On-the-Job Search and Local Labor Markets

	Mean	S.D.	P10	P25	P50	P75	P90
Firm level (N = 5,958)							
Poaching Share	0.51	0.13	0.35	0.44	0.52	0.60	0.63
Share of local EE	0.70	0.17	0.47	0.62	0.73	0.81	0.88
Share of local UE	0.56	0.22	0.31	0.42	0.54	0.70	0.83
Commuting-zone level (N = 252)							
Poaching Share	0.49	0.05	0.42	0.46	0.48	0.52	0.54
Share of local EE	0.69	0.09	0.57	0.64	0.70	0.76	0.79
Share of local UE	0.58	0.11	0.45	0.52	0.60	0.66	0.69

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

Table SA.2: Spatial Heterogeneity: Distribution of Key Statistics

	Mean	S.D.	P10	P25	P50	P75	P90
Average Wages	3,133	401	2,616	2,849	3,093	3,364	3,662
Average Value Added	4,640	687	3,903	4,202	4,523	4,872	5,518
Average Firm Size	11	2	9	10	11	12	13
Share Emp. Top 10%	0.56	0.06	0.49	0.52	0.55	0.59	0.63
Population Density	292	422	83	110	165	272	589
Population	317,149	420,183	92,979	127,139	190,745	325,078	596,006

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

SA.6 Counterfactuals and Policy Exercise: Technical Details

SA.6.1 The No-Sorting Counterfactual

We adjust $\tilde{b}(\ell)$ so that the reservation wage in each ℓ remains the same as in the baseline model, i.e., $w^R(\ell) = A(\ell)\underline{y}$, see (A.25). We also keep the estimated schedules (A, B, h) from the baseline model. But, without spatial firm sorting, F_ℓ (and thus Γ_ℓ), (λ^U, λ^E) and d all differ from the baseline model.

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

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

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

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

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

SA.6.2 The Role of Endogenous Firm Sorting

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

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

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

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

where the value of search is again calculated assuming $w^R(\ell) = \tilde{A}(\ell)\underline{y}$, supported by adjusting $\tilde{b}(\ell)$;[59] and where we set the new value of search, ρV^U, to achieve consistency with the total population size from the baseline economy, $\bar{L} = \int L(\ell)dR(\ell)$. Based on unemployed workers' welfare equalization, we obtain $\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)$ and therefore $\lambda^E(\ell;\kappa,\Gamma_\ell,\rho V^U) = \kappa\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U)$. With λ^U for each ℓ in hand, we can also compute $\lambda^F(\ell;\kappa,\Gamma_\ell,\rho V^U) = \mathcal{A}^{\frac{1}{\alpha}}(\lambda^U(\ell;\kappa,\Gamma_\ell,\rho V^U))^{1-\frac{1}{\alpha}}$, as well as the *match value* of a firm type p and location ℓ,

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

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

SA.7 Quantitative Results: Additional Robustness

SA.7.1 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.

[59] In particular, $\tilde{b}(\ell)$ is defined as in (A.25) but using $\tilde{A}(\ell)$, F_ℓ (which we can compute based on the postulated Γ_ℓ) and $(\lambda^U(\ell), \lambda^E(\ell))$ obtained above.

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

SA.7.2 Estimation on an Alternative Data Source (FDZ)

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

SA.7.3 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. As in the baseline model, we assume that firms must first purchase one unit of land and then search for workers. Upon receiving job offers, unemployed workers accept them only when productivity exceeds an endogenous region-specific cutoff $\underline{y}(\ell) \in [\underline{y}, \overline{y})$. Note that a firm of productivity $\underline{y}(\ell)$ makes zero profit, and $A(\ell)\underline{y}(\ell) = w^R(\ell)$. We further assume that the unemployment benefit equals $\widehat{b} > 0$, which we choose such that $\underline{y}(\underline{\ell}) = \underline{y}$.

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

$$A(\ell)\underline{y}(\ell) = \widehat{b} + (1-\kappa)\varphi^U(\ell) \int_{\underline{y}(\ell)}^{\overline{y}} (1-\Gamma_\ell(t)) 2\varphi^E(\ell)\gamma_\ell(y) \int_{\underline{y}(\ell)}^{\overline{y}} \frac{A(\ell)}{(1+\varphi^E(\ell)(1-\Gamma_\ell(t)))^2} dtdy. \quad \text{(SA.21)}$$

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:

$$\text{jfr}(\ell) = \lambda^U(\ell)(1 - \Gamma_\ell(\underline{y}(\ell))) \tag{SA.22}$$

$$u(\ell) = \frac{\text{jfr}(\ell)}{\text{jfr}(\ell) + \delta(\ell)}. \tag{SA.23}$$

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

$$\lambda^U(\ell) = \mathcal{A}(u(\ell) + \kappa(1 - u(\ell)))^{-1/2} \left(\frac{\bar{l}(\ell)}{1 - u(\ell)}\right)^{-1/2}, \tag{SA.24}$$

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

$$\mathbb{E}[z(y,\ell)|\ell] = A(\ell)\frac{1}{1 - \Gamma_\ell(\underline{y}(\ell))}\int_{\underline{y}(\ell)}^{\overline{y}} y g_\ell(y)dy, \tag{SA.25}$$

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

Identification. When the cutoff, $\underline{y}(\ell)$, is greater than $\underline{y}$ (which is the case we focus on), we can identify the firm type $p = \mu(\ell)$ that settled in location ℓ from its labor share in the same way as in the baseline model. For the remaining parameters, first note that we can express $\lambda^U(\ell)$ as a function of $\underline{y}(\ell)$ by combining (SA.22), (SA.23), and (SA.24)

$$\lambda^U(\ell) = \mathcal{A}\left(\frac{\delta(\ell) + \kappa\lambda^U(\ell)(1 - \Gamma_\ell(\underline{y}(\ell)))}{\lambda^U(\ell)(1 - \Gamma_\ell(\underline{y}(\ell)))}\bar{l}(\ell)\right)^{-1/2}. \tag{SA.26}$$

We then jointly identify $(\underline{y}(\ell), \lambda^U(\ell), A(\ell))$ for each ℓ using three equations, (SA.21), (SA.25) and (SA.26), along with separation rate $\delta(\ell)$, average firm size $\bar{l}(\ell)$, mean value added $\mathbb{E}[z(y,\ell)|\ell]$, the overall matching efficiency $\mathcal{A}$, and the relative matching efficiency κ from the baseline estimation. And we recover job finding rate from (SA.22).

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

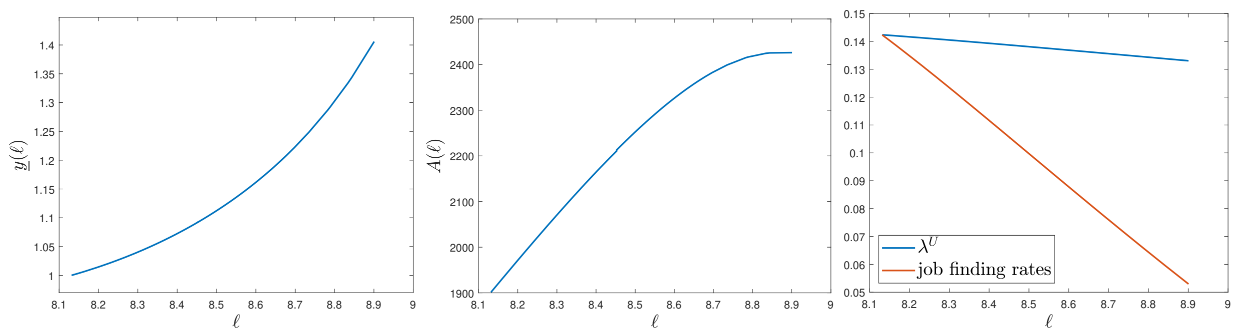
(21) to endogenize firm selection, and (ii) housing market clearing. They are respectively given by:

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

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

Results. We report the estimation results of the main model objects graphically, focusing on the changes compared to the baseline model. The left panel of Figure SA.4 shows the key new object, productivity cutoff $\underline{y}(\ell)$, which increases in ℓ. This pattern indicates that workers are more selective in prosperous locations, where high-productivity firms are concentrated. As a result, the job finding rate decreases in ℓ (see orange line, right panel). Despite the more pronounced changes in firm composition across space, estimated local TFP is still increasing in ℓ (middle panel).

Figure SA.4: 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.

SA.7.4 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:

$$\frac{L(\ell)}{\overline{L}} = \frac{(V^U(\ell))^\nu}{\sum_k (V^U(k))^\nu}. \tag{SA.27}$$

Choice probabilities (SA.27) 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 (21), the elasticity of the local population size with respect to local average wages can be approximated as (ignoring a general equilibrium constant)

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

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

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

In the literature, the elasticity of migration flows with respect to income commonly ranges between 2 and 4 (e.g., Allen and Donaldson, 2020). We are interested in long-term effects, so we take a value at the higher end. Since we are considering long-run migration, we assume an elasticity of 4, which implies $\nu = 5.42$. Finally, we pin down amenity schedule $B(\cdot)$ by matching the spatial population distribution in the data. We do so by taking the ratio of the population size of two regions and plugging in the value of search (21), which yields the following equation:

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

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

Counterfactual. To compute the counterfactual without firm sorting, i.e., $\Gamma_\ell(y) = \Gamma(y)$, we find the schedules $\{L(\ell), \lambda^U(\ell), d(\ell)\}$ that satisfy (SA.27), 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.