Firm Sorting and Agglomeration

Cecile Gaubert*
University of California, Berkeley

January 2017

Abstract

To account for the uneven distribution of economic activity in space, I propose a theory of the location choices of heterogeneous firms in a variety of sectors across cities. In this spatial equilibrium, the distribution of city sizes and the sorting patterns of firms are uniquely determined and affect both aggregate TFP and welfare. I estimate the model using firm-level data and find that nearly half of the productivity advantage of large cities is due to firm sorting. I also quantify the general equilibrium effects of place-based policies. Policies that subsidize smaller cities have negative aggregate effects.

*cecile.gaubert@berkeley.edu. I am extremely grateful to Esteban Rossi-Hansberg, Steve Redding and Gene Grossman for continuous support and guidance. I also thank Treb Allen, Costas Arkolakis, Thomas Chaney, Don Davis, Jan De Loecker, Gilles Duranton, Pablo Fajgelbaum, Joshua Gottlieb, Pat Kline, Oleg Itskhoki, Kiminori Matsuyama, Eduardo Morales, Dávid Nagy, Andrés Rodríguez-Clare, Felix Tintelnot and Jonathan Vogel as well as numerous seminar and conference participants for helpful discussions and comments. This research benefited from financial support from the International Economics Section at Princeton University.

1. Introduction

A striking and widely studied fact in economics is that the distribution of economic activity across space is very uneven. Location choices made by firms arguably play an important role in shaping these spatial disparities: some areas boom, driven by the presence of large and productive firms, while other areas barely attract any. Yet, the uneven distribution of economic activity is traditionally studied under the assumption that all firms are identical and indifferent between locations. In this paper, in contrast, I allow for heterogeneity of firms and study how they sort across a variety of cities within a country.

It is important to account for the fact that heterogeneous firms make different location choices. First, it matters for understanding aggregate productivity. Aggregate productivity depends on the productivity of individual firms and on the way factors are allocated to them (Hsieh and Klenow (2009)). Location shapes both dimensions: local agglomeration externalities can enhance a firm's productivity, and labor and real estate costs vary greatly over space, shaping the allocation of these inputs across firms. Second, the location choices of heterogeneous firms have direct policy implications. A range of spatial policies are put in place in order to attract firms to specific areas – in general, to the less developed ones. By changing the spatial allocation of firms, these policies impact aggregate productivity. They also impact spatial disparities. For example, if subsidies attract low-productivity firms, while leaving high-productivity firms in booming regions, they can reinforce existing disparities. Third, accounting for the spatial sorting of heterogenous firms matters in the measurement of agglomeration externalities. The observed higher productivity of firms in larger cities can be driven both by the strength of agglomeration externalities caused by a greater concentration of economic activity, and by the sorting of more productive firms into larger cities. This second channel is often overlooked, confounding the measurement of agglomeration externalities. In contrast, the approach taken here explicitly accounts for firm sorting.

In the model, cities form endogenously, on sites that are ex-ante identical. They grow in population as firms choose to locate there and increase local labor demand. Cities are the locus of agglomeration externalities such as thick labor markets or knowledge spillovers (Duranton and Puga (2003)). Firms are heterogeneous in productivity and produce in a variety of sectors. They sort across cities of different sizes. This sorting is driven by a trade-off between gains in productivity through local externalities, and higher labor costs. I assume that more efficient firms benefit relatively more from these local externalities (Combes et al. (2012)). This generates positive assortative matching: more efficient firms locate in larger cities, reinforcing their initial edge. Finally, city developers compete to attract firms to their city. They act as a coordinating device in the economy, leading to a unique spatial equilibrium.

Using firm-level data, I show that the model is able to reproduce salient stylized facts about French firms. First, in the model, initial differences in productivity between firms induce sorting across city sizes. This, in turn, reinforces firm heterogeneity, as firms in large cities benefit from stronger agglomeration forces. The resulting firm-size distribution is more thick-tailed for sectors that tend to locate in larger cities. Second, across sectors, firm sorting is shaped by the intensity of input use. In labor-intensive sectors, firms locate more in small cities, where wages are lower. Third, within sectors, firms have higher revenues in larger cities. They may have lower employment, though, since they face higher labor costs. Finally, the model can also match the striking empirical regularities that govern both the firm-size and the city-size distributions in the data.[1]

I structurally estimate the model using firm-level data, and recover a model-based estimate of the shape of agglomeration externalities. This allows me to disentangle the roles played by agglomeration forces on the one hand, and firm sorting on the other hand, in explaining the productivity difference between cities of different sizes. I estimate that the magnitude of the productivity advantage of large cities is 4.2%, in line with existing measures of agglomeration externalities in the literature as reported by Rosenthal and Strange (2004). Using a counterfactual analysis, I estimate that nearly half of this measured productivity advantage of large cities comes from the sorting of firms based on their efficiency.

Finally, I analyze the general equilibrium impact of spatial policies that aim to influence the location choice of firms. These policies are pervasive.[2] Federal programs in the United States and in Europe provide generous tax breaks to firms that elect to locate in less developed areas. Because they induce a complex reallocation of factors across space, the aggregate impact of such programs is a priori ambiguous. First, positive local impacts may be counterbalanced by undesirable effects in other regions. Second, subsidies may attract low-productivity firms to the targeted zones while leaving high-productivity firms in the more developed regions, thereby reinforcing spatial disparities. I use the estimated model to quantify the effect of such spatial policies. Specifically, I simulate the new spatial equilibrium that results from two types of programs: a tax-relief scheme targeted at firms locating in smaller cities, and the removal of regulations that hamper city growth, such as zoning or building-height regulations.[3] The productive efficiency of the new equilibrium depends in particular on the new city-size distribution: this distribution drives the extent of agglomeration externalities leveraged by firms in the economy. I find that a policy that subsidizes less productive areas has negative aggregate effects on TFP and welfare. In contrast, a policy that favors the growth of cities leads to a new spatial equilibrium that is significantly more productive, by endogenously creating agglomeration externalities and reducing the impact of market failures.

The paper is related to several strands of the literature. The main contribution of the paper is to propose a tractable model of spatial equilibrium that features freely mobile and heterogenous firms, and has a unique equilibrium. The literature that studies systems of cities, pioneered by Henderson (1974), has traditionally focused on homogeneous firms. Recent contributions have introduced richer heterogeneity in the spatial setting.[4] In a seminal contribution, Behrens et al. (2014) study the spatial sorting of entrepreneurs who produce non-tradable intermediates. I study the polar case of producers of perfectly tradable goods, and show that the setup is tractable enough that it can be extended to feature trade costs. Furthermore, an important difference for policy analysis is the uniqueness of the

[1] The distributions of city sizes and firm sizes have been shown to follow Zipf's law with a remarkable regularity; see for example Gabaix (2009) for a review.

[2] Kline and Moretti (2013) report that an estimated \$95 billion are spent annually in the United States to attract firms to certain locations.

[3] Glaeser and Gottlieb (2008) in particular advocate in favor of reducing this type of regulation.

[4] Early studies of heterogenous firms in the spatial context include Nocke (2006) and Baldwin and Okubo (2006). They predict that more productive firms self-select into larger markets. These results are obtained in a setting where the size of regions is fixed and exogenously given. In a recent contribution, Serrato and Zidar (2014) study the incidence of the corporate tax in a spatial setting with heterogenous firms. Cities are taken as given and the paper abstracts from agglomeration externalities. In contrast, in this research, city sizes respond endogenously to the location choice of firms and the intensity of agglomeration externalities. Another strand of the literature studies the sorting of heterogenous firms within a given urban area (Brinkman et al. (forthcoming)).

equilibrium I obtain here. Another closely related strand of the literature (Eeckhout et al. (2010), Davis and Dingel (2012) and Davis and Dingel (2013)) studies the spatial sorting of workers who differ in skill level, to shed light on patterns of wage inequality and on the spatial distribution of skills. My research uses similar conceptual tools, borrowed from the assignment literature, to focus on how *firms* sort and impact local labor demand, motivated in particular by the fact that firms are directly targeted by place-based policies.[5] My research is motivated by the empirical finding of Combes et al. (2012). They show that the productivity advantage of firms in large cities is not driven by tougher competition hence stronger selection in larger cities, but by agglomeration effects. Moreover, they find that the most efficient firms are disproportionately more efficient in large cities, indicating potential complementarities between firm productivity and city size. I build on this result and integrate this complementarity in a spatial equilibrium model with mobile and heterogenous firms, a feature absent from their approach. Duranton and Puga (2001) develop a lifecycle model of firm location, in order to explicitly tackle the topic of urban diversity. They propose that diversified cities serve as incubators for new ideas, which are then implemented in specialized cities. Contrary to this research, firms do not differ in productivity types, nor is there heterogeneity in city sizes.

As in Desmet and Rossi-Hansberg (2013) and Behrens et al. (2013), I use structural estimation of a model of a system of cities to assess the welfare implications of the spatial equilibrium. The focus of the analysis is different, since they do not explicitly account for sorting by heterogeneous firms. The paper also contributes to the literature that measures agglomeration externalities, as reviewed in Rosenthal and Strange (2004). There is some empirical evidence that sorting across space matters to understand the wage distribution This literature uses detailed data on workers' characteristics or a fixed effect approach to control for worker heterogeneity and sorting in a reduced form analysis (Combes et al. (2008), Mion and Naticchioni (2009), Matano and Naticchioni (2011)). I use a structural approach to explicitly account for the sorting of firms.

Finally, the counterfactual policy analysis offers a complementary approach to research that assesses the impact of specific place-based policies. The empirical literature has traditionally focused on estimating the local effects of these policies. A notable exception is Kline and Moretti (2013), who develop a methodology to estimate their aggregate effects.[6] They estimate that, following a local productivity boost, additional positive local effects due to the endogenous creation of agglomeration externalities are offset by losses in other parts of the country. My approach explicitly models the reaction of mobile firms to financial incentives, and I find a negative aggregate effect of policies attracting firms to the smallest cities. In contrast, a desirable policy is one that subsidizes firms to choose larger cities, as the decentralized equilibrium features a suboptimal creation of agglomeration externalities. Finally, Glaeser and Gottlieb (2008) study theoretically the economic impact of place-based policies. My analysis brings in heterogeneous firms and the general equilibrium effect of place-based policies on the productive efficiency

[5]The model borrows insights from the assignment model literature, such as Costinot and Vogel (2010), and in particular Eeckhout and Kircher (2012) and Sampson (forthcoming), who focus on the matching of heterogenous firms to heterogenous workers. Here, firms match with heterogenous city sizes.

[6]Albouy (2012) focuses on a related question. He argues that federal taxes impose a de facto unequal geographic burden since they do not account for differences in local cost of living, and estimates the corresponding welfare cost. On the measurement of local effects, see, for example, Busso et al. (2013) for the US, Mayer et al. (2012) for France and Criscuolo et al. (2012) for the UK.

of the country.

The paper is organized as follows. Section 2 presents the model and its predictions. Section 3 details the empirical analysis. I show salient features from French firm-level data that are consistent with the forces at play in the model. I then structurally estimate the model using simulated method of moments. In section 4, I conduct a counterfactual policy analysis using the estimated model. Section 5 concludes.

2. A Model of the Location Choice of Heterogeneous Firms

Consider an economy in which production takes place in locations that I call cities. Cities are constrained in land supply, which acts as a congestion force. The economy is composed of a variety of sectors. Within sectors, firms are heterogeneous in productivity. They produce, in cities, using local labor and traded capital. Non-market interactions within cities give rise to positive agglomeration externalities. I assume that they have heterogeneous effects on firms, in the sense that more efficient firms are more able to leverage local externalities. Firms' choice of city results from a trade-off between the strength of local externalities, the local level of input prices and, possibly, the existence of local subsidies. Heterogeneous firms face different incentives, which yields heterogeneity in their choice. I follow Henderson (1974) and postulate the existence of a class of city developers. In each potential city site, a developer represents local landowners and competes against other sites to attract firms. City developers play a coordination role in the creation of cities, which leads to a unique equilibrium of the economy. The model describes a long-run steady state of the economy and abstracts from dynamics. All derivations and proofs are detailed in the supplemental material. Extensions of the model that feature costly trade, imperfect sorting and no city developers are discussed at the end of the analysis.

2.1. Set-up and agents' problem

2.1.1. Cities

Each potential city site has a given stock of land, normalized to 1. Sites are identical ex-ante. Cities emerge endogenously on these sites, potentially with different population levels L. In what follows, I index cities and all the relevant city-level parameters by city size L. City size is sufficient to characterize all the economic forces at play, in the tradition of models of systems of cities pioneered by Henderson (1974). In particular the distance between two cities plays no role as goods produced in the economy are either freely traded between cities within the country, or are, in the case of housing, non tradable.

Land is used to build housing, which is divisible and consumed by workers. Atomistic landowners construct housing h^S by combining their land γ with local labor ℓ, according to the housing production function

$$h^S = \gamma^b \Big(\frac{\ell}{1-b}\Big)^{1-b}. \tag{1}$$

Landowners compete in the housing market, taking both the housing price $p_H(L)$ and the local wage $w(L)$ as given.

2.1.2. Workers

Set-up There is a mass N of identical workers. Each worker is endowed with one unit of labor. A worker lives in the city of his choosing, consumes a bundle of traded goods and housing, and is paid the local wage $w(L)$. Workers' utility is

$$U = \left(\frac{c}{\eta}\right)^{\eta} \left(\frac{h}{1-\eta}\right)^{1-\eta}, \tag{2}$$

where h denotes housing and c is a Cobb-Douglas bundle of goods across S sectors and a CES bundle of varieties within sector, defined as

$$c = \prod_{j=1}^{j=S} c_j^{\xi_j} \ , \ \text{with} \quad \sum_{j=1}^{j=S} \xi_j = 1 \text{ and } \ c_j = \left[\int c_j(i)^{\frac{\sigma_j - 1}{\sigma_j}} di\right]^{\frac{\sigma_j}{\sigma_j - 1}}.$$

I denote by $P = \left[\prod_{j=1}^{S} (\frac{P_j}{\xi_j})^{-\xi_j}\right]^{-1}$ the aggregate price index for the composite good c. Since goods are freely tradable, the price index is the same across cities. Workers are perfectly mobile and ex ante identical.

Workers' problem Workers in city L consume $c(L)$ units of the good and $h(L)$ units of housing to maximize their utility (2), under the budget constraint $P\, c(L) + p_H(L)h(L) = w(L)$. Given (1) and the housing market clearing condition, the quantity of housing consumed by each worker in equilibrium in city L is

$$h(L) = (1-\eta)^{1-b} L^{-b}. \tag{3}$$

Housing consumption is lower in more populous cities because cities are constrained in space. This congestion force counterbalances the agglomeration-inducing effects of positive production externalities in cities and prevents the economy from complete agglomeration into one city.

Since workers are freely mobile, their utility must be equalized in equilibrium across all inhabited locations to a level $\bar{U}$. In equilibrium, wages must increase with city size to compensate workers for congestion costs, according to

$$w(L) = \bar{w}((1-\eta)\, L)^{b\frac{1-\eta}{\eta}}, \tag{4}$$

where $\bar{w} = \bar{U}^{\frac{1}{\eta}} P$ is an economy-wide constant to be determined in the general equilibrium.

2.1.3. Firms

Production The economy consists of S sectors that manufacture differentiated tradable products. Sectors are indexed by $j = 1, ..., S$. Firms produce varieties using two factors of production that have the following key characteristics. One has a price that increases with city size; the other has a constant price across cities. For simplicity, I consider only one factor whose price depends on city size: labor. In particular, I do not consider land directly in the firm production function. I call the other factor capital, as a shorthand for freely tradable inputs. Capital is provided competitively by absentee capitalists. The price of capital is fixed exogenously in international markets, and the stock of capital in the country adjusts to the demand of firms.[7]

Within their sectors, firms differ exogenously in efficiency z. A firm of efficiency z in sector j and city of size L produces output according to the following Cobb-Douglas production function

$$y_j(z,L) \;=\; \psi(z,L,s_j)\; k^{\alpha_j}\; \ell^{1-\alpha_j},$$

where ℓ and k denote labor and capital inputs, α_j is the capital intensity of all firms in sector j and $\psi(z,L,s_j)$ is a firm-specific Hicks-neutral productivity shifter, as detailed below. It is determined by the firm's 'raw' efficiency, the extent of the local agglomeration externalities and a sector-specific parameter s_j.

Firms engage in monopolistic competition. Varieties produced by firms are freely tradable across space: there is a sectoral price index that is constant through space. Firms take it as given. What matters for location choice is the trade-off between production externalities and costs of production. The relative input price varies with city size. Wages increase with L (equation (4)), whereas capital has a uniform price. Therefore, the factor intensity of a firm shapes, in part, its location decision. A more labor-intensive firm faces, all else equal, a greater incentive to locate in a smaller city where wages are lower.

Productivity and agglomeration The productivity of a firm $\psi(z,L,s_j)$ increases with its own 'raw' efficiency z and with local agglomeration externalities that depend on city size L. The productivity function is also indexed by a sector-specific parameter s_j. I explain the roles of these parameters in turn.

A key assumption of the model is that the productivity of a firm $\psi(z,L,s_j)$ exhibits a strong complementarity between local externalities and the 'raw' efficiency of the firm. This assumption is driven by the findings of Combes et al. (2012), who study a wide set of French industries and provide evidence that more efficient firms are disproportionately more productive in larger cities, pointing to such a complementarity as a potential explanation for this fact. Moreover, in Section 3, I present a set of stylized facts on French firms' location and production patterns. They are consistent with sorting, a consequence of the assumed complementarity.

Knowledge spillovers can arguably exhibit this type of complementarity. More efficient firms can better leverage the local information they obtain. A similar idea, though for individual agents, is provided by Davis and Dingel (2012). In their model, more able individuals optimally spend less time producing and more time leveraging local knowledge, which increases their productivity, leading to such a complementarity.

In what follows, I remain agnostic on the source of agglomeration externalities and their specific func-

[7]Featuring this input to production allows the model to capture that input-use intensity is one of the determinants of location choice of firms. Beyond allowing to capture these type of effects, capital is not necessary to build the equilibrium of the model.

tional form. This allows me to highlight the generic features of an economy with such complementarities. I let the productivity $\psi(z, L, s)$ have the following properties:

Assumption A *$\psi(z, L, s)$ is log-supermodular in city size L, firm raw efficiency z and sectoral characteristic s, and is twice differentiable. In addition, $\psi(z, L, s)$ is strictly log-supermodular in (z, L). That is,*

$$\frac{\partial^2 log\, \psi(z, L, s)}{\partial L \partial z} > 0, \quad \frac{\partial^2 log\, \psi(z, L, s)}{\partial L \partial s} \geq 0, \quad and \quad \frac{\partial^2 log\, \psi(z, L, s)}{\partial z \partial s} \geq 0.$$

I introduce a sector-specific parameter s_j that allows sectors to vary in the way they benefit from local urbanization externalities. Rosenthal and Strange (2004) note that empirical studies suggest that the force and scope of agglomeration externalities vary across industries. More specifically, Audretsch and Feldman (1996) suggest that the benefits from agglomeration externalities are shaped by an industry's life-cycle and that highly innovative sectors benefit more strongly from local externalities than mature industries. I index industries such that, in high s sectors, firms benefit from stronger agglomeration forces, for a given city size. In the estimation of the model, I allow for parameter values that shut down the heterogeneous effect between agglomeration externalities and firm efficiency. The specification I retain for ψ nests the typical specification considered in the literature, where only agglomeration forces of the form $\psi = zL^s$ are at play.[8]

Finally, I restrict the analysis to productivity functions $\psi(z, L, s)$ for which the firms' problem is well defined and concave, absent any local subsidies, for all firms. In other words, I assume that the positive effects of agglomeration externalities are not too strong compared to the congestion forces. A sufficient condition for this, given that the congestion forces increase with city size with a constant elasticity, is that agglomeration externalities have decreasing elasticity with respect to city size for any firm type z.

Entry and location choice There is an infinite supply of potential entrants who can enter the sector of their choosing. Firms pay a sunk cost f_{Ej} in terms of the final good to enter sector j, then draw a raw efficiency level z from a distribution $F_j(.)$.[9] Once firms discover their raw efficiency, they choose the size of the city where they want to produce. Contrary to the setting in Melitz (2003), the model abstracts from any selection of firms at entry, since there is no fixed cost to produce. I focus instead on *where* firms decide to produce once they discover their efficiency, and how this shapes the spatial equilibrium of the economy. That is, rather than selection on entry, I focus on selection on city size.

Firms' problem A firms' choice of city size is influenced by three factors. First, relative input prices vary by city size. Second, firm productivity increases with city size, through greater agglomeration externalities. Third, local city developers compete to attract firms to their cities by subsidizing profits at rate $T_j(L)$, which varies by city-size and sector.[10] The firm's problem can be solved recursively. For

[8] In that case $\frac{\partial^2 log\, \psi(z,L,s)}{\partial L \partial z} = 0$, $\frac{\partial^2 log\, \psi(z,L,s)}{\partial s \partial z} = 0$ and $\frac{\partial^2 log\, \psi(z,L,s)}{\partial L \partial s} > 0$.

[9] I assume that this distribution is an interval (possibly unbounded) on the real line. This assumption is made for tractability; the results carry through without it, although the notation is more cumbersome.

[10] The subsidy offered by city developers $T_j(L)$ may in principle differ across cities for cities of the same size. As I show in the next section, this is not the case in equilibrium. Anticipating this, I abstract from introducing cumbersome notation denoting different cities of the same size.

a given city size, the problem of the firm is to hire labor and capital and set prices to maximize profits, taking as given the size of the city (and hence the size of the externality term), input prices, and subsidies. Then, firms choose location to maximize this optimized profit.[11]

Consider a firm of efficiency z producing in sector j and in a city of size L. Firms hire optimally labor and capital, given the relative factor prices $\frac{w(L)}{\rho}$ – where ρ denotes the cost of capital – and their local productivity $\psi(z, L, s_j)$. This choice is not distorted by local subsidies to firms' profits. Firms treat local productivity as exogenous, so that the agglomeration economies take the form of external economies of scale. Given the CES preferences and the monopolistic competition, firms set constant markups over their marginal cost. This yields optimized profits for firm z in sector j as a function of city size L

$$\pi_j(z, L) = \kappa_{1j}\big(1 + T_j(L)\big)\left(\frac{\psi(z, L, s_j)}{w(L)^{1-\alpha_j}}\right)^{\sigma_j - 1} R_j\, P_j^{\sigma_j - 1}, \tag{5}$$

where P_j is the sectoral price index, R_j is the aggregate spending on goods from sector j and $\kappa_{1j} = \frac{((\sigma_j-1)\alpha_j^{\alpha_j}(1-\alpha_j)^{1-\alpha_j}(\rho P)^{-\alpha_j})^{\sigma_j-1}}{\sigma_j^{\sigma_j}}$ is a sector-specific constant.

Note that firm employment, conditional on being in a city of size L, is given by

$$\ell_j(z, L) = (1 - \alpha_j)(\sigma_j - 1)\frac{\pi_j(z, L)}{w(L)\big(1 + T_j(L)\big)}. \tag{6}$$

The proportionality between profits - net of subsidies - and the wage bill is a direct consequence of constant factor shares, implied by the Cobb-Douglas production function, and of constant markup pricing. The problem of the firm thus is to choose the city size L to maximize (5).

2.1.4. City developers

Set-up There is one city-developer for each potential city site. City developers fully tax local landowners. They are therefore the residual claimants on local land value. Their objective is to maximize these revenues, net of the cost of the policies they put in place to maximize local land value. They compete to attract firms to their city by subsidizing firms' profits. Absent these developers, there would be a coordination failure as atomistic agents alone - firms, workers or landowners - cannot create a new city. City developers are, in contrast, large players at the city level. As in Henderson (1974), city developers act as a coordinating device that allows a unique equilibrium to emerge in terms of city-size distribution. There is perfect competition and free entry among city developers, which drives their profits to zero in equilibrium.

City developers' problem Each city developer announces a subsidy to local firms' profits in sector j, $T_j(L)$, which may depend on city size L.[12] Developers are funded by fully taxing profits made on the housing market. As the housing market clears in each city, aggregate landowner profits at the city level

[11] In reality, there are two types of sorting: ex-post sorting - that is, firms that are already established and decide to change location - and ex-ante sorting - that is, new firms, or new establishments of an existing firm, being created somewhere. Since the model is static, it conflates both types of sorting.

[12] Alternatively, one could consider firm-specific subsidies rather than ad-valorem subsidies. The equilibrium I find here is still an equilibrium in this case, as shown in online Appendix B.3.

are

$$\pi_H(L) = b(1-\eta)Lw(L). \tag{7}$$

It will prove useful when solving for the equilibrium to note that a constant share of the local labor force is hired to build housing, namely

$$\ell_H(L) = (1-b)(1-\eta)L. \tag{8}$$

A city developer developing a city of size L faces the following problem:

$$\begin{aligned} &\max_{\{T_j(L)\}_{j\in 1,\dots,S}} \Pi_L = b(1-\eta)w(L)L - \sum_{j=1}^{S}\int_z T_j(L)\frac{\pi_j(z,L)}{1+T_j^i(L)}\,\mathbb{1}_j(z,L)dF_j(z), \\ &\text{such that} \\ &\qquad \mathbb{1}_j(z,L) = 1 \qquad \text{if firm } z \text{ chooses their city },\\ &\qquad \mathbb{1}_j(z,L) = 0 \qquad \text{otherwise.} \end{aligned} \tag{9}$$

In this expression, $F_j(.)$ is the distribution of firm's raw efficiencies in sector j and $\pi_j(z,L)$ is the local profit of a firm of efficiency z in sector j, as defined in (5).

2.2. Spatial equilibrium

Having set up the problems of workers, firms, landowners and city developers, I am now ready to solve for the equilibrium of the economy. I show that this equilibrium exists and is unique.

2.2.1. Equilibrium definition

Definition 1 *An equilibrium is a set of cities $\mathcal{L}$ characterized by a city-size distribution $f_L(.)$, a wage schedule $w(L)$, a housing-price $p_H(L)$ and for each sector $j = 1,...,S$ a location function $L_j(z)$, an employment function $\ell_j(z)$, a capital-use function $k_j(z)$, a production function $y_j(z)$, a price index P_j and a mass of firms M_j such that*

(i) workers maximize utility (equation (2)) given $w(L), p_H(L)$ and P_j,

(ii) utility is equalized across all inhabited cities,

(iii) firms maximize profits (equation (5)) given $w(L), \rho$ and P_j,

(iv) landowners maximize profits given $w(L)$ and $p_H(L)$,

(v) city developers choose $T_j(L)$ to maximize profits (equation (9)) given $w(L)$ and the firm problem,

(vi) factors, goods and housing markets clear; in particular, the labor market clears in each city,

(vii) capital is optimally allocated, and

(viii) firms and city developers earn zero profits.

In what follows, I present a constructive proof of the existence of a such an equilibrium. Furthermore, I show that the equilibrium is unique, and stable. As is standard in the literature, I allow for the possibility of a non-integer number of cities of any given size (see Abdel-Rahman and Anas (2004) for a review and more recently Rossi-Hansberg and Wright (2007b) or Behrens et al. (2014)).

Proposition 1 *There exists a unique equilibrium of this economy.*

The equilibrium is unique in terms of distribution of outcomes, such as firm-size distribution, city-size distribution and matching functions between firms and city sizes. It is not, of course, unique in terms of *which site* is occupied by a city of a given size, as all sites are identical ex ante.

2.2.2. Constructing the spatial equilibrium

The equilibrium is constructed in four steps. First, I solve for the equilibrium subsidy offered by city developers. Second, I show that it pins down how firms match with city sizes, as well as the set of city sizes generated in equilibrium by city developers. Third, general equilibrium quantities are determined by market clearing conditions and free entry conditions in the traded goods sectors, once we know the equilibrium matching function from step 2. Finally, the city-size distribution is determined by these quantities, using labor-market clearing conditions. In each step, the relevant functions and quantities are uniquely determined; hence, the equilibrium is unique.

Step 1: Equilibrium subsidy

Lemma 2 *In equilibrium, city developers offer and firms take-up a constant subsidy to firms' profit* $T_j^* = \frac{b(1-\eta)(1-\alpha_j)(\sigma_j-1)}{1-(1-\eta)(1-b)}$ *for firms in sector* j*, irrespective of city size* L *or firm type* z*.*

Formally, each city developer announces a subsidy $T_j^*\delta(L - L_i)$ for a city size L_i where $\delta(0) = 1$, and $\delta(x) = 0$ for $x \neq 0$. A city developer is indifferent between attracting firms from one or many sectors. Therefore, there is an indeterminacy in equilibrium as to which sector(s) each city developer targets.

I sketch the proof in the case of an economy with only one traded goods sector. The formal proof with S sectors follows the same logic and is given in online Appendix C. City developers face perfect competition, which drives their profits down to zero in equilibrium. Their revenues correspond to the profits made in the housing sector (equation (7)), which are proportional to the aggregate wage bill in the city $w(L)L$. They compete to attract firms by subsidizing their profits. In equilibrium, irrespective of which firms choose to locate in city L, these profits will also be proportional to the sectoral wage bill $w(L)N$, where N is the labor force hired in the traded goods sector locally, as can be seen from equation (6). Finally, the local labor force works either in the housing sector (equation (8)) or the traded goods sector, so that $N = L(1 - (1 - b)(1 - \eta))$. Profits given by (9) simplify to $b(1 - \eta)w(L)L - T\frac{(1-(1-b)(1-\eta))}{(\sigma-1)(1-\alpha)}w(L)L$. The choice of city size is irrelevant, and T^* is the only subsidy consistent with zero profits. City developers that offer lower subsidies will not attract any firm, hence will not create cities. City developers that offer higher subsidies attract firms but make negative profits. Note that the equilibrium subsidy does not depend on city size, hence equilibrium subsidies do not alter the location choice of firms compared to a world without subsidy, conditional on the same cities existing. Therefore, the properties of the equilibrium in terms of the distribution of firms' outcomes, which I detail later, are valid both in the equilibrium with city developers, and in one without developers where cities would be exogenously given.

Step 2: Equilibrium city sizes and the matching function The city developers' problem determines the equilibrium city sizes generated in the economy. Cities are opened up when there is an incentive for city developers to do so, i.e. when there exists a set of firms and workers that would be better off choosing this city size. Workers are indifferent between all locations, but firms are not, since their profits vary with city size. Given the equilibrium subsidy T_j^* offered by city developers, the profit function of firm z in sector j is:

$$\pi_j^*(z, L) = \kappa_{1j}\big(1 + T_j^*\big)\left(\frac{\psi(z, L, s_j)}{w(L)^{1-\alpha_j}}\right)^{\sigma_j - 1} R_j\, P_j^{\sigma_j - 1} \tag{10}$$

The problem of the firm is simply summarized by the following first-order condition, given the wage in equation (4) and writing $\psi_2(z, L, s_j) = \frac{\partial\psi(z,L,s_j)}{\partial L}$:

$$\frac{\psi_2(z, L, s_j)\, L}{\psi(z, L, s_j)} = (1 - \alpha_j)\, b\, \frac{1 - \eta}{\eta} \tag{11}$$

There is a unique profit-maximizing city size for a firm of type z in sector j, under the regularity conditions I have assumed. Define the optimal city size as follows

$$L_j^{**}(z) = \arg\max_{L \geq 0} \pi_j^*(z, L). \tag{12}$$

Assume that, for some firm type z and sector j, no city of size $L_j^{**}(z)$ exists. There is then a profitable deviation for a city developer on an unoccupied site to open up this city. It will attract the corresponding firms and workers, and city developers will make a positive profit by subsidizing firms at a rate marginally smaller than T_j^*. The number of such cities adjusts so that each city has the right size in equilibrium. This leads to the following lemma, letting $\mathcal{L}$ denote the set of city sizes in equilibrium:

Lemma 3 *The set of city sizes $\mathcal{L}$ in equilibrium is the optimal set of city sizes for firms.*

Given this set of city sizes, the optimal choice of each firm is fully determined. Define the matching function

$$L_j^*(z) = \arg\max_{L \in \mathcal{L}} \pi_j^*(z, L). \tag{13}$$

It is readily seen that the profit function of the firm (equation (10)) inherits the strict log-supermodularity of the productivity function in z and L. Therefore, the following lemma holds.

Lemma 4 *The matching function $L_j^*(z)$ is increasing in z.*

This result comes from a classic theorem in monotone comparative statics (Topkis (1998)). The benefit to being in larger cities is greater for more productive firms and only they are willing in equilibrium to pay the higher factor prices there. Furthermore, the matching function is fully determined by the firm maximization problem, conditional on the set of city sizes $\mathcal{L}$. As seen from equation (10), this optimal choice does not depend on general equilibrium quantities that enter the profit function proportionally for all city sizes. Finally, under the regularity assumptions made on ψ as well as on the distribution of z, $F_j(.)$, the optimal set of city sizes for firms in a given sector is an interval (possibly unbounded). The sectoral matching function is invertible over this support. For a given sector, I use the notation $z_j^*(L)$ to denote the inverse of $L_j^*(z)$. It is increasing in L. The set of city sizes $\mathcal{L}$ available in equilibrium is the union of the sector-by-sector intervals.

Step 3: General equilibrium quantities The equilibrium has been constructed up to the determination of the following general equilibrium values. The reference level of wages $\bar{w}$ defined in equation (4) is taken as the numeraire. The remaining unknowns are the aggregate revenues in the traded goods sector R, the mass of firms M_j and the sectoral price indexes P_j. I use the following notation:

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

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.[13] 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 (14)), the goods market clearing conditions (equation (15)), and the national labor market clearing condition (workers works either in one of the traded goods sectors or in the construction sector, equation (16)). This leads to the following system of equations:

$$f_{Ej}\, P = (1 + T_j^*)\kappa_{1j}\, S_j\, \xi_j R\, P_j^{\sigma_j - 1}, \text{ for all } j \in \{1, ..., S\}, \tag{14}$$

$$1 = \sigma_j \kappa_{1j}\, M_j\, S_j\, P_j^{\sigma_j - 1}, \text{ for all } j \in \{1, ..., S\}, \tag{15}$$

$$Nv = \sum_{j=1}^{S} \kappa_{2j} E_j\, M_j\, \xi_j R\, P_j^{\sigma_j - 1} + N(1-b)(1-\eta), \tag{16}$$

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 (8). Inverting this system of $2S+1$ equations gives unique the $2S+1$ unknowns P_j, M_j for all $j \in \{1, ..., S\}$ and R, the aggregate revenues in the traded goods sector, as detailed in online Appendix E.

Step 4: Equilibrium city-size distribution The city developers' problem and the firms' problem jointly characterize (1) the set of city sizes that necessarily exist in equilibrium and (2) the matching

[13] 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 $\kappa_{2j} = \kappa_{1j}(1-\alpha_j)(\sigma_j - 1)$.

function between firm type and city size. Given these, the city-size distribution is pinned down by the labor market clearing conditions. The population living in a city of size smaller than any L must equal the number of workers employed by firms that have chosen to locate in these same cities, plus the workers hired to build housing. That is,$\forall L > L_{min}$,:

$$\int_{L_{min}}^{L} u f_L(u)\ du = \sum_{j=1}^{S} M_j \int_{z_j^*(L_{min})}^{z_j^*(L)} \ell_j(z, L_j^*(z))\ dF_j(z) + (1-\eta)(1-b) \int_{L_{min}}^{L} u f_L(u)\ du,$$

where $L_{min} = \inf(\mathcal{L})$ the smallest city size in the equilibrium.

Differentiating this with respect to L and dividing by L on both sides gives the city size density ($f_L(L)$ is not normalized to sum to 1)

$$f_L(L) = \kappa_4 \frac{\sum_{j=1}^{S} M_j \mathbb{1}_j(L)\ \ell_j(z_j^*(L))\ f_j(z_j^*(L))\ \frac{dz_j^*(L)}{dL}}{L}, \tag{17}$$

where $\kappa_4 = \frac{1}{1-(1-\eta)(1-b)}$ and $\mathbb{1}_j(L) = 1$ if sector j has firms in cities L, and 0 otherwise. The equilibrium distribution of city sizes $f_L(.)$ is uniquely determined by equation (17), hence the following lemma:

Lemma 5 $f_L(.)$ *is the unique equilibrium of this economy in terms of the distribution of city sizes.*

Several remarks are in order here. First, the city-size distribution is shaped by the distribution of firm efficiency and by the sorting mechanism. This offers a static view of the determination of the city-size distribution, driven by heterogeneity in firm types. In the empirical exercise, I compute the city-size distribution obtained with equation (17), where firm heterogeneity is estimated from French firm-level data but the city-size distribution is not used in the estimation. It exhibits Zipf's law, consistent with the data on cities. I show in Appendix C.5 that under some parametric restrictions for ψ, consistent with the ones used in the empirical section, the city-size distribution follows Zipf's law if the firm-size distributions follow Zipf's law.

Second, for each city size, the share of employment in each sector can be computed using the same method, now sector by sector. For a given city size, the average sectoral composition over all cities of a given size size L is determined by the model. On the other hand, the model is silent on the sectoral composition of any *individual* city of size L, which is irrelevant for aggregate outcomes. City developers in particular are indifferent to the sectoral composition of their city.[14]

Third, the equilibrium features cities that host a variety of sectors, and sectors that spread out over a range of cities of different sizes. This is contrast to classic urban models that rely on homogenous firms and hence generally feature fully specialized cities and a single city size for each sector.

Finally, I verify in the online Appendix D that this equilibrium is stable and provide there a detailed discussion of stability. This step completes the full characterization of the unique equilibrium of the economy.

[14]This city-level indeterminacy comes from the fact that agglomeration externalities depend on the overall size of the city, and not on its sectoral composition. To lift this, the model could easily be extended to accommodate *localization* externalities. The agglomeration externality depends in that case on the size of a given (set of) sector(s), and not the overall city size. Cities would then be perfectly specialized in that sector(s), since the congestion costs depend on the overall city size, but the benefits are sector(s)-specific. This would not change any other characterization of the equilibrium. In particular, the city-size distribution defined in equation (17) and lemma 5 would still hold.

2.2.3. Model Extensions

I detail several extensions to the model in the supplemental material, and show that the general features of the equilibrium are robust to these extensions. In particular, I show that the model can be extended to feature *costly trade* between cities. Even with costly trade and unequal price indexes across cities, I show that city size L remains a sufficient statistic for the local economic conditions, and that the result that more efficient firms sort into larger cities still obtains.

Second, I describe the characteristics of equilibria without city developers. If the proof of uniqueness of equilibrium relies on their existence, the properties of the equilibrium - and in particular the results on firm sorting – do not. That is, if one takes as exogenous a given set of city sizes $\mathcal{L}$ and assumes away city-developers, then more productive firms still sort into larger cities, and the equilibrium of this economy is still characterized by the propositions of section 2.3.

Finally, I examine the properties of the model in the presence of imperfect sorting as hypothesized in the empirical specification of section 3. The properties of equilibrium described below are either unchanged, or hold true on average, rather than systematically, in that case.

2.3. Properties of the equilibrium

2.3.1. Within-sector patterns

Within a given sector j, the revenue, production and employment distributions are all determined by the matching function $L_j^*(z)$. In the sorting equilibrium, for a firm of efficiency z, productivity, revenues and employment are given by

$$\psi_j^*(z) = \psi(z, L^*(z), s_j),$$

$$r_j^*(z) = \sigma_j \kappa_{1j} \left(\frac{\psi(z, L^*(z), s_j)}{w(L_j^*(z))^{1-\alpha_j}} \right)^{\sigma_j - 1} P_j^{\sigma-1} R_j, \tag{18}$$

$$\ell_j^*(z) = \kappa_{2j} \frac{\psi(z, L^*(z), s_j)^{\sigma-1}}{w(L_j^*(z))^{(\sigma-1)(1-\alpha_j)+1}} P_j^{\sigma_j - 1} R_j, \tag{19}$$

where the starred variables denote the outcomes in the sorting equilibrium. Since there is positive assortative matching between a firm's raw efficiency and city size (lemma 4), firm-level observables also exhibit complementarities with city size. Let $\mathcal{L}$ denote the set of city sizes in the economy.

Proposition 6 *In equilibrium, within each sector, firm revenues, profits and productivity increase with city size, in the following sense. For any $L_H, L_L \in \mathcal{L}$ such that $L_H > L_L$, take z_H such that $L_j^*(z_H) = L_H$ and $L_j^*(z_L) = L_L$. Then, $r_j^*(z_H) > r_j^*(z_L)$, $\pi_j^*(z_H) > \pi_j^*(z_L)$, and $\psi_j^*(z_H) > \psi_j^*(z_L)$.*

These strong predictions on the ranking of the size of firms (in revenues or productivity) vis a vis the city size are a direct consequence of the perfect sorting of firms. In contrast, employment can be either positively or negatively associated with city size through the effect of wages. Within a sector, $\ell^*(z) \propto \frac{r^*(z)}{w(L^*(z))}$, where both revenues and wages increase with city size. Firms may have lower employment in larger cities, even though they are more productive and profitable. More precisely, if $\epsilon_l = \frac{d \log \bar{\ell}^*(L)}{d \log L}$ and

$\epsilon_r = \frac{d \log \bar{r^*}(L)}{d \log L}$ are the elasticities of mean employment and mean revenues with respect to city size in equilibrium, then

$$\epsilon_l = \epsilon_r - b\frac{1-\eta}{\eta}, \tag{20}$$

so that ϵ_l is not necessarily positive.

2.3.2. Comparative statics across sectors

I now compare the predicted distribution of firm outcomes across sectors. Sectors differ in their capital intensity α_j and in the strength of their benefit from agglomeration externalities s_j. Both impact the sorting process, leading in turn to differences in observed outcomes. The following comparative statics exercises examine how the geographic and size distribution of firms in a sector vary with each parameter holding all other sectoral characteristics constant, in particular the distribution of raw efficiencies $F(.)$.[15]

Geographic distribution Define the *geographic distribution* of firms in a sector as the probability that a firm from the sector is in a city of size smaller than L. That is, let

$$\tilde{F}(L; \alpha_j, s_j) = P(\text{firm from sector } (\alpha_j, s_j) \text{ is in a city of size smaller that } L).$$

Proposition 7 *The geographic distribution $\tilde{F}_j$ of a high α_j sector first-order stochastically dominates that of a lower α_k sector, all else equal. The geographic distribution $\tilde{F}_j$ of a high s_j sector first-order stochastically dominates that of a lower s_k sector, all else equal.*

These results stem from the following observation. As shown before, the matching function $L_j^*(z)$ is always increasing, but its slope and absolute level depend on the capital intensity α_j and the strength of agglomeration externalities s_j in the sector. In labor-intensive sectors, the weight of the wage effect is heavier in the trade-off between the benefits of agglomeration externalities and labor costs. This pushes the matching function down, towards smaller cities. For any city size threshold, there are more firms from a labor-intensive sector that choose to locate in a city smaller than the threshold. In contrast, in sectors with strong agglomeration externalities, firms benefit more from a given city size, which pushes the matching function up for all firms. All else equal, they locate more in larger cities.

Firm-size distribution The intensity of sorting, which reinforces initial differences between firms, impacts the dispersion of the observed sectoral firm-size distribution. Let $Q_j(p)$ denote the p-th quantile of the firm revenue distribution in sector j.

Proposition 8 *All else equal, if $(\alpha_j, s_j) \geq (\alpha_i, s_i)$, the observed firm-size distribution in revenues is more spread in Sector 2 than in Sector 1. For any $p_1 < p_2 \in (0,1)$, $\frac{Q_i(p_2)}{Q_i(p_1)} \leq \frac{Q_j(p_2)}{Q_j(p_1)}$.*

[15] Note that if firms where not heterogeneous, both these distributions would be degenerate. In every sector, there would be one firm size, and firms would be located in one city size.

In other words, if one normalizes the median of the revenue distribution to a common level, all higher quantiles in the revenue distribution of Sector 2 are strictly higher than in Sector 1, and all lower quantiles are below. This comes from the fact that the distribution of firm revenues is shaped not only by the distribution of raw efficiencies (held constant across sectors in this comparative statics exercise), but also by the complementarity between z and city size, whose choice is endogeneous. In higher-s sectors for example, there is more to be gained by more productive firms to locate in larger cities. As a consequence, the difference in city size choice, and in turn in firm revenues, is larger between high- and low-z firms in a higher-s sector compared to a lower-s sector. The distribution of firms outcomes is more unequal.

In particular, higher α_j or higher s_j sectors have thicker upper-tails in their firm-size distributions. This leads to a characterization that will prove useful empirically. Firm-size distributions are empirically well approximated by power law distributions, in their right tail. The exponent of this distribution characterizes the thickness of the tail of the distribution. Assume that the revenue distribution of firms in two sectors 1 and 2 can be approximated by a power law distribution in the right tail, with respective exponents ζ_1 and ζ_2. Then the following corollary holds

Corollary 9 *Let $(\alpha_2, s_2) \geq (\alpha_1, s_1)$. The tail of the firm-size distribution in Sector 2 is thicker than the tail of the firm-size distribution in sector 1: $\zeta_2 \leq \zeta_1$.*

2.4. Welfare analysis

Having characterized the positive properties of the equilibrium, I now turn to studying its welfare implications. This theoretical analysis is complemented in section 4 by a set of quantitative counterfactual policy analyses, in which I quantify the welfare implications of typical spatial policies.

I consider the problem of a benevolent social planner who freely chooses allocations in this economy so as to maximize total welfare in spatial equilibrium, i.e. under the constraint of free mobility of workers.[16] In order to focus on the inefficiencies that arise in the traded good sector, I take the housing sector as given in what follows, i.e. a constant fraction of the local labor force is used to build housing as in the competitive equilibrium. Welfare could potentially be improved beyond what is laid out here through an intervention on the housing market.[17]

The problem of the social planner, formally stated and solved in online Appendix F, is to choose allocations of firms and workers - in particular, she chooses firm's location, firm's employment and firm's production, as well as the consumption and location of workers, so as to maximize welfare. In her choice of city sizes she faces a trade-off between increasing productive efficiency by creating larger cities to leverage agglomeration externalities on the one hand, and limiting the disutility from congestion borne by workers on the other hand. The first-order condition for the location choice of firm z in the social planner's problem writes:

[16] Total welfare could be further improved if this constraint of equal utility of workers across inhabited cities was lifted. I do not consider this case since this equilibrium would not be stable. Some workers would always have an incentive to move to increase their utility. This alternative equilibrium would also raises equity issues as identical workers would have different levels of utility in equilibrium.

[17] A benefit of this approach is that these results hold irrespective of the source of congestion that I consider, as long as the congestion force increases log-linearly with city size.

$$\frac{\psi_2(z,L,s)L}{\psi(z,L,s)} = b\frac{1-\eta}{\eta}(1-\alpha)\chi(z), \tag{21}$$

where $\chi(z) < 1$ is a wedge between the first-order condition in the market equilibrium (11) and the one of the social planner. Firms choose cities that are too small in the market equilibrium relative to the optimal spatial equilibrium.

Proposition 10 *The equilibrium is suboptimal. Firms locate in cities that are too small. To reach the optimum, the first-best policy taxes wages offered by firms in smaller cities and subsidizes firms' wages in larger cities, according to a tax/subsidy schedule that varies monotonically with city size.*

The intuition for this result is as follows. The social marginal benefit of choosing a larger city is higher than the private benefit perceived by firms through their profit function. The benefit that is not internalized by firms is that by choosing a larger city they increase the productivity of the economy as a whole which decreases the cost of entry for all firms. Fostering entry increases welfare, by the love of variety effect. Firms ignore the effect of their choice of city size on the cost of entry, and therefore choose cities that are too small compared to the optimum. This general equilibrium cross-city effect is not internalized by firms nor by city developers who, despite being large local players, are still atomistic at the national level.

To align firms' incentives to the solution to the social planner's problem, wages have to be subsidized/taxed so the wage schedule paid by firms is

$$w(L) \propto (L^{b\frac{1-\eta}{\eta}} + A) \tag{22}$$

where A is a constant, as opposed to $w(L) \propto L^{b\frac{1-\eta}{\eta}}$ in the competitive equilibrium.

This analysis helps see that "threshold-type" spatial policies, in which firms are subsidized when they locate in cities smaller than a given size, are not intuitively welfare enhancing. They tend to distort the choice of city size in the wrong direction. They attract firms to cities that are smaller (rather than larger) than the one they choose in the competitive equilibrium. Desirable policies on the other hand are ones that tend to flatten out the wage schedule, making the wage increase less fast with city size than it would otherwise. I explore these points further, quantitatively, in section 4.

3. Estimation of the Model

I now take the model to the data, in order to be able to perform a quantitative policy analysis. Using French firm-level data, I first show that sectors display location patterns and firm-size distribution characteristics that are consistent with the theoretical predictions. I then structurally estimate the model.

3.1. Data

The firm-level data set of French firms that I use contains information on the balance sheets of French firms, declared for tax purposes. All firms with revenues over 730,000 euros are included. It reports information on employment, capital, value added, production, and 3-digit industry classification. It is matched with establishment-level data, which indicate the geographical location at the postal code level of each establishment of a given firm-year. As is standard in the literature, the geographical areas I use to measure city size are the 314 French commuting zones, or *"Zones d'emploi"* (employment zones), within metropolitan France. They are defined with respect to the observed commuting patterns of workers and cover all of France. They are designed to capture local labor markets and are better suited than administrative areas, which they abstract from, to capturing the economic forces at play in the model.[18] To measure the size of the city, I use the total local employment of the area, since I need a proxy for externalities such as knowledge spillovers or labor market pooling that depend on the size of the workforce. I use the data for the year 2000 in the estimation procedure.

I retain only tradable sectors in the analysis, consistent with the assumptions of the model. The set of industries is the one considered in Combes et al. (2012), i.e., manufacturing sectors and business services, excluding finance and insurance. I trim the bottom and top 1% of the data. This leaves me with 157,070 firms. Summary statistics are reported in Table 1.

3.2. Descriptive evidence on sorting

Before proceeding to the structural estimation of the model, I present a first look at the raw data. My objective is to check that the comparative statics of the model are broadly consistent with the patterns exhibited in the data. Recall that in the model, the complementarity between firm efficiency and agglomeration forces leads to the sorting of firms across cities of different sizes. This impacts the elasticity of firm-level observables with respect to city size within industries (prop. 6). Furthermore, firm sorting is shaped by two key sectoral parameters, namely, the sectoral strength of agglomeration externalities s_j as well as the sectoral intensity of use of traded inputs α_j. The model shows how these parameters shape (i) the location patterns of firms in a given industry (prop. 7) and (ii) the dispersion of the sectoral firm-size distribution (prop. 8). I turn to examining the raw data in these dimensions.

To do so, I use the most disaggregated level of industry available in the data. I keep sectors with more than 200 observations, for a total of 146 industries, and present correlations between different sectoral characteristics, guided by the theory. These correlations could be driven by explanations alternative to the ones I propose in the model. To mitigate these concerns, I check that the patterns I find are robust to a set of sectoral controls that I detail below. The broad consistency of the data with the salient features of the model are only suggestive evidence that sorting forces may be at play.

I first investigate how, in each sector, average firm value added and average firm employment change as city size increases.[19] In the model, the elasticity of firm revenues to city size is positive within industries whereas the elasticity of employment to city size is strictly lower and possibly negative. Empirically, I compute the average firm-level value added and employment by industry and city and compute their

[18] They are presented as areas where *"most workers live and work, and where establishments can find most of their workforce"*. The previous definition of these zones was constrained by some administrative borders. I use a new definition of these zones published by INSEE in 2011 that abstracts from these constraints. These zones are a collection of towns ("code commune"). I use a concordance table between these and postal codes to classify the firm data in terms of zone d'emploi.

[19] Because the model does not feature the use of intermediates, I use value added as the measure of firm output.

elasticity with respect to city size.[20] Figure 1 plots the distributions of these elasticities. The elasticity of employment to city size almost always lies below the elasticity of value added to city size. For value added, it is positive for 85% of industries, corresponding to 93% of firms, and is significantly negative for only one industry, the manufacture of kitchen furniture. This is broadly consistent with the intuition of the model.

Second, the model suggests that firm location choices are linked to sectoral characteristics and, in particular, the intensity of input use, which I can measure in the data. To proxy for inputs whose price does not systematically increase with city size, I use a measure of "tradable capital" defined as total capital net of real estate assets. I measure tradable capital intensity, α^K, as the Cobb-Douglas share of this tradable capital in value added. I then run the following regression:

$$share_j = \beta_0 + \beta_1 \alpha_j^K + \beta_2 X_j + \epsilon_j,$$

where j indexes sectors, $share_j$ is the share of establishments in sector j located in large cities (i.e., the largest cities that hosts half of the population) and X_j is a set of control variables varying at the industry level. Table 2 reports the coefficient estimates. It shows that industries that use more tradable capital are significantly more likely to be located in larger cities. However, these industries could also be the ones with higher skill intensity, driven to larger cities in search of skilled workers. To control for that, I use an auxiliary data set to measure industry-level skill intensity.[21] Specification (III) in Table 2 shows that controlling for industry-level skill intensity does not affect the results. Specification (IV) runs the same regression, limiting the sample to export-intensive industries. This control aims at mitigating the concern that location choice may be driven by demand-side explanation, whereas the model focuses on the supply side. Again, the results are robust to using this reduced sample. Overall, Table 2 is consistent with the idea that firms location choices are shaped by the intensity of input use in their industry.[22]

Third, the model predicts that firms that locate in large cities benefit disproportionally from agglomeration externalities. As a consequence, the sectoral firm-size distribution is more fat-tailed for industries located in larger cities. Table 3 correlates the thickness of the industry-level firm-size distribution, summarized by its shape parameter ζ_j, with the share of establishments located in large cities in industry j.[23] Table 3 shows a negative correlation between ζ_j and the fraction of establishments in industry j located in large cities (defined as in Table 2). In other words, industries that locate more in large cities also have thicker-tailed firm-size distributions. This negative correlation is robust to controlling for the number

[20]For this measure, I restrict the sample to single-establishment firms as the data on value added is only available at the firm level for multi-establishment firms. Single establishment account for 83% of firms and 44% of employment in the sample. I verify using the full sample that, reassuringly, the employment/city size relationship is not systematically biaised for multi-establishment firms compared to mono-establishment firms.

[21]I use the random sample of 1/12 of the French workforce published by INSEE. It contains information on workers' skill level, salary and industry. For each industry, I measure the share of the labor force that is high-skilled. I define a dummy variable that equals one for sectors with above-median skill-intensity.

[22]I further check that these results are robust to alternative specifications of what constitute a large city (top cities hosting 25% and 33% of the workforce), as well as to dropping business services, which are arguably less tradable, altogether from the analysis.

[23]The shape parameter ζ_j is estimated by running the following regression, following Gabaix and Ibragimov (2011): $\log\left(rank_{ij} - \frac{1}{2}\right) = \alpha_j - \zeta_j \log(\text{value added}_{ij}) + \epsilon_{ij}$, where j indexes industries, i indexes firms and $rank_{ji}$ is the rank of firm i in industry j in terms of value added.

of firms and the average value added in industry i (Specification II), as well as reducing the sample to export-intensive industries (Specification III).

Finally, the model predicts that more efficient firms self-select into larger cities. I investigate this question by focusing on the relocation pattern of movers, i.e. mono-establishment firms that change location from one year to the next. The nature of this question leads me to extend the sample period to 1999-2006. There is no direct way to measure a firm's raw efficiency from the data. However, in the model, within a city-industry pair, firm revenues increase with firm efficiency. I thus compute the following firm-level residual ω_{ijt} and use it to proxy for firm efficiency:

$$\log(\text{value added}^{c}_{ijt}) = \delta^c + \delta^t + \delta^j + \omega_{ijt},$$

where δ^c, δ^t and δ^j are sets of, respectively, city, year and industry fixed-effects, and i is a firm in industry j located in city c in year t. For all firms relocating from year t to $t+1$, I define $\Delta_t \text{City Size}_i$ as $\log(L_{i,t+1}/L_{i,t})$, where $L_{i,t}$ is the size of the city where firm i is located in year t. I then estimate:

$$\Delta_t \text{City Size}_i = \alpha + \beta\omega_{ijt} + X_{it} + \epsilon_{it}$$

where X_{it} includes an industry fixed effect and the logarithm of $L_{i,t}$ or a set of initial city fixed effects.[24] Table 4 shows that, conditional on moving, firms that are initially larger tend to move into larger cities. Similar results obtain when I drop firms that switch industry when they move. I emphasize that this result is a simple correlation and cannot be interpreted causally in the absence of a valid instrument for the selection into the sample of movers. Table 4 simply shows that, among the set of movers, there exists a positive correlation between initial firm size and the size of the city the firm moves into, a correlation pattern that is consistent with sorting.

3.3. Structural estimation

I now turn to the estimation of the model. The model is estimated industry by industry, on 23 aggregated industries. I minimize the distance between moments of the data and their simulated counterparts to estimate the sectoral parameters that govern the model.

3.3.1. Model specification

I first lay out the econometric specification of the model. The literature has traditionally assumed that agglomeration externalities were of the form $\psi(z, L, s_j) = zL^{a_j}$, where a_j measures the strength of externalities. In such a framework, firm productivity is not log-supermodular in z and L. In contrast, I postulate the following functional form of the productivity function, for a firm i of raw efficiency z_i operating in sector $j \in 1 \dots S$:

[24] These controls absorb the mechanical relationship by which firms in large (resp. small) cities are more likely, conditional on moving, to move to smaller (resp. larger) cities.

Assumption B

$$\log(\phi_j(z_i, L, s_j)) = a_j \log L + \log(z_i)(1 + \log \frac{L}{L_o})^{s_j} + \epsilon_{i,L} \quad \textit{for } \log(z_i) \geq 0 \textit{ and } L \geq L_o \tag{23}$$

$$\log(\phi_j(z_i, L, s_j)) = 0 \quad \textit{for } L < L_o$$

The parameter a_j measures the classic log-linear agglomeration externality. The strength of the complementarity between agglomeration externalities and firm efficiency is captured by s_j. When $s_j = 0$, the model nests the traditional model of agglomeration externalities without complementarity. L_o measures the minimum city size below which a city is too small for a firm to produce in. In what follows, I write $\tilde{L} = \frac{L}{L_o}$, and $\mathcal{L}$ the set of normalized city sizes in the simulated economy.[25] I assume that $\log(z)$ is distributed according to a normal distribution with variance $\nu_{Z,j}$, truncated at its mean to prevent $\log(z)$ from being negative. This restriction is needed for the productivity of firms to be increasing in city size in equation (B). I introduce an error structure by assuming that firms draw idiosyncratic productivity shocks $\epsilon_{i,L}$ for each city size, where $\epsilon_{i,L}$ is i.i.d. across city sizes and firms. It is distributed as a type-I extreme value, with mean zero and variance $\nu_{R,j}$. This shock captures the fact that an entrepreneur has idiosyncratic motives for choosing a specific location: for example, he could decide to locate in a city where he has a lot of personal connections that make him more efficient at developing his business. These idiosyncratic motives for location generate imperfect sorting. The predictions of the theory are still relevant to the case of imperfect sorting, once adapted to reflect the fact that they hold for firms *on average* within a city size rather than systematically for all firms in a given city, as shown formally in online Appendix B.

I assume that idiosyncratic shocks are city-size specific, with mean zero and a constant variance across city size bins, and not city-specific as would perhaps be more natural. Still, these shocks can themselves represent the maximum of shocks at a more disaggregated level (e.g., at the city level). The maximum of a finite number of independent draws from a type-I extreme value distribution is also distributed as a type-I EV, with the same variance. Aggregating at the city-size level does not impact the estimation of the variance of the draws. I normalize the mean to be zero. If the model is misspecified and in reality, there is a systematic difference in mean idiosyncratic shocks across different city-size bins, this mean value is not separately identified from the log-linear agglomeration externality term a_j, which will capture both in the estimation.

Finally, note that I estimate the model under the assumption, made in section 2, that city sites are all ex ante identical. In reality, sites differ in their natural amenities. This can contaminate the estimation of the model if there is a correlation in the data between these amenities, local firms' productivities and city sizes. Under the assumption that these natural advantages benefit all firms in the same way, this correlation will be captured in estimation by the log-linear agglomeration externality term a_j. This will tend to bias upward the coefficient a_j but importantly does not affect the estimation of the log-supermodular term s_j. Furthermore, the bias is likely to be small. Combes et al. (2008) show that the role played by natural endowments on the productivity of French cities is quite limited. Michaels

[25] L_o is a normalization parameter in levels that changes proportionally the size of all cities but does not affect the estimation, which relies on *relative* measures. $\tilde{L}$ is the relevant measure for firm choices. L_o is calibrated to match the actual level of city sizes in the data.

and Rauch (2013) argue that French cities locations are strongly path dependent. They were efficiently chosen at the time of the Roman Empire, but remained largely unchanged over time. Cities' locations are unlikely to reflect strong exogenous comparative advantage from the perspective of modern technologies.

3.3.2. Estimation procedure

The estimation is conducted in two stages. In the first stage, I start by calibrating for each industry its capital intensity α_j and elasticity of substitution σ_j. The capital intensities are calibrated to the share of capital in sectoral Cobb-Douglas production functions, and the elasticity of substitution is calibrated to match the average revenue to cost margin in each sector.[26] I then calibrate the composite parameter $b\frac{1-\eta}{\eta}$, equal to the elasticity of wages to city size in the model. To do so, I follow equation (20) and use the difference between the elasticities of average firm revenue to city size and the elasticities of average firm employment to city size across all sectors. Finally, I calibrate the Cobb-Douglas share of each industry ξ_j by measuring its share of value-added produced.

In the second stage, I use simulated method of moments (SMM) to back out the quadruple $\theta_0^j = (a^j, s^j, \nu_R^j, \nu_Z^j)$ for each sector $j \in (1..S)$. Firms make a discrete choice of (normalized) city size, according to the following equation

$$\log \tilde{L}_j^*(z_i) = \underset{\log \tilde{L} \in \mathcal{L}}{\arg\max} \quad \log(z_i)\big(1 + \log\, \tilde{L}\big)^{s_j} + (a_j - b(1-\alpha_j)\frac{1-\eta}{\eta}) \log \tilde{L} + \epsilon_{i,L}, \tag{24}$$

which is the empirical counterpart of equation (13). Because the choice equation involves unobservable heterogeneity across firms and is non-linear, I have to use a simulation method (Gouriéroux and Monfort (1997)) to recover the model primitives. The SMM method is carried through sector by sector. I retain a rather aggregated definition of sectors, corresponding to 23 industries of the French NAF classification, in order to limit the computing requirements of the procedure. The general approach is close to the one in Eaton et al. (2011). The estimate θ_{II}^j minimizes the loss function

$$\|m_j - \hat{m}_j(\theta)\|_{W_j^2} = \big(m_j - \hat{m}_j(\theta)\big)' \, W_j \, \big(m_j - \hat{m}_j(\theta)\big), \tag{25}$$

where m_j is a vector stacking a set of moments constructed using firm data, as detailed below; $\hat{m}_j(\theta)$ is the vector for the corresponding moments constructed from the simulated economy for parameter value θ; W_j is a matrix of weights.[27] Details of the estimation procedure are reported in the online appendix.

3.3.3. Identification and choice of moments

I use three sets of non-parametric moments, for each sector, to characterize the economy. The first set of moments describes non-parametrically how average firm value-added increases with city size, sector by sector. I use 4 moments for each sector, capturing average firm size for each quartile of city size.

[26] In each sector, $\hat{\sigma_j}$ and $\hat{\alpha_j}$ are calibrated using $\frac{\hat{\sigma_j}}{\hat{\sigma_j}-1} = mean(\frac{\text{v.a.}}{\text{costs}})$ where costs exclude the cost of intermediate inputs, and $\hat{\alpha_j} = \alpha_j^{CD} \frac{\sigma_j}{\sigma_j - 1}$ where α_j^{CD} is the sectoral revenue-based Cobb-Douglas share of capital.

[27] The weighting matrix W_j for sector j is a generalized inverse of the estimated variance-covariance matrix Ω_j of the moments calculated from the data m_j.

Intuitively, how fast firm size increases with city size helps pin down the agglomeration parameters a and s. The parameters a and s both impact firm productivity and value-added, 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. Online Appendix G shows that, if the variance of idiosyncratic shocks ν_R was fixed, the distribution of value added conditional on city size would formally identify a and s. It would also identify ν_z, the distribution of raw efficiencies, which interacts with s to generate sorting.

To help identify the parameters that govern firm-level heterogeneity ν_z and ν_R, I also use 5 moments that characterize non-parametrically the firm-size distribution in value added.[28] If the distribution of value added conditional on city size does not allow me to identify separately ν_R from ν_Z, these moments do. The parameter ν_R governs the variance of the noise in a classic discrete choice setting problem (equation (24)). This parameter is usually normalized, as it cannot be inferred from simply observing the choice of the agent. Here, in contrast to a classic discrete choice setting, I observe not only the choice of city size made by firms, but also additional outcomes that are impacted by this choice, for instance, firm value-added. This last part is unusual, and these additional moments allow me to identify the variance of idiosyncratic shocks separately form the variance of firm's raw efficiency.

In addition, I use moments that describe non parametrically the distribution of sectoral value-added across city sizes. I measure the share of value-added in a given sector that falls into one of 4 bins of city sizes.[29] These moments summarize the geographic distribution of economic activity within a sector. Together with the first set of moments, they give information on the density of firms located in different city sizes. Therefore, they contribute to identifying the distribution of raw efficiencies, conditional on the agglomeration parameters.

3.3.4. Model fit

Figures in Appendix I report the model fit for the set of moments targeted by the estimation procedure. Specifically, the way firm average value-added increases with city size is generally well captured by the estimated model (Figure I.1). Further, the estimation relies on five moments of the sectoral firm-size distribution in revenues. To get a sense of the fit of the model fit in this dimension, I show in figure I.2 how the *whole* firm-size distribution compares in the data and in the model. In general, the fit is better for the upper tail than the lower tail, which is intuitive since the estimation focuses on upper-tail quantiles and the initial distribution of z is truncated to the left. Finally, the estimation relies on the share of sectoral value-added in four given city-size bins. I compute more generally for each sector the share of employment by decile of city size and represent the simulated vs. actual shares on Figure I.3. The model accurately captures the cross-sectoral heterogeneity in location patterns. The within-sector patterns are noisier, but still follow well the overall trends in the data.

Finally, I focus on a moment not directly targeted in the estimation, which is the city-size distribution.

[28] These bins are defined by the 25, 50, 75 and 90th percentiles of the distribution in the data, normalized by the median. As in Eaton et al. (2011), higher quantiles are emphasized in the procedure, since they capture most of the value added, and the bottom quantiles are noisier.

[29] 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.

The estimation is made on a grid of possible city sizes that have the same maximum to minimum range as in the data. I make, however, no assumption on the number of cities in each size bin, i.e., on the city-size distribution. Armed with sectoral estimations, I can solve for the general equilibrium of the model and in particular compute the city-size distribution that clears labor markets at the estimated parameter values (see Section 2). The estimated city-size distribution exhibits Zipf's law and follows quite well the actual city-size distribution measured here in total local employment of the city, consistent with the data used in estimation. The fit is shown in Figure 2, where the city-size distribution is plotted for the simulated data and the actual data.[30]

3.3.5. Analysis of the parameter estimates

The estimated parameters industry by industry are reported in table 5. The sectoral estimate of s_j, the parameter that governs the strength of the complementarity between firm efficiency and agglomeration externalities, is positive for a vast majority of industries. It is negative for two industries, the shoes and leather industry and metallurgy, which are relatively mature industries.[31] That more mature industries tend to exhibit different agglomeration forces is reminiscent of the argument in Audretsch and Feldman (1996), who argues that the nature of agglomeration forces depend on the life cycle of industries and show that agglomeration forces tend to decline as industries get more mature and less innovative.

Together, the agglomeration parameters and the variance parameters *jointly* determine the distribution of the realized productivity of firms and, crucially, the productivity gains associated with city size in equilibrium. These gains have been used in the literature as a proxy to measure agglomeration externalities. Here, the productivity gains associated with city size depend not only on the strength of agglomeration externalities, but on the sorting of firms, and on their selection on local idiosyncratic productivity shocks. In what follows, I present direct and counterfactual measures of the elasticity of firm productivity to city size to highlight how these forces interplay and understand how the parameter estimates translate into economic forces. For simplicity, I present average measures across sectors.

A first raw measure of the observed elasticity of firm productivity to city size can be computed by running the following simple OLS regression:

$$\log \psi_{i,j} = \beta_0 + \beta_1 \log\, L_i + \delta_j + \mu_i, \tag{26}$$

where $\psi_{i,j}$ is the equilibrium productivity of simulated firm i with efficiency z_i in industry j, $L_i = L_j^*(z_i)$ is the size of the city where firm i has chosen to produce and δ_j is an industry fixed effect. The OLS estimate of β_1, the elasticity of observed firm productivity to city size, is 4.2%. Interestingly, this measure falls within the range of existing measures of agglomeration externalities, as reported in Rosenthal and Strange (2004). They typically range from 3% to 8%. Rosenthal and Strange (2004) note that most studies do not account for sorting or selection effects when estimating the economic gains to density -

[30]The fitted lines correspond to a log rank-log size regression run on each of these distributions. Parallel slopes indicate that both distributions have the same tail. The levels are arbitrary and chosen so that the figure is readable.

[31]In the food manufacturing sector, the estimation backs out a degenerate distribution for firms productivity, so that the log-supermodular coefficient is not defined for that sector.

they are therefore broadly comparable, in scope, to the OLS estimation of β_1 in equation (26).[32] In the estimated model, these productivity gains are driven only in part by the existence of agglomeration externalities. Part of these gains come from the sorting of more efficient firms into larger cities, which I examine now.

To measure the contribution of firm sorting in the observed economic gains to density, I conduct the following counterfactual analysis. I recalibrate the model with firms constrained to choose their city size as if they all had the average efficiency in their sector, thereby shutting down systematic sorting. In this exercise, the difference in firms' location choice is only driven by firm-city size specific iid productivity shocks. I find that the relationship between firm-level productivity and city size is flatter in the counterfactual simulation than in the baseline model. Estimating equation (26) on this counterfactual data leads to an elasticity of firm productivity to city size of 2.3%. By this account, firm sorting accounts for almost half of the productivity gains measured in equilibrium between cities of different sizes.

Finally, to gauge the importance of the sorting forces emphasized in the model I decompose, as detailed in online Appendix G.5, the variance of productivity between the contribution of the systematic component of productivity on which firms sort, and the contribution of the idiosyncratic part. I find that they contribute equally to explaining firm productivity in the estimated model, which points at sorting as an important mechanism to rationalize the data.

4. The Aggregate Impact of Place-Based Policies

Equipped with the estimates of the model's parameters, I finally turn to the evaluation of the general equilibrium impact of a set of place-based policies. I use the real wage, constant across space, as a measure of welfare. Details of the implementation are reported in the online appendix.

4.1. Local tax incentives

I first study policies that subsidize firms locating in less developed cities or regions. This type of federal program aims at reducing spatial disparities and is also advocated for reason of efficiency. The case for increased efficiency relies on the idea that in the presence of agglomeration externalities, jump-starting a local area by attracting more economic activity can locally create more agglomeration externalities, enhancing local TFP and attracting even more firms. This argument, however, needs to be refined. As has been pointed out in the literature (Glaeser and Gottlieb (2008), Kline and Moretti (2013)) this effect depends in particular on the overall shape of agglomeration externalities. While smaller cities may in fact benefit from these policies, larger cities marginally lose some resources – and therefore benefit from less agglomeration economies. The net effect on the overall economy is a priori ambiguous. Turning to spatial disparities, since utility is equalized across all workers, there is no welfare inequality in equilibrium in the model. Nevertheless, the economy is characterized by other spatial disparities, in city size or GDP per capita for example, that may matter for political economy reasons. Place-based policies impact these

[32] An exception is Combes et al. (2008), who estimate agglomeration externalities using detailed French worker-level data and control for the sorting of workers across locations. They find an estimate of 3.7% of the elasticity of productivity to employment density.

spatial differences. They tend to benefit the targeted areas, but the extent to which they reduce aggregate measures of spatial differences depends on the overall reallocation of economic activity in space, which I examine in the quantitative exercise below.

To evaluate these policies in the context of my model, I consider a set of counterfactuals in which firms are subsidized to locate in the smallest cities of the country, which are also the least productive ones. To calibrate the intensity of the simulated policy, I choose as a reference point an example of a specific policy put in place in France, which targets disadvantaged neighborhoods that cover an overall population of 1.5M or 2.3% of the French population, the French "ZFU" program (*Zones Franches Urbaines*). I simulate a scheme that subsidizes firms locating in the smallest cities corresponding to 2.3% of the population in the simulated data. I implement a subsidy of 12% of firms' profits in these areas, paid for by a lump-sum tax levied on all firms in the country.[33]

Local effects The model predicts large effects of subsidizing small cities on the targeted areas. In targeted cities, the number of establishments grows by 19%. The corresponding local increase in population is, however, only 4%. This is because the firms attracted by the policy in these areas are small and have low productivity. These results are roughly consistent with the order of magnitude estimated in Mayer et al. (2012) on the effect of the French ZFU; Mayer et al. (2012) find a 31% increase in the entry rate of establishments in the three years following the policy's implementation and note that these new establishments are small relative to existing establishments. Of course, this is just a "plausibility check" since the two exercises are not directly comparable - the ZFU targets sub-areas smaller than the cities of my model, and the model does not have dynamic effects.

Aggregate effects Beyond evaluating the effects of the policy on the targeted cities, the counterfactual exercise allows me to compute the general equilibrium effect of this type of policy. I compute the aggregate TFP and welfare effects of the policy for different levels of the subsidy, holding constant the targeted areas. The welfare measure I use is the real income of the representative worker. It does not account for other elements that can arguably be in the objective function of the decision maker, such as measures of equity, and which often motivate these policies. In that sense, the (negative) welfare effects reported here can be seen as costs of this policy, to be weighted against potential benefits that are outside of the model.

The simulation shows that these place-based policies have negative long-run effects, both on the productive efficiency of the economy and in terms of welfare. A subsidy to smaller cities that amounts to 1% of GDP leads to a loss of 1.05% in TFP in the aggregate, and a loss of 1.4% in welfare. While such a policy allows to decrease congestion overall, the welfare gain from decreasing congestion is largely dominated by the negative TFP effect.[34]

[33] The French "ZFU" program (*Zones Franches Urbaines*) is a policy similar to to the Empowerment Zone program in the US. This policy costs 500 to 600 million euros in a typical year, corresponding to extensive tax breaks given to local establishments. The gross cost of this subsidy in the model is 0.04% of GDP, which matches the one reported for France for the ZFU program. A 12% subsidy on profits correspond to a 36% tax break on the French corporate tax, whereas the French ZFU program offers full corporate tax breaks for 5 years as well as other generous labor tax and property tax relief.

[34] TFP has a magnified impact on welfare as capital flows in and out of the economy in response to the TFP shock. The corresponding formulas are explicited in the supplemental material (H.2).

I then use the counterfactual economy to study the impact of these place-based policies on the dispersion of spatial outcomes, by measuring how the Gini coefficient for the distribution of GDP per capita in the economy reacts to the policy. A reason to focus on such a measure is that policy makers may want to smooth out this type of disparity across cities. Perhaps surprisingly, the type of place-based policies I study leads to an increase in spatial disparities as measured by this Gini index. The intuition for this result is as follows. The counterfactual equilibrium is characterized by (1) growth in the size of smaller cities, (2) a decrease in the population of mid-size cities, and (3) an increase in the population of the largest cities. This change in city size distribution is plotted in figure 3. That larger cities grow in the counterfactual economy comes from the fact that, as mid-size cities lose population in favor of smaller cities, they offer less agglomeration externalities. As a consequence, these mid-size cities become less attractive than larger cities for a set of firms that were previously indifferent between these mid-size cities and *larger* cities. Small and large cities thus expand at the expense of mid-size cities. Quantitatively, this leads to a rise in the Gini coefficient.

According to these results, place-based policies may have general equilibrium effects that run counter to their rationale.

4.2. Land-use regulation

Glaeser and Gottlieb (2008) forcefully argue against policies that limit the growth of cities and constrain the available housing supply. Zoning regulations or regulations on the type or height of buildings that can be built within a city constitute examples of such land-use regulations. A rationale for these restrictions on land-use development is that they may increase the quality of life for existing residents. On the other hand, by constraining the housing supply and limiting the size of cities, they may dampen the agglomeration effects at play in the economy.

I model the loosening of local land-use regulation that could be mandated by a federal government by decreasing the land-use intensity parameter b in the housing production function (equation (1)).[35] Decreasing this parameter increases the elasticity of housing supply. To quantify the impact of land-use regulation policies, I first need to calibrate b. To do so, I assume that b is such that the housing supply elasticity is at the median measure across US cities, as estimated by Saiz (2010).[36] I then compare the aggregate TFP and welfare of two counterfactual economies: one where the housing supply elasticity is set at the 25th percentile of the housing supply elasticity distribution in Saiz (2010) and one where it is set at the 75th percentile.

Increasing the housing supply elasticity has two separate effects on welfare. First, a direct – mechanical – effect on utility. All else equal, as the housing sector becomes more productive and the housing supply elasticity increases, the housing units available to households increase, which directly raise their utility. This mechanical effect is not the focus here. Beyond this direct effect, an increase in the housing supply elasticity flattens out the wage schedule (see equation (4)), which leads firms in the heterogeneous goods

[35]Land-use regulation have been largely delegated to local municipalities in France over the past 30 years, but the national government can still decide on general rules that apply everywhere (RNU, *Reglement National d'Urbanisme*).

[36]For France, Combes et al. (2016) propose a range of estimates for the (mean) price elasticity of housing with respect to city size. The measure I use is well within this range.

sectors to locate in larger cities. This indirect effect enhances the productive efficiency of these sectors.[37]

Figure I.5 reports TFP and welfare, relative to the reference equilibrium, for various levels of the housing supply elasticity. An overall increase in the housing supply elasticity from the 25th to the 75th percentile leads to a 1.6% gain in TFP and a 1.8% indirect gain in welfare.

This policy experiment illustrates how increasing housing supply in cities can have positive effects beyond directly reducing congestion costs. They allow for a more efficient spatial organization of production in the differentiated goods sectors by endogenously creating agglomeration externalities and enhancing the way labor is allocated to heterogeneous firms in the economy.

5. Conclusion

I offer a new general equilibrium model of heterogeneous firms that are freely mobile within a country and can choose the size of the city where they produce. I show that the way firms sort across cities of different sizes is relevant to understanding aggregate outcomes. This sorting shapes the productivity of each firm and the amount of agglomeration externalities in the economy. It also shapes the allocation of factors across firms: cities are a medium through which the production of heterogeneous firms is organized within a country. The sorting of firms, mediated by the existence of city developers who act as a coordinating device for the creation of cities, leads to a unique spatial equilibrium of this economy. Therefore, the model can be used to conduct policy analysis. It allows the quantification of the complex spatial equilibrium effects of spatial policies. This complements the existing literature, which has traditionally focused on estimating local effects of such policies. Using the structure of the model, I estimate the general equilibrium effects of two types of place-based policies. A policy that explicitly targets firms locating in the least productive cities tends to hamper the productivity of the economy as a whole. For the specific policy I study, spending 1% of GDP on local tax relief leads to an aggregate welfare loss of 1.4% and does not reduce observed spatial dispersions that may matter for political economy reasons. On the other hand, policies that encourage the growth of all cities - not just the smallest ones - can enhance equilibrium productivity and welfare: moving the housing-supply elasticity from the 25th to the 75th percentile of housing-supply elasticity leads to a 1.6% gain in TFP and 1.8% in welfare through a spatial reorganization of production.

References

Abdel-Rahman, Hesham and Alex Anas, "Theories of systems of cities," *Handbook of Regional and Urban Economics*, 2004, *4*, 2293–2339.

Albouy, David, "Evaluating the efficiency and equity of federal fiscal equalization," *Journal of Public Economics*, 2012, *96* (9-10), 824–839.

[37]To focus on this indirect effect, I control for the direct effect on utility of an increase in housing supply as follows. For each value of the housing supply elasticity, I simulate the equilibrium of the economy as described above. To measure welfare per capita, I take into account the spatial reallocation of economic activity, but hold constant b, hence the price of housing, in the utility of workers. Fixing the price of housing mutes the mechanical welfare effect coming from an increase in housing supply. The aggregate welfare gain, including both the direct and the indirect effects of an increase in housing supply, is 20.5%

Allen, Treb and Costas Arkolakis, “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, 2014, *129* (3), 1085–1140.

Athey, Susan, “Monotone comparative statics under uncertainty,” *quarterly Journal of Economics*, 2002, pp. 187–223.

Audretsch, David and Maryann Feldman, “Innovative clusters and the industry life cycle,” *Review of Industrial Organization*, 1996, *11* (2), 253–273.

Baldwin, Richard and Toshihiro Okubo, “Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting,” *Journal of Economic Geography*, June 2006, *6* (3), 323–346.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud, “Productive cities: Sorting, selection, and agglomeration,” *Journal of Political Economy*, 2014, *122* (3), 507–553.

_ , Giordano Mion, Yasusada Murata, and Jens Südekum, “Spatial frictions,” Technical Report, Discussion Paper Series, Forschungsinstitut zur Zukunft der Arbeit 2013.

Brinkman, Jeff, Daniele Coen-Pirani, and Holger Sieg, “Firm Dynamics within an Urban Economy,” *International Economic Review*, forthcoming.

Busso, Matias, Jesse Gregory, and Patrick Kline, “Assessing the Incidence and Efficiency of a Prominent Place Based Policy,” *American Economic Review*, April 2013, *103* (2), 897–947.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon, “Spatial wage disparities: Sorting matters!,” *Journal of Urban Economics*, 2008, *63* (2), 723–742.

_ , _ , and _ , “The costs of agglomeration: Land prices in French cities,” Technical Report 2016.

_ , _ , _ , Diego Puga, and Sébastien Roux, “The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection,” *Econometrica*, November 2012, *80* (6), 2543–2594.

Costinot, Arnaud, “An Elementary Theory of Comparative Advantage,” *Econometrica*, 07 2009, *77* (4), 1165–1192.

_ and Jonathan Vogel, “Matching and Inequality in the World Economy,” *Journal of Political Economy*, 08 2010, *118* (4), 747–786.

Criscuolo, Chiara, Ralf Martin, Henry Overman, and John Van Reenen, “The Causal Effects of an Industrial Policy,” CEP Discussion Papers dp01113, Centre for Economic Performance, LSE January 2012.

Davis, Donald and Jonathan Dingel, “A Spatial Knowledge Economy,” NBER Working Papers 18188, National Bureau of Economic Research June 2012.

_ and _ , “The Comparative Advantage of Cities,” Working Paper 2013.

Desmet, Klaus and Esteban Rossi-Hansberg, “Urban Accounting and Welfare,” *American Economic Review*, 2013, *103* (6), 2296–2327.

Duranton, Gilles and Diego Puga, “Nursery cities: Urban diversity, process innovation, and the life cycle of products,” *American Economic Review*, 2001, pp. 1454–1477.

_ and _ , “Micro-foundation of urban agglomeration economies, in Henderson V. J. and Thisse JF.(eds) “Handbook of Regional and Urban Economics Vol. 4 Cities and Geography,” 2003.

Eaton, Jonathan, Samuel Kortum, and Francis Kramarz, “An anatomy of international trade: Evidence from French firms,” *Econometrica*, 2011, *79* (5), 1453–1498.

Eeckhout, Jan and Philipp Kircher, “Assortative matching with large firms: Span of control over more versus better workers,” Working Paper 2012.

_ , Roberto Pinheiro, and Kurt Schmidheiny, “Spatial sorting: Why New York, Los Angeles and

Detroit attract the greatest minds as well as the unskilled," Working Paper 2010.

Gabaix, Xavier, "Power Laws in Economics and Finance," *Annual Review of Economics*, 2009, *1* (1), 255–294.

_ **and Rustam Ibragimov**, "Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents," *Journal of Business & Economic Statistics*, 2011, *29* (1), 24–39.

Glaeser, Edward and Joshua Gottlieb, "The economics of place-making policies," *Brookings Papers on Economic Activity*, 2008, *2008* (1), 155–239.

Gourieroux, Christian, Alain Monfort, and Eric Renault, "Indirect inference," *Journal of applied econometrics*, 1993, *8* (S1), S85–S118.

Gouriéroux, Christian and Alain Monfort, *Simulation-based Econometric Methods* OUP/CORE Lecture Series, OUP Oxford, 1997.

Henderson, Vernon, "The Sizes and Types of Cities," *American Economic Review*, September 1974, *64* (4), 640–56.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and manufacturing TFP in China and India," *Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

Kline, Patrick and Enrico Moretti, "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority," August 2013.

Matano, Alessia and Paolo Naticchioni, "Wage distribution and the spatial sorting of workers," *Journal of Economic Geography*, 2011, *12* (2), 379.

Mayer, Thierry, Florian Mayneris, and Loriane Py, "The Impact of Urban Enterprise Zones on Establishments' Location Decisions: Evidence from French ZFUs," 2012.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, November 2003, *71* (6), 1695–1725.

Michaels, Guy and Ferdinand Rauch, "Resetting the Urban Network: 117-2012," Technical Report, CEPR Discussion Papers 2013.

Milgrom, Paul R, "Good news and bad news: Representation theorems and applications," *The Bell Journal of Economics*, 1981, pp. 380–391.

Mion, Giordano and Paolo Naticchioni, "Triage et arrimage des compétences et des entreprises dans l'espace.," *Canadian Journal of Economics/Revue canadienne d'économique*, 2009, *42* (1), 28–55.

Nocke, Volker, "A Gap for Me: Entrepreneurs and Entry," *Journal of the European Economic Association*, 09 2006, *4* (5), 929–956.

Rosenthal, Stuart and William Strange, "Evidence on the nature and sources of agglomeration economies," *Handbook of Regional and Urban Economics*, 2004, *4*, 2119–2171.

Rossi-Hansberg, Esteban and Mark Wright, "Urban structure and growth," *Review of Economic Studies*, 2007b, *74* (2), 597–624.

Saiz, Albert, "The geographic determinants of housing supply," *Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Sampson, Thomas, "Selection into trade and wage inequality," *American Economic Journal: Microeconomics*, forthcoming.

Serrato, Juan Carlos Suárez and Owen Zidar, "Who benefits from state corporate tax cuts? A local labor markets approach with heterogeneous firms," Technical Report, Manuscript 2014.

Topkis, D, *Supermodularity and complementarity*, Princeton University Press, 1998.

Table 1: Summary statistics.

	log value added			log employment			N
	mean	p25	p75	mean	p25	p75	
Manufacture of food products and beverages	7.57	5.81	8.43	2.30	1.39	3.00	14,102
Manufacture of textiles	8.06	6.10	9.12	2.80	1.95	3.66	2,955
Manufacture of wearing apparel	7.50	5.36	8.57	2.47	1.61	3.43	3,219
Manufacture of leather goods and footwear, leather tanning	7.90	5.88	8.95	2.79	1.79	3.78	878
Manufacture and products of wood, except furniture	7.77	6.20	8.55	2.43	1.79	3.09	3,688
Manufacture of pulp, paper and paper products	8.66	6.55	9.69	3.15	2.20	4.04	1,284
Publishing, printing and reproduction of recorded media	7.20	4.96	8.26	1.94	1.10	2.71	11,238
Manufacture of chemicals and chemical products	8.77	6.03	10.33	3.11	1.79	4.30	2,647
Manufacture of rubber and plastic products	8.41	6.50	9.40	2.94	2.08	3.81	3,563
Manufacture of glass, ceramic, brick and cement products	7.99	6.12	8.89	2.55	1.61	3.30	3,143
Manufacture of basic metals	9.06	6.90	10.18	3.52	2.30	4.51	834
Manufacture of fabricated metal products, except machinery	8.02	6.50	8.81	2.50	1.79	3.22	16,160
Manufacture of machinery	8.00	6.16	8.96	2.48	1.61	3.33	7,689
Manufacture of office machinery and computers	7.91	5.52	9.04	2.54	1.61	3.40	312
Manufacture of electrical machinery	8.14	6.22	9.09	2.70	1.61	3.56	2,273
Manufacture of radio, television and communication equipment	8.17	5.88	9.25	2.74	1.61	3.69	1,544
Manufacture of medical, precision and optical instruments	7.70	5.89	8.55	2.18	1.39	2.94	4,235
Manufacture of motor vehicles	8.48	6.39	9.42	3.06	1.95	3.83	1,346
Manufacture of other transport equipment	7.86	5.45	8.98	2.60	1.39	3.56	1,007
Manufacture of furniture	7.28	5.24	8.30	2.14	1.10	3.00	5,269
Recycling	7.62	5.89	8.45	2.11	1.39	2.77	1,394
Information technology services	7.24	4.84	8.48	1.95	0.69	2.83	10,617
Business services, non I.T.	6.93	4.79	7.95	1.55	0.69	2.20	57,673

Table 2: Share of establishment in larger cities and tradable capital intensity.

Dep. variable	Share of establishments in large cities			
	I	II	III	IV
Sample		*all tradables*		*export intensive*
Tradable capital intensity	0.479**	0.613**	0.596**	0.551**
	(0.152)	(0.147)	(0.168)	(0.200)
High skill intensity=1			0.058**	0.042
			(0.029)	(0.033)
Nb firms	no	yes	yes	yes
Mean va	no	yes	yes	yes
R-squared	0.065	0.219	0.174	0.139
Observations	146	146	117	84

(*)$p < 0.10$, (**)$p < 0.05$. Tradable capital intensity: share of capital net of real estate assets in a Cobb-Douglas production function with labor, tradable capital and non tradable capital. Large cities: larger cities representing 50% of workers. Nb firms: number of firms. Mean va: average value added per firm. High skill intensity are sectors above median of skill intensity. Export intensive: industry above median for all sectors in the economy in export intensity, proxied by the ratio of export to domestic sales.

Table 3: Tail of the firm-size distribution (ζ) vs sector location.

	ζ, tail of firm-size distribution		
	I	II	III
Sample		*all tradables*	*export intensive*
Share in large cities	-0.544**	-0.686**	-0.263**
	(0.142)	(0.122)	(0.115)
Nb firms	no	yes	yes
Mean va	no	yes	yes
R-squared	0.092	0.578	0.487
Observations	146	146	89

(*)$p < 0.10$, (**)$p < 0.05$. Pareto Shape: ζ estimated by $Log(Rank_i - 1/2) = a - \zeta Log(va_i) + \epsilon_i$, on firms above median size, for industries with more than 200 firms. Nb firms: number of firms. Mean va: average value added per firm. Export intensive: industry above median for all sectors in the economy in export intensity, proxied by the ratio of export to domestic sales. Heteroskedasticity-robust standard errors.

Table 4: Movers.

	Δ_t City Size			
	I	II	III	IV
Sample	*all tradables*		*export intensive*	
log(firm size)	0.089**	0.073**	0.090**	0.080**
	(0.020)	(0.019)	(0.026)	(0.025)
Initial City Size	-0.987**		-0.986**	
	(0.055)		(0.048)	
Constant	12.296**			
	(0.651)			
Industry F.E.	yes	yes	yes	yes
Initial city F.E.		yes		yes
R-squared	0.537	0.629	0.540	0.635
Observations	6103	6103	3675	3675

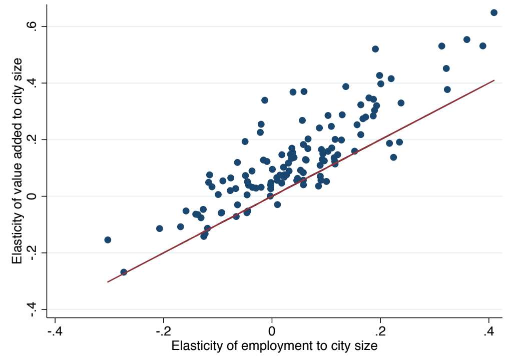
(*)$p < 0.10$, (**)$p < 0.05$. Set of mono-establishment firms which move between 2 years, between 1999 and 2005. Δ_t City Size= $log(\frac{L_{t+1}}{L_t})$, where L_t is the size of the city where the firm locates at time t. Size is measured by the firm value added relative to other firms in the same sector-year-city, as the residual of $log(VA)_i = DS_i + DT_i + DC_i + \epsilon_i$ where DS is a sector fixed effect, DT a year fixed effect, DC a city fixed effect. Export intensive: industry above median for all sectors in the economy in export intensity, proxied by the ratio of export to domestic sales.

Table 5: Estimated parameters.

	$\hat{s}$	$\hat{\nu_R}$	$\hat{\nu_z}$	$\hat{a}$
Manufacture of food products and beverages	n.d	0.476	0.000	0.142
	n.d.	(0.016)	(0.089)	(0.058)
Manufacture of textiles	0.038	0.294	0.274	0.009
	(0.788)	(0.036)	(0.297)	(0.080)
Manufacture of wearing apparel	0.147	0.219	0.252	0.040
	(0.233)	(0.020)	(0.138)	(0.034)
Manufacture of leather goods and footwear, leather tanning	-0.102	0.162	0.465	0.033
	(0.068)	(0.037)	(0.027)	(0.006)
Manufacture and products of wood, except furniture	0.043	0.176	0.397	-0.020
	(0.021)	(0.009)	(0.021)	(0.004)
Manufacture of pulp, paper and paper products	0.049	0.243	0.589	0.019
	(0.014)	(0.005)	(0.013)	(0.008)
Publishing, printing and reproduction of recorded media	0.210	0.408	0.407	0.171
	(0.639)	(0.403)	(0.958)	(0.220)
Manufacture of chemicals and chemical products	0.217	0.430	0.977	-0.001
	(0.305)	(0.205)	(0.694)	(0.029)
Manufacture of rubber and plastic products	0.001	0.137	0.738	0.021
	(0.003)	(0.006)	(0.005)	(0.001)
Manufacture of glass, ceramic, brick and cement products	0.056	0.172	0.741	-0.019
	(0.010)	(0.016)	(0.021)	(0.005)
Manufacture of basic metals	-0.037	0.172	0.790	0.027
	(0.007)	(0.005)	(0.009)	(0.004)
Manufacture of fabricated metal products, except machinery	0.065	0.178	0.317	0.027
	(0.017)	(0.006)	(0.011)	(0.003)
Manufacture of machinery	0.070	0.137	0.496	0.024
	(0.006)	(0.006)	(0.012)	(0.002)
Manufacture of computers and office machinery	-0.009	0.176	0.529	0.123
	(0.076)	(0.022)	(0.088)	(0.022)
Manufacture of electrical machinery	0.033	0.071	0.552	0.034
	(0.005)	(0.004)	(0.006)	(0.001)
Manufacture of radio, television and communication equipment	0.060	0.191	0.536	0.051
	(0.058)	(0.143)	(0.235)	(0.024)
Manufacture of medical, precision and optical instruments	0.138	0.196	0.432	0.044
	(0.009)	(0.010)	(0.013)	(0.004)
Manufacture of motor vehicles	0.743	0.281	0.147	-0.062
	(0.019)	(0.008)	(0.010)	(0.031)
Manufacture of other transport equipment	0.045	0.201	0.536	0.103
	(0.021)	(0.011)	(0.023)	(0.015)
Manufacture of furniture	0.553	0.340	0.091	0.013
	(0.244)	(0.015)	(0.028)	(0.065)
Recycling	0.178	0.482	0.567	0.052
	(0.039)	(0.015)	(0.042)	(0.016)
Information technology services	0.426	0.280	0.301	0.152
	(0.086)	(0.020)	(0.068)	(0.053)
Business services, non I.T.	0.058	0.300	0.548	0.199
	(0.021)	(0.037)	(0.063)	(0.017)

Note: s is the log-supermodular agglomeration coefficient, a the log-linear agglomeration coefficient, ν_R the variance of iid shocks, ν_z the variance of firms raw efficiency. The log-supermodular coefficient s_1 is not defined for the first sector, as the estimation backs out a degenerate distribution for firms productivity in that sector.

Figure 1: Elasticity of mean value added and employment with city size.



Note: This figure plots for β in the regression: $\log \text{mean va}(L_j) = \alpha + \beta \log L_j + \epsilon_j$ against β in the regression β: $\log \text{mean empl}(L_j) = \alpha + \beta \log L_j + \epsilon_j$, ran sector by sector at the NAF600 level for industries with more than 200 mono-establishment firms.

Figure 2: City size distribution, model and data.

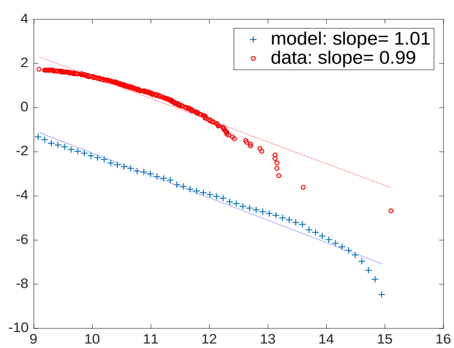
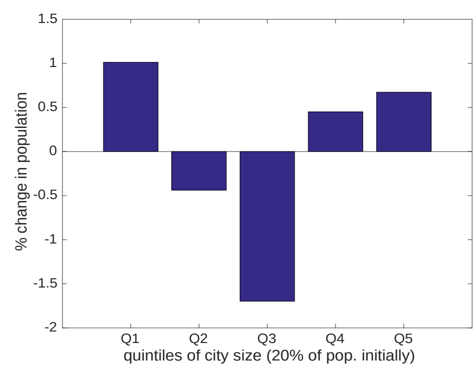


Figure 3: Change in city size distribution, after policy.



Note: This figure plots the change in city size distribution after implementing the policy described in section 4.1. Smallest cities corresponding to 2.3% of pop. are subsidized ; the subsidy amounts to 1% of GDP.

Online appendix - Not for publication

A. Housing market

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

$$H(L) = \left(\frac{p_H(L)}{w(L)}\right)^{\frac{1-b}{b}}, \tag{A.27}$$

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

The aggregate local demand for housing is

$$H(L) = \frac{(1-\eta)w(L)L}{p_H(L)}. \tag{A.28}$$

Equations (A.27) and (A.28) pin down prices and quantities of housing produced. Housing supply (equation (A.27)) and demand (equation (A.28)) 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:

$$\ell_H(L) = (1-b)(1-\eta)L, \quad \text{and} \tag{A.29}$$

$$\pi_H(L) = b(1-\eta)Lw(L). \tag{A.30}$$

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. Extensions of the model

B.1. Model with imperfect sorting

I examine the properties of the model in the presence of imperfect sorting as hypothesized in the empirical specification of section 3. 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 (23) 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:

$$V_j(z, L) = \kappa_{1j}\left(\frac{\psi(z, L, s_j)}{L^{b\frac{1-\eta}{\eta}(1-\alpha_j)}}\right)^{\sigma_j - 1} R_j\, P_j^{\sigma_j-1}, \tag{B.31}$$

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 4 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 4': 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:

$$p(L|z) = \frac{V(z,L)^{\frac{\nu_R}{\sigma-1}}}{\sum_{L'} V(z,L')^{\frac{\nu_R}{\sigma-1}}} \tag{B.32}$$

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 6 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 6': 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 (24) 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:

$$\begin{aligned} p(\pi^*|L^*) &= \int_z p(\pi^* \mid L^*, z) p_j(z|L^*)\, dz \\ &= \int_z p(\pi^* \mid z) p_j(z|L^*)\, dz \end{aligned} \tag{B.33}$$

Therefore, firm profits in city L are on average:

$$E[\pi^*|L^*] = \int_{\pi^*}\int_z \pi^* p(\pi^* \mid z)p(z|L^*)\,d\pi^*\,dz \tag{B.34}$$

$$= \int_z \left[\int_{\pi^*} \pi^* p(\pi^* \mid z)d\pi^*\right] p(z|L^*)\,dz \tag{B.35}$$

$$= \int_z E[\pi^* \mid z]\,p(z|L^*)\,dz$$

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:

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

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

$$E[r^*|L_2^*] > E[r^*|L_1^*]$$

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 :

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

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 $E[\pi^*(z, L^*)|L^*]$. Furthermore, $L^{*(1-\alpha_j)\frac{1-\eta}{\eta}b}$ also increases with L^*. Therefore, $E[\phi|L^*]$ increases with L^* . ∎

Third, proposition 7 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 7': 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 (B.31) 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 8 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 8': 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 (24) 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. ∎

B.2. 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:

$$P_i = \left[\int_{j \in \mathcal{C}} \left(\frac{\tau_{ji} w_j}{\bar{\psi}_j} \right)^{1-\sigma} dj \right]^{\frac{1}{1-\sigma}}. \tag{B.36}$$

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:

$$MA_i = \int_{j \in \mathcal{C}} \tau_{ij}^{1-\sigma} w_j L_j P_j^{\sigma-1} dj. \tag{B.37}$$

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

[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.

in equation (2) of the main text combined with the housing production equation (3), is:

$$U = U_i = \kappa_0 \left(\frac{w_i}{P_i}\right)^{\eta} L_i^{-b(1-\eta)}, \tag{B.38}$$

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

$$w_i^{1-\sigma} L_i^{-b(1-\sigma)\left(\frac{1-\eta}{\eta}\right)} = \widetilde{U}^{-1} \int_{j\in\mathcal{C}} (\tau_{ji} w_j)^{1-\sigma}\, \bar{\psi}_j^{\sigma-1} dj, \tag{B.39}$$

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 (B.38) and simplifying, it follows that:

$$w_i^{\sigma} L_i \bar{\psi}_i^{1-\sigma} = \widetilde{U}^{-1} \int_j {\tau_{ij}}^{1-\sigma} {w_j}^{\sigma} L_j^{1-b(\sigma-1)\frac{1-\eta}{\eta}} dj \tag{B.40}$$

This system of 2N equations (B.39)-(B.40) 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:

$${w_i}^{2\sigma-1} {L_i}^{1+b\frac{1-\eta}{\eta}(1-\sigma)} {\bar{\psi}_i}^{1-\sigma} = \Gamma \tag{B.41}$$

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 (B.38), 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.

To conclude, this shows that, in an extension of the model with trade costs, city size L is still a sufficient statistic for the local economic conditions in equilibrium, and there is still positive assortative matching between firm type and city size summarized by L.

B.3. 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 (13) 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 :

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

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:

$$S_2\geq \widetilde{\pi}(z,L^*)+T^*\widetilde{\pi}(z,L^*)-\widetilde{\pi}(z,L_2), \tag{B.42}$$

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:

$$S_2N_2\leq b(1-\eta)w(L_2)L_2,$$

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: $N_2=\frac{L_2(1-(1-\eta)(1-b))}{\ell_2(z,L_2)}=\frac{L_2w(L_2)(1-(1-\eta)(1-b))}{(1-\alpha)(\sigma-1)\widetilde{\pi}(z,L_2)}$. 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 (B.42) . 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.

B.4. Equilibrium without city developers

I describe here the characteristics of equilibria without city developers. If the proof of uniqueness of equilibrium relies on their existence, the properties of the equilibrium - and in particular the results on firm sorting – do not. That is, if one takes as exogenous a given set of city sizes $\mathcal{L}$ and assumes away city-developers, then more productive firms still sort into larger cities, and the equilibrium of this economy is still characterized by the propositions of section 2.3.

Call $\mathcal{L}^*$ the support of city sizes in the equilibrium with city developers. Consider now an equilibrium in the absence of city developers, and call $\widetilde{\mathcal{L}}$ the support of city sizes present in equilibrium. The set $\widetilde{\mathcal{L}}$ does not have to be an interval, and it could be in particular a discrete collection of city sizes. Necessarily, firms are located such that their choice of city size maximizes their profit function, defined in equation (10) of the main text:

$$\widetilde{L}_j(z) = \arg\max_{L \in \widetilde{\mathcal{L}}} \pi_j^*(z,L).$$

Firms choose their city size - among the ones present in equilibrium - optimally. In the main text, we show that when $\widetilde{\mathcal{L}} = \mathcal{L}^*$, then $\widetilde{L}_j(z)$ is an increasing function. I show here that if $\widetilde{\mathcal{L}} \neq \mathcal{L}^*$, then $\widetilde{L}_j(z)$ is rather a non-decreasing function. To show this, I use the same argument as in the proof of Lemma 4. The proof of monotonicity of $\widetilde{L}_j(z)$ does not rely on the characteristics of $\widetilde{\mathcal{L}}$, but only on the strict log-supermodularity in z and L of the profit function of the firm (equation (10)). 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, $\widetilde{L}_j(z_1) \geq \widetilde{L}_j(z_2)$, and $\widetilde{L}_j(z)$ is a non-decreasing function.

I describe here for completeness an equilibrium when the set $\widetilde{\mathcal{L}}$ is not convex. Consider a non convex set of city sizes $\widetilde{\mathcal{L}}$ 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 $z_i = \widetilde{L}_j^{-1}(a_i)$. 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 $\widetilde{L}_j(z)$ is increasing on $[z_1,z_2]$ and $[z_3,z_4]$ respectively. Then, pick $z \in (z_2,z_3)$. We know that $\widetilde{L}_j(z_2) \leq \widetilde{L}_j(z) \leq \widetilde{L}_j(z_3)$ since $\widetilde{L}_j(z)$ is non decreasing, hence $a_2 \leq \widetilde{L}_j(z) \leq a_3$. Since $L^*(z) \in \mathcal{L}$, this means that $\widetilde{L}_j(z) = a_2$, or $\widetilde{L}_j(z) = a_3$. By monotonicity of $\widetilde{L}_j(z)$, there exists a threshold $\tilde{z} \in (z_2,z_3)$ such that if $z \in (z_2,\tilde{z})$, $\widetilde{L}_j(z) = a_2$ and if $z \in (\tilde{z},z_3)$ $\widetilde{L}_j(z) = a_3$. 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.

Overall, equilibria that do not correspond to the case of city developers are not very different from the one studied in the main text. There is still positive assortative matching. Because of the bunching of firms that cannot access their optimal city size, some firms coexist in the same city size with firms that have a different, but "close", z. This bunching still preserves the monotonicity of the matching function $L_j^*(.)$.

Beyond Lemma 4, Propositions 6, 7, 8 and 9 hold exactly as is in any equilibrium without city developers. Nothing in their proof relies on the strict monotonicity of $L^*(z)$, (weak) monotonicity suffices.

C. Proofs of section 2

C.1. Lemma 2

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 (9) then simplifies to

$$\max_{L,\{T_j(L)\}_{j\in 1,\dots,S}} \Pi_L = b\,(1-\eta)w(L)L - \sum_{j=1}^{S} \frac{M_j w(L)}{(1-\alpha_j)(\sigma_j-1)} \int_z T_j(L)\ell_j(z,L)\,\mathbb{1}_j(z,L,i)dF_j(z) \tag{C.43}$$

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 (8).

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 $\Pi^i < 0$. ∎

C.2. Lemma 3

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. ∎

C.3. Lemma 4

Proof Fix 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, $L^*(z_1) \geq L^*(z_2)$.

Moreover, under the technical assumptions made here, $L_j^*(z)$ is a strictly increasing function. Since the set of z is convex, 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 $\frac{dL_j^*(z)}{dz} = -\frac{\frac{\partial(\frac{\psi_2 L}{\psi})}{\partial z}(z,L_j^*(z),s)}{\frac{\partial(\frac{\psi_2 L}{\psi})}{\partial L}(z,L_j^*(z),s)}$. ∎

C.4. Lemma 5

The proof is in the main text.

C.5. City size distribution

Lemma 11 *: The city-size distribution follows Zipf's law whenever the firm size distribution also follows Zipf's law and firm revenues increase with constant elasticity with respect to city size.*

Proof There is a bijection between z and firm level observable at equilibrium r^*, ℓ^*, L^*. By an abuse of notation, this functional relationship will be denoted $z(r), z(L), \ell(z)$. All of these relationship pertain to the sorting equilibrium, but I omit the star to keep the notations light.

Step 1: Write $g_j(r)$ the distribution of revenues in sector j. The firm-size distribution in revenues is readily computed through a change of variable, starting from the raw efficiency distribution:

$$g_j(r) = f_j(z(r))\frac{dz}{dr}(r) \tag{C.44}$$

Step 2: Assume that in all sectors - or at least for the ones that will locate in the largest cities in equilibrium -, firm-size distribution in revenues is well approximated by a Pareto distribution with Pareto shape ζ_j, close to 1. Write $g_j(r)$ the distribution of revenues in sector j:

$$\exists r_{j,o} \ : \ \forall \ r > r_{j,o}, \ g_j(r) \propto r^{-\zeta_j - 1} \tag{C.45}$$

Step 3: The city size distribution $f_L(.)$ as defined by equation 17 can be written as $f_L(L) = \sum_{j=1}^{S} M_j \tilde{f^j}(L)$, where $\tilde{f^j}(L) = \frac{\ell_j(z_j^*(L))\, f_j(z_j^*(L)) \, \frac{dz_j^*(L)}{dL}}{L}$. Then,

$$\tilde{f^j}(L) \propto \ g_j(r^*(L))\frac{dr_j^*}{dL}\frac{r_j(L)}{L^{b\frac{\eta}{1-\eta}+1}} \propto \frac{r_J(L)^{-\zeta_j+1}}{L^{b\frac{\eta}{1-\eta}+2}}$$

The first inequality uses (C.44), $\ell_j \propto \frac{r_j(L)}{w(L)}$ and $w(L) \propto L^{b\frac{\eta}{1-\eta}}$. The second uses the assumption that r^* has constant elasticity with respect to L at the sorting equilibrium, and (C.45). As $\zeta \sim 1$,

$$\tilde{f^j}(L) \propto \frac{1}{L^{b\frac{\eta}{1-\eta}+2}} \ \text{ and } \ f_L(L) \propto \frac{1}{L^{b\frac{\eta}{1-\eta}+2}}.$$

Finally, $b\frac{\eta}{1-\eta} << 1$, as it measures the elasticity of wages with city size, on the order of magnitude of 5%. The tail of the city size distribution is well approximated by a Pareto of shape close to 1. City size distribution follows Zipf's law. ∎

C.6. Proposition 6

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, $\pi(z_H, L_H, s_j) \geq \pi(z_H, L_L, s_j)$ 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 $\epsilon_w = b\frac{\eta}{1-\eta}$ ∎

C.7. Proposition 7

Proof The proof covers both the case of the main assumptions of the model (continuity and convexity of the support of z and L), and the case where the set of city sizes is exogenously given (as detailed above in B.4.), and in particular discrete. Let $\mathcal{Z} : \mathcal{L}$ x A x $E \to Z$ 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 under the assumptions made in the main text (see proof of Lemma 4).) 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 12 *$log\, \pi$ 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 13 *$\bar{z}(L, \alpha, s)$ 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 $(z_H, \alpha_H, s_H) \geq (z_L, \alpha_L, s_L) \Rightarrow L^*(z_H, \alpha_H, s_H) \geq L^*(z_L, \alpha_L, s_L)$. Note that everywhere, the $\geq$ 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 $L^*(z, \alpha_H, s) \geq L^*(z, \alpha_L, s)$. In particular, fix a given L and s and write using shorter notation: $\bar{z}_{\alpha_L} = \bar{z}(L, \alpha_L, s)$. Then $L^*(\bar{z}_{\alpha_L}, \alpha_H, s) \geq L^*(\bar{z}_{\alpha_L}, \alpha_L, s) = L$. Because $L^*(z, \alpha_H, s)$ is increasing in z, it follows that:

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

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. ∎

C.8. Proposition 8

Proof The proof covers both the case of the main assumptions of the model (continuity and convexity of the support of z and L), and the case where the set of city sizes is exogenously given (as detailed above in B.4.), and in particular discrete. 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. ∎

C.9. Corollary 9

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:

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

$$\frac{r}{r_j} \approx (1 + p_j - p)^{-\frac{1}{\zeta_j}}$$

Letting $p_0 = \max(p_1, p_2)$ and writing $r_j = Q_j(p_0)$ for $j = 1, 2$, and using proposition (8) gives:

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

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

D. Stability

I verify here that the equilibrium described in section 2 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

$$\sum_j n_j K_j \frac{\psi(z_j, L, s_j)^{\sigma_j - 1}}{w(L)^{(1-\alpha_j)(\sigma_j - 1)+1}} \left[(\sigma_j - 1)\frac{\psi_2}{\psi} - ((1-\alpha_j)(\sigma_j - 1) + 1)\frac{w'(L)}{w(L)} \right] = .1 \tag{D.46}$$

Using in equation (11) 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 (D.46) 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:

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

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:

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

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. ∎

E. General Equilibrium

I solve here the system of $2S+1$ equations in $2S+1$ unknowns $(R, M_j, P_j)_{j=1\ldots S}$ of the main text (equations 14-16).First, the national labor market clearing condition (16) together with equation (15) leads to the aggregate revenues in manufacturing,

$$R = N\frac{1-(1-b)(1-\eta)}{\sum_{j=1}^{S}\frac{(1-\alpha_j)(\sigma_j-1)}{\sigma_j}\xi_j\frac{E_j}{S_j}} \tag{E.47}$$

This pins down uniquely the general equilibrium quantity R. Second, I combine equations (E.49) and (14) 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.[2] This leads to :

$$P_j^{\frac{1}{\alpha_j}} \prod_{k=1}^{S} (\frac{P_k}{\xi_k})^{-\xi_k} = \left(\frac{1}{\tilde{\kappa}_{1j}(1+T_j^*)\, S_j\, \xi_j R} \right)^{\frac{1}{\alpha_j(\sigma_j-1)}}, \tag{E.48}$$

where I have used that $P = \prod_{j=1}^{S} (\frac{P_j}{\xi_j})^{\xi_j}$. Note that the matrix $\begin{bmatrix} -\frac{1}{\alpha_1}+\xi_1 & \xi_2 & \dots & \xi_n \\ \xi_1 & -\frac{1}{\alpha_2}+\xi_2 & \dots & \xi_n \\ \vdots & \vdots & & \vdots \\ \xi_1 & \xi_2 & \dots & -\frac{1}{\alpha_n}+\xi_n \end{bmatrix}$ has full rank and is invertible. Therefore, equation (E.48) has a unique solution in $\{P_j\}_{j=1..N}$. This pins down P in turn. Finally, equations (15) leads to the sectoral mass of firms:

$$M_j = \frac{P^{\alpha_j(\sigma_j-1)}}{\sigma_j \widetilde{\kappa}_{1j}\, S_j\, P_j^{\sigma_j-1}}. \tag{E.49}$$

Therefore, equations 14-16 have a unique solution $(R, M_j, P_j)_{j=1...S}$.

F. Welfare analysis

F.1. Social planner's problem

The utility function is as follows[3]:

$$U(L) = c(L) L^{-\frac{b(1-\eta)}{\eta}}. \tag{F.50}$$

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 $G(L)$
(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. $C = Q - M f_E - M\rho \int k(z) dF(z)$; $Q = \left(\int M q(z, L(z))^{\frac{\sigma-1}{\sigma}} dF(z) \right)^{\frac{\sigma}{\sigma-1}}$ 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)

[2] 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.

[3] 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 (F.50) to the power $\frac{1}{\eta}$.

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[4]:

$$dG(L) = \frac{M \mathbb{1}(L(z)) \, \ell(z)}{L(z)} dF(z) \tag{F.51}$$

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:

$$\max_{L(z),\ell(z),k(z),M} \frac{C}{\Gamma} \tag{F.52}$$

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

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

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

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:

$$\frac{\psi_2(z, L, s) L}{\psi(z, L, s)} = b\frac{1-\eta}{\eta}(1-\alpha)\chi(z), \tag{F.53}$$

where[5]

$$\chi(z) = \left(\frac{\tilde{Q} - f_E}{\tilde{Q}} \right) \frac{\ell(z) L(z)^{b\frac{1-\eta}{\eta}}}{\int \ell(z) L(z)^{b\frac{1-\eta}{\eta}}} \frac{\int \tilde{q}(z)^{\phi} \, dF(z)}{\tilde{q}_j^{\phi}}. \tag{F.54}$$

The first order condition with respect to M yields

$$\frac{1}{(\sigma-1)(1-\alpha)} \left(\frac{\tilde{Q}}{\tilde{Q} - f_E} \right) = \lambda N \tag{F.55}$$

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

$$\left(\frac{\tilde{Q}}{\tilde{Q} - f_E} \right) \frac{\tilde{q}_j^{\phi}}{\int \tilde{q}(z)^{\phi} \, dF(z)} - \frac{\ell(z) L(z)^{b\frac{1-\eta}{\eta}}}{\int \ell(z) L(z)^{b\frac{1-\eta}{\eta}} dF(z)} = \lambda M \ell(z) \tag{F.56}$$

In particular, summing this over all types of firms and using (F.55) and the labor market clearing condition lead to:

$$\frac{f_E}{\tilde{Q}} = \frac{1}{(\sigma-1)(1-\alpha)}, \tag{F.57}$$

[4]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.

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

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 (F.56) and using (F.55) gives the following expression for $\ell(z)$:

$$\ell(z) = \left(\frac{\psi(z, L, s)}{\left(\int \tilde{q}(z)^{\phi}\right)^{\frac{1}{\phi}}}\right)^{\sigma-1} \left(\frac{\tilde{Q} - f_E}{\tilde{Q}} \frac{L(z)^{b\frac{1-\eta}{\eta}}}{\int \ell(z)L(z)^{b\frac{1-\eta}{\eta}}\, dF(z)} + \frac{M}{N}\frac{1}{(\sigma-1)(1-\alpha)}\right)^{\alpha\sigma-\alpha-\sigma}. \tag{F.58}$$

F.2. Comparison with the competitive equilbrium

Rearranging equation (F.54), using (F.58) and (F.57), leads to

$$\chi(z) = \frac{L(z)^{b\frac{1-\eta}{\eta}}}{L(z)^{b\frac{1-\eta}{\eta}} + \frac{\Gamma}{N}\frac{1}{(\sigma-1)(1-\alpha)-1}}, \tag{F.59}$$

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 (11)) and the social planner problem (equation (F.53)). 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.

F.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

$$w(L) \propto (L^{b\frac{1-\eta}{\eta}} + A) \tag{F.60}$$

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.[6]

G. Estimation

G.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

[6] 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 $M = \frac{1}{(\sigma-1)(1-\alpha)}\frac{Q}{f_E}$

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:

$$E(\pi|z) = \Gamma(\nu)\left[\sum_L V(z,L)^\nu\right]^{\frac{1}{\nu}}, \tag{G.61}$$

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:

$$p(L|z) = \frac{V(z,L)^\nu}{\sum_{L'} V(z,L')^\nu} \tag{G.62}$$

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 (B.31) and the definition of productivity in equation (23) 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:

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

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[7] 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:

$$log(E[r_i|L_i]) = C'' - \log\left(f_L(L_i)\right) + \beta_1 \log L_i + \beta_2\left(1+\log\,L_i\right)^{\mathbf{2s}} + \log\left[\Phi((\sigma_j-1)\,S(L_i)\nu_z)\right], \tag{G.63}$$

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.

G.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 (G.63) as the sample mean M_q of $\log(\bar{r}_j(L_i))$. The targeted moments are $\{M_q\}_{q=1,2,3,4}$.

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

[7] 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,

i is $s_i^{L,j} = \frac{\sum\limits_{t_i^L \leq L < t_{i+1}^L} \int r_j^*(z) \mathbb{1}_{L_j^*(z)=L} \, dF_j(z)}{\int r_j^*(z) dF_j(z)}$, 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 $s_i^{r,j} = \frac{\int \mathbb{1}_{t_i^{r,j} \leq \tilde{r}_j(z) < t_{i+1}^{r,j}} dF_j(z)}{\int dF_j(z)}$, where $\tilde{r}_j(z)$ is the normalized value added of firm z in sector j.

G.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.[8] 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 (24).

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 (25)) 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.[9]

G.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:

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

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.

[8] As pointed out in the theory section and developed above in B.4., the characterizations of the economy provided in Section 2 hold if the set of possible city sizes is exogenously given.

[9] 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 (18) and (19)). Normalized by its median value, the distribution of firm sizes within a sector is fully determined by the matching function.

G.5. Variance decomposition

The estimation then allows me to quantify the importance of the systematic (portable) vs iid (non-portable) component of productivity that allows to match the data best. Armed with the parameter estimates of the model, I quantify the contribution of both components of firm productivity as follows. I run the following variance decomposition regressions to quantify the part of the variance in productivity coming from the systematic vs the random component of productivity in each sector:

$$\log(z_i)(1+\log\frac{L}{L_o})^{s_j} = \beta_{j,systematic}\log(\widetilde{\phi}_i) + \nu_i$$
$$\epsilon_{iL} = \beta_{j,random}\log(\widetilde{\phi}_i) + \nu_i,$$

where $\log(\widetilde{\phi}_i) = \log(z_i)(1+\log\frac{L}{L_o})^{s_j} + \epsilon_{iL}$. Mechanically, this variance decomposition procedure yields β coefficients that sum to one and give us a metric for the relative importance of sorting vs random shocks to shape the distribution of firms' productivities. I find that, on average across sectors, $\beta_{systematic}$ =51%. In that sense, roughly half of the variation in firm productivity comes from the portable component of its productivity (reinforced by sorting), and the other half comes from the other dimension of firm location choice.

H. Policy analysis

H.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.[10]

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.

H.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.49) and (E.48), 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.

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

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.

I. Additional figures

I.1. Model fit

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

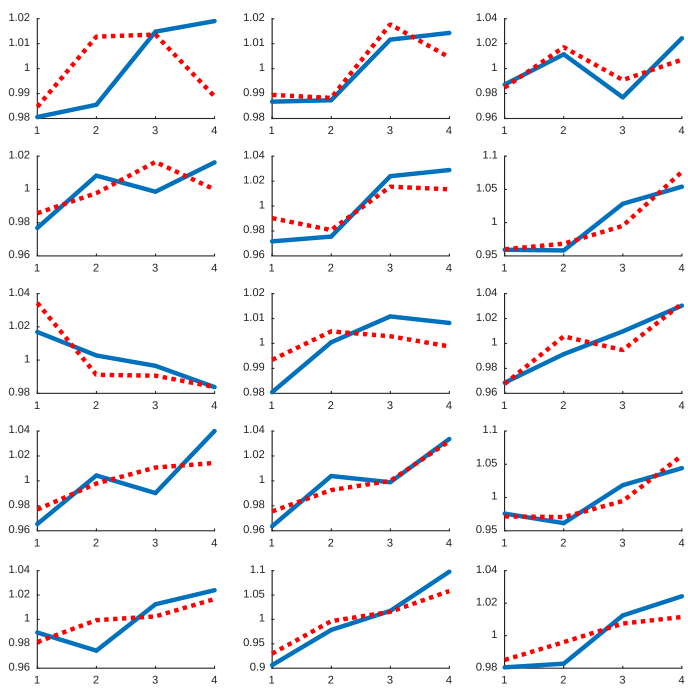


Figure I.2: Sectoral distribution of firms revenues, model (blue) and data (red).

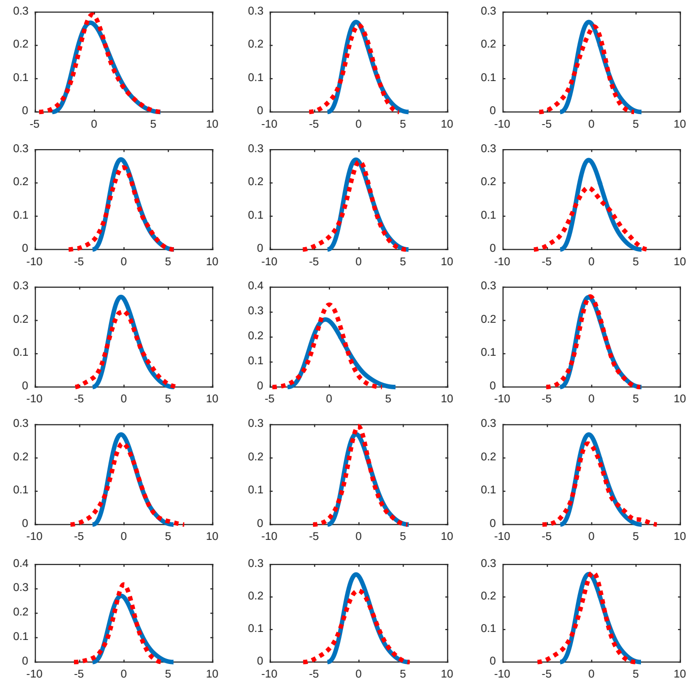
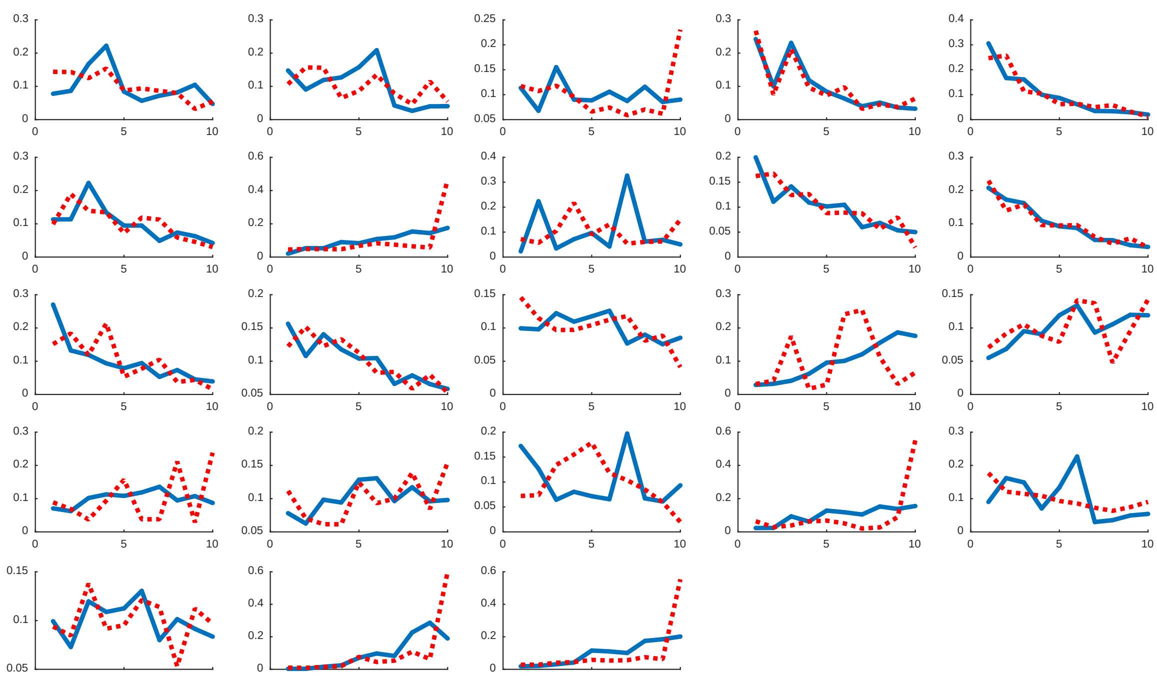
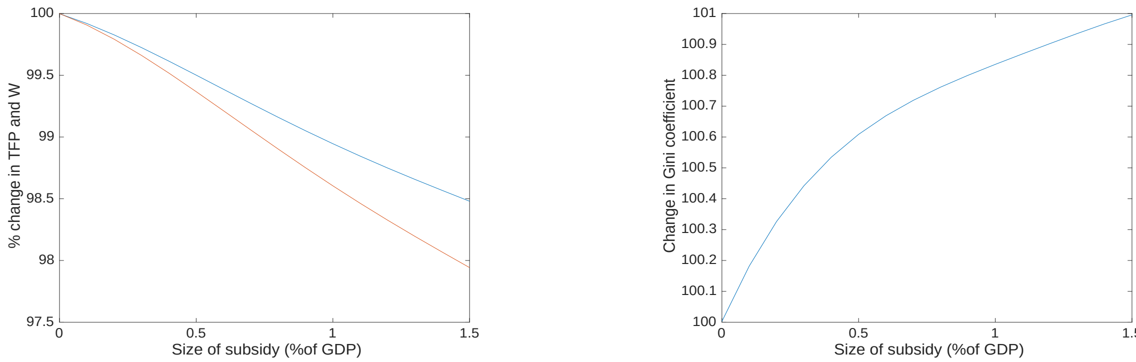


Figure I.3: Employment share by decile of city size, model (blue) and data (red).



I.2. Impact of policies

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

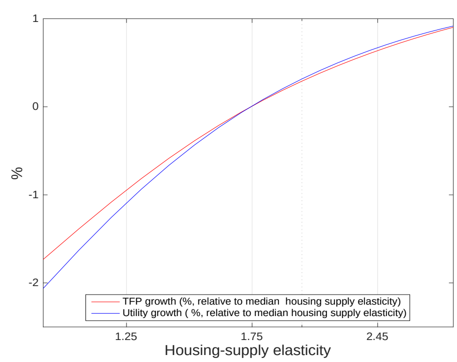


A. TFP (red) and welfare impact (blue), relative to the reference equilibrium

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 I.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.