Firm Sorting and Agglomeration, Cecile Gaubert
Online appendix

A. Housing market

Since there is a fixed total supply of land equal to 1 in the city, the housing supply equation is

$$H(L) = \big(\frac{p_H(L)}{w(L)}\big)^{\frac{1-b}{b}}, \tag{A.24}$$

where $H(L)$ is the total quantity of housing supplied in a city of size L.

The aggregate local demand for housing is

$$H(L) = \frac{(1-\eta)w(L)L}{p_H(L)}. \tag{A.25}$$

Equations (A.24) and (A.25) pin down prices and quantities of housing produced. Housing supply (equation (A.24)) and demand (equation (A.25)) equate so that $p_H(L) = (1-\eta)^b\, w(L)\, L^b$. This yields the following labor use and profits in the housing sector:

$$\ell_H(L) = (1-b)(1-\eta)L, \quad \text{and} \tag{A.26}$$
$$\pi_H(L) = b(1-\eta)Lw(L). \tag{A.27}$$

The housing supply elasticity is given by $\frac{d\log H(L)}{d\log p_H(L)} = \eta\frac{1-b}{b}$. Anticipating on the policy discussion, note that a decrease in b increases the housing supply elasticity and also leads to a flatter wage schedule across city sizes, as $\frac{d\log w(L)}{d\log L} = b\frac{1-\eta}{\eta}$.

B. Proofs of section 3.3

B.1. Lemma 1

Proof Fix the sector s. Since $\pi(z,L,s)$ is strictly LSM in (z,L), if follows that for all $z_1 > z_2$ and $L_1 > L_2$, $\frac{\pi(z_1,L_1,s)}{\pi(z_1,L_2,s)} > \frac{\pi(z_2,L_1,s)}{\pi(z_2,L_2,s)}$. So if z_2 has higher profits in L_1 than in L_2, so does z_1. Necessarily, $L_j^*(z_1) \geq L_j^*(z_2)$, and $L_j^*(z)$ is a non-decreasing function. ∎

Moreover, in the case where the support of city sizes $\mathcal{L}$ is convex, then $L_j^*(z)$ is a strictly increasing function. Since the set of z is convex by assumption, and $\psi(z,L,s)$ is such that the profit maximization problem is concave for all firms, the optimal set of city sizes is itself convex. It follow that $L_j^*(z)$ is invertible. It is locally differentiable (using in addition that $\psi(z,L,s)$ is differentiable), as the implicit function theorem applies and $\frac{dL_j^*(z)}{dz} = -\frac{\frac{\partial(\frac{\psi_2 L}{\psi})}{\partial z}(z,L_j^*(z),s)}{\frac{\partial(\frac{\psi_2 L}{\psi})}{\partial L}(z,L_j^*(z),s)}$.

B.2. Equilibrium when $\mathcal{L}$ is not convex

I describe here for completeness an equilibrium when the set $\mathcal{L}$ is not convex. Consider a non convex set of city sizes $\mathcal{L}$ that I write it as a union of intervals on $\mathbb{R}^+$. : $\widetilde{\mathcal{L}} = \bigcup_{i\sim odd}[a_i, a_{i+1}]$. This nests in particular the case of a discrete number of city sizes. I focus on the case where these intervals closed, but the proof is similar if some intervals are open. Consider $[a_1,a_2]$ and $[a_3,a_4]$ with $a_3 < a_4$ two such intervals, without any city available in-between. Consider firms whose unconstrained city choice would fall between a_1 and a_4, which correspond to a closed interval: $[z_1,z_4] = L_j^{*-1}([a_1,a_4])$ (it is well defined, given that $\widetilde{L_j}$ is continuous and invertible). Write $z_i = L_j^{*-1}(a_i)$. By construction, for all $z \in [z_1,z_2]\bigcup[z_3,z_4]$, we get that $L_j^*(z) = \widetilde{L_j}(z)$, hence $L_j^*(z)$ is increasing on

$[z_1, z_2]$ and $[z_3, z_4]$ respectively. Then, pick $z \in (z_2, z_3)$. We know that $L_j^*(z_2) \le L_j^*(z) \le L_j^*(z_3)$ since $L_j^*(z)$ is non decreasing, hence $a_2 \le L_j^*(z) \le a_3$. Since $L^*(z) \in \mathcal{L}$, this means that $L_j^*(z) = a_2$, or $L_j^*(z) = a_3$. By monotonicity of $L_j^*(z)$, there exists a threshold $\tilde{z} \in (z_2, z_3)$ such that if $z \in (z_2, \tilde{z})$, $L_j^*(z) = a_2$ and if $z \in (\tilde{z}, z_3)$ $L_j^*(z) = a_3$. Firms "bunch" at the city size closest to their optimal unconstrained choice (either the one to the left or to the right), with a higher-z firm choosing a city size at least as large as a lower-z firm. This bunching preserves the monotonicity of the matching function $L_j^*(.)$.

B.3. Proposition 2

Proof Fix j. For productivity, the results comes from the facts that (1) $L_j^*(z)$ is non decreasing in z and (2) that $\psi(z, L, s_j)$ is increasing in L. Revenues are proportional to profits ($r_j * (z) = \frac{\sigma_j}{1+T_j^*}\pi_j^*(z)$). The proof for profits is as follows. $\psi(z_H, L_L, s_j) > \psi(z_L, L_L, s_j)$ as ψ is increasing in z, which leads to $\pi(z_H, L_L) > \pi(z_L, L_L)$, as firms face the same wage in the same city. Finally, $\pi(z_H, L_H, s_j) \ge \pi(z_H, L_L, s_j)$ as L_H is the profit maximizing choice for z_H. Therefore, $\pi(z_H, L_H, s_j) > \pi(z_L, L_L, s_j)$. ∎

In addition, $\epsilon_l = \epsilon_r - (1-\alpha)\frac{1-\eta}{\eta}$.

Proof For a given city size L and a given sector j, $\bar{r_j^*}(L) = \sum_{z\ in\ L} r_j^*(z) \propto \sum_{z\ in\ L} w(L)\ell_j^*(z) \propto w(L)\bar{\ell_j^*}(L)$, where their proportion is constant across city sizes. Therefore $\frac{d\log\ \bar{\ell_j^*}(L)}{d\log\ L} = \frac{d\log\ \bar{r_j^*}(L)}{d\log\ L} - \epsilon_w$, where the elasticity of wages with respect to city sizes is $\epsilon_w = b\frac{\eta}{1-\eta}$ ∎

B.4. Proposition 3

Proof Let $\mathcal{Z} : \mathcal{L}$ x A x $E \rightarrow Z$ be the correspondence that assigns to any $L \in \mathcal{L}$ and $\alpha \in A$ a set of z that chooses L at equilibrium. (It is a function when $\mathcal{L}$ is convex (see proof of Lemma 1).) Define $\bar{z}(L, \alpha, s) = \max_z\{z \in \mathcal{Z}(L, \alpha, s)\}$ as the maximum efficiency level of a firm that chooses city size L in a sector characterized by the parameters (α, s). I will use the following lemmas:

Lemma 11 *$log\,\pi$ is supermodular with respect to the triple* (z, L, α)

It is readily seen that: $\frac{\partial^2 log\,\pi(z,L,\alpha,s)}{\partial z \partial L} > 0$, $\frac{\partial^2 log\,\pi(z,L,\alpha,s)}{\partial z \partial \alpha} = 0$ and $\frac{\partial^2 log\,\pi(z,L,\alpha,s)}{\partial L \partial \alpha} = \frac{(\sigma-1)b(1-\eta)}{\eta L} > 0$. This result does not rely on an assumption on the convexity of $\mathcal{L}$. Checking the cross partials are sufficient to prove the supermodularity even if L is taken from a discrete set, as π can be extended straightforwardly to a convex domain, the convex hull of $\mathcal{L}$.

Lemma 12 $\bar{z}(L, \alpha, s)$ *is non decreasing in* α, s.

The lemma is a direct consequence of the supermodularity of $log\,\pi$ with respect to the quadruple (z, L, α, s). Using a classical theorem in monotone comparative statics, if $log\,\pi(z, L, \alpha, s)$ is supermodular in (z, L, α, s), and $L^*(z, \alpha, s) = \max_L log\,\pi(z, L, \alpha, s)$ then $(z_H, \alpha_H, s_H) \ge (z_L, \alpha_L, s_L) \Rightarrow L^*(z_H, \alpha_H, s_H) \ge L^*(z_L, \alpha_L, s_L)$. Note that everywhere, the $\ge$ sign denotes the lattice order on $\mathcal{R}^3$ (all elements are greater or equal than).

Coming back to the proof of the main proposition, we can now write:

$$\begin{aligned}\tilde{F}(L; \alpha, s) &= P(\text{firm from sector}(\alpha, s) \text{ is in a city of size smaller that } L)\\ &= F(\bar{z}(L, \alpha, s))\end{aligned}$$

where $F(.)$ the the raw efficiency distribution of the firms in the industry. Let $\alpha_H > \alpha_L$.

For any $z \in Z$, the previous lemma ensures that $L^*(z, \alpha_H, s) \ge L^*(z, \alpha_L, s)$. In particular, fix a given L and s and write using shorter notation: $\bar{z}_{\alpha_L} = \bar{z}(L, \alpha_L, s)$. Then $L^*(\bar{z}_{\alpha_L}, \alpha_H, s) \ge L^*(\bar{z}_{\alpha_L}, \alpha_L, s) = L$. Because $L^*(z, \alpha_H, s)$ is increasing in z, it follows that:

$$z \in \mathcal{Z}(L, \alpha_H, s) \Rightarrow z \le \bar{z}_{\alpha_L}$$

and therefore $\bar{z}_{\alpha_H} \leq \bar{z}_{\alpha_L}$ or using the long notation: $\bar{z}(L, \alpha_H, s) \leq \bar{z}(L, \alpha_L, s)$

It follows that $F(\bar{z}(L, \alpha_H)) \leq F(\bar{z}(L, \alpha_L))$ and that $F(L; \alpha, s)$ is decreasing in α. This completes the proof of the first order stochastic dominance of the geographic distribution of a high α sector vs that of a lower α.

The proof is exactly the same for the comparative statics in s, we just have to verify that $\pi(z, L, s)$ is log supermodular in (z, L, s). Since $\pi(z, L, s_j) = \kappa \left(\frac{\psi(z, L, s_j)}{w(L)^{1-\alpha}} \right)^{\sigma-1} \frac{R_j}{P_j^{1-\sigma}}$ and $w(L)$ doesn't depend on s, $\pi(z, L, s)$ directly inherits the log supermodularity of $\psi(z, L, s)$ in its parameters. ∎

B.5. Proposition 4

Proof Within sectors, the revenue function $r_j^*(z)$ at the sorting equilibrium is an increasing function for any j. Let $p_1 < p_2 \in (0, 1)$. Under the assumption, maintained throughout the comparative statics exercise, that sectors draw z from the same distribution, there $\exists\ z_1 < z_2$ such that $Q_{j_1}(p_1) = r_{j_1}^*(z_1)$ and $Q_{j_2}(p_1) = r_{j_2}^*(z_1)$ (same thing for z_2 and p_2), ie. the quantiles of the r_{j1}^* and r_{j2}^* distributions correspond to the same quantile of the z distribution. This yields $\frac{Q_{j_1}(p_2)}{Q_{j_1}(p_1)} = \frac{r_{j1}^*(z_2)}{r_{j1}^*(z_1)}$, and $\frac{Q_{j_2}(p_2)}{Q_{j_2}(p_1)} = \frac{r_{j2}^*(z_2)}{r_{j2}^*(z_1)}$.

Finally, it is a classic result in monotone comparative statics (Topkis (1998)) that if $\pi(z, L, \alpha)$ is log-supermodular in (z, L, α), then $\pi^*(z, \alpha) = \max_L \pi(z, L, \alpha)$ is log supermodular in (z, α), or $\frac{\pi_{j2}^*(z_2)}{\pi_{j2}^*(z_1)} \geq \frac{\pi_{j1}^*(z_2)}{\pi_{j1}^*(z_1)}$. Revenues are proportional to profits within sectors, which completes the proof. The same proof applies for s. ∎

B.6. Corollary 5

Proof Let $p_j \in (0, 1)$ be a threshold above which the distribution is well approximated by a Pareto distribution in sector j, and r_j the corresponding quantile of the distribution. The distribution of r conditional on being larger than r_j is:

$$\forall r > r_j,\ H_j(r \,|\, r \geq r_j) \approx 1 - (\frac{r}{r_j})^{-\zeta_j},$$

where ζ_j is the shape parameter of the Pareto distribution for sector j. Thus, if $F_j(r) = p$, one can write:

$$\forall p > p_j, \quad p \approx F_j(r_j) + H_j(r) \approx p_j + 1 - (\frac{r}{r_j})^{-\zeta_j}$$
$$\frac{r}{r_j} \approx (1 + p_j - p)^{-\frac{1}{\zeta_j}}$$

Letting $p_0 = \max(p_1, p_2)$ and writing $r_j = Q_j(p_0)$ for $j = 1, 2$, and using proposition (4) gives:

$$\frac{Q_{j1}(p)}{Q_{j1}(p_0)} \leq \frac{Q_{j2}(p)}{Q_{j2}(p_0)}$$
$$(1 + p_0 - p)^{-\frac{1}{\zeta_1}} \leq (1 + p_0 - p)^{-\frac{1}{\zeta_2}} \qquad \text{for all } p > p_0 \text{ and } p < 1$$
$$\zeta_1 \geq \zeta_2,$$

where the last inequality comes from $1 + p_0 - p \in (0, 1)$. ∎

C. Extensions of the model

C.1. Extension with costly trade

In this extension of the model, I consider an economy with a more realistic geography. Call $\mathcal{C}$ the set of sites where firms and workers can locate. To ship goods from site i to site $j \in \mathcal{C}^2$, firms incur an iceberg trade cost τ_{ij}. To simplify the exposition, I consider an economy with only one sector and where firms only use labor as an input. Extension to the cases with several sectors and the use of capital is straightforward. In the presence of trade costs, local price indexes P_i for the traded good produced by firms are not equalized between cities and depend on the whole distribution of firms across space:

$$P_i = \left[\int_{j \in \mathcal{C}} \int_{z \in \mathcal{Z}(j)} \left(\frac{\tau_{ji} w_j}{\psi(z, L_j)} \right)^{1-\sigma} dF_j(z) dj \right]^{\frac{1}{1-\sigma}}.$$

In this expression, the set $\mathcal{Z}(j)$ is the (endogenous) set of firms that are located on site j, and $F_j(z)$ the corresponding distribution of firm types. It is convenient to define the average city-level productivity $\bar{\psi}_j$ for any city j:

$$\bar{\psi}_j = \left[\int_{z \in \mathcal{Z}(j)} \psi(z, L_j)^{\sigma-1} dF_j(z) \right]^{\frac{1}{\sigma-1}}$$

We can then rewrite the price index simply as:

$$P_i = \left[\int_{j \in \mathcal{C}} \left(\frac{\tau_{ji} w_j}{\bar{\psi}_j} \right)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}}. \tag{C.28}$$

A firm of type z located in site i has marginal costs $\frac{\tau_{ij} w_i}{\psi(z, L_i)}$ when serving city j. This firm's demand from city j (where total demand is $w_j L_j$ and demand across goods is CES) is therefore:

$$r_{ij}(z) = \left(\frac{\tau_{ij} w_i}{\psi(z, L_i)} \right)^{1-\sigma} w_j L_j P_j^{\sigma-1}$$

Firms' profits, if located in i, are therefore $\pi(z, i) = \frac{1}{\sigma} \int_{j \in \mathcal{C}} \left(\frac{\tau_{ij} w_i}{\psi(z, L_i)} \right)^{1-\sigma} w_j L_j P_j^{\sigma-1} dj$. Define the city i's market access as:

$$MA_i = \int_{j \in \mathcal{C}} \tau_{ij}^{1-\sigma} w_j L_j P_j^{\sigma-1} dj. \tag{C.29}$$

Then firm's profits are simply:

$$\pi(z, i) = \frac{1}{\sigma} \psi(z, L_i)^{\sigma-1} w_i^{1-\sigma} MA_i.$$

From this expression, we see already that $\pi(z, i)$ is log-supermodular in z and city size L_i. Therefore, for a given equilibrium distribution of wages, market access and city sizes, more productive firms necessarily choose larger cities L_i: there is positive assortative matching between firm type and city size.[1]

Furthermore, city size L is still a sufficient statistic for the economic condition of a city, like in the version without trade costs. To show this, we first use the free mobility condition. The utility of a worker in city i , defined in equation (2) of the main text combined with the housing production equation (3), is:

$$U = U_i = \kappa_0 \left(\frac{w_i}{P_i} \right)^{\eta} L_i^{-b(1-\eta)}, \tag{C.30}$$

where $\kappa_0 = \eta^{-\eta} (1 - \eta)^{-b(1-\eta)}$ is an economy-wide constant. Using the free mobility condition and plugging in the

[1] Proof:Assume that it was not the case, that is that there are two firms $z_1 < z_2$ that choose city i_1 and i_2 with $L(i_1) > L(i_2)$. This means, by revealed preferences, that: $\frac{\pi(z_1, L(i_1))}{\pi(z_1, L(i_2))} > 1$. Now, by log-supermodularity of ψ :

$$\frac{\psi(z_2, L(i_1))}{\psi(z_2, L(i_2))} > \frac{\psi(z_1, L(i_1))}{\psi(z_1, L(i_2))}$$

Taking this to the power $\sigma - 1$ and multiplying both sides by the positive number $w_{i1}^{1-\sigma} MA_{i1} / w_{i2}^{1-\sigma} MA_{i2}$ leads to:

$$\frac{\pi(z_2, L(i_1))}{\pi(z_2, L(i_2))} > \frac{\pi(z_1, L(i_1))}{\pi(z_1, L(i_2))} > 1$$

Therefore i_2 cannot be the optimal choice of firm z_2. This proves that firms choose cities whose size is increasing with z.

expression for the price index lead to:

$$w_i^{1-\sigma} L_i^{-b(1-\sigma)\left(\frac{1-\eta}{\eta}\right)} = \widetilde{U}^{-1} \int_{j\in\mathcal{C}} (\tau_{ji} w_j)^{1-\sigma}\, \bar{\psi}_j^{\sigma-1} dj, \tag{C.31}$$

where the economy-wide constant $\widetilde{U}$ is defined by $\widetilde{U}^{-1} = (\kappa_0/U)^{(\sigma-1)/\eta}$. Second, the goods market clearing condition writes:

$$w_i L_i = \int_j \left(\frac{\tau_{ij} w_i}{\bar{\psi}_i}\right)^{1-\sigma} w_j L_j {P_j}^{\sigma-1} dj,$$

hence, using the expression for the price index implicitly given by (C.30) and simplifying, it follows that:

$$w_i^\sigma L_i \bar{\psi}_i^{1-\sigma} = \widetilde{U}^{-1} \int_j {\tau_{ij}}^{1-\sigma} {w_j}^\sigma L_j^{1-b(\sigma-1)\frac{1-\eta}{\eta}} dj \tag{C.32}$$

This system of 2N equations (C.31)-(C.32) corresponds to the one in Allen and Arkolakis (2014), where the congestion force is ${L_i}^{-b\frac{1-\eta}{\eta}}$, and the local productivities are for now taken to be fixed at $\bar{\psi}_i$. Therefore, for a given vector $\bar{\psi}_i$, and assuming that trade costs are symmetric ($\tau_{ij} = \tau_{ji}$), we can invoke theorem 2 in Allen and Arkolakis (2014) to show that there exists a unique vector of L_i and w_i such that this system of equation holds and that, the following holds for some endogenous constant Γ: $w_i^\sigma L_i \bar{\psi}_i^{1-\sigma} = \Gamma w_i^{1-\sigma} L_i^{-b(1-\sigma)\left(\frac{1-\eta}{\eta}\right)}$. This can be rewritten as:

$${w_i}^{2\sigma-1} {L_i}^{1+b\frac{1-\eta}{\eta}(1-\sigma)} {\bar{\psi}_i}^{1-\sigma} = \Gamma \tag{C.33}$$

Recombining equations lead to the following expression for market access:

$$MA_i = w_i^\sigma L_i \bar{\psi}_i^{1-\sigma} = \Gamma w_i^{1-\sigma} L_i^{-b(1-\sigma)\left(\frac{1-\eta}{\eta}\right)}.$$

Firm profits are therefore:

$$\pi(z,i) = \frac{\Gamma}{\sigma} \left(\frac{\psi(z,L_i) L_i^{b\left(\frac{1-\eta}{\eta}\right)}}{w_i^2}\right)^{\sigma-1}.$$

It follows from this expression that in equilibrium, two cities with the same size L cannot have different wages w - otherwise, firms that choose a city of that size L would only go to the city with the lowest wage. Furthermore, equilibrium wages must be increasing function of city size $w(L)$, since firm profits are increasing in L but decreasing in w (no firm would choose a city with a lower L if it came with a higher wage). In turn, the local price index P_i can be simply expressed, in equilibrium, a function of city size by (C.30), using the fact that wages are (in equilibrium) a function of local population, $w(L)$.

Despite the introduction of costly trade in the model, it is still the case that, in equilibrium, price and wages are a function of city size only, that is : city size is again a sufficient statistic to describe the equilibrium in terms of firms and consumer choices.

Furthermore, firms profits can therefore be written:

$$\pi(z,L,s) = \frac{\Gamma}{\sigma} \left(\frac{\psi(z,L,s) L^{b\left(\frac{1-\eta}{\eta}\right)}}{w(L)^2}\right)^{\sigma-1}.$$

from which it is readily seen that the profit function is log-supermodular in (z, L, s) and (z, L, α).The proofs of lemma 1 and propositions 2, 3 and 4 - which are based on this property - carry through to that case.

C.2. Extension with imperfect sorting

I examine the properties of the model in the presence of imperfect sorting as hypothesized in the empirical specification of section 4. The properties of equilibrium described in section 2.3. of the main text either hold true on average, rather than systematically, in that case, or are unchanged.

Set up with imperfect sorting Productivity in city size L is given by equation (20) in the main text. The idiosyncratic productivity shocks $\epsilon_{i,L}$ for each city size are i.i.d. across city sizes and firms and distributed as a type-I extreme value, with mean zero and variance ν_R. Therefore, writing ψ the non-stochastic part of firm's productivity $\psi(z_i, L, s_j) = \exp\left(a_j \log L + \log(z_i)(1 + \log \frac{L}{L_o})^{s_j}\right)$ leads to the following expression for firm i's profit:

$$\pi_j(z_i, L) = \kappa_{1j} \left(\frac{\psi(z_i, L, s_j) e^{\epsilon_{i,L}}}{L^{b\frac{1-\eta}{\eta}(1-\alpha_j)}} \right)^{\sigma_j - 1} R_j \, P_j^{\sigma_j - 1}.$$

It will prove useful to write $V_j(z, L)$ the non-stochastic part of firm profits:

$$V_j(z, L) = \kappa_{1j} \left(\frac{\psi(z, L, s_j)}{L^{b\frac{1-\eta}{\eta}(1-\alpha_j)}} \right)^{\sigma_j - 1} R_j \, P_j^{\sigma_j - 1}, \tag{C.34}$$

and to note that the multiplicative random term $e^{(\sigma_j - 1)\epsilon_{i,L}}$ is distributed Frechet, with shape parameter $\frac{\nu_{R_j}}{\sigma_j - 1}$. The firm's discrete choice problem is then:

$$L_j^*(z_i) = \underset{L \in \mathcal{L}}{\arg\max} \quad V_j(z_i, L) e^{(\sigma_j - 1)\epsilon_{i,L}}$$

Given this setup, the following characterizations of the equilibrium hold in the case of imperfect sorting.

Characterizations with imperfect sorting First, Lemma 1 in the main text states that, within each sector, the matching function is increasing: high-z firms are *systematically* found in larger cities than lower-z firms. With imperfect sorting, we can prove a related result:

Lemma 1': Take $z_1 < z_2$. Within a given sector, the distribution of city sizes for z_2-firms first order statistically dominates the one for z_1 firms. That is, defining $F(.|z)$ the CDF of the distribution of city sizes chosen by firms of type z :

$$z_1 < z_2 \Rightarrow F(L \mid z_2) \leq F(L \mid z_1)$$

High-z firms are *more likely* to be found in large cities than lower-z firms.

Proof The firm seeks to maximize profits. Given the properties of the Frechet distribution, the probability that a firm of type z in sector j chooses city size L is:

$$p(L|z) = \frac{V(z, L)^{\frac{\nu_R}{\sigma - 1}}}{\sum_{L'} V(z, L')^{\frac{\nu_R}{\sigma - 1}}} \tag{C.35}$$

Since $\psi(z, L, s_j)$ is LSM in (z, L), $V(z, L)^{\frac{\nu_R}{\sigma - 1}}$ is LSM in (z, L). In turn, $\frac{p(L_2|z_2)}{p(L_1|z_2)} = \left(\frac{V(z_2,L_2)}{V(z_2,L_1)}\right)^{\frac{\nu}{\sigma-1}} > \left(\frac{V(z_1,L_2)}{V(z_1,L_1)}\right)^{\frac{\nu}{\sigma-1}} = \frac{p(L_2|z_1)}{p(L_1|z_1)}$. This means that $p(L|z)$ has the monotone likelihood ratio property (Milgrom (1981); Costinot (2009)). Hence, in particular, the distribution of L for a high z first-order stochastically dominates the distribution of L for a low z. Furthermore, it follows that $p(z|L_2)$ first order stochastically dominates $p(z|L_1)$ when $L_2 > L_1$, because:

$$\frac{p(z_2, L_2)}{p(z_2, L_1)} = \frac{p(L_2|z_2) f(z_2)}{p(L_1|z_2) f(z_2)} = \frac{p(L_2|z_2)}{p(L_1|z_2)} > \frac{p(L_2|z_1)}{p(L_1|z_1)} = \frac{p(z_1, L_2)}{p(z_1, L_1)}.$$

Within a sector, firms' raw efficiencies z are higher in larger cities in the sense of first order stochastic dominance. ∎

Second, proposition p:obs of the main text states that, within each sector, firm profits, revenues and productivities increase in equilibrium with city size. With imperfect sorting, the results hold true on average over firms located in a given city:

Proposition p:obs': Within each sector, average firm profits, revenue and productivity increase in equilibrium with city size.

Proof A firm chooses city size to maximize profits following (21) in the main text. Given the properties of the Frechet distribution, the distribution of realized profits $\pi^*$of a firm of type z, once it has optimally chosen its location, is independent on its endogenous choice of city size L^*. That is, $F(\pi^*|z) = F(\pi^*|z, L^*)$. Therefore, the distribution of profits of firms located in a given city of size L^* is given by:

$$\begin{aligned} p(\pi^*|L^*) &= \int_z p(\pi^* \mid L^*, z) p_j(z|L^*)\, dz \\ &= \int_z p(\pi^* \mid z) p_j(z|L^*)\, dz \end{aligned}$$

Therefore, firm profits in city L are on average:

$$\begin{aligned} E\left[\pi^*|L^*\right] &= \int_{\pi^*} \int_z \pi^* p(\pi^* \mid z) p(z|L^*)\, d\pi^*\, dz \\ &= \int_z \left[\int_{\pi^*} \pi^* p(\pi^* \mid z) d\pi^* \right] p(z|L^*)\, dz \\ &= \int_z E\left[\pi^* \mid z\right] p(z|L^*)\, dz \end{aligned}$$

We know that $\pi(z, L)$ increases in z for all L, therefore $E(\pi^* \mid z)$ is also increasing in z. Given that $p(z|L_2)$ first order stochastically dominates $p(z|L_1)$ if $L_2 > L_1$, it follows that:

$$E\left[\pi^*|L_2^*\right] > E\left[\pi^*|L_1^*\right]$$

Firms are monopolistically competitive and demand is CES, so that profits are a constant proportion of revenues within a sector. It follows that:

$$E\left[r^*|L_2^*\right] > E\left[r^*|L_1^*\right]$$

Finally, the productivity of firm i in its chosen city size L^* is $\phi_i = \psi(z_i, L^*, s_j) e^{\epsilon_{i,L^*}}$, where:

$$\pi(z_i, L^*) = \kappa_{1j} \left(\frac{\phi_i}{L^{*(1-\alpha_j)\frac{1-\eta}{\eta}b}} \right)^{\sigma_j - 1} R_j\, P_j^{\sigma_j - 1}$$

therefore

$$\phi_i = \frac{\pi(z_i, L^*)^{\frac{1}{\sigma_j - 1}}\ L^{*(1-\alpha_j)\frac{1-\eta}{\eta}b}}{(\kappa_{1j} R_j)^{\frac{1}{\sigma_j - 1}}\ P_j}$$

Then, within sector j :

$$E\left[\phi|L^*\right] = E\left[\pi^*(z, L^*)^{\frac{1}{\sigma_j - 1}}\, |L^*\right] \frac{L^{*(1-\alpha_j)\frac{1-\eta}{\eta}b}}{(\kappa_{1j} R_j)^{\frac{1}{\sigma_j - 1}}\ P_j}$$

The term $E\left[\pi^*(z, L^*)^{\frac{1}{\sigma_j - 1}}\, |L^*\right]$ increases with L^*, by argument similar to the one made above for $E\left[\pi^*(z, L^*)|L^*\right]$. Furthermore, $L^{*(1-\alpha_j)\frac{1-\eta}{\eta}b}$ also increases with L^*. Therefore, $E\left[\phi|L^*\right]$ increases with L^* . ∎

Third, proposition p:sorting of the main text states that the geographic distribution of a high α (resp. s) sector first-order stochastically dominates that of a lower α (resp. s) sector. This statement is unchanged in the case of imperfect sorting.

Proposition p:sorting': The geographic distribution of a high α (resp. s) sector first-order stochastically dominates that of a lower α (resp. s) sector.

Proof I make here explicit the dependency of V on sectoral parameters s and α, and write expression (C.34) as $V(z,L,s,\alpha)$. We know that $V(z,L,s,\alpha)$ is log-supermodular (LSM) in (z,L,s,α), as the properties of the non stochastic part of productivity $\psi(z,L,s)$ are the same than in the main text. Therefore, $V(z,L,s,\alpha)^{\frac{\nu_R}{\sigma-1}}$ is also LSM in (z,L,s,α). For any $t \geq 0$, define the auxiliary function $\mathbb{1}_{[0,t]}(L)$ equal to 1 if $L \in [0,t]$ and 0 otherwise. This function is LSM in (z,L,s,α,t), by lemma 3 in Athey (2002). Define

$$G(z,s,\alpha,t) = \int_0^\infty V(z,L,s,\alpha)^{\frac{\nu_R}{\sigma-1}} \mathbb{1}_{[0,t]}(L)\, dF_L(L) = \int_0^t V(z,L,s,\alpha)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L).$$

where $F_L(L)$ is the economy-wide city size distribution. By lemma 4 in Athey (2002), we get that $G(z,s,\alpha,t)$ is LSM in (z,s,α,t). The probability that a firm of a given type z chooses a city size smaller than t is:

$$p\left(\text{firm z chooses city size } L \leq t|\alpha,s\right) = \frac{\int_0^t V(z,L,\alpha,s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}{\int_0^\infty V(z,L,\alpha,s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}$$

By log-supermodularity of $\int_0^t V(z,L,s,\alpha)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)$ the following comparative statics follow if $s \leq s'$:

$$\frac{\int_0^t V(z,L,\alpha,s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}{\int_0^\infty V(z,L',\alpha,s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)} \leq \frac{\int_0^t V(z,L,\alpha,s')^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}{\int_0^\infty V(z,L',\alpha,s')^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}$$

and similarly if $\alpha \leq \alpha'$:

$$\frac{\int_0^t V(z,L,\alpha,s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}{\int_0^\infty V(z,L',\alpha,s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)} \leq \frac{\int_0^t V(z,L,\alpha',s)^{\frac{\nu_R}{\sigma-1}}\, dF_L(L)}{\int_0^\infty V(z,L',\alpha',s')\, dF_L(L)}$$

Therefore, the conditional probability $p\left(\text{firm z chooses city size } L \leq t|\alpha,s\right)$ increases with s (resp. with α),that is: the geographic distribution of a high α (resp. high s) sector - all else equal - first order stochastically dominates that of a lower α (resp. lower s) sector. ∎

Fourth, proposition 4 in the main text states that the firm size distribution in revenues of a high α (resp. s) sector is more spread out than that of a lower α (resp. s) sector. With imperfect sorting, the following characterization holds:

Proposition 4 ': Normalize the distribution of firm revenues across sectors by their mean. Then, the distribution of log-revenues of firms in a high α (resp. s) sector is a mean-preserving spread of that of a lower α (resp. s) sector.

Proof Given the discrete choice problem (21) in the main text, a firm of type z has a distribution of optimized profits π (resp. revenues r) that is distributed Frechet, with location parameter $T(z,s,\alpha) = \left(\sum_{L'} V(z,L',s,\alpha)^{\frac{\nu_R}{\sigma-1}}\right)^{\frac{\sigma-1}{\nu_R}}$ and shape parameter $\frac{\nu_R}{\sigma-1}$ (common to all firm types z). The distribution of log-revenues in a given sector depends therefore on the distribution of raw efficiency z and of a shock ϵ according to:

$$log(r(z,\epsilon;s,\alpha)) = \kappa + \log T(z,s,\alpha) + \epsilon,$$

where κ is a sectoral constant, ϵ is distributed type-1 EV, with location parameter 0 and shape parameter $\kappa = \frac{\nu_R}{\sigma-1}$, and is independent of z. Let $s_1 < s_2$. Define the constant $K_s = E_z\left[\log T(z,s_1,\alpha)\right] - E_z\left[\log T(z,s_2)\right]$. The distributions of $log(r(z,\epsilon;s_1,\alpha))$ and $log(r(z,\epsilon;s_2,\alpha)) + K_s$ have the same mean.

The location parameter $T(z,s,\alpha)$ is LSM in (z,s) and (z,α). To see this, note that $T(z,s,\alpha) = E\left[V(z,L',s,\alpha)^{\frac{\nu}{\sigma-1}}\right]^{\frac{\sigma-1}{\nu_R}}$, where the expectation is taken over the economy-wide distribution of city sizes. Since $V(z,L,s,\alpha)$ is log-supermodular in (z,s), $V(z,L,s,\alpha)^{\frac{\nu}{\sigma-1}}$ is also LSM in in (z,s), then $E_L\left[V(z,L',s,\alpha)^{\frac{\nu}{\sigma-1}}\right]$ is LSM (Athey (2002) shows that the expectation of a LSM function is LSM) and finally $T(z,s,\alpha)$ is LSM in (z,s). By a similar reasoning, it is also LSM in (z,α).

Fix α. The function $\log T(z,\alpha,s)$ is supermodular in z and s and increasing in z, so $\log T(z,\alpha,s_1)$ and $\log T(z,\alpha,s_2)+K_s$ cross (at most) once as functions of z. For z above that point we have $\log T(z,\alpha,s_2)+K_s > \log T(z,\alpha,s_1)$. The opposite is true below this point. Writing $h(z)=\log T(z,\alpha,s_1)-\log T(z,\alpha,s_2)-K_s$,we get that $E_z\left[h(z)\right]=0$ by definition of K_s. Given that $h(z)$ is first positive then negative, it follows that:

$$\int_0^Z h(z)dF(z)\geq 0$$

for all Z. This proves that the distribution $\log T(z,\alpha,s_2)+K_s$ second-order stochastically dominates the distribution $\log T(z,\alpha,s_1)$. Since $\log T(z,\alpha,s_1)$ and ϵ are independent, we get in turn that, $log(r(z,\epsilon;s_2,\alpha))+K_s$ second-order stochastically dominates $log(r(z,\epsilon;s_1,\alpha))$. Therefore, $log(r(z,\epsilon;s_2,\alpha))$ is, once de-meaned, a mean-preserving spread of $log(r(z,\epsilon;s_1,\alpha))$. The same proof is readily adaptable to the case of α, now holding s fixed. ∎

D. Proofs of section 3.3

D.1. Lemma 7

Proof Consider a given city of size L developed by city developer i. Equation (6) shows that, for a given city size and a given sector, labor hired by local firms is proportionate to the ratio of firms profit to the (common) local wage. Using this relationship, the city developers problem (14) then simplifies to

$$\max_{L,\{T_j(L)\}_{j\in 1,\dots,S}} \Pi_L = b\,(1-\eta)w(L)L - \sum_{j=1}^{S}\frac{M_j w(L)}{(1-\alpha_j)(\sigma_j-1)}\int_z T_j(L)\ell_j(z,L)\,\mathbb{1}_j(z,L,i)dF_j(z) \tag{D.36}$$

Let $N_j(L,i)=\int_z \ell_j(z,L)\mathbb{1}(z,L,i)M_j dF_j(z)$ denote the number of workers working in sector j in this specific city i. It follows that $\sum_{j=1}^{S} N_k(L,i)=L-\ell_H(L)=L\big(1-(1-\eta)(1-b)\big)$ where $\ell_H(L)$ is the labor force hired in the construction sector and the second equality uses (13).

The problem is akin to a Bertrand game. Consider a given city size L. Free entry pushes the profit of city developers to zero in equilibrium. I prove now that this drives $T_j(L,z)$ to the common level $T_j^*=\frac{b(1-\eta)(1-\alpha_j)(\sigma_j-1)}{1-(1-\eta)(1-b)}$. First, assume that for a given (z,j), the maximum subsidy offered is strictly less than T_j^*. New city developers could offer T_j^* for (z,j) and 0 for all other sectors, attract all (z,j) firms for whom this subsidy is more attractive, and make exactly zero profit, as $M_j\int_z \ell_j(z,L)\mathbb{1}(z,L,i)dF_j(z)=L\big(1-(1-\eta)(1-b)\big)$. Second, assume a city developer offers a subsidy $T_j(L,z)>T_j^*$ for a couple (z,j). This leads to negative profits. To see this, consider all cities of size L, and take the one that offers the highest subsidy city to (z,j) firms. Call this city i. From the first step of the proof, we know that in any given city, for all sectors k, either $T_k(L,z)\geq T_k^*$ and $N_k\geq 0$ or $T_k<T_k^*$ and $N_k=0$. Therefore,

$$\begin{aligned}\sum_{k=1}^{S}\frac{M_k w(L)T_k^*}{(1-\alpha_k)(\sigma_k-1)}\int_z \ell_k(z,L)\mathbb{1}(z,L,i)dF_k(z) &= \sum_{k=1}^{S}\frac{w(L)T_k^*}{(1-\alpha_k)(\sigma_k-1)}N_k\\ &> b\,(1-\eta)w(L)L\end{aligned}$$

so that $\Pi^i<0$. ∎

D.2. Extension with Specific Subsidies

I examine here the case where land developers can observe firm types z and offer specific subsidies that are z-industry-city specific, rather than ad-valorem and constant within industry in the baseline model. Specifically, land developers offer a specific subsidy $S_j(z;L)$ to each firm of type z in industry j coming to their city of size L. I show here that the same outcome as in the baseline model is still an equilibrium. That is, the following is an equilibrium:

- A city developer targets a city size L_0 and announces a fixed subsidy $S_j(L_0)\delta(z-z_j^*(L_0))\delta(L-L_0)$ where $\delta(0)=1$, and $\delta(x)=0$ for $x\neq 0$. This subsidy is targeted to firms for which L_0 is the best choice of city absent any subsidy, ie. the ones for which $z=z_j^*(L)$, where $z_j^*(L)$ is the inverse of $L_j^*(z)$ defined in equation (8) in the main text. The subsidy is 0 for other firms. The subsidy does not vary with the profit of the firm, but instead is fixed to the same level as what is effectively paid in the baseline equilibrium: $S_j(L_0) = T_j^* \widetilde{\pi}_j(z_j^*(L_0), L_0)$, where I write $\widetilde{\pi}_j(z_j^*(L), L) = \kappa_{1j} \left(\frac{\psi(z,L,s_j)}{w(L)^{1-\alpha_j}} \right)^{\sigma_j - 1} R_j \, P_j^{\sigma_j-1}$ the profit of the firm absent any subsidy.

- All cities in the optimal set $\mathcal{L}$ are announced by developers
- Firms of type $z^*(L)$ choose cities of size L.

The proof that this is an equilibrium is as follows. Given these subsidies offered by developers, a firm of type z chooses its optimal location as follows:

$$\max_L \widetilde{\pi}_j(z, L) + \mathbb{1}_{L=L_j^*(z)} S_j(z; L_j^*(z)).$$

Given that $\max_L \widetilde{\pi}_j(z, L) = L_j^*(z)$, the optimal choice of the firm with subsidy is also $L_j^*(z)$. A developer makes the following profit in his city, where I write N_j the number of firms in sector j that end up in this city :

$$\begin{aligned}
\Pi_L &= b(1-\eta)w(L)L - \sum_j S_j N_j \\
&= b(1-\eta)w(L)L - \sum_j N_j T_j^* \widetilde{\pi}_j(z_j^*(L), L) \\
&= b(1-\eta)w(L)L - \sum_j N_j T_j^* \frac{w(L)\ell_j(z_j^*(L), L)}{(1-\alpha_j)(\sigma_j - 1)} \\
&= b(1-\eta)w(L)L - \sum_j N_j \frac{b(1-\eta)}{1-(1-\eta)(1-b)} w(L)\ell_j(z_j^*(L), L) \\
&= b(1-\eta)w(L)L - \frac{b(1-\eta)}{1-(1-\eta)(1-b)} L(1-(1-b)(1-\eta)) \\
&= 0
\end{aligned}$$

where the second equality comes from the definition of the subsidy, the third equality comes from equation (6) in the main text, the fourth equality uses the definition of T_j^* from the main text, and the last one uses the local labor market clearing condition: $L(1-(1-b)(1-\eta)) = \sum_j N_j \ell_j(z_j^*(L), L)$. I finally show that there is no profitable deviation for a developer. First, it is clear that no developer wants to offer a higher subsidy for firms for the same city size (that is, for $z = z_j^*(L)$ in city size L), since it would lead to negative profits given that the current subsidies yield 0 profit. Also, lower subsidies for the same city size would not attract any firms. We need to check whether a developer want and can attract a firm of type z in a city that is not the unconstrained optimal choice $L^*(z) of the firm$. The proof is by contradiction. Assume that a developer targets firms z in cities of size $L_2 \neq L^*(z)$ and offers them a specific subsidy S_2. For the subsidy to be attractive for firms, it has to be that:

$$S_2 \geq \widetilde{\pi}(z, L^*) + T^* \widetilde{\pi}(z, L^*) - \widetilde{\pi}(z, L_2), \tag{D.37}$$

since the alternative for firm z is to choose city $L^*(z)$ – simply written L^* here – and get a profit of $\widetilde{\pi}(z, L^*)$ plus a subsidy $T^* \widetilde{\pi}(z, L^*)$. For the subsidy to generate positive profits for the developer, it has to be that:

$$S_2 N_2 \leq b(1-\eta)w(L_2)L_2,$$

where N_2 is the number of firms of type z that populate a city L_2 such that the local labor market clears, that is: $N_2 = \frac{L_2(1-(1-\eta)(1-b))}{\ell_2(z,L_2)} = \frac{L_2 w(L_2)(1-(1-\eta)(1-b))}{(1-\alpha)(\sigma-1)\widetilde{\pi}(z,L_2)}$. Therefore the condition for positive profits becomes: $S_2 \leq \frac{b(1-\eta)}{1-(1-\eta)(1-b)}(1-\alpha)(\sigma-1)\widetilde{\pi}(z, L_2)$, which is precisely $T^* \widetilde{\pi}(z, L_2)$. Finally, note that by optimality of L^*,

$$\widetilde{\pi}(z, L^*) + T^* \widetilde{\pi}(z, L^*) \geq \widetilde{\pi}(z, L_2) + T^* \widetilde{\pi}(z, L_2).$$

Therefore, $T^* \widetilde{\pi}(z, L_2) \leq \widetilde{\pi}(z, L^*) + T^* \widetilde{\pi}(z, L^*) - \widetilde{\pi}(z, L_2)$ and S_2 cannot at the same time satisfy $S_2 \leq T^* \widetilde{\pi}(z, L_2)$ and condition (D.37) . This contradiction means that there is no profitable deviation for a land developer. This conclude the proof that the distribution of firms and cities of the baseline model is still an equilibrium of the model with specific subsidy for type-z firms.

D.3. Lemma 8

Proof Let L_o denote the suboptimal city size where firms of type (z, j) are located. They get profit $\pi_j^*(z, L_o)$. Denote $\Delta = \frac{\pi_j^*(z_j, L^*(z))}{\pi_j^*(z, L_o)} - 1 > 0$. A city developer can open a city of size $L^*(z)$ by offering a subsidy $\tilde{T}_j = \frac{1+\frac{\Delta}{2}}{1+\Delta}(1+T_j^*) - 1$, which will attract firms as they make a higher profit than at L_o, and allows the city developer to make positive profits. City size distribution adjusts in equilibrium to determine the number of such cities. ∎

D.4. Lemma 9

The proof is in the main text.

E. General Equilibrium

I use the following notation:

$$E_j = \int \frac{\psi(z, L_j^*(z), s_j)^{\sigma_j - 1}}{\left[(1-\eta)L_j^*(z)\right]^{\frac{b(1-\eta)(1+(1-\alpha_j)(\sigma_j-1))}{\eta}}} dF_j(z) \quad \text{, and} \quad S_j = \int \left(\frac{\psi(z, L_j^*(z), s_j)}{\left[(1-\eta)L_j^*(z)\right]^{\frac{b(1-\eta)(1-\alpha_j)}{\eta}}}\right)^{\sigma_j - 1} dF_j(z),$$

where E_j and S_j are sectoral quantities that are fully determined by the matching functions $L_j^*(z)$ for each sector j. They are normalized measures of employment and sales in each sector.[2] To find general equilibrium quantities P_j, M_j for all $j \in \{1, ..., S\}$ and R, the aggregate revenues in the traded goods sector, I write the free entry conditions for firms (equation (E.38)), the goods market clearing conditions (equation (E.39)), and the national labor market clearing condition (workers works either in one of the traded goods sectors or in the construction sector, equation (E.40)). This leads to the following system of equations:

$$f_{Ej}\, P = (1 + T_j^*)\kappa_{1j}\, S_j\, \xi_j R\, P_j^{\sigma_j - 1}, \text{ for all } j \in \{1, ..., S\}, \tag{E.38}$$

$$1 = \sigma_j \kappa_{1j}\, M_j\, S_j\, P_j^{\sigma_j - 1}, \text{ for all } j \in \{1, ..., S\}, \tag{E.39}$$

$$Nv = \sum_{j=1}^{S} \kappa_{2j} E_j\, M_j\, \xi_j R\, P_j^{\sigma_j - 1} + N(1-b)(1-\eta), \tag{E.40}$$

where f_{Ej} is the units of final goods used up in the sunk cost of entry, P is the aggregate price index, and the last term derives from equation (13). First, the national labor market clearing condition (E.40) together with equation (E.39) leads to the aggregate revenues in manufacturing,

$$R = N \frac{1 - (1-b)(1-\eta)}{\sum_{j=1}^{S} \frac{(1-\alpha_j)(\sigma_j - 1)}{\sigma_j} \xi_j \frac{E_j}{S_j}} \tag{E.41}$$

This pins down uniquely the general equilibrium quantity R. Second, I combine equations (E.43) and (E.38) and write $\tilde{\kappa}_{1j} = \kappa_{1j} P^{\alpha_j(\sigma_j - 1)}$. This is a constant parameter, whereas κ_{1j} depended on the GE quantity P.[3] This leads

[2] Given the wage equation (4) and the expression for operating profits (5), aggregate operating profits in sector j are $\kappa_{1j}\, M_j\, S_j\, R_j\, P_j^{\sigma_j - 1}(1 + T_j^*)$. Similarly, aggregate revenues in sector j are $\sigma_j\, \kappa_{1j}\, M_j S_j R_j P_j^{\sigma_j - 1}$ and aggregate employment in sector j is $\kappa_{2j}\, M_j\, E_j\, R_j\, P_j^{\sigma_j - 1}$, where the sectoral constant κ_{2j} is $\kappa_{2j} = \kappa_{1j}(1 - \alpha_j)(\sigma_j - 1)$.

[3] This is because κ_{1j} depends on the price of capital which is constant, fixed in international markets in units of the internationally traded good. Since the price of the traded good is not taken as the numeraire here, the cost of capital if ρP in terms of the numeraire, with ρ a constant.

to :

$$P_j^{\frac{1}{\alpha_j}} \prod_{k=1}^{S} (\frac{P_k}{\xi_k})^{-\xi_k} = \left(\frac{1}{\tilde{\kappa}_{1j}(1+T_j^*)\, S_j\, \xi_j R} \right)^{\frac{1}{\alpha_j(\sigma_j-1)}}, \tag{E.42}$$

where I have used that $P = \prod_{j=1}^{S} (\frac{P_j}{\xi_j})^{\xi_j}$. Note that the matrix $\begin{bmatrix} -\frac{1}{\alpha_1}+\xi_1 & \xi_2 & ... & \xi_n \\ \xi_1 & -\frac{1}{\alpha_2}+\xi_2 & ... & \xi_n \\ \vdots & \vdots & & \vdots \\ \xi_1 & \xi_2 & ... & -\frac{1}{\alpha_n}+\xi_n \end{bmatrix}$ has full rank and is invertible. Therefore, equation (E.42) has a unique solution in $\{P_j\}_{j=1..N}$. This pins down P in turn. Finally, equations (E.39) leads to the sectoral mass of firms:

$$M_j = \frac{P^{\alpha_j(\sigma_j-1)}}{\sigma_j \widetilde{\kappa}_{1j}\, S_j\, P_j^{\sigma_j-1}}. \tag{E.43}$$

Therefore, equations E.38-E.40 have a unique solution $(R, M_j, P_j)_{j=1...S}$.

F. Stability

I verify here that the equilibrium described in section 3 is stable. First, I study the reaction of the economy to a perturbation of the equilibrium where only workers's location or firms' location are perturbed. Second, I examine a perturbation of both firms' and workers' location.

It is straightfoward to see that the equilibrium is stable to a small perturbation of the location of firms, holding workers location constant. No firm has an incentive to deviate from the initial equilibrium, as they all choose their profit maximizing city size in the first place. The equilibrium is also stable to a small perturbation of the location of workers, holding firms location constant. To see this, fix the set of equilibrium cities as well as the set of firms located in each cities. Consider city i. In equilibrium its population is L, and it has n_j firms of raw productivity z_j from sector j. Labor demand for each firm is $\ell_j = K_j \frac{\psi(z_j,L,s_j)^{\sigma_j-1}}{w(L)^{(1-\alpha_j)(\sigma_j-1)+1}}$ with K_j a set of general equilibrium. The local labor market clearing condition is $\sum_j n_j K_j \frac{\psi(z_j,L,s_j)^{\sigma_j-1}}{w(L)^{(1-\alpha_j)(\sigma_j-1)+1}} = L$. This implicitly pins down the wage $w(L)$ as a function of L if workers move to the city. Workers' utility in this city is is $U(L) = w(L)L^{\frac{b(1-\eta)}{\eta}}$. I now show by contradiction that this level of utility decreases with L. Since $\frac{\partial logu(L)}{\partial logL} = \frac{w'(L)L}{w(L)} - \frac{b(1-\eta)}{\eta}$, assume that $\frac{w'(L)L}{w(L)} > \frac{b(1-\eta)}{\eta}$. Differentiating the local labor market clearing condition leads to

$$\sum_j n_j K_j \frac{\psi(z_j,L,s_j)^{\sigma_j-1}}{w(L)^{(1-\alpha_j)(\sigma_j-1)+1}} \left[(\sigma_j-1)\frac{\psi_2}{\psi} - ((1-\alpha_j)(\sigma_j-1)+1)\frac{w'(L)}{w(L)} \right] = .1 \tag{F.44}$$

Using in equation (7) leads to $L\left[(\sigma_j-1)\frac{\psi_2}{\psi} - ((1-\alpha_j)(\sigma_j-1)+1)\frac{w'(L)}{w(L)}\right] < -\frac{b(1-\eta)}{\eta} < 0$,, so that (F.44) is contradicted. Hence $\frac{\partial\, logu(L)}{\partial logL} < 0$. ∎

Second, I study the reaction of the economy to a perturbation of the equilibrium where *both* workers and firms' location are perturbed. I show here that the economy converges back to the initial equilibrium. In the initial equilibrium, land developers on these sites had posted a subsidy schedule $T_j^*\delta(L-L_i)$, which was the one compatible with the initial equilibrium with city size distribution $f_L^*(L)$ (see main text, section 2.2.2.). Sites were initially populated with the posted number of workers (L_i for developer i), and firms which chose these sites got subsidy T_j^*, but this is not necessarily the case anymore. If their population has changed following the perturbation, then firms earn 0 subsidy in these cities, and land developer make strictly positive profits in these cities.

To study the stability of the initial equilibrium to this perturbation, I assume that the game is played sequentially. Land developers play first, in decreasing order of their current profit (for example). They announce a new subsidy scheme. Once all of the current land developers have spoken, potential entrants can also announce a subsidy scheme. Then, firms and workers can choose to relocate if they want to, taking these subsidy as given. If necessary, the game repeats until an equilibrium is reached. But I show here that the equilibrium is reached after one iteration, because the optimal subsidy schedule $T_j^*\delta(L-L_i)$ will be posted by land developers. The proof is by contradiction. Let us first take the subsidy distribution as given, and study how firms and workers sort across space. Necessarily, workers choose cities such that $U(L)=\tilde{U}$ for some value $\tilde{U}$. Otherwise, the workers would move away from cities with lower utility and into cities with higher utility. This location choice of workers leads to a set of city sizes $\tilde{\mathcal{L}}$, and pins down the wage schedule up to a constant (see equation (4)): $w(L)=\tilde{w}\,L^{b\frac{1-\eta}{\eta}}$. Necessarily, firms choose the city that maximizes their profit, that is:

$$\begin{cases} \tilde{\pi}_j(z,L)=\tilde{\kappa}_j\big(1+\widetilde{T_j}(L)\big)\left(\frac{\psi(z,L,s_j)}{L^{b\frac{1-\eta}{\eta}(1-\alpha_j)}}\right)^{\sigma_j-1} R_j\,P_j^{\sigma_j-1} \\ L_j^*(z)=\arg\max_{L\in\tilde{\mathcal{L}}}\tilde{\pi}_j(z,L). \end{cases}$$

Finally, land developers make the following profit: $\left[b(1-\eta)-T\frac{(1-(1-b)(1-\eta))}{(\sigma-1)(1-\alpha)}\right]w(L)L$.

First, assume first that for some city size L_0 , a city developer makes positive profits (ie the effective subsidy there is $T<\frac{b(1-\eta)(\sigma-1)(1-\alpha)}{1-(1-b)(1-\eta)}$). This is not compatible with all city developers maximizing profit. Indeed, a city developer with unused land, anticipating this, would have offered a subsidy $(T+\epsilon)\,\delta(L-L_0)$, with $T+\epsilon<T^*$, that would have attracted the same firms and generated profits for the developer. Therefore, it must be that no city developer makes positive profits after that round. In other words, effective subsidies collected by firms are necessarily $T=T^*$. Therefore, firms chose:

$$\begin{cases} \tilde{\pi}_j(z,L)=\tilde{\kappa}_j\big(1+T_j^*\big)\left(\frac{\psi(z,L,s_j)}{L^{b\frac{1-\eta}{\eta}(1-\alpha_j)}}\right)^{\sigma_j-1} R_j\,P_j^{\sigma_j-1} \\ L_j^*(z)=\arg\max_{L\in\tilde{\mathcal{L}}}\tilde{\pi}_j(z,L). \end{cases}$$

Second, assume that some firms are not back to their optimal city size $L^*(z)$. That is, not all city size in $\mathcal{L}^*$ are offered in $\tilde{\mathcal{L}}$. There exists a city size L_0 for which $f_L(L_0)>0$ in the baseline equilibrium, but no developer has offered the subsidy scheme $T_j^*\delta(L-L_0)$. Absent this option, the corresponding firms $z_0=L^{*-1}(L_0)$ must have chosen a suboptimal city L_1 with subsidy T^*. These firms make a profit $\pi_j^*(z_0,L_1)<\pi_j^*(z_0,L_0)$. A city developer with no city, anticipating this, would have offered a subsidy $T_j^*\delta(L-L_0)-\epsilon$ (with $\epsilon>0$ arbitrarily small), that would have attracted the same firms, as it strictly improves their profits.

We have thereby shown by contradiction that it must be that all optimal city sizes are announced by developers with a subsidy T^*. Therefore, the economy converges back to the intial equilibrium, which is stable to a small perturbation of both firms' and workers' locations. ∎

G. Welfare analysis

G.1. Social planner's problem

The utility function is as follows[4]:

$$U(L)=c(L)L^{-\frac{b(1-\eta)}{\eta}}. \tag{G.45}$$

I report here the results for a single-sector economy, for simplicity. The intuitions are unchanged in a multi-sector setup. The problem of the planner is to choose allocations optimally, namely
(1) for each firm z, its level of input $\ell(z)$ and $k(z)$ and its city size $L(z)$
(2) the mass of firms M and the distribution of city sizes $G(L)$

[4] As in the competitive equilibrium, a constant fraction of the local labor force is used to build housing. In this reduced-form utility function, congestion increases log-linearly with city size. The following results therefore hold irrespective of the source of congestion in the economy, as long as it increases log-linearly with city size. Utility has been renormalized by a constant and by taking utility in (G.45) to the power $\frac{1}{\eta}$.

(3) the share $\gamma(L)$ of total consumption C consumed by a worker living in a city of size L, in order to maximize:

$$U(L) = \frac{\gamma(L)C}{L^{b\frac{1-\eta}{\eta}}},$$

such that:

1. $U(L) = \bar{U}$ if $g(L) > 0$ (free mobility of workers)
2. $C = Q - Mf_E - M\rho \int k(z)dF(z)$; $Q = \left(\int Mq(z, L(z))^{\frac{\sigma-1}{\sigma}} dF(z)\right)^{\frac{\sigma}{\sigma-1}}$ and $q(z, L) = \ \psi(z, L, s)\ k(z)^{\alpha}\ \ell(z)^{1-\alpha}$ (production technology)
3. $\int \gamma(L)L\ dG(L) \leq 1$ (workers consume at most C)
4. $[1-(1-\eta)(1-b)] \int_0^L u\, dG(u) = M \int_{z^*(0)}^{z^*(L)} \ell(z)\ dF(z)$ (local labor markets clear)
5. $\int L\ dG(L) = N$ (aggregate labor market clears)

Combining the constraints lead to the following:

$$\int \gamma(L)L\ dG(L) = \frac{\bar{U}}{C} \int L^{b\frac{1-\eta}{\eta}+1}\ dG(L)$$
$$\bar{U} = \frac{C}{\int L^{b\frac{1-\eta}{\eta}+1}\ dG(L)}.$$

The local labor market clearing condition for cities of size L_i yields[5]:

$$dG(L) = \frac{M\mathbb{1}(L(z))\ \ell(z)}{L(z)} dF(z) \tag{G.46}$$

Define $\Gamma = M \int_z L(z)^{b\frac{1-\eta}{\eta}} \ell(z)\ dF(z) = \int L^{b\frac{1-\eta}{\eta}+1}\ dG(L)$ the aggregate congestion in the economy. The problem of the social planner reduces to:

$$\max_{L(z), \ell(z), k(z), M} \frac{C}{\Gamma} \tag{G.47}$$

such that $M \int_z \ell(z)\ dF(z) = N$, with $C = \left(\int M \left[\psi(z, L, s)\ k(z)^{\alpha}\ \ell(z)^{1-\alpha}\right]^{\frac{\sigma-1}{\sigma}} dF(z)\right)^{\frac{\sigma}{\sigma-1}} - Mf_E - M\rho \int k(z)dF(z)$.

The city size distribution $G(L)$ does not directly enter the objective function. It adjusts such that the local labor markets clearing condition holds in equilibrium.

Taking the first order conditions with respect to $k(z)$ and solving out for $k(z)$ leads to

$$C = \kappa^* M^{1+\frac{1}{(1-\alpha)(\sigma-1)}} \left[\int (\psi(z, L, s)\ \ell(z)^{1-\alpha})^{\phi}\ dF(z)\right]^{\frac{1}{\phi(1-\alpha)}} - Mf_E,$$

where $\kappa^* = \left(\frac{\alpha}{\rho}\right)^{\frac{\alpha}{1-\alpha}} (1-\alpha)$ and $\phi = \frac{\sigma-1}{\sigma+\alpha-\alpha\sigma}$.

Taking the first order condition with respect to $L(z)$ leads to:

$$\frac{\psi_2(z, L, s)L}{\psi(z, L, s)} = b\frac{1-\eta}{\eta}(1-\alpha)\chi(z), \tag{G.48}$$

where[6]

$$\chi(z) = \left(\frac{\tilde{Q} - f_E}{\tilde{Q}}\right) \frac{\ell(z)L(z)^{b\frac{1-\eta}{\eta}}}{\int \ell(z)L(z)^{b\frac{1-\eta}{\eta}}} \frac{\int \tilde{q}(z)^{\phi}\ dF(z)}{\tilde{q}_j^{\phi}}. \tag{G.49}$$

[5]In particular, this yields the distribution of city sizes $G(L)$ once $M, k(z), \ell(z)$ and $L(z)$ are known for all firms.

[6]I use the notations $\tilde{q}(z) = \psi(z, L, s)\ell(z)^{1-\alpha}$, $\quad \tilde{Q} = \kappa^* M^{\frac{1}{(1-\alpha)(\sigma-1)}} \left[\int (\tilde{q}(z))^{\phi}\ dF(z)\right]^{\frac{1}{\phi(1-\alpha)}}$.

The first order condition with respect to M yields

$$\frac{1}{(\sigma-1)(1-\alpha)}\left(\frac{\tilde{Q}}{\tilde{Q}-f_E}\right)=\lambda N \tag{G.50}$$

Taking the first order condition with respect to $\ell(z)$ leads to:

$$\left(\frac{\tilde{Q}}{\tilde{Q}-f_E}\right)\frac{\tilde{q}_j^{\phi}}{\int\tilde{q}(z)^{\phi}\,dF(z)}-\frac{\ell(z)L(z)^{b\frac{1-\eta}{\eta}}}{\int\ell(z)L(z)^{b\frac{1-\eta}{\eta}}\,dF(z)}=\lambda M\ell(z) \tag{G.51}$$

In particular, summing this over all types of firms and using (G.50) and the labor market clearing condition lead to:

$$\frac{f_E}{\tilde{Q}}=\frac{1}{(\sigma-1)(1-\alpha)}, \tag{G.52}$$

and $\lambda=\frac{1}{N}\frac{1}{(\sigma-1)(1-\alpha)-1}$.

Plugging in $\tilde{q}_j=\psi(z_j,L(z_j))\ell(z)^{1-\alpha}$ into equation (G.51) and using (G.50) gives the following expression for $\ell(z)$:

$$\ell(z)=\left(\frac{\psi(z,L,s)}{\left(\int\tilde{q}(z)^{\phi}\right)^{\frac{1}{\phi}}}\right)^{\sigma-1}\left(\frac{\tilde{Q}-f_E}{\tilde{Q}}\frac{L(z)^{b\frac{1-\eta}{\eta}}}{\int\ell(z)L(z)^{b\frac{1-\eta}{\eta}}dF(z)}+\frac{M}{N}\frac{1}{(\sigma-1)(1-\alpha)}\right)^{\alpha\sigma-\alpha-\sigma}. \tag{G.53}$$

G.2. Comparison with the competitive equilbrium

Rearranging equation (G.49), using (G.53) and (G.52), leads to

$$\chi(z)=\frac{L(z)^{b\frac{1-\eta}{\eta}}}{L(z)^{b\frac{1-\eta}{\eta}}+\frac{\Gamma}{N}\frac{1}{(\sigma-1)(1-\alpha)-1}}, \tag{G.54}$$

where $\Gamma=M\int_z L(z)^{b\frac{1-\eta}{\eta}}\ell(z)\;dF(z)$ is a measure of "aggregate congestion" in the economy. Therefore, in particular, $\chi(z)<1$ for all j. There is a wedge in the incentives of location choice between the competitive equilibrium (equation (7)) and the social planner problem (equation (G.48)). Since $\frac{\psi_2(z,L,s)L}{\psi(z,L,s)}$ is decreasing in L by assumption (which ensures the concavity of firm's profit function), this means that firms choose cities that are too small in the decentralized equilibrium.

G.3. Implementing first best

To align firms' incentives in the competitive to the solution to the social planner's problem, firms have to see a wage of the form

$$w(L)\propto(L^{b\frac{1-\eta}{\eta}}+A) \tag{G.55}$$

where $A=\frac{\Gamma}{N}\frac{1}{(\sigma-1)(1-\alpha)-1}$. This is in contrast to $w(L)\propto L^{b\frac{1-\eta}{\eta}}$ in the decentralized equilibrium, set by the free mobility assumption. This allows both the size of the workforce and the choice of city size to be aligned in the two equilibria. Finally, the mass of entrants is suboptimal in the competitive equilibrium (after correcting for these effects). This effect is classic in monopolistic competition framework, and is not of direct interest here as it does not interact with the choice of city sizes.[7]

[7] In the competitive equilibrium, the mass of firm is given by $M=\frac{(1+T^*)}{\sigma}\frac{Q}{f_E}$, whereas in the social planner's problem it is given by $M=\frac{1}{(\sigma-1)(1-\alpha)}\frac{Q}{f_E}$

H. Estimation

H.1. Identification

To guide intuition on identification, I derive the distribution of firm value-added across cities of different sizes. The setup is the one developed to study imperfect sorting in section B.1.

Note that firm value added is proportional to profits: $r_j(z_i, L) = \sigma_j \pi_j(z_i, L)$ I focus from now on on one sector and omit the sectoral subscript for simplicity. The distribution of value added across cities of different sizes is:

$$
\begin{aligned}
E(r|L) = & \qquad E_{z|L}\left[E(r|L,z)\right] \\
&= \int_z E(r|L,z)\, p(z|L) dz \\
&= \int_z p(L|z) \frac{f(z)}{f_L(L)} E(r|L,z)\, dz \\
&= \int_z p(L|z) \frac{f(z)}{f_L(L)} E(r|z)\, dz
\end{aligned}
$$

The first equality uses the law of iterated expectation. The last equality uses a property of the Frechet distribution: the expectation of the profits are the same for a firm of a given type z irrespective of which city size the firm has chosen. That is, $E(\pi|z) = E(\pi|z, L^*)$ so that in turn, since profits are proportional to value added, $E(r|z) = E(r|z, L^*)$. Furthermore, using again the properties of the Frechet distribution, this expectation is:

$$
E(\pi|z) = \Gamma(\nu) \left[\sum_L V(z,L)^\nu\right]^{\frac{1}{\nu}}, \tag{H.56}
$$

where we write $\nu = \frac{\nu_R}{\sigma-1}$ the shape parameter of the Frechet distribution relevant for profits. We also know that the probability that a firm of type $z$chooses a city of size L is:

$$
p(L|z) = \frac{V(z,L)^\nu}{\sum_{L'} V(z,L')^\nu} \tag{H.57}
$$

We can therefore write that:

$$
\begin{aligned}
E(r|L) &= \frac{1}{f_L(L)} \int_z p(L|z) E(r|z) f(z)\, dz \\
&= \frac{C}{f_L(L)} \int_z V(z,L)^\nu E(\pi|z)^{1-\nu} f(z)\, dz.
\end{aligned}
$$

where C is a sectoral constant. One case that helps understand the intuition behind the identification is when $\nu_R = \sigma - 1$. In that case, we can readily see that the distribution of value added across cities of different sizes simplifies to:

$$
\begin{aligned}
E(r|L) &= \frac{C}{f_L(L)} \int_z V(z,L) f(z)\, dz \\
&= \frac{C}{f_L(L)} L^{(\sigma-1)\left[a - b\frac{1-\eta}{\eta}(1-\alpha)\right]} \int_z exp\left((\sigma-1)\left\{\log z \left(1+\log \tilde{L}\right)^s\right\}\right) f(z)\, dz,
\end{aligned}
$$

where we have used the value of $V(z, L)$ from equation (C.34) and the definition of productivity in equation (20) of the main text. Given that $z$is (truncated-) log normally distributed, that is, $\log z$ is distributed like a mean-0 normal truncated at its mean, this integral can be computed as follows. Note $S(L) = \left(1 + \log \tilde{L}\right)^s$.Then, we get that:

$$E(r|L) = \frac{C}{f_L(L)} L^{(\sigma-1)\left[a-b\frac{1-\eta}{\eta}(1-\alpha)\right]} \int_z z^{(\sigma-1)S(L)} f(z)\, dz$$
$$= \frac{C}{f_L(L)} L^{(\sigma-1)\left[a-b\frac{1-\eta}{\eta}(1-\alpha)\right]} E_z\left[z^{(\sigma-1)S(L)}\right]$$

If z was log normally distributed without truncation, we would simply get that $E_{\log\mathcal{N}}\left[z^{(\sigma_j-1)S(L)}\right] = exp(\frac{S(L)^2(\sigma-1)^2\nu_z^2}{2})$, so that:

$$E(r|L) = \frac{C}{f_L(L)} L^{(\sigma-1)\left[a-b\frac{1-\eta}{\eta}(1-\alpha)\right]} \exp\left(\frac{\left(1+\log\,\tilde{L}\right)^{2s}(\sigma-1)^2\nu_z^2}{2}\right)$$

Taking into account that z is truncated (at the mean of the normal) we get an additional term[8]so that $E_z\left[z^{(\sigma-1)S(L)}\right] = \frac{exp(\frac{S(L)^2(\sigma-1)^2\nu_z^2}{2})\Phi((\sigma-1)S(L)\nu_z)}{1/2}$ and:

$$E(r|L) = \frac{C'}{f_L(L)} L^{(\sigma-1)\left[a-b\frac{1-\eta}{\eta}(1-\alpha)\right]} \exp\left(\frac{\left(1+\log\,\tilde{L}\right)^{2s}(\sigma_j-1)^2\nu_z^2}{2}\right)\Phi((\sigma_j-1)\,S(L)\nu_z),$$

where C' is a sectoral constant. Finally, taking logs, this equation gives us the relationship between average value added and city size (within a sector) in a (non linear) regression format, and thus helps us understand what variation in the data help identify the parameter:

$$log(E[r_i|L_i]) = C'' - \log\left(f_L(L_i)\right) + \beta_1 \log L_i + \beta_2\left(1+\log\,L_i\right)^{2\mathrm{s}} + \log\left[\Phi((\sigma_j-1)\,S(L_i)\nu_z)\right], \tag{H.58}$$

where

$$\beta_1 = (\sigma-1)\left[a - b\frac{1-\eta}{\eta}(1-\alpha)\right]$$
$$\beta_2 = \frac{(\sigma-1)^2\nu_z^2}{2}$$

The parameters $\sigma, b\frac{1-\eta}{\eta}$ and α are calibrated in the first stage, the parameters (a, ν_z, s) are the ones we aim to estimate. We can see from this expression that we can identify β_1,β_2 and s (hence a, s and ν_z) from a non linear least square regression of r on functions of L. The parameters a and s both impact firm productivity and profits, but a impacts them log-linearly with city size, and s impacts them more than log-linearly because it entails the sorting of more productive (high z) firms into larger cities. The shape of the distribution of firm value added with respect to city size pins down the agglomeration parameters. The log-linear term identifies the classic agglomeration economies forces a, and the convex term identifies the sorting forces, that is the interaction of ν_z and s. To identify in addition the parameter ν_R, I bring in additional moments that characterize the firm-size distribution and the sectoral distribution of activity.

H.2. Moments

Distribution of average value-aded by city size. The distribution of average firm value-added as a function of city size is computed as follows in sector j. Define $\bar{r}_j(L) = \frac{\int r_j^*(z)\mathbb{1}_{L_j^*(z)=L}dF_j(z)}{\int \mathbb{1}_{L_j^*(z)=L}dF_j(z)}$ the average value-added of sector j firms that locate in city L. Normalize firms value-added within a given sector by their median value. Group cities by quartile of city sizes, call them $q = 1...4$. For each quartile, compute the data counterpart of $E(\log(\bar{r}_j|L_i)))$ in (H.58) as the sample mean M_q of $\log(\bar{r}_j(L_i))$. The targeted moments are $\{M_q\}_{q=1,2,3,4}$.

[8]If Z is distributed log normal, where the normal has mean μ and variance ν_z^2, then:

$$E(Z|Z > e^0) = \frac{\int_{e^0}^{\infty} zg(z)dz}{1-\Phi(0)} = \frac{e^{\mu+\nu_z^2/2}\Phi\left(\frac{\mu+\nu_z^2-ln(e^0)}{\nu_z}\right)}{1-\Phi(0)},$$

where Φ is the CDF of the standard normal distribution,

Distribution of total value-added by city size I order cities in the data by size and create bins using as thresholds cities with less than 25%, 50% and 75% of the overall workforce I normalize city sizes by the size of the smallest city, and call t_i^L the city size these thresholds. I compute the fraction of value-added for each sector in each of the city bins, both in the data and in the simulated sample. The corresponding moment for sector j and bin i is $s_i^{L,j} = \frac{\sum\limits_{t_i^L \le L < t_{i+1}^L} \int r_j^*(z) \mathbb{1}_{L_j^*(z)=L} \, dF_j(z)}{\int r_j^*(z) dF_j(z)}$, where $r_j^*(z)$ is the value-added of firm z and $\mathbb{1}_{L_j^*(z)=L}$ is a characteristic function which equals 1 if and only if firm z in sector j chooses to locate in city size L.

Firm-size distribution. I retrieve from the data the 25, 50, 75 and 90th percentiles of the distribution of firms' normalized value added and denote them $t_i^{r,j}$. These percentiles define 5 bins of normalized value-added. I then count the fraction of firms that fall into each bin $s_i^{r,j} = \frac{\int \mathbb{1}_{t_i^{r,j} \le \tilde{r}_j(z) < t_{i+1}^{r,j}} dF_j(z)}{\int dF_j(z)}$, where $\tilde{r}_j(z)$ is the normalized value added of firm z in sector j.

H.3. Simulation and estimation procedure

I simulate an economy with 100,000 firms and 200 city sizes. I follow the literature in using a number of draws that is much larger than the actual number of firms in each sector, to minimize the simulation error. I use a grid of 200 normalized city sizes $\tilde{L}$, ranging from 1 to M where M is the ratio of the size of the largest city to the size of the smallest city among the 314 cities observed in the French data. This set of city-sizes $\mathcal{L}$ is taken as exogenously given.[9] In contrast, the corresponding city-size distribution is not given a priori: the number of cities of each size will adjust to firm choices in general equilibrium to satisfy the labor-market clearing conditions.

The algorithm I use to simulate the economy and estimate the parameters for each sector is as follows:

Step 1: I draw, once and for all, a set of 100,000 random seeds and a set of 100,000 $\times$ 200 random seeds from a uniform distribution on $(0, 1)$.

Step 2: For given parameter values of ν_R and ν_z, I transform these seeds into the relevant distribution for firm efficiency and firm-city size shocks.

Step 3: For given parameter values of a and s, I compute the optimal city size choice of firms according to equation (21).

Step 4: I compute the 13 targeted moments described below.

Step 5: I find the parameters that minimize the distance between the simulated moments and the targeted moments from the data (equation (22)) using the simulated annealing algorithm.

The estimation is made in partial equilibrium, given the choice set of normalized city-sizes $\mathcal{L}$. It relies on measures that are independent of general equilibrium quantities, namely the sectoral matching function between firm efficiency and city size, and relative measures of firm size within a sector.[10]

H.4. Standard errors

Following Gourieroux et al. (1993), the matrix of variance-covariance V_j of the parameter estimates in sector j is computed as follows:

$$V_j = (1 + \frac{1}{N_s})(G_j' W_j G_j)^{-1}(G_j' W_j \Omega_j W_j G_j')(G_j' W_j G_j)^{-1}, \text{ where}$$
$$G_j = E\left[\frac{\partial m_j(\theta_{j0})}{\partial \theta}\right] \quad , \quad \Omega_j = E\left[m_j(\theta_{j0}) m_j(\theta_{j0})'\right],$$

[9] As pointed out in the theory section and developed above in B.4., the characterizations of the economy provided in Section 3 hold if the set of possible city sizes is exogenously given.

[10] Specifically, as detailed in the theoretical model, the optimal choice of city size by a firm depends only on its productivity function and on the elasticity of wages with respect to city size. It does not depend on general equilibrium quantities. The sizes of all firms in a given sector depend proportionally on a sector-level constant determined in general equilibrium (see equations (9) and (10)). Normalized by its median value, the distribution of firm sizes within a sector is fully determined by the matching function.

N_s is the number of simulation draws and W_j is the variance-covariance matrix of the data moments used in estimation. The reported standard errors are the square-root of the diagonal of V_j.

I. Policy analysis

I.1. Computing new equilibria in response to policy change

To compute the counterfactual equilibrium, I proceed as follows.

Step 1: I start from the equilibrium estimated in the data. I hold fixed the number of workers in the economy, the real price of capital, the set of idiosyncratic productivity shocks for each firm and city-size bin, and the distribution of firms' initial raw efficiencies.

Step 2: I recompute the optimal choice of city-size by firms, taking into account the altered incentives they face in the presence of the subsidy.

Step 3: Because the composition of firms within a given city-size bin changes, total labor demand in a city-size bin is modified. I hold constant the number of cities in each bin and allow the city size to grow (or shrink) so that the labor market clears within each city-size bin. This methodology captures the idea that these policies are intended to"push" or jump-start local areas, which in addition grow through agglomeration effects.[11]

Step 4: As city sizes change, the agglomeration economies and wage schedules are modified, which feeds back into firms' location choice.

Step 5: I iterate this procedure from step 2, using the interim city-size distribution.

The fixed point of this procedure constitutes the new counterfactual equilibrium.

I.2. Decomposition

Welfare is measured by worker's real income, constant across space. It is given by $\bar{U} = \frac{w}{P^{\eta} p_H^{1-\eta}}$, where p_H is the local housing cost. Plugging in the values of w and p_H as functions of L, this can be simply reexpressed as $\bar{U} = \left(\frac{\bar{w}}{P}\right)^{\eta} = \left(\frac{1}{P}\right)^{\eta}$, given the choice of numeraire $\bar{w}$. From (E.43) and (E.42), one gets the expression for the aggregate price index, which leads to

$$\bar{U} \propto \left(\prod_{j=1}^{S} TFP_j^{\xi_j}\right)^{\frac{\eta}{1-\bar{\alpha}}} \left(\prod_{j=1}^{S} (\frac{S_j}{E_j})^{\xi_j(1-\alpha_j)}\right)^{-\frac{\eta}{1-\bar{\alpha}}},$$

where $\bar{\alpha} = \sum_{j=1}^{S} \alpha_j \xi_j$ is an aggregate measure of the capital intensity of the economy.

The term $\prod_{j=1}^{S} TFP_j^{\xi_j}$ is a model-based measure of aggregate productivity. Take the example of a policy that increases TFP by pushing firms to larger cities. It has a direct positive impact on welfare, magnified by the term $\frac{1}{1-\bar{\alpha}}$ that captures the fact that capital flows in response to the increased TFP in the economy, making workers more productive. This effect is dampened by the second term, which captures the congestion effects that are at play in the economy. Wages increase to compensate workers for increased congestion costs in larger cities. Here, $\frac{S_j}{E_j}$ measures the ratio of the average sales of firms to their average employment in a given sector. It is a model-based measure of the representative wage in the economy, since $\frac{r_j^*(z)}{\ell_j^*(z)} \propto w(L^*(z))$ for each firm. A policy that tends to push firms into larger cities will also tend to increase aggregate congestion in the economy by pushing workers more into larger cities. Individual workers are compensated for this congestion by increased wages, in relative terms across cities, so that all workers are indifferent across city sizes. But the level of congestion borne by the *representative* worker depends on how workers are distributed across city sizes. It increases as the economy is pushed toward larger cities. This negative effect is captured by the second term in the welfare expression that decreases with the representative wage.

[11] I maintain the subsidy to the cities initially targeted as they grow.

J. Additional figures

J.1. Model fit

Figure J.1: Average value added by quartile of city size, model (blue) and data (red).

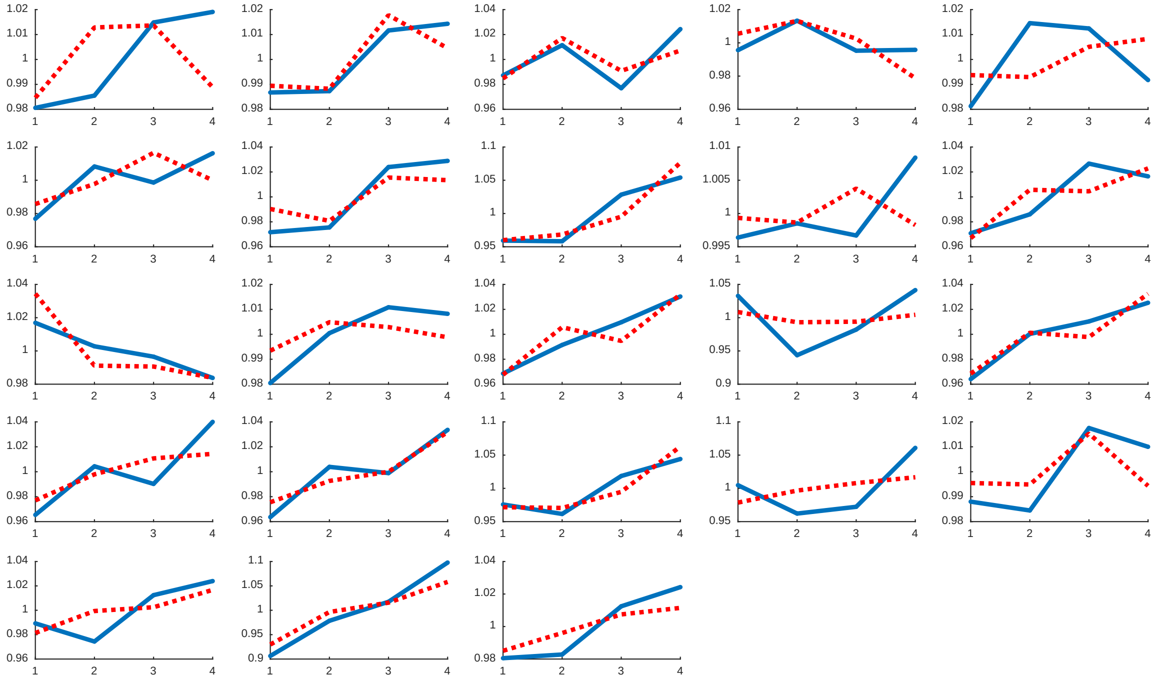


Figure J.2: Sectoral distribution of firms revenues, model (blue) and data (red).

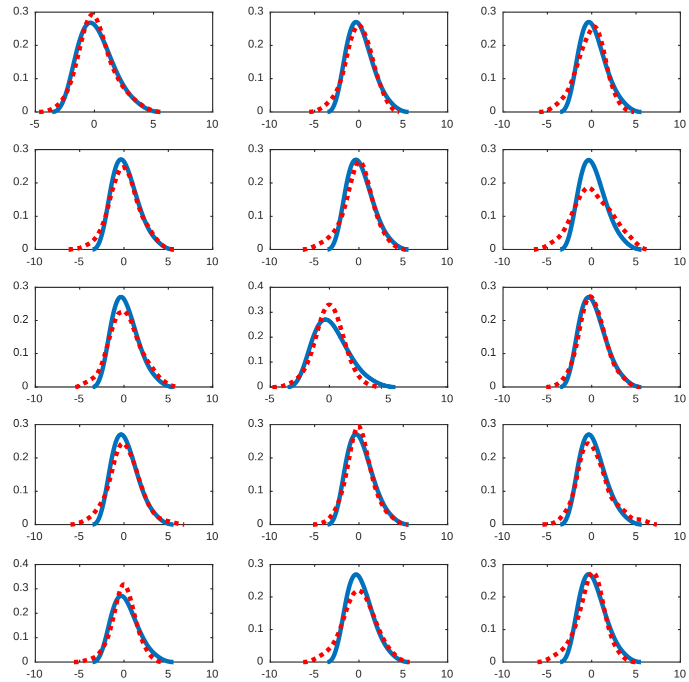
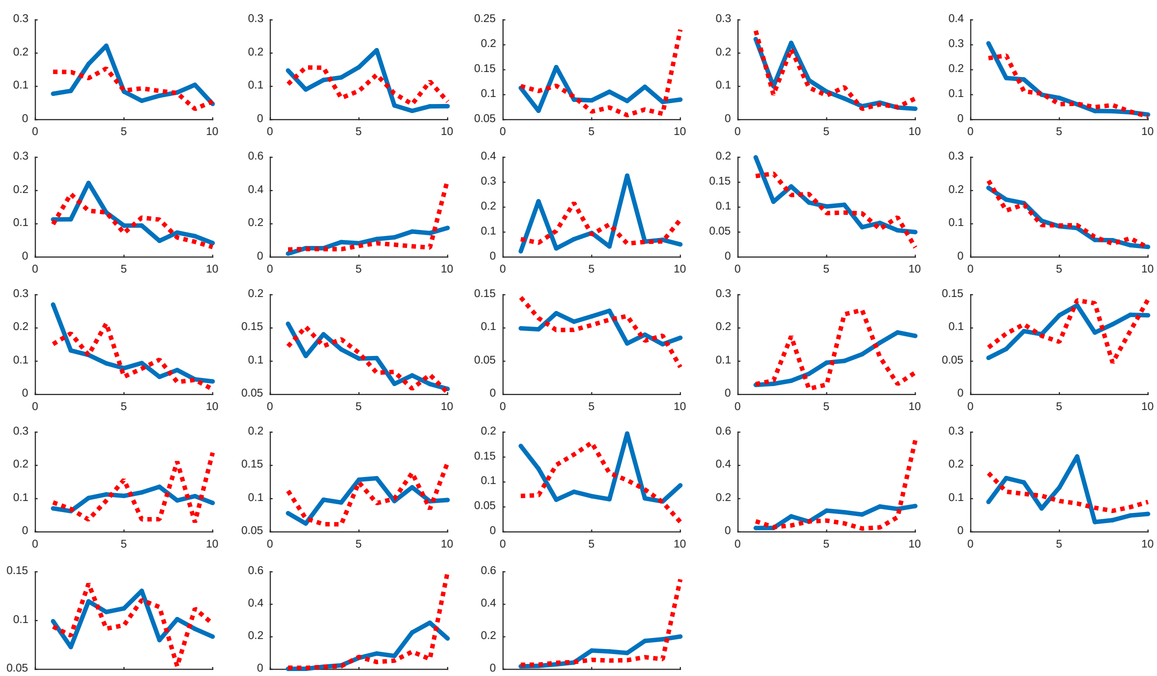
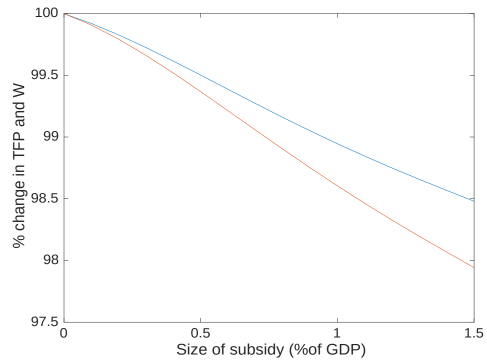


Figure J.3: Employment share by decile of city size, model (blue) and data (red).

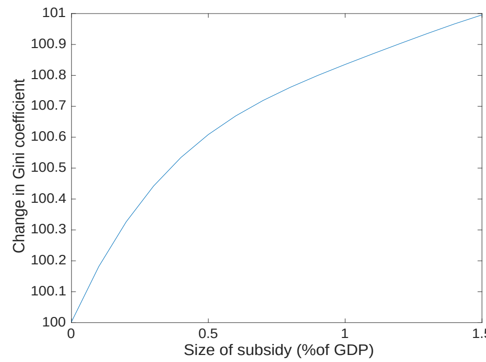


J.2. Impact of policies

Figure J.4: Aggregate impact of local subsidies, as a function of the cost of the policy (% of GDP).



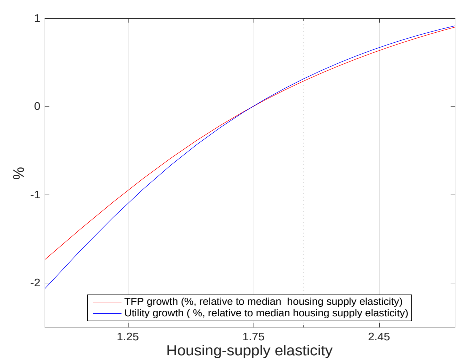
A. TFP (red) and welfare impact (blue), relative to the reference equilibrium



B. Change in the Gini coefficients for real wage inequality and city production inequality

Note: The x axis represent the cost of the policy in percentage of GDP. Firms profits are subsidized when they locate in the smallest cities of the reference equilibrium. The targeted area represents 2.3% of the population. The policy is financed by a lump-sum tax on firms.

Figure J.5: TFP and indirect welfare effects of increasing housing-supply elasticity.



The horizontal axis measures housing supply elasticity in the economy $\frac{dlogH}{dlogp_H}$. Saiz (2010) reports that median elasticity of housing supply is 1.75, the 25th percentile is at 2.45 and the 75th percentile at 1.25.