Spatial frictions

Kristian Behrens* Giordano Mion† Yasusada Murata‡ Jens Südekum§

September 8, 2011

Abstract

The world is replete with spatial frictions. Shipping goods across cities entails trade frictions. Commuting within cities causes urban frictions. How important are these frictions in shaping the spatial economy? We develop and quantify a novel framework to address this question at three different levels: Do spatial frictions matter for the city-size distribution? Do they affect individual city sizes? Do they contribute to the productivity advantage of large cities and the nature of competition in cities? The short answers are: no, yes, and it depends.

Keywords: trade frictions; urban frictions; productivity; city-size distribution; markups

JEL Classification: F12; R12

*Canada Research Chair in Regional Impacts of Globalization, Department of Economics, Université du Québec à Montréal (UQAM), Canada; CIRPÉE, Canada; and CEPR, UK. E-mail: behrens.kristian@uqam.ca.

†Department of Geography and Environment, London School of Economics (LSE), UK; CEP, UK; and CEPR, UK. E-mail: g.mion@lse.ac.uk

‡Advanced Research Institute for the Sciences and Humanities (ARISH), Nihon University, Tokyo, Japan; and Nihon University Population Research Institute (NUPRI), Japan. E-mail: murata.yasusada@nihon-u.ac.jp

§Mercator School of Management, Universität Duisburg-Essen, Germany; Ruhr Graduate School of Economics, Germany; and IZA, Germany. E-mail: jens.suedekum@uni-due.de

1 Introduction

The world is replete with spatial frictions. Trade frictions for shipping goods across cities induce consumers and firms to spatially concentrate to take advantage of large local markets. Yet, such a concentration generates urban frictions within cities – people spend a lot of time on commuting and pay high land rents. Economists have studied this fundamental trade-off for decades, analyzing how firms and workers choose their locations depending on the magnitudes of – and changes in – spatial frictions (Fujita *et al.*, 1999; Fujita and Thisse, 2002). Still, little is known to date about how important urban and trade frictions are in shaping the spatial economy. How would the US economic geography look like if there were no spatial frictions? More specifically, we focus on the following three questions: Do spatial frictions matter for the city-size distribution? Do they affect individual city sizes? Do they contribute to the productivity advantage of large cities and the nature of competition in cities?

To address these questions, we develop a novel multi-city general equilibrium model with urban and trade frictions, where city sizes, productivity, and competition are all endogenous. Using data for 356 US metropolitan statistical areas (MSAs) in 2007, we structurally estimate the model and conduct two counterfactual experiments taking into account all market and spatial equilibrium conditions. We first explore what would happen in a hypothetical world where commuting within cities is costless. We then analyze the other counterfactual scenario where consumers face the same trade costs for local and non-local products. By comparing the actual and the counterfactual equilibria in both cases, we can assess the importance of urban and trade frictions for the city-size distribution, individual city sizes, as well as productivity and competition.

How important are spatial frictions in shaping the spatial economy? First, we find that neither type of frictions significantly affects the US city-size distribution. Even in a world without urban or trade frictions, that distribution would follow the rank-size rule – also known as Zipf's law – fairly well. Second, we find that eliminating spatial frictions would change individual city sizes within the stable distribution. Without urban frictions, large congested cities like New York or cities close by (e.g., New Haven-Milford, CT) would gain population, while small isolated cities (e.g., Casper, WY) would lose population. In contrast, without trade frictions, large cities would shrink compared to small cities as local market access no longer matters. In total, about 4 million people would move in the former and around 10 million people would move in the latter case. Last, turning to productivity and competition, eliminating trade frictions would lead to aggregate productivity gains of 68% and markup reductions of 40%, both of which are unevenly distributed across MSAs. Eliminating urban frictions would generate smaller productivity gains of less than 1%, but still lead to a notable markup reduction of about 10%. In a nutshell, spatial frictions do not matter for the city-size distribution, they do matter for individual city sizes, and they matter for productivity and competition to a different extent depending on the type of frictions we consider.

Our findings have clear-cut implications for future spatial modeling. As far as the city-size distribution is concerned, our results suggest that we can abstract from either urban or trade frictions without loss of generality. Hence, the recent modeling strategies taken by Gabaix (1999), Eeckhout (2004), Duranton (2007) and Rossi-Hansberg and Wright (2007), where trade frictions are assumed away, provide good approximations. However, to explain the rise and fall of individual cities within the stable distribution requires a model that takes both types of spatial frictions into account. Our results also suggest that such a model may be needed to understand productivity and markup differences across cities.

What ingredients are required in our framework? Obviously, we need a system of cities as in Henderson (1974), extended to include spatial frictions within *and* across cities. Both urban and trade frictions are introduced in standard ways. For urban frictions, we use a monocentric city model with commuting costs and land rents as in Alonso (1964) and Fujita (1989). To capture trade frictions, we rely on a monopolistic competition model with trade costs as in the new trade theory and the new economic geography (NEG). However, workhorse constant elasticity of substitution (CES) models such as Krugman (1980, 1991) do not account for the empirical facts that large cities are more productive, more competitive, and allow for greater consumption diversity (see Syverson, 2004; Handbury and Weinstein, 2011).[1] We incorporate all these aspects into a single framework by building on recent developments in the heterogeneous firms literature. The two prominent approaches, however, have limitations for our purpose: in Melitz (2003) the CES specification implies constant markups so that spatial frictions do not matter for competition; whereas in Melitz and Ottaviano (2008) the quasi-linear specification rules out income effects of demand for varieties and, more importantly, imposes restrictions on feasible city size differences.[2] The latter feature is not well suited to urban settings where observed city sizes substantially differ and counterfactual city sizes are a priori unknown.

To overcome those limitations, we develop a novel multi-city monopolistic competition framework that allows for the joint determination of city sizes, productivities, markups, wages, consumption diversity, and the number and size distribution of firms.[3] City sizes are determined by aggregating individual location decisions based on wages, rents, and prices, which in turn, are influenced by spatial frictions and amenities. We model these location decisions by using discrete choice theory as in McFadden (1974), and embed the choices into spatial equilibrium conditions following Tabuchi and Thisse (2002) and Murata (2003).

Our multi-city framework features multiple margins of adjustment to shocks in spatial frictions. Given the distribution of population, changes in spatial frictions directly affect the productivity advantage of cities and the nature of competition in cities. Such changes in productivity and

[1]Early work by Krugman (1979, Section 3.3) sheds light on the latter two issues, using an aspatial model with variable elasticity of substitution (VES). Ottaviano *et al.* (2002) develop a NEG model featuring VES in which large markets are more competitive and have lower markups.

[2]More specifically, the quasi-linear framework requires that market size differences are bounded to maintain an equilibrium with incomplete specialization (see Melitz and Ottaviano, 2008, footnote 18).

[3]Holmes *et al.* (2010) also depart from the CES and quasi-linear frameworks and develop a two-region NEG model building on Bernard *et al.* (2003) to explore the issues of productivity and regional agglomeration from a theoretical perspective.

competition, in turn, induce changes in indirect utility differences across cities – through changes in wages, rents, and prices – thereby affecting individual location decisions. Put differently, shocks to spatial frictions are absorbed into: productivity and competition, as in the heterogeneous firms literature; and population movements, as in the urban economics and NEG literatures. Despite the richness of our setting, we can derive clear comparative static results with two cities. We show that, for a given population distribution, firms in the larger city face the higher wage and tougher selection to offset their advantage of having the larger local market. At the same time, commuting costs and land rents are higher in the larger city, which reduces its attractiveness. Ceteris paribus, eliminating urban frictions favors agglomeration by increasing the number of people who choose the larger city, while eliminating trade frictions induces dispersion by making the smaller city more attractive.

With these qualitative results in hand, we quantify the multi-city model for the US. We first use MSA-level data on population, commuting time, and hours worked to compute city-specific measures of urban frictions. Then we estimate a gravity equation for trade flows to obtain a measure of trade frictions. The friction parameters thus obtained are used in the market equilibrium conditions to back out unobserved MSA-level technological possibilities. This allows us to structurally identify the parameters of firms' productivity distributions by matching predicted and observed firm size distributions. Finally, we use the spatial equilibrium conditions to perfectly fit the observed US city-size distribution. In doing so, we pin down the relative weight of economic variables and observed amenities in determining individual location decisions, and back out measures of unobserved amenities at the MSA level.[4] We pay particular attention to model fit and verify that our framework can reproduce several empirical features at the MSA and firm levels. For example, it fairly well replicates the observed patterns of aggregate land rents that are linked to urban frictions. It also replicates reasonably well the distribution of average wages across MSAs and matches available micro-evidence on the spatial structure of US firms' shipments (Hummels and Hilberry, 2008; Holmes and Stevens, 2010) that are linked to trade frictions.

Our quantitative analysis contributes to both the recent empirical NEG and urban economics literatures. Although the former literature has made some important progress recently (e.g., Hanson, 2005; Redding and Sturm, 2008; Redding, 2010; Combes and Lafourcade, 2011), NEG models have still been confronted with data mostly in a reduced-form manner. It is fair to say that few attempts have been made to conduct comprehensive counterfactual experiments. One notable exception in the urban economics literature is the recent paper by Desmet and Rossi-Hansberg (2010), who investigate the contribution of different wedges to the observed US city-size distribution. Unlike the NEG literature, however, their framework builds on a perfect competition model and abstracts from trade between cities. Hence, it is not suited to investigate how trade frictions affect city sizes, productivity, and competition.

The rest of the paper is organized as follows. In Section 2, we develop a single-city model

[4]Contrary to more conventional hedonic approaches (e.g., Roback, 1982; Albouy, 2008), unobserved amenities and technological possibilities are obtained here from a model that encompasses both trade and urban frictions.

to highlight some basic properties. In Section 3, we extend it to a multi-city framework and provide comparative static results for the case with two cities. Section 4 describes our quantification procedure and discusses the model fit. We then turn to our counterfactual experiments in Section 5. Section 6 provides some extensions and discussion of our main results. Section 7 concludes.

2 Basic model: Single city

We consider a mass $L > 0$ of identical consumers/workers and a large amount of land that stretches out on a two-dimensional featureless plane. Labor is the only factor of production and land is used for housing only. Each agent consumes inelastically one unit of land, and the amount of land available at each location is set to one. All firms in the city are located at an exogenously given and dimensionless Central Business District (CBD). A monocentric city of size L then covers the surface of a disk with radius $\bar{x} \equiv \sqrt{L/\pi}$, with the CBD located in the middle of that disk and the workers evenly distributed within it.

We introduce *urban frictions* in a standard way into our model by assuming that agents have to commute to the CBD for work. In particular, we assume that each individual is endowed with $\overline{h}$ hours of time, which is the gross labor supply per capita (in terms of hours) including commuting time. Commuting costs are of the 'iceberg' type: the *effective* labor supply of a worker living at a distance $x \leq \bar{x}$ from the CBD is given by

$$s(x) = \overline{h}\mathrm{e}^{-\theta x}, \tag{1}$$

where $\theta \geq 0$ captures the efficiency loss due to commuting within the city.[5] The *total effective labor supply* at the CBD is then given by

$$S = \int_0^{\bar{x}} 2\pi x s(x)\mathrm{d}x = \frac{2\pi\overline{h}}{\theta^2}\left[1 - \left(1 + \theta\sqrt{L/\pi}\right)\mathrm{e}^{-\theta\sqrt{L/\pi}}\right]. \tag{2}$$

In what follows, it will be useful to define the *effective labor supply per capita* as $h \equiv S/L$, which is the average number of hours worked in the city. It directly follows from (2) that S is positive and increasing in L, while h is decreasing in L. That is, given gross labor supply per capita $\overline{h}$ and commuting technology $\theta > 0$, the effective labor supply per capita is lower in a larger city. We can further show that h is decreasing in θ, which captures urban frictions. With $\theta = 0$, we would have $h = \overline{h}$ regardless of the city size L.

Let w stand for the wage rate paid to the workers by the firms at the CBD. Then, the wage income net of commuting costs earned by a worker residing at the city edge is $ws(\bar{x}) = w\overline{h}\mathrm{e}^{-\theta\bar{x}}$.

[5] We use an exponential rather than a linear iceberg commuting cost (as in, e.g., Murata and Thisse, 2005) since the linear specification is subject to a boundary condition to ensure positive effective labor supply at each location in the city. Keeping track of this condition becomes tedious with multiple cities and intercity movements of people. The negative exponential specification has been used extensively in the literature (e.g., Lucas and Rossi-Hansberg, 2002), and the convexity of the efficiency loss with respect to distance from the CBD can also be justified in a modal choice framework of intra-city transportation (e.g., Glaeser, 2008, pp.24–25).

Because workers are identical, the wages net of commuting costs and land rents are equalized across all locations in the city: $ws(x) - R(x) = ws(\bar{x}) - R(\bar{x})$, where $R(x)$ is the land rent at a distance x from the CBD. We normalize the opportunity cost of land at the urban fringe to zero, i.e., $R(\bar{x}) = 0$. The equilibrium land rent schedule in the city is then given by $R^*(x) = w(\mathrm{e}^{-\theta x} - \mathrm{e}^{-\theta\bar{x}})\overline{h}$, which yields the following aggregate land rents:

$$\mathrm{ALR} = \int_0^{\bar{x}} 2\pi x R^*(x)\mathrm{d}x = \frac{2\pi w\overline{h}}{\theta^2}\left[1 - \left(1 + \theta\sqrt{L/\pi} + \frac{\theta^2 L}{2\pi}\right)\mathrm{e}^{-\theta\sqrt{L/\pi}}\right]. \tag{3}$$

In what follows, we assume that each worker owns an equal share of the land in the city and has an equal claim to firms' profits. Accordingly, in addition to the wage net of commuting costs and land rent, each worker receives an equal share of aggregate land rents ALR, and an equal share of aggregate profits Π. The expenditure per capita is then given by $E = w\overline{h}\mathrm{e}^{-\theta\sqrt{L/\pi}} + (\mathrm{ALR} + \Pi)/L$.

2.1 Preferences and demands

We assume that there is a continuum of horizontally differentiated varieties available for consumption. Denote by Ω the endogenously determined set of varieties, with measure N. All consumers have identical preferences that display 'love of variety' and give rise to demands with variable elasticity. Following Behrens and Murata (2007), the utility maximization problem of a representative consumer is given by:

$$\max_{q(j),\, j\in\Omega} \; U \equiv \int_\Omega \left[1 - \mathrm{e}^{-\alpha q(j)}\right]\mathrm{d}j \qquad \text{s.t.} \quad \int_\Omega p(j)q(j)\mathrm{d}j = E, \tag{4}$$

where $p(j) > 0$ and $q(j) \geq 0$ stand for the price and the per capita consumption of variety j; and $\alpha > 0$ is a parameter. Solving (4) yields the following demand functions:

$$q(i) = \frac{E}{N\overline{p}} - \frac{1}{\alpha}\left\{\ln\left[\frac{p(i)}{N\overline{p}}\right] + \eta\right\}, \quad \forall i \in \Omega, \tag{5}$$

where

$$\overline{p} \equiv \frac{1}{N}\int_\Omega p(j)\mathrm{d}j \quad \text{and} \quad \eta \equiv -\int_\Omega \ln\left[\frac{p(j)}{N\overline{p}}\right]\frac{p(j)}{N\overline{p}}\mathrm{d}j$$

denote the average price and the differential entropy of the price distribution, respectively.[6] Since marginal utility at zero consumption is bounded, the demand for a variety need not be positive. Indeed, as can be seen from (5), the demand for variety i is positive if and only if its price is lower than the *reservation price* p^d. Formally,

$$q(i) > 0 \quad \Longleftrightarrow \quad p(i) < p^d \equiv N\overline{p}\,\mathrm{e}^{\frac{\alpha E}{N\overline{p}} - \eta}. \tag{6}$$

[6] As shown in Reza (1994, pp.278–279), the differential entropy η takes its maximum value when there is no price dispersion, i.e., $p(i) = \overline{p}$ for all $i \in \Omega$. In that case, we would observe $\eta = -\ln(1/N)$ and thus $q(i) = E/(N\overline{p})$ by (5). Behrens and Murata (2007) entirely focus on such a symmetric case. By contrast, this paper considers firm heterogeneity so that not only the average price $\overline{p}$, but also the entire price distribution matters for the demand $q(i)$. The differential entropy η in (5) does capture the dispersion of the price distribution.

Note that the reservation price p^d is a function of the price aggregates $\overline{p}$ and η. Combining expressions (5) and (6) allows us to rewrite the demand for variety i concisely as follows:

$$q(i) = \frac{1}{\alpha}\ln\left[\frac{p^d}{p(i)}\right]. \tag{7}$$

Observe that the price elasticity of demand for variety i is given by $1/[\alpha q(i)]$. Thus, if individuals consume more of this variety (which is, e.g., the case when their expenditure increases), they become less price sensitive. Last, since $\mathrm{e}^{-\alpha q(i)} = p(i)/p^d$, the indirect utility is given by

$$U = N - \int_\Omega \frac{p(i)}{p^d}\mathrm{d}i = N\left(1 - \frac{\overline{p}}{p^d}\right). \tag{8}$$

2.2 Technology and market structure

Each variety is produced by a single firm. The labor market is perfectly competitive so that all firms at the CBD take the wage rate w as given. Prior to production, each firm enters the market by engaging in research and development, which requires a fixed amount F of labor paid at the market wage. Each firm i discovers its marginal labor requirement $m(i) \geq 0$, expressed in terms of hours of labor required per unit of output, only after making this irreversible entry decision. We assume that $m(i)$ is drawn from a common and known, continuously differentiable distribution G. Since F is sunk, an entrant will stay in the market provided it can charge a price $p(i)$ above marginal cost $m(i)w$. Each surviving firm sets its price to maximize operating profits

$$\pi(i) = L\big[p(i) - m(i)w\big]q(i), \tag{9}$$

where $q(i)$ is given by (7). Since there is a continuum of firms, no individual firm has any impact on the reservation price. All firms therefore take p^d as given, so that the first-order conditions for (operating) profit maximization are:

$$\ln\left[\frac{p^d}{p(i)}\right] = \frac{p(i) - m(i)w}{p(i)}, \quad \forall i \in \Omega. \tag{10}$$

A price distribution satisfying (10) is called a *price equilibrium*. Equations (7) and (10) imply that $q(i) = (1/\alpha)[1 - m(i)w/p(i)]$. Thus, the minimum output is given by $q(i) = 0$ at $p(i) = m(i)w$ which, by (10), implies that $p(i) = p^d$. The cutoff marginal labor requirement for surviving in the market is then defined as $m^d \equiv p^d/w$. All firms that draw $m \geq m^d$ choose not to produce, whereas all firms with a draw $m < m^d$ will operate in equilibrium. Hence, given a mass of entrants N^E, only a fraction $G(m^d)$ of them will have positive output. The mass of surviving firms, which is identical to the mass of varieties consumed in the single city case, is then given by $N = N^E G(m^d)$.

Since firms differ by their marginal labor requirement only, we can write down all firm-level variables in terms of m. Solving (10) by using the Lambert W function, defined as $\varphi = W(\varphi)\mathrm{e}^{W(\varphi)}$, the profit-maximizing prices and quantities, as well as operating profits, can be expressed as follows:[7]

$$p(m) = \frac{mw}{W}, \quad q(m) = \frac{1}{\alpha}(1 - W), \quad \pi(m) = \frac{Lmw}{\alpha}\left(W^{-1} + W - 2\right), \tag{11}$$

where we suppress the argument $\mathrm{e}m/m^d$ of W to alleviate notation. It is readily verified that $W' > 0$ for all non-negative arguments and that $W(0) = 0$ and $W(\mathrm{e}) = 1$. Hence, $0 \leq W \leq 1$ if $0 \leq m \leq m^d$. The expressions in (11) then show that a firm with a draw m^d charges a price equal to marginal cost, faces zero demand, and earns zero operating profit. Since $W' > 0$, we readily obtain $\partial p(m)/\partial m > 0$, $\partial q(m)/\partial m < 0$ and $\partial \pi(m)/\partial m < 0$. In words, firms with better productivity draws charge lower prices, sell larger quantities, and earn higher operating profits as in Melitz (2003). Yet, our specification with variable demand elasticity also features higher markups for more productive firms. Indeed,

$$\Lambda(m) \equiv \frac{p(m)}{mw} = \frac{1}{W} \tag{12}$$

implies that $\partial \Lambda(m)/\partial m < 0$. Unlike Melitz and Ottaviano (2008), who use quasi-linear preferences, we incorporate this feature into a full-fledged general equilibrium model with income effects for varieties and without restrictions on feasible city size differences. Since $\partial W/\partial m^d < 0$, firm-level markups are also smaller in tougher markets, which is in line with firm-level evidence (see Syverson, 2004).

2.3 Equilibrium

The equilibrium conditions in the single city case consist of zero expected profits and labor market clearing. These two conditions can be solved for the cutoff m^d and the mass of entrants N^E. Using expression (9), the zero expected profit condition for each firm is given by:

$$L \int_0^{m^d} [p(m) - mw]\, q(m) \mathrm{d}G(m) = Fw. \tag{13}$$

This expression, combined with (11), can be rewritten as a function of m^d only:

$$\frac{L}{\alpha} \int_0^{m^d} m \left(W^{-1} + W - 2\right) \mathrm{d}G(m) = F, \tag{14}$$

which yields a unique equilibrium cutoff because the left-hand side of (14) is strictly increasing in m^d from 0 to ∞ (see Appendix A.1). Furthermore, the labor market clearing condition is given by:[8]

$$N^E \left[L \int_0^{m^d} mq(m) \mathrm{d}G(m) + F \right] = S \tag{15}$$

[7] Further details about the Lambert W function, the technical properties of which are key to making our model tractable, can be found in Appendix A.2 of Behrens *et al.* (2009).

[8] From (13) and $N^E \int_0^{m^d} p(m)q(m)\mathrm{d}G(m) = E$, we obtain $EL/(wN^E) = L \int_0^{m^d} mq(m)\mathrm{d}G(m) + F$ which, together with (15), yields $E = (S/L)w = hw$ in equilibrium. The expenditure of the representative consumer thus depends only on the effective labor supply per capita and the wage rate.

which, combined with (11), can be rewritten as a function of m^d and N^E:

$$N^E \left[\frac{L}{\alpha} \int_0^{m^d} m\,(1-W)\,\mathrm{d}G(m) + F \right] = S. \tag{16}$$

Given the equilibrium cutoff m^d from (14), equation (16) can be uniquely solved for N^E. As in Melitz and Ottaviano (2008) and many other existing studies, we choose a particular distribution function for firms' productivity draws, $1/m$, namely a Pareto distribution:

$$G(m) = \left(\frac{m}{m^{\max}}\right)^k,$$

where $m^{\max} > 0$ and $k \geq 1$ are the upper bound and the shape parameter, respectively. This distribution is useful for deriving analytical results and taking the model to the data. In particular, we obtain the following closed-form solutions for the equilibrium cutoff and the mass of entrants:

$$m^d = \left(\frac{\mu^{\max}}{L}\right)^{\frac{1}{k+1}} \quad \text{and} \quad N^E = \frac{\kappa_2}{\kappa_1+\kappa_2}\frac{S}{F}, \tag{17}$$

where κ_1 and κ_2 are positive constants that solely depend on k (see Appendices B.1 and B.2 for details). The term $\mu^{\max} \equiv \left[\alpha F (m^{\max})^k\right]/\kappa_2$ can be interpreted as an inverse measure of *technological possibilities*: the lower is the fixed labor requirement for entry, F, or the lower is the upper bound, $m^{\max}$, the lower is $\mu^{\max}$.

How do population size and technological possibilities affect entry and selection? Recall from (2) that the total effective labor supply, S, is increasing in population L. The second expression in (17) then shows that there are more entrants in a larger city. The first expression in (17), in turn, shows that a larger L or a smaller $\mu^{\max}$ entail a smaller cutoff and, thus, a lower survival probability $G(m^d)$ of entrants. This tougher selection maps into higher average productivity, $1/\overline{m}$, since $\overline{m} \equiv (1/N)\int_\Omega m(i)\mathrm{d}i = [k/(k+1)]m^d$ under a Pareto distribution. Observe that for now in our model, larger cities are more productive because of tougher selection, but not because of technological externalities associated with agglomeration. In line with an abundant empirical literature (e.g., Rosenthal and Strange, 2004), we extend our framework to allow for such agglomeration economies in Section 6. All of our theoretical and quantitative key insights are robust to that extension.

We can also study the mass of surviving firms and the average markup faced by the consumers in the city. Using $N = N^E G(m^d)$, the mass of surviving firms is equal to

$$N = \frac{\alpha}{\kappa_1+\kappa_2}\frac{h}{m^d} = \frac{\alpha h}{\kappa_1+\kappa_2}\left(\frac{L}{\mu^{\max}}\right)^{\frac{1}{k+1}}. \tag{18}$$

Since those firms are heterogeneous and have different markups and market shares, the simple (unweighted) average of markups is not an adequate measure of consumers' exposure to market power. Using (11) and (12), we hence define the (expenditure share) weighted average of firm-level markups as follows:

$$\overline{\Lambda} \equiv \frac{1}{G(m^d)}\int_0^{m^d} \frac{p(m)q(m)}{E}\Lambda(m)\mathrm{d}G(m) = \frac{\kappa_3}{\alpha}\frac{m^d}{h}, \tag{19}$$

where κ_3 is a positive constant that solely depends on k (see Appendix B.3).[9] Note that the weighted average of markups is proportional to the cutoff. It thus follows from (18) and (19) that our model displays pro-competitive effects since $\overline{\Lambda} = [\kappa_3/(\kappa_1 + \kappa_2)](1/N)$ decreases with the mass of firms competing in the city.

Furthermore, we show in Appendix A.2 that the indirect utility can be expressed as

$$U = \alpha \left[\frac{1}{(k+1)(\kappa_1 + \kappa_2)} - 1 \right] \frac{h}{m^d} = \left[\frac{1}{(k+1)(\kappa_1 + \kappa_2)} - 1 \right] \frac{\kappa_3}{\overline{\Lambda}}, \tag{20}$$

where the term in square brackets is, by construction of the utility function, positive for all $k \geq 1$. Alternatively, the indirect utility can be rewritten as $U = [1/(k+1) - (\kappa_1 + \kappa_2)]N$. Hence, as can be seen from expressions (17)–(20), a city with better technological possibilities allows for higher utility because of tougher selection, tougher competition, and greater consumption diversity.

Finally, the impact of city size on consumption diversity, on the weighted average of markups, and on welfare can be established as follows. Using (2) and (17), we can rewrite indirect utility as

$$U = \alpha \left[\frac{1}{(k+1)(\kappa_1 + \kappa_2)} - 1 \right] \left\{ \frac{2\pi \overline{h}}{\theta^2 L} \left[1 - \left(1 + \theta \sqrt{L/\pi} \right) \mathrm{e}^{-\theta \sqrt{L/\pi}} \right] \right\} \left(\frac{L}{\mu^{\max}} \right)^{\frac{1}{k+1}}. \tag{21}$$

The term in braces in (21) equals the effective labor supply per capita, $h = S/L$, which decreases with L. The last term in (21) captures the cutoff productivity level, $1/m^d$, which increases with L. The net effect of an increase in L on the indirect utility U is thus ambiguous, highlighting the trade-off between a dispersion force (urban frictions) and an agglomeration force (tougher firm selection) inherent in our model. Yet, it can be shown that U is single-peaked with respect to L (see Appendix A.2). Since the indirect utility is proportional to N, it immediately follows that consumption diversity also exhibits a $\cap$-shaped pattern, while the weighted average of markups $\overline{\Lambda}$ is $\cup$-shaped with respect to L.

3 Urban system: Multiple cities

We now turn to the case with multiple cities and endogenous location decisions. The timing of events is as follows. First, workers/consumers choose their locations. Second, firm entry, selection and production take place.[10] We start the analysis by describing preferences and technology, as well as trade frictions, for this urban system with K asymmetric cities. We then derive the market equilibrium conditions, given city sizes, and define the spatial equilibrium where individuals endogenously choose their locations. Finally, we analyze the two-city case to build intuition for our counterfactual experiments. The internal structure of cities is analogous to that in the previous section, but cities may differ in their commuting technologies θ_r and gross labor supplies $\overline{h}_r$. Preferences and technology are also analogous, and we indicate changes wherever needed.

[9] Recent work by Feenstra and Weinstein (2010) uses a similar (expenditure share) weighted average of markups in a translog framework to quantify the impacts of international trade on the US price level.

[10] This timing simplifies our model because we need not specify which types of firms relocate as workers move across cities. The spatial sorting of firms or workers is not the topic of the present paper.

3.1 Preferences and demands

Let $p_{sr}(i)$ and $q_{sr}(i)$ denote the price and the per capita consumption of variety i produced in city s and consumed in city r. The utility maximization problem of a consumer in city r is then given by:

$$\max_{q_{sr}(j),\, j\in\Omega_{sr}} \quad U_r \equiv \sum_s \int_{\Omega_{sr}} \left[1 - \mathrm{e}^{-\alpha q_{sr}(j)}\right]\mathrm{d}j \qquad \text{s.t.} \quad \sum_s \int_{\Omega_{sr}} p_{sr}(j) q_{sr}(j)\mathrm{d}j = E_r, \tag{22}$$

where Ω_{sr} denotes the set of varieties produced in city s and consumed in city r.[11] It is readily verified that the demand functions are given as follows:

$$q_{sr}(i) = \frac{E_r}{N_r^c \overline{p}_r} - \frac{1}{\alpha}\left\{\ln\left[\frac{p_{sr}(i)}{N_r^c\overline{p}_r}\right] + \eta_r\right\}, \quad \forall i \in \Omega_{sr},$$

where N_r^c is the mass of varieties consumed in city r, and

$$\overline{p}_r \equiv \frac{1}{N_r^c}\sum_s \int_{\Omega_{sr}} p_{sr}(j)\mathrm{d}j \quad \text{and} \quad \eta_r \equiv -\sum_s \int_{\Omega_{sr}} \ln\left[\frac{p_{sr}(j)}{N_r^c\overline{p}_r}\right]\frac{p_{sr}(j)}{N_r^c\overline{p}_r}\mathrm{d}j$$

denote the (unweighted) average price and the differential entropy of the price distribution of all varieties consumed in city r, respectively. As in the case of a single city, the demand for a locally produced variety i (resp., a non-locally produced variety j) is positive if and only if the price of variety i (resp., of variety j) is lower than the reservation price p_r^d. Formally,

$$q_{rr}(i) > 0 \iff p_{rr}(i) < p_r^d \quad \text{and} \quad q_{sr}(j) > 0 \iff p_{sr}(j) < p_r^d,$$

where $p_r^d \equiv N_r^c\overline{p}_r \mathrm{e}^{\alpha E_r/(N_r^c\overline{p}_r)-\eta_r}$ is a function of the price aggregates $\overline{p}_r$ and η_r. The demands for local and non-local varieties can then be concisely expressed as follows:

$$q_{rr}(i) = \frac{1}{\alpha}\ln\left[\frac{p_r^d}{p_{rr}(i)}\right] \quad \text{and} \quad q_{sr}(j) = \frac{1}{\alpha}\ln\left[\frac{p_r^d}{p_{sr}(j)}\right]. \tag{23}$$

Since $\mathrm{e}^{-\alpha q_{sr}(j)} = p_{sr}(j)/p_r^d$, the indirect utility is given by

$$U_r = N_r^c - \sum_s \int_{\Omega_{sr}} \frac{p_{sr}(j)}{p_r^d}\mathrm{d}j = N_r^c\left(1 - \frac{\overline{p}_r}{p_r^d}\right). \tag{24}$$

3.2 Technology and market structure

We consider segmented markets, where resale or third-party arbitrage are sufficiently costly, and assume that firms are free to price discriminate between markets. Firms in city r independently draw their value of m from a city-specific distribution G_r. We introduce *trade frictions* into our

[11] We assume that land is collectively owned in each city, and that every resident has an equal claim to aggregate land rents in that city. As firms make zero aggregate profits, this implies that $E_r = (S_r/L_r)w_r = h_r w_r$.

model by assuming that shipments from r to s are subject to costs $\tau_{rs} > 1$ for all r and s, which firms incur in terms of labor.[12] The operating profit of firm i in r is then given by:

$$\pi_r(i) = \sum_s \pi_{rs}(i) = \sum_s L_s q_{rs}(i) \left[p_{rs}(i) - \tau_{rs} m_r(i) w_r\right]. \tag{25}$$

Each firm maximizes (25) with respect to its prices $p_{rs}(i)$ separately. Since it has no impact on the price aggregates and on the wages, the first-order conditions are given by:

$$\ln\left[\frac{p_s^d}{p_{rs}(i)}\right] = \frac{p_{rs}(i) - \tau_{rs} m_r(i) w_r}{p_{rs}(i)}, \quad \forall i \in \Omega_{rs}. \tag{26}$$

Equations (23) and (26) imply that $q_{rs}(i) = (1/\alpha)[1 - \tau_{rs} m_r(i) w_r / p_{rs}(i)]$, which shows that $q_{rs}(i) = 0$ at $p_{rs}(i) = \tau_{rs} m_r(i) w_r$. It then follows from (26) that $p_{rs}(i) = p_s^d$. Hence, a firm located in r with draw $m_{rs}^x \equiv p_s^d / (\tau_{rs} w_r)$ is just indifferent between selling and not selling in city s. All firms in r with draws below m_{rs}^x are productive enough to sell to city s.[13] In what follows, we refer to $m_{ss}^x \equiv m_s^d$ as the *internal cutoff* in city s, whereas m_{rs}^x with $r \neq s$ is the *external cutoff* for selling to city s when located in city r. External and internal cutoffs are linked as follows:

$$m_{rs}^x = \frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d. \tag{27}$$

Expression (27) reveals the key relationship between trade costs, wages, and productivity. In particular, it shows how trade costs and wage differences affect firms' ability to break into market s. When wages are equalized ($w_r = w_s$) and local trade is less costly than external trade ($\tau_{ss} < \tau_{rs}$), all external cutoffs must fall short of the internal cutoffs. Breaking into market s is then always harder for firms in $r \neq s$ than for firms in s, which is the standard case considered in the literature (e.g., Melitz, 2003; Melitz and Ottaviano, 2008). However, consider a case where $w_s > w_r$. In that case, firms from the low wage city r may have a cost advantage in selling to the high wage market s compared to the local competitors there. Surviving in market s is then easier for firms selling from r than for local firms in s, i.e., $m_{rs}^x > m_s^d$. More generally, in the presence of wage differences and trade costs, the usual ranking $m_{rs}^x < m_s^d$ prevails only when $\tau_{ss} w_s < \tau_{rs} w_r$.

Given a mass of entrants N_r^E and external cutoffs m_{rs}^x, only $N_r^p = N_r^E G_r(\max_s \{m_{rs}^x\})$ firms survive in city r, namely those which are productive enough to sell at least in one market (which need not be the local market in our model because of wage differences across cities). The mass of varieties consumed in city r is given by

$$N_r^c = \sum_s N_s^E G_s(m_{sr}^x), \tag{28}$$

[12] We allow for internal trade costs $\tau_{rr} > 1$ in order to capture the empirical fact that firms also incur shipping and distribution costs in their local markets.

[13] Unlike in the CES model by Melitz (2003), we need not assume fixed costs for 'exporting' to explain why some firms do not ship to some cities. The reason is that, for each variety, marginal utility at zero consumption is finite in our model. While fixed costs of exporting are certainly pervasive in an international context, they appear much less plausible at the city or the zip code level within a country (also see Hillberry and Hummels, 2008).

which is the sum of all firms that are productive enough to serve market r.

As in the case of a single city, the first-order conditions (26) can be solved by using the Lambert W function. The profit-maximizing prices and quantities, as well as operating profits, are given by:

$$p_{rs}(m) = \frac{\tau_{rs} m w_r}{W}, \quad q_{rs}(m) = \frac{1}{\alpha}\left(1 - W\right), \quad \pi_{rs} = \frac{L_s \tau_{rs} m w_r}{\alpha}(W^{-1} + W - 2), \tag{29}$$

where we suppress the argument $\mathrm{e}\tau_{rs} m w_r / p_s^d$ of W. It can be readily verified that more productive firms again charge lower prices, sell larger quantities, and earn higher operating profits than less productive firms. Markups, defined as $\Lambda_{rs}(m) \equiv p_{rs}(m)/(\tau_{rs} m w_r) = 1/W$, are also higher the smaller m is.

3.3 Equilibrium

3.3.1 Market equilibrium

We first examine the market equilibrium for given population sizes in the general case with K asymmetric cities. We assume that productivity draws $1/m$ follow Pareto distributions with identical shape parameters $k \geq 1$, but the upper bounds are allowed to vary across cities, i.e., $G_r(m) = (m/m_r^{\max})^k$. Given that assumption, the equilibrium conditions – zero expected profits, labor market clearing, and the trade balance – are as follows (see Appendix C for details). First, labor market clearing requires that

$$N_r^E \left[\frac{\kappa_1}{\alpha \left(m_r^{\max}\right)^k} \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d \right)^{k+1} + F \right] = S_r. \tag{30}$$

Second, zero expected profits imply that

$$\mu_r^{\max} = \sum_s L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d \right)^{k+1}, \tag{31}$$

where $\mu_r^{\max} \equiv [\alpha F \left(m_r^{\max}\right)^k]/\kappa_2$ denotes city r's technological possibilities. Last, balanced trade requires that

$$\frac{N_r^E w_r}{\left(m_r^{\max}\right)^k} \sum_{s \neq r} L_s \tau_{rs} \left(\frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} m_s^d \right)^{k+1} = L_r \sum_{s \neq r} \tau_{sr} \frac{N_s^E w_s}{\left(m_s^{\max}\right)^k} \left(\frac{\tau_{rr}}{\tau_{sr}} \frac{w_r}{w_s} m_r^d \right)^{k+1}. \tag{32}$$

The $3 \times K$ conditions (30)–(32) depend on $3 \times K$ unknowns: the wages, w_r, the masses of entrants, N_r^E, and the internal cutoffs, m_r^d. The external cutoffs, m_{rs}^x, can then be recovered from (27). Combining (30) and (31) immediately shows that

$$N_r^E = \frac{\kappa_2}{\kappa_1 + \kappa_2} \frac{S_r}{F}. \tag{33}$$

Thus, the mass of entrants in city r still depends positively on that city's effective labor supply S_r and negatively on the fixed labor requirement F. Adding the term in r that is missing on both sides of (32), and using (31) and (33), we obtain the following equilibrium relationship:

$$\frac{h_r}{(m_r^d)^{k+1}} = \sum_s S_s \tau_{rr} \left(\frac{\tau_{rr}}{\tau_{sr}} \frac{w_r}{w_s} \right)^k \frac{1}{\mu_s^{\max}}. \tag{34}$$

Expressions (31) and (34) jointly summarize how wages, technological possibilites, cutoffs, trade costs, population sizes, and effective labor supplies are related in the market equilibrium.

Using the foregoing expressions, we can show that the mass of varieties consumed in a city is inversely proportional to the internal cutoff in that city (see Appendix A.3 for the derivation):

$$N_r^c = \frac{\alpha}{(\kappa_1 + \kappa_2)\tau_{rr}} \frac{h_r}{m_r^d}. \tag{35}$$

Furthermore, the (expenditure share) weighted average of markups that consumers face in city r can be expressed as follows (see Appendix A.4 for the derivation):

$$\overline{\Lambda}_r \equiv \frac{\sum_s N_s^E \int_0^{m_{sr}^x} \frac{p_{sr}(m) q_{sr}(m)}{E_r} \Lambda_{sr}(m) \mathrm{d}G_s(m)}{\sum_s N_s^E G_s(m_{sr}^x)} = \frac{\kappa_3 \tau_{rr}}{\alpha} \frac{m_r^d}{h_r}. \tag{36}$$

Hence, it immediately follows from (35) and (36) that there are pro-competitive effects, since $\overline{\Lambda}_r$ decreases with the mass N_r^c of firms competing in city r as $\overline{\Lambda}_r = [\kappa_3/(\kappa_1 + \kappa_2)](1/N_r^c)$. Last, we show in Appendix A.5 that the indirect utility is given by

$$U_r = \frac{\alpha}{\tau_{rr}} \left[\frac{1}{(k+1)(\kappa_1 + \kappa_2)} - 1 \right] \frac{h_r}{m_r^d} = \left[\frac{1}{(k+1)(\kappa_1 + \kappa_2)} - 1 \right] \frac{\kappa_3}{\overline{\Lambda}_r}, \tag{37}$$

which implies that greater effective labor supply per capita, tougher selection, and a lower weighted average of markups in city r translate into higher welfare. Alternatively, the indirect utility can be rewritten as $U_r = [1/(k+1) - (\kappa_1 + \kappa_2)] N_r^c$, i.e., it is proportional to the mass of varieties consumed in a city.

3.3.2 Spatial equilibrium

We now introduce city-specific amenities and taste heterogeneity in residential location into our model. This is done for two reasons. First, individuals in reality choose their location not only based on wages, prices and productivities that result from market interactions, but also based on non-market features such as amenities (e.g., climate or landscape). Second, individuals do not necessarily react in the same way to regional gaps in wages and cost-of-living (Tabuchi and Thisse, 2002; Murata, 2003). Such taste heterogeneity tends to offset the extreme outcome that often arises in typical NEG models, namely that *all* mobile economic activity concentrates in a single city. When we take our model to data, taste heterogeneity is thus useful for capturing an observed non-degenerate equilibrium distribution of city sizes.

We assume that the location choice of an individual ℓ is described by linear random utility as $V_r^\ell = U_r + A_r + \xi_r^\ell$, where U_r is given by (37) and A_r subsumes city-specific amenities that are equally valued by all individuals. For the empirical implementation, we assume that $A_r \equiv A(A_r^o, A_r^u)$, where A_r^o refers to observed amenities such as costal location and A_r^u to the unobserved part. Finally, the random variable ξ_r^ℓ captures idiosyncratic taste differences in residential location. Following McFadden (1974), we assume that the ξ_r^ℓ are i.i.d. across individuals and cities according to a double exponential distribution with zero mean and variance equal to $\pi^2\beta^2/6$, where β is a positive constant. Since β has a positive relationship with variance, the larger the value of β, the more heterogeneous are the workers' attachments to each city. Given the population distribution, an individual's probability of choosing city r can then be expressed as a logit form:

$$\mathbb{P}_r = \Pr\left(V_r^\ell > \max_{s \neq r} V_s^\ell\right) = \frac{\exp((U_r + A_r)/\beta)}{\sum_{s=1}^K \exp((U_s + A_s)/\beta)}. \tag{38}$$

If $\beta \to 0$, people choose their location based only on $U_r + A_r$. This corresponds to the case without taste heterogeneity, i.e., they choose a city with the highest $U_r + A_r$ with probability one. By contrast, if $\beta \to \infty$, individuals choose their location with equal probability $1/K$. In that case, taste for residential location is extremely heterogeneous, so that $U_r + A_r$ does not affect location decisions at all. We finally define a spatial equilibrium as a city-size distribution satisfying

$$\mathbb{P}_r = \frac{L_r}{L}, \quad \forall r. \tag{39}$$

In words, the choice probability of each city is equal to the population share of that city.

3.4 Some analytical results

To build intuition for our counterfactual experiments, we now illustrate how spatial frictions affect the fundamental trade-off between agglomeration and dispersion forces in the special case with two cities. We assume that trade costs are symmetric ($\tau_{12} = \tau_{21} = \tau$ and $\tau_{11} = \tau_{22} = t$), and that intra-city trade is less costly than inter-city trade ($t \leq \tau$). For given city sizes L_1 and L_2, the market equilibrium is given by a system of three equations (31)–(33) with three unknowns (the two internal cutoffs m_1^d and m_2^d, and the relative wage $\omega \equiv w_1/w_2$) as follows:

$$\mu_1^{\max} = L_1 t \left(m_1^d\right)^{k+1} + L_2 \tau \left(\frac{t}{\tau}\frac{1}{\omega} m_2^d\right)^{k+1} \tag{40}$$

$$\mu_2^{\max} = L_2 t \left(m_2^d\right)^{k+1} + L_1 \tau \left(\frac{t}{\tau}\omega m_1^d\right)^{k+1} \tag{41}$$

$$\omega^{2k+1} = \frac{\rho}{\sigma}\left(\frac{m_2^d}{m_1^d}\right)^{k+1}, \tag{42}$$

where $\rho \equiv \mu_2^{\max}/\mu_1^{\max}$ and $\sigma \equiv h_2/h_1 = (S_2/L_2)/(S_1/L_1)$. When $t < \tau$, equations (40) and (41) can be uniquely solved for the cutoffs as a function of ω:

$$(m_1^d)^{k+1} = \frac{\mu_1^{\max}}{L_1 t}\,\frac{1 - \rho(t/\tau)^k \omega^{-(k+1)}}{1 - (t/\tau)^{2k}} \quad \text{and} \quad (m_2^d)^{k+1} = \frac{\mu_2^{\max}}{L_2 t}\,\frac{1 - \rho^{-1}(t/\tau)^k \omega^{k+1}}{1 - (t/\tau)^{2k}}, \tag{43}$$

and substituting the cutoffs (43) into (42) yields, after some simplification, the following expression:

$$\text{LHS} \equiv \omega^k = \rho\,\frac{S_1}{S_2}\,\frac{\rho - (t/\tau)^k\,\omega^{k+1}}{\omega^{k+1} - \rho\,(t/\tau)^k} \equiv \text{RHS}. \tag{44}$$

When $t < \tau$, the RHS of (44) is non-negative if and only if $\underline{\omega} < \omega < \overline{\omega}$, where $\underline{\omega} \equiv \rho^{1/(k+1)}\,(t/\tau)^{k/(k+1)}$ and $\overline{\omega} \equiv \rho^{1/(k+1)}\,(\tau/t)^{k/(k+1)}$. Furthermore, the RHS is strictly decreasing in $\omega \in (\underline{\omega}, \overline{\omega})$ with $\lim_{\omega \to \underline{\omega}+} \text{RHS} = \infty$ and $\lim_{\omega \to \overline{\omega}-} \text{RHS} = 0$. Since the LHS of (44) is strictly increasing in $\omega \in (0, \infty)$, there exists a unique equilibrium relative wage $\omega^* \in (\underline{\omega}, \overline{\omega})$.

Consider two cities that differ in size but are identical with respect to their gross labor supplies, commuting technologies, and technological possibilities ($\overline{h}_1 = \overline{h}_2 = \overline{h}$, $\theta_1 = \theta_2 = \theta$, and $\rho = 1$). Then, *the larger city has the higher wage and the lower cutoff.* To see this, observe first that $L_1/L_2 = 1$ implies $S_1/S_2 = 1$, so that the unique equilibrium wage is $\omega^* = 1$ by (44). Now suppose that city 1 is larger than city 2, i.e., $L_1/L_2 > 1$, which implies $S_1/S_2 > 1$. Then, the equilibrium relative wage satisfies $\omega^* > 1$ because an increase in S_1/S_2 raises the RHS of (44) without affecting the LHS. Finally, expression (42), together with the foregoing assumption, yields $\omega^{2k+1} = (1/\sigma)\left(m_2^d/m_1^d\right)^{k+1}$. As $L_1 > L_2$ implies $\omega > 1$ and $\sigma > 1$ (recall that $h \equiv S/L$ is decreasing in L), it follows that $m_1^d < m_2^d$. Hence, the larger city has the lower cutoff. The intuition is that the larger city has an advantage in terms of the larger local market due to trade frictions, and that this advantage must be offset by the higher wage and tougher selection.

As can be seen from (37), the higher productivity (lower cutoff) constitutes an agglomeration force as it raises indirect utility in the larger city. Yet, the larger city also has lower effective labor supply per capita $h_r = S_r/L_r$ due to urban frictions, which negatively affects indirect utility, thus representing a dispersion force.[14] In the case of two cities, the choice probabilities (38) that constitute the spatial equilibrium can be rewritten as

$$\mathbb{P}_1 = \frac{\exp(\Upsilon/\beta)}{\exp(\Upsilon/\beta) + 1} \quad \text{and} \quad \mathbb{P}_2 = \frac{1}{\exp(\Upsilon/\beta) + 1},$$

where $\Upsilon \equiv (U_1 - U_2) + (A_1 - A_2)$. Hence, $\mathbb{P}_1$ is increasing and $\mathbb{P}_2$ is decreasing in Υ. Plugging (37) into the definition of Υ, we readily obtain

$$\Upsilon = \left(\frac{\alpha}{t}\right)\left[\frac{1}{(k+1)(\kappa_1 + \kappa_2)} - 1\right]\left(\frac{h_1}{m_1^d} - \frac{h_2}{m_2^d}\right) + (A_1 - A_2). \tag{45}$$

In what follows, we focus on the case where $L_1 > L_2$ and $A_1 = A_2$. Then, by (39) the spatial equilibrium is such that $\mathbb{P}_1 > \mathbb{P}_2$, which implies $\Upsilon > 0$ and $h_1/m_1^d > h_2/m_2^d$ by (45). The larger city then has greater consumption diversity according to (35) and a lower average markup according to (36) than the smaller city.

[14] Recall that the gross labor supply, $\overline{h}_r$, is exogenous in our model. When quantifying the model in Section 4, we use data on $\overline{h}_r$ across MSAs, which shows that $\overline{h}_r$ is higher in big cities like New York. In this subsection, we abstract from such an "urban rat race" that would work against the effect of urban frictions by raising the effective labor supply per capita, h_r, in the larger city. A better commuting technology (lower θ_r) in the larger city would also work in the same direction.

3.4.1 No urban frictions

Our first counterfactual experiment will be to eliminate urban frictions while leaving trade frictions unchanged. This is equivalent to setting $\theta = 0$, holding τ and t constant. In what follows, we consider – starting from the initial spatial equilibrium – how Υ is affected by such a change. This allows us to study if the elimination of urban frictions involves more agglomeration (larger $\mathbb{P}_1$) or more dispersion (smaller $\mathbb{P}_1$). Let $\widetilde{\Upsilon}$ be the value of Υ in the counterfactual scenario, keeping city sizes fixed at their initial levels. Other counterfactual values are also denoted with a tilde. Observing that $\widetilde{h}_1 = \widetilde{h}_2 = \overline{h}$ when $\theta = 0$, we have

$$\text{sign}\left\{\widetilde{\Upsilon} - \Upsilon\right\} = \text{sign}\left\{\frac{1}{\widetilde{m}_1^d}(\overline{h} - h_1) - \frac{1}{\widetilde{m}_2^d}(\overline{h} - h_2) + h_1\left(\frac{1}{\widetilde{m}_1^d} - \frac{1}{m_1^d}\right) - h_2\left(\frac{1}{\widetilde{m}_2^d} - \frac{1}{m_2^d}\right)\right\}, \quad (46)$$

where the first two terms stand for the direct effects of eliminating urban frictions and the second two terms capture the indirect effects through the cutoffs. In the initial situation where $\theta > 0$, we know that $h_1 < h_2 < \overline{h}$ as urban frictions are greater in the larger city. We also know that $m_1^d < m_2^d$ holds even without urban frictions as $L_1 > L_2$, so that $\widetilde{m}_1^d < \widetilde{m}_2^d$. Hence, the first positive term always dominates the second negative term, thus showing that the direct effects favor the large city by increasing the probability $\mathbb{P}_1$ of choosing city 1. However, the indirect effects through the cutoffs work against this. To see this, notice that the reduction in θ from any given positive value to zero raises S_1/S_2 (see Appendix A.6) and thus the relative wage ω by shifting up the RHS of (44). We thus observe *wage divergence*. The expressions in (43) then show that the cutoff *increases* in the large city to offset the cost disadvantage, whereas it *decreases* in the small city. In other words, we have $m_1^d < \widetilde{m}_1^d < \widetilde{m}_2^d < m_2^d$. Hence, both the third and fourth terms are negative, so that the indirect effects through the cutoffs work against agglomeration.

If the direct effects dominate the indirect effects, we have $\widetilde{\Upsilon} > \Upsilon$ so that $\mathbb{P}_1$ increases and the large city becomes even larger as urban frictions are eliminated. The increase in population then leads to a productivity gain, which may offset the productivity drop at a given population size ($m_1^d < \widetilde{m}_1^d$). As we show below, such a pattern indeed emerges in the quantified multi-city model (see Figures 5, 7 and 12): Big cities like New York become even larger. Given the initial population, productivity goes down in New York while it goes up once we take population changes into account (see also Section 6.1 below). By the same argument, small cities may end up with a lower productivity due to the loss in population. Hence, the elimination of urban frictions makes the overall productivity change ambiguous.

3.4.2 No trade frictions

Our second counterfactual experiment will be to eliminate trade frictions while leaving urban frictions unchanged. More specifically, we consider a scenario where consumers face the same trade costs for local and non-local varieties. This is equivalent to setting $\tau = t$, holding θ constant. As before, let $\widetilde{\Upsilon}$ be the value of Υ in the counterfactual scenario, while keeping city sizes fixed at the initial level. Other counterfactual values, given the initial population distribution, are also denoted with a tilde. Noting that h_1 and h_2 remain constant, the change in Υ can now be written as

$$\operatorname{sign}\left\{\widetilde{\Upsilon} - \Upsilon\right\} = \operatorname{sign}\left\{h_1\left(\frac{1}{\widetilde{m}_1^d} - \frac{1}{m_1^d}\right) - h_2\left(\frac{1}{\widetilde{m}_2^d} - \frac{1}{m_2^d}\right)\right\}. \tag{47}$$

It can be shown that now *both* cutoffs decrease for given population sizes, i.e., $\widetilde{m}_1^d < m_1^d$ and $\widetilde{m}_2^d < m_2^d$ (see Appendix A.6). Both cities therefore experience a productivity gain. The first term in the square brackets in (47) is thus positive, while the second term is negative. We can show that, when switching to a world without trade frictions, $\widetilde{\Upsilon} < \Upsilon$ holds if $\rho^{1/(k+1)} \leq \sigma$ (see Appendix A.6). In other words, the large city becomes smaller if the two cities are initially not too different in terms of their technological possibilities. In contrast, the small city becomes larger and, consequently, experiences a further productivity gain. We show below that these analytical results are consistent with the pattern that emerges in our quantified multi-city model (see Figures 9 and 11): small cities tend to gain population, and they experience stronger gains in productivity than large cities.[15]

4 Quantification

We now take our multi-city model to the data by estimating or calibrating its parameters (see Appendix D for the data description). Our procedure can be summarized in the following 6 steps:

1. Using the definition of total effective labor supply and data on commuting time, hours worked, and city size at the MSA level, we obtain the city-specific commuting technology parameters $\widehat{\theta}_r$ that constitute *urban frictions*.

2. Using the specification $\tau_{rs} \equiv d_{rs}^{\gamma}$, where d_{rs} is the distance from r to s, we estimate a gravity equation that relates the value of bilateral trade flows to distance. For a given value of the Pareto shape parameter k, we obtain the distance elasticity $\widehat{\gamma}$ that constitutes *trade frictions*.

3. The estimated distance elasticity, together with data on labor supply, value added per worker, and city size, allows us to back out the set of city-specific technological possibilities $\widehat{\mu}_r^{\max}$ and wages $\widehat{w}_r$ that are consistent with the *market equilibrium conditions*.

4. Using the set of city-specific technological possibilities thus obtained, we draw a large sample of firms from within the model to compute the difference between the simulated and observed establishment size distributions.

[15] Other two-region NEG models with commuting costs (Tabuchi, 1998; Murata and Thisse, 2005) would come to qualitatively similar conclusions about how falling transport or commuting costs affect the spatial equilibrium. Helpman (1998) considers a fixed supply of land instead of commuting, but his model would also display a similar pattern as falling transport costs are dispersive, while greater abundance of land is agglomerative. Though useful for illustrative purposes, such two-region examples do not deliver a sense of magnitude about the quantitative importance of spatial frictions in practice, however. They are also silent on productivity.

5. Iterating through steps 2 to 4, we search over the parameter space to find the value of the Pareto shape parameter k that minimizes the sum of squared differences between the simulated and observed establishment size distributions.

6. Using the *spatial equilibrium conditions*, the expression of indirect utility, and data on natural amenities, we obtain a measure of unobservable consumption amenities and the relative weight of economic factors and amenities that are consistent with the observed city-size distribution.

In what follows, we explain each step in more detail.

4.1 Urban frictions θ_r

To obtain the city-specific commuting technology parameters $\widehat{\theta}_r$ that constitute urban frictions, we rewrite equation (2) as

$$L_r \frac{h_r}{\overline{h}_r} = \frac{2\pi}{\theta_r^2} \left[1 - \left(1 + \theta_r \sqrt{L_r/\pi} \right) \mathrm{e}^{-\theta_r \sqrt{L_r/\pi}} \right], \tag{48}$$

where we use $S_r = L_r h_r$. We compute h_r as the average number of hours worked per week in MSA r. The gross labor supply per capita, $\overline{h}_r$, which is the endowment of hours available for work and commuting, is constructed as the sum of h_r and hours per week spent by workers in each MSA for travel-to-work commuting in 2007. Given h_r, $\overline{h}_r$, as well as city size L_r, the above equation can be uniquely solved for the city-specific commuting parameter $\widehat{\theta}_r$. Table 1 provides their values for the 356 MSAs, which are further discussed in Section 4.6.

4.2 Trade frictions τ_{rs}

To estimate the distance elasticity $\widehat{\gamma}$ that constitutes trade frictions, we consider the value of sales from r to s:

$$X_{rs} = N_r^E L_s \int_0^{m_{rs}^x} p_{rs}(m) q_{rs}(m) \mathrm{d}G_r(m). \tag{49}$$

Using (27), (29), (33), and the result from Appendix B.4, we then obtain the gravity equation

$$X_{rs} = S_r L_s \tau_{rs}^{-k} \tau_{ss}^{k+1} (w_s/w_r)^{k+1} w_r \left(m_s^d \right)^{k+1} (\mu_r^{\max})^{-1}.$$

Turning to the specification of trade costs τ_{rs}, we stick to standard practice and assume that $\tau_{rs} \equiv d_{rs}^{\gamma}$, where d_{rs} stands for the distance from r to s (measured in kilometers and computed using the great circle formula). The gravity equation can then be rewritten in log-linear stochastic form as

$$\ln X_{rs} = \text{const.} - k\gamma \ln d_{rs} + I_{rs}^0 + \zeta_r^1 + \zeta_s^2 + \varepsilon_{rs}, \tag{50}$$

where all terms specific to the origin and the destination are collapsed into fixed effects ζ_r^1 and ζ_s^2, where I_{rs}^0 is a zero-flow dummy, and ε_{rs} is an error term with the usual properties for OLS consistency.[16] Using aggregate bilateral trade flows X_{rs} in 2007 for the 48 contiguous US states that cover all MSAs used in the subsequent analysis, we estimate the gravity equation on state-to-state trade flows.[17] Given a value of k, we thus obtain an estimate of the distance elasticity $\widehat{\gamma}$.

4.3 Market equilibrium conditions $(w_r, \mu_r^{\max})$

We next turn to the market equilibrium conditions. Observe that expressions (31) and (34) can be rewritten as:

$$\mu_r^{\max} = \sum_s L_s \tau_{rs} \left(m_s^d \frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r} \right)^{k+1} \tag{51}$$

$$\frac{S_r}{L_r} \frac{1}{(m_r^d)^{k+1}} = \sum_s S_s \tau_{rr} \left(\frac{\tau_{sr}}{\tau_{rr}} \frac{w_s}{w_r} \right)^{-k} \frac{1}{\mu_s^{\max}}. \tag{52}$$

Ideally, we would like to use data on technological possibilities $\mu_r^{\max}$ to solve for the wages and cutoffs. Yet, $\mu_r^{\max}$ is unobservable. We thus solve for wages and technological possibilities $(\widehat{w}_r, \widehat{\mu}_r^{\max})$ by using the values of m_r^d that are obtained as follows. Under the Pareto distribution, we have $(1/\overline{m}_r) = [k/(k+1)](1/m_r^d)$, where $1/\overline{m}_r$ is the average productivity in MSA r. The latter can be computed as GDP per employee, using data on GDP of MSA r and the total number of hours worked in that MSA (hours worked per week times total employment). Given an estimate of $1/\overline{m}_r$ and the value of k, we can compute the cutoffs m_r^d. Using the value of k, the cutoffs m_r^d, the city-specific commuting technologies $\widehat{\theta}_r$, the observed MSA populations L_r, as well as trade frictions $\widehat{\tau}_{rs} = d_{rs}^{\widehat{\gamma}}$, we can solve (51) and (52) for the wages and unobserved technological possibilities $(\widehat{w}_r, \widehat{\mu}_r^{\max})$ that are consistent with the market equilibrium. We compare in Section 4.7 the predicted wages $\widehat{w}_r$ with observed wages at the MSA level to assess how well our model fits the data.

4.4 Firm size distribution and Pareto shape parameter k

The quantification procedure described thus far has assumed a given value of the shape parameter k. In order to estimate k structurally, we proceed as follows. First, given a value of k, we can compute trade frictions $\widehat{\tau}_{rs}$ and the wages and cutoffs $(\widehat{w}_r, \widehat{\mu}_r^{\max})$ as described in Sections 4.2 and 4.3. This, together with the internal cutoff m_r^d computed from data as described in Section 4.3, yields the

[16] There are 179 'zero flows' out of 2,304 in the data, i.e., 7.7% of the sample. We control for them by using a standard dummy-variable approach, where I_{rs}^0 takes value 1 if $X_{rs} = 0$ and 0 otherwise. Note that these flows are not true zeros as we exclude Alaska, Hawaii, and Washington DC (see the 2007 Commodity Flow Survey (CFS) data). Rather, they are unreported observations because of lack of statistical precision, so that a Heckman-type correction procedure is not warranted.

[17] We work at the state level since MSA trade flows from the CFS public files can only be meaningfully exploited for a relatively small sample of large 'CFS regions'. Duranton *et al.* (2011, p.10) work with aggregate trade flows for "65 CFS regions organized around the core county of a US metropolitan area" to estimate the distance elasticity. We used their estimate as a robustness check, and our results are little sensitive.

external cutoffs $\widehat{m}_{rs}^x$ by (27). With that information in hand, we can compute the share $\widehat{\nu}_r$ of surviving firms in each MSA as follows:

$$\widehat{\nu}_r \equiv \frac{\widehat{N}_r^p}{\sum_s \widehat{N}_s^p}, \quad \text{where} \quad \widehat{N}_r^p = \widehat{N}_r^E G_r\left(\max_s \widehat{m}_{rs}^x\right) = \frac{\alpha}{\kappa_1+\kappa_2} S_r\left(\widehat{\mu}_r^{\max}\right)^{-1}\left(\max_s \widehat{m}_{rs}^x\right)^k$$

denotes the number of firms operating in MSA r. The total effective labor supply S_r is computed as in Section 4.1. Note that $\widehat{\nu}_r$ is independent of the unobservable constant scaling $\alpha/(\kappa_1+\kappa_2)$ that multiplies the number of firms. The distribution of marginal labor requirements of surviving firms in city r is $\widehat{G}_r(m) = \left(m/m_r^d\right)^k$.

Second, we draw a large sample of firms from our calibrated MSA-level productivity distributions $\widehat{G}_r$. For that sample to be representative, we draw firms in MSA r in proportion to its share $\widehat{\nu}_r$. For each sampled firm with marginal labor requirement m in MSA r, we can compute its employment as follows:[18]

$$\text{employment}_r(m) = m\sum_s \widehat{\chi}_{rs} L_s q_{rs}(m) = \frac{m}{\alpha}\sum_s \widehat{\chi}_{rs} L_s\left[1 - W\left(\mathrm{e}\frac{m}{\widehat{m}_{rs}^x}\right)\right],$$

where $\widehat{\chi}_{rs} = 1$ if $m < \widehat{m}_{rs}^x$ (the establishment can sell to MSA s) and zero otherwise (the establishment cannot sell to MSA s). Since we can identify employment only up to some positive scaling coefficient (which depends on the unobservable α) we choose, without loss of generality, that coefficient such that the average employment per firm in our sample of establishments matches the observed average employment in the 2007 CBP. Doing so allows us to readily compare the generated and observed data as we can sort the sampled firms into the same size bins as those used for the observed firms. We use four standard employment size bins from the CBP: $\iota = \{1\text{–}19,\ 20\text{–}99,\ 100\text{–}499,\ 500+\}$ employees. Let $N_{(\iota)}^{\text{SIM}}$ and $N_{(\iota)}^{\text{CBP}}$ denote the number of firms in each size bin ι in our sample and in the CBP, respectively. Let also N^{SIM} and N^{CBP} denote our sample size and the observed number of establishments in the CBP. Given a value of k, the following statistic is a natural measure of the goodness-of-fit of the simulated establishment-size distribution:

$$\text{SS}(k) = \sum_{\iota=1}^{4}\left[\frac{N_{(\iota)}^{\text{SIM}}}{N^{SIM}} - \frac{N_{(\iota)}^{\text{CBP}}}{N^{\text{CBP}}}\right]^2, \tag{53}$$

the value of which depends on the chosen k. It is clear from (53) that we can choose any large sample size N^{SIM} since it would not affect the ratio $N_{(\iota)}^{\text{SIM}}/N^{\text{SIM}}$. Without loss of generality, we choose the sample size such that the total number of simulated firms operating matches the observed total number of establishments ($N^{\text{SIM}} = N^{\text{CBP}}$). There are 6,431,884 establishments across our 356 MSAs in the 2007 CBP, and we sample the same number of firms from our quantified model.[19] We finally choose k by minimizing $\text{SS}(k)$.

[18] We exclude the labor used for shipping goods and the sunk initial labor requirement.

[19] Doing so allows for a direct comparison of $N_{(\iota)}^{\text{SIM}}$ and $N_{(\iota)}^{\text{CBP}}$ for each ι. The very small differences in the aggregate numbers reported in Tables 2 and 3 are due to rounding as the number of firms in each MSA has to be an integer.

4.5 Spatial equilibrium conditions A_r

Our final step is to take into account the spatial equilibrium conditions. Recall that the choice probabilities are given by (38). Setting $U_1 + A_1 \equiv 0$ as a normalization, and using the observed values of L_r for the 356 MSAs, the spatial equilibrium conditions $\mathbb{P}_r = L_r/L$ for $r = 2, 3, \ldots, K$ can be uniquely solved for $(U_r + A_r)/\beta$.[20] We thus obtain the values of $(U_r + A_r)/\beta$ that replicate the observed city-size distribution as a spatial equilibrium. Let $\widehat{D}_r$ denote the solution satisfying

$$\mathbb{P}_r = \frac{\exp(\widehat{D}_r)}{\sum_{s=1}^{K} \exp(\widehat{D}_s)} = \frac{L_r}{L}, \quad \widehat{D}_1 = 0. \tag{54}$$

We then regress $\widehat{D}_r$ on our measure of indirect utility $\widehat{U}_r$ and data on natural amenities A_r^o to gauge their relative importance:

$$\widehat{D}_r = \alpha_0 + \alpha_1 \widehat{U}_r + \alpha_2 A_r^o + \varepsilon_r, \tag{55}$$

where $\widehat{U}_r$ is obtained from (37) using our measures of L_r, S_r, and m_r^d, as well as the estimate of $\widehat{\tau}_{rr}$.[21] Estimating the coefficients on indirect utility $\widehat{U}_r$ and natural amenities A_r^o allows us to solve the issue of how to weight these two terms appropriately in consumers' location choices. The fitted residuals $\widehat{\varepsilon}_r$ can be interpreted as the implicit measure of the unobserved part of the MSA amenities. We hence let $\widehat{A}_r^u \equiv \widehat{\varepsilon}_r$. By construction, that measure is uncorrelated with A_r^o and does not capture natural amenities such as climate or access to the sea that are subsumed by A_r^o.

4.6 Quantification results

Concerning the Pareto shape parameter, our iterative procedure yields $\widehat{k} = 6.4$ that minimizes the sum of squared differences between the observed and computed firm size distributions by size bins. Columns 2 and 3 of Table 2 show that, despite having only a single degree of freedom, the fit to the observed distribution is rather good, with the model somewhat under-predicting the number of small establishments and over-predicting the number of large establishments.

Insert Table 1 about here.

Turning to spatial frictions, we first use (48) to obtain an estimate for the commuting technology parameters that constitute urban frictions for each MSA. As can be seen from Table 1, the values of $\widehat{\theta}_r$ range from 0.0708 (Los Angeles-Long Beach-Santa Ana and New York-Northern New Jersey-Long Island) and 0.0867 (Chicago-Naperville-Joliet) to 0.9995 (Yuba City, CA) and 1.4824 (Hinesville-Fort Stewart, GA). Thus, big cities tend to have better commuting technologies.[22]

[20] Since $\sum_{r=1}^{K} \mathbb{P}_r = 1$, the above equations are not independent. We drop the first one without loss of generality.

[21] Due to the specification in (55), neglecting multiplicative constants in (37) does not affect our results.

[22] For any given distance x from the CBD, a smaller θ implies that people spend less time to commute to the CBD. However, this does not necessarily mean that average commuting time is smaller in larger cities because of longer commuting distances.

We then use (50) to obtain an estimate for the distance elasticity that constitutes trade frictions. Our fixed effects estimation of the gravity equation yields $\widehat{\gamma k} = 1.2918$ (with standard error 0.0271) which, given $\widehat{k} = 6.4$, implies $\widehat{\gamma} = 0.2018$. Our estimate $\widehat{\gamma k}$ for the year 2007 closely matches the value of 1.348 reported by Hillberry and Hummels (2008) from estimation of a gravity equation at the 3-digit zip code level using the 1997 confidential CFS microdata. It is larger than the value of $\gamma k = 0.82$ reported by Duranton *et al.* (2011) which is obtained from a small sample of large CFS regions. Our subsequent results do not change qualitatively and change little quantitatively when using their estimate of the distance elasticity as a robustness check.

Having solved equations (54) for $\widehat{D}_r$, we run a simple OLS estimation of (55), which yields:

$$\widehat{D}_r = \underset{(0.2644)}{-0.2194} + \underset{(0.5289)}{1.7481}^{***}\widehat{U}_r + \underset{(0.0199)}{0.0652}^{***}A^o_r + \widehat{\varepsilon}_r. \tag{56}$$

Consistent with theory, both indirect utility and natural amenities significantly influence the spatial distribution of population across MSAs, with both coefficients being positive as expected. Table 1 further reports the observed MSA populations (scaled by their mean, i.e., $L_r/\overline{L}$), average productivities ($1/\overline{m}_r$) and observed amenity scores A^o_r, as well as the estimated/calibrated values of technological possibilities $\widehat{\mu}^{\max}_r$ and unobserved consumption amenities $\widehat{A}^u_r \equiv \widehat{\varepsilon}_r$.

Insert Figures 1 and 2 about here.

Figures 1 and 2 depict the spatial distribution of natural amenities, unobserved amenities, technological possibilities, and commuting technologies. There are several points worth emphasizing here. First, our quantified model yields detailed spatial patterns of unobserved consumption amenities and technological possibilities, the latter of which may be viewed as a measure for MSA-level production amenities. In contrast to, e.g., Roback (1982) and Albouy (2008), our amenity measures are derived from a framework where geography matters as trade frictions are explicitly taken into account. Both natural and unobserved amenities are positively correlated with city size, the correlation being stronger for the latter (0.7023) than for the former (0.1334). Larger cities thus tend to have better unobserved consumption amenities. Second, while the correlation between natural and unobserved amenities is zero by construction, there is also little correlation between technological possibilities and the two types of consumption amenities (-0.0867 and 0.0713 for A^o_r and $\widehat{A}^u_r$, respectively). This is consistent with the results by Chen and Rosenthal (2008) who find that good business locations in the US often have low consumption amenities, and vice versa.

4.7 MSA- and firm-level model fit

Our model can replicate several features of the data, both at the MSA and firm levels, that have not been used in the quantification procedure. We first compute the correlation between actual relative wages and those predicted by our model (see Appendix D for details on the data). The correlation is 0.7379 and thus reasonably high. We can also replicate the distribution of establishments across MSAs. Table 2 reports the mean, standard deviation, minimum, and maximum of the number of establishments (top part) and average establishment size (bottom part) at the MSA level, and the number of establishments is further broken down by employment size. The last column of Table 2 reports the correlation between the observed and our simulated data, which shows that the simple cross-MSA correlation between the observed and simulated total number of establishments is 0.7253, with a slightly larger rank correlation of 0.733. Again, these are reasonably large numbers. Turning to each size class, the fit is less good for small firms (size class 1–19) with a correlation of 0.3824. However, our model replicates fairly well the numbers of medium-sized and large establishments (size classes 20–99, 100–499 and 500+) across MSAs. This can be seen from the mean across MSAs, the corresponding standard deviations, and the minimum and maximum values. Furthermore, the correlations between the observed and predicted numbers of establishments across MSAs are fairly high (between 0.889 and 0.9412).

Insert Tables 2 and 3 about here.

Since our key objective is to investigate the importance of urban and trade frictions, having an idea of how well our model captures these frictions is very important. We hence assess our model's ability to replicate observed measures and proxies of these frictions.

Urban frictions. First, we consider urban frictions by comparing the 'model-based' and observed aggregate land rents. The former can be obtained as follows:

$$\widehat{\mathrm{ALR}}_r = \frac{2\pi w_r \overline{h}_r}{\widehat{\theta}_r^2}\left[1 - \left(1 + \widehat{\theta}_r\sqrt{L_r/\pi} + \frac{\widehat{\theta}_r^2 L_r}{2\pi}\right)\mathrm{e}^{-\widehat{\theta}_r\sqrt{L_r/\pi}}\right],$$

where we use our computed $\widehat{\theta}_r$ and the data on the wage w_r, the gross labor supply per capita $\overline{h}_r$, and city size L_r. The observed aggregate land rent is, in turn, obtained by $\mathrm{ALR}_r = \mathrm{GMR}_r/(1 - \mathrm{ownershare}_r)$, where GMR is the (aggregate) gross monthly rent.[23] The simple correlation between the model-based and observed aggregate land rents across MSAs is 0.9805, while the Spearman rank correlation is 0.9379. Alternatively, we can use $\mathrm{ALR}_r = \mathrm{ERV}_r/(\mathrm{ownershare}_r)$, where ERV_r is the equivalent rent value for houses that are owned. Under this alternative formula, the correlation between the model-based and observed aggregate land rents becomes 0.9624, while the Spearman rank correlation is 0.9129. In all cases, the correlations are high, thus suggesting that our model does a good job in capturing urban frictions across MSAs.

One might argue that our simple monocentric city model is not the most appropriate specification as large MSAs are usually polycentric. To see how urban frictions relate to polycentricity, we compute a simple correlation between $\widehat{\theta}_r$ and the number of employment centers in each MSA for the year

[23] The formula can be obtained as follows. First, the total amount of expenditure in housing services (ALR) is given by the sum of the gross monthly rent (GMR) and the equivalent rent value for houses that are owned (ERV). Data on GMR, which can be decomposed as (average rent) × (number of houses that are rented), is available. Now assume that GMR/(number of houses rented) = ERV/(number of houses owned) holds in each city at equilibrium by arbitrage. Under this hypothesis, we obtain ALR = GMR/(1 − share of houses that are owned).

2000 as identified by Arribas-Bel and Sanz Gracia (2010). The correlation is -0.4282, while the Spearman rank correlation is -0.5643, thus suggesting that our monocentric model with city-specific commuting technology captures the tendency that larger cities are more efficient for commuting as they allow for more employment centers, thereby reducing the average commuting distance.

Trade frictions. We next assess to what extent our model can cope with the existing micro evidence on the spatial structure of shipping patterns. To this end, we consider that the value of sales from an establishment in city r to city s represents one shipment (characterized by an origin MSA, a destination MSA, a shipping value, a unit price, and a shipping distance). We then draw a representative sample of 40,000 establishments from all MSAs, which yields a total of $40,000 \times 356^2$ potential shipments.[24] Most of the shipments do of course not occur, and there are only 243,784 positive shipments in our sample. Figure 3, which is analogous to Figures 1–3 in Hillberry and Hummels (2008), reports kernel regressions of various shipment characteristics on distance.[25] As one can see, both aggregate shipment values and the number of shipments fall off very quickly with distance – becoming very small beyond a threshold of about 200 miles – whereas price per unit first rises with distance and average shipment values do not display a clear pattern. These results are in line with the micro evidence from the CFS data provided by Hillberry and Hummels (2008).

Insert Figure 3 about here.

Table 3 further summarizes the observed and predicted shipping shares and shipping distances by establishment size class. The latter are obtained as follows. First, for each establishment with labor requirement m in MSA r, we compute the value of its sales:

$$\text{sales}_r(m) = \sum_s \chi_{rs} L_s p_{rs}(m) q_{rs}(m) = \frac{\widehat{w}_r m}{\alpha} \sum_s \chi_{rs} L_s d_{rs}^{\widehat{\gamma}} [W(\text{e}m/\widehat{m}_{rs}^x)^{-1} - 1].$$

We then classify all 6,431,886 establishments in our sample by employment size class, and disaggregate the value of sales for each establishment by distance shipped to compute the shares reported in Table 3.[26] The observed patterns in Table 3 come from Holmes and Stevens (2010) who use confidential CFS microdata from 1997 to compute the shares of shipping values by distance as well as average shipping distances. As can be seen from Table 3, our model can qualitatively reproduce the observed shipment shares, although it over- (under-) predicts the share of shipments within a short distance for small (large) establishments while it under- (over-) predicts the share of ship-

[24] As in Section 4.4, the sample size is immaterial for our results provided that it is large enough. Given that the number of shipments is substantially larger than the number of firms, drawing a large sample of 6.5 million firms as before proves computationally infeasible.

[25] As in Hillberry and Hummels (2008), we use a Gaussian kernel with optimal bandwidth and calculated on 100 points. We report results for distances greater than about 10 miles (the minimum in our sample) and up to slightly below 3,000 miles (the maximum in our sample). Note that we have less variation in distances than Hillberry and Hummels (2008) who use either 3-digit or 5-digit zip code level data instead of MSA data.

[26] Since we work with shares, the unobservable scaling parameter α does not affect our results.

ments within a long distance for small (large) establishments. Finally, our model can also explain the tendency that the mean distance shipped increases with establishment size (columns 10–12).

5 Counterfactuals

Having shown that our quantified model performs well in replicating several features of the data, we now use it for counterfactual analysis. Our aim is, in particular, to assess how the US city-size distribution, the sizes of individual cities, as well as the distributions of productivity and markups across MSAs would change if either urban frictions or trade frictions were eliminated.

5.1 Numerical procedure

We first explain in some detail the procedure used for running counterfactuals in our framework. In our first counterfactual experiment (which we call 'no urban frictions'), we set all commuting-related frictions – and hence all land rents – to zero ($\widehat{\theta}_r = 0$ for all r) while keeping trade frictions $\widehat{\tau}_{rs}$, technological possibilities $\widehat{\mu}_r^{\max}$, and amenities (A_r^o and $\widehat{A}_r^u$) constant.[27] This corresponds to a hypothetical world where only goods are costly to transport while living in cities does not impose any urban costs. In our second counterfactual experiment (which we call 'no trade frictions'), we set external trade costs from s to r equal to internal trade costs in r ($\tau_{sr} = \tau_{rr}$ for all r and s) while holding urban frictions $\widehat{\theta}_r$, technological possibilities $\widehat{\mu}_r^{\max}$, and amenities (A_r^o and $\widehat{A}_r^u$) constant. This corresponds to a hypothetical world where consumers face the same trade costs for local and non-local varieties.[28] For the sake of brevity, we explain the procedure for the case without urban frictions only as it works analogously for the case without trade frictions.

First, we let $\widehat{\theta}_r = 0$ for all r and keep the initial population distribution fixed. This parameter change induces changes in the indirect utility levels. Let $\widetilde{U}_r^0$ denote the new counterfactual utility in MSA r, evaluated at the initial population and $\widehat{\theta}_r = 0$. Second, we replace $\widehat{U}_r$ with its new counterfactual value $\widetilde{U}_r^0$ to obtain $\widetilde{D}_r^0 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{U}_r^0 + \widehat{\alpha}_2 A_r^o + \widehat{A}_r^u$. The spatial equilibrium conditions (54) will then, in general, no longer be satisfied, and hence city sizes must change. We thus consider the following iterative adjustment procedure to find the new counterfactual spatial equilibrium:

1. Consider the new choice probabilities

$$\widetilde{\mathbb{P}}_r^0 = \frac{\exp(\widetilde{D}_r^0)}{\sum_s \exp(\widetilde{D}_s^0)} \tag{57}$$

[27] Although workers are mobile in our model, we can set urban frictions to zero without having degenerate equilibria with full agglomeration in a single city. The reason is that, as explained before, consumers' location choice probabilities are expressed as a logit so that no city can completely disappear.

[28] We have also experimented with setting $\tau_{rs} = \tau_{rr}$ for all r and s, which corresponds to a hypothetical world where goods are as costly to trade between MSAs as within MSAs from the firms' perspective. Although the magnitudes delivered by this alternative counterfactual scenario are slightly larger, there are no qualitative changes.

induced by the change in spatial frictions, which yield a new population distribution $\widetilde{L}_r^0 = L\widetilde{\mathbb{P}}_r^0$ for all $r = 1, ..., K$.

2. Given the intial $\widehat{\mu}_r^{\max}$, the new population distribution $\widetilde{L}_r^0$ for all $r = 1, ..., K$, as well as the counterfactual value for the commuting technology parameter $\widehat{\theta}_r = 0$, the market equilibrium conditions (51) and (52) generate new wages $\widetilde{w}_r^1$ and cutoffs $(\widetilde{m}_r^d)^1$. Expression (37) then yields new utility levels $\widetilde{U}_r^1$.

3. Using $\widetilde{D}_r^1 = \widehat{\alpha}_0 + \widehat{\alpha}_1 \widetilde{U}_r^1 + \widehat{\alpha}_2 A_r^o + \widehat{A}_r^u$, the choice probabilities can be updated as in (57), which yields a new population distribution $\widetilde{L}_r^1 = L\widetilde{\mathbb{P}}_r^1$ for all $r = 1, ..., K$.

4. We iterate through steps 2–3 until convergence of the population distribution to obtain $\{\widetilde{L}_r, \widetilde{w}_r, \widetilde{m}_r^d\}$ for all $r = 1, ..., K$.

5.2 No urban frictions

How would the US economic geography look like without urban frictions? In this subsection, we focus on counterfactual changes in population, productivity, and markups. Starting with city sizes, eliminating urban frictions leads to (gross) cross-MSA population movements of about 4 million people, i.e., 1.6% of the total MSA population in our sample. These population changes are unevenly spread across MSAs. New York, for example, gains about 8.5% and some MSAs close to New York and Boston gain even more (New Haven-Milford, CT, gains about 12.1% and Bridgeport-Stamford-Norwalk, CT, about 15.9%). Consistent with the comparative static results of Section 3.4, large cities on average gain population, whereas small- and medium-sized cities tend to lose. These results are depicted in Figure 5, which plots percentage changes in MSA population against the initial log MSA population. Further insights are provided by the the top panel of Figure 6, which depicts the distribution of counterfactual percentage changes in L_r. As there are many more small cities that lose population than large cities that gain population, the implied distribution of percentage changes is skewed to the left. Last, these population changes follow a rich spatial pattern, as depicted in the top panel of Figure 7. Although individual city sizes would be substantially affected by the fall in urban frictions, the city-size distribution remains fairly stable as shown in Figure 4. A standard rank-size rule regression reveals that the coefficient on log size rises slightly from -0.9249 to -0.9178, the change being however statistically insignificant.[29] We will discuss this stability in greater depth in Section 6.3.

Insert Figures 4 and 5 about here.

Turning to changes in average productivity, observe that most MSAs actually lose when urban frictions are eliminated (see the middle panel of Figure 6). Indeed, as shown in Figures 6 and 7, productivity changes can go either way. For example, Monroe, MI (a smaller MSA) experiences a

[29] We follow Gabaix and Ibragimov (2011) and adjust the rank by subtracting 1/2.

productivity decrease of 0.9%, whereas New York sees its productivity increase by 0.76%. This is consistent with our results from Section 3.4: as small MSAs lose population, local market size and, thereby, average productivity deteriorate; in contrast, large MSAs and cities close by see their market size expand, which raises productivity as trade frictions are unchanged. Interestingly, smaller cities near New York, like Bridgeport-Stamford-Norwalk, CT, and Trenton-Ewing, NJ, see their productivity increase by about 1.4% and 0.9%, respectively, which even exceeds the productivity gain in New York itself. Computing the nation-wide productivity change, weighted by MSA population shares in the initial equilibrium, we find that eliminating urban frictions would increase average productivity by a mere 0.04%.

Insert Figures 6 and 7 about here.

As for markups, the bottom panels of Figures 6 and 7 reveal that this is the dimension where the largest changes take place. Markups would decrease everywhere, with reductions ranging from 5.3% to about 16%, but the more so for the most populated areas of the East and West coasts. As can be seen from (36), the reason for these large changes is twofold. First, eliminating urban frictions increases the effective labor supply per capita $h_r = S_r/L_r$ everywhere, which allows for more firms in each MSA and, therefore, for more competition. Second, there is an effect going through the cutoffs. Some places see their cutoffs fall, which puts additional pressure on markups. While cutoffs may increase in cities that lose population, the second effect is always dominated by the first one, so that markups fall in all MSAs.

To summarize, even without urban frictions, the city-size distribution would remain fairly stable, despite the fact that larger cities tend to grow and smaller cities tend to shrink. Furthermore, the 'no urban frictions' case supports more firms, which reduces markups and expands product diversity, though firms are not on average much more productive than in a world with urban frictions. The productivity gap between large and small cities would, however, widen.

5.3 No trade frictions

What would happen to individual city sizes, to the city-size distribution, and to productivity and markups in a world where consumers face the same trade costs for local and non-local varieties? To investigate this issue, we set $\widehat{\tau}_{sr} = \widehat{\tau}_{rr}$ for all r and s.[30] Starting with city sizes, eliminating trade frictions would lead to significant (gross) cross-MSA population movements of about 10.2 million people, i.e., 4.08% of the total MSA population in our sample. Some small cities would gain substantially. For example, the population of Casper, WY, would grow by about 105% and that of Hinesville-Fort Stewart, GA, by about 99.4%. Figure 9 plots the percentage changes in MSA population against the initial log MSA population. Consistent with the comparative static results of Section 3.4, in a world without trade frictions larger cities lose ground and agents move, on average,

[30]Eaton and Kortum (2002) consider a similar counterfactual scenario in the context of international trade, yet without considering induced changes in the population distribution and with a fixed mass of varieties.

to smaller cities to relax urban costs. These changes are depicted in the top panel of Figure 11. Although individual cities would be substantially affected by the fall in trade frictions, the city-size distribution remains again quite stable, as can be seen from Figure 8. The coefficient on log size drops from -0.9249 to -0.9392, yet this change is again statistically insignificant.

Insert Figures 8 and 9 about here.

Concerning the changes in average productivity, observe first that all MSAs gain. Yet, as can be seen from the middle panels of Figures 10 and 11, the gains are unevenly spread across MSAs. Whereas some small cities gain substantially (e.g., an increase of about 125.5% in Great Falls, MT), large cities gain significantly less: 41.18% in New York, 48.08% in Los Angeles, and 55.71% in Chicago. The first reason is linked to market access. Indeed, the more populated areas, e.g., those centered around California and New England, would be those gaining the least from a reduction of trade frictions, as they already provide firms with a good access to a large local market. The second reason is that, as stated above, large cities tend to lose population, thereby reducing the productivity gains brought about by the fall in trade frictions. Computing the nation-wide productivity change, weighted by MSA population shares in the initial equilibrium, we find that eliminating trade frictions would increase average productivity by 67.59%. Thus, reducing spatial frictions for shipping goods would entail substantial aggregate productivity gains.

Insert Figures 10 and 11 about here.

As for markups, the bottom panels of Figures 10 and 11 reveal that they would decrease considerably in a world without trade frictions, with reductions ranging from 29% to 55%. Such reductions are particularly strong in MSAs with poor market access, i.e., the center of the US and the areas close to the borders. Observe that the changes in markups – though substantial – are more compressed than the changes in productivity (the coefficient of variation for productivity changes is 0.18, while that for changes in markups is 0.09). The reason is the following. Eliminating trade frictions reduces cutoffs in all MSAs, but especially in small and remote ones. This puts downward pressure on markups. Yet, there is also an indirect effect through changes in effective labor supply h_r. An increase in h_r, which occurs in big cities that lose population, reduces markups more strongly than what is implied by the direct change only, while the decrease in h_r that occurs in small and remote cities gaining population works in the opposite direction and dampens the markup reductions.

To summarize, even without trade frictions, the city-size distribution would remain fairly stable, despite the fact that larger cities tend to shrink and smaller cities tend to grow. Furthermore, the 'no trade frictions' case allows for higher average productivity and lower markups by intensifying competition in all MSAs, and especially in small and remote ones. The productivity gap between large and small cities would, therefore, shrink.

6 Extensions and discussion

6.1 Short- vs long-run impacts

The main insights from our two counterfactual experiments are summarized in the top panel of Table 4. These results refer to the *long-run* impacts of eliminating urban or trade frictions as they include the effects of population movements. To gauge the contribution of labor mobility to these overall impacts of spatial frictions in the US, it is useful to disentangle the *short-run* effects, before the population reshuffling has taken place, from the long-run effects.

We now consider the same counterfactual experiments as in the previous section, yet we do not allow for labor mobility and hold city sizes fixed at their initial levels. The margins of adjustment are then productivity, markups and wages. Key results are given in the middle panel of Table 4. As one can see by comparing the short-run and the long-run figures, the bulk of changes takes place already in the short-run.

Insert Table 4 and Figure 12 about here.

One noticeable exception is productivity changes whose sign gets reversed between the short- and long-run in the no urban frictions case. This decomposition of the short- and long-run effects can also be related to the comparative static results of Section 3.4. There, we have shown that the instantaneous impact of reducing urban frictions – keeping L_r fixed – is to raise m_r^d (i.e., to lower productivity) in the large city and to raise productivity in the small city. The quantitative findings summarized in the top panel of Figure 12 are consistent with this prediction, as they show that the cutoffs m_r^d rise, on average, in larger cities when urban frictions are eliminated while population is held fixed. However, as can be seen from the bottom panel of Figure 12, the subsequent movement of population (which flows toward the larger cities), more than offsets this initial change, thereby generating larger productivity gains in the bigger cities in the long-run equilibrium.[31] Summing up, whereas short-run impacts play a key role in the overall adjustments to spatial frictions, population mobility is crucial for understanding productivity changes.

6.2 Agglomeration economies

There is a large body of literature showing that agglomeration economies, i.e., productivity gains due to larger or denser urban areas, are a prevalent feature of the spatial economy (Rosenthal and Strange, 2004; Melo *et al.*, 2010). We have so far focused entirely on one channel: larger cities are more productive because of tougher firm selection. Yet, larger or denser cities can become more productive for various other reasons such as sharing–matching–learning externalities (Duranton and Puga, 2004), and sorting by human capital (Combes *et al.*, 2008; Behrens *et al.*, 2010). Although

[31]Some simple OLS regressions of the change in m_r^d in the short- and in the long-run on inital population yield: $\Delta m_r^d = -0.0821^{***} + 0.0127^{***} L_r$ in the short-run, and $\Delta m_r^d = 0.0817^{***} - 0.0194^{***} L_r$ in the long-run, thus showing the switch in the results depending on whether or not population is considered mobile.

some recent studies attempt to assess the relative importance of firm selection and more conventional agglomeration economies in explaining the productivity advantage of large cities, it is fair to say that the issue is not settled yet (see, e.g., Combes *et al.*, 2010; Holmes *et al.*, 2010).

In this subsection, we illustrate a simple way to extend our framework to include agglomeration economies. Specifically, we allow the upper bound in each MSA ($m_r^{\max}$) to be a function of the density of that MSA. Agglomeration economies are thus modeled as a right-shift in the *ex ante* productivity distribution: upon entry, a firm in a denser MSA has a higher probability of getting a better productivity draw.[32] Starting from the baseline model, assume that technological possibilities $\mu_r^{\max}$ can be expressed as $\mu_r^{\max} = c \cdot \text{density}_r^{-k\xi} \cdot \psi_r^{\max}$, where $\text{density}_r \equiv L_r/\text{surface}_r$, ξ is the elasticity of the *ex ante* upper bound of the marginal labor requirement with respect to density, and $\psi_r^{\max}$ is an idiosyncratic measure of technological possibilities that is purged from agglomeration effects. We can then estimate the ex ante productivity advantage of large cities by running a simple log-log regression of $\widehat{\mu}_r^{\max}$ on MSA population densities and a constant, which yields:

$$\ln(\widehat{\mu}_r^{\max}) = \underset{(0.3566)}{2.6898^{***}} - \underset{(0.0813)}{0.1889^{**}} \ln(\text{density}_r).$$

Since in the model $\ln \mu_r^{\max} = k \ln m_r^{\max}$ plus a constant, the elasticity ξ of $m_r^{\max}$ with respect to density is given by $-0.1889/\widehat{k} = 0.0295$, which is the value we use in what follows. In words, doubling MSA density reduces the upper bound (and, equivalently, the mean by the properties of the Pareto distribution) of the *ex ante* marginal labor requirement of entrants by 2.95%. That figure, though computed for the ex ante distribution, lies within the consensus range of previous elasticity estimates for agglomeration economies measured using ex post productivity distributions (see Melo *et al.*, 2010). Note that this effect is independent of the subsequent truncation of the ex post productivity distribution, thus disentangling agglomeration from selection.

We compute $\widehat{\mu}_r^{\max}$ in the initial equilibrium. Call it $\widehat{\mu}_r^{\max,0}$. Assume now that the population of MSA r changes from L_r^0 to L_r^1. The new $\widehat{\mu}_r^{\max}$ is then given by $\widehat{\mu}_r^{\max,1} = c \cdot (L_r^1/\text{surface}_r)^{-k\xi} \cdot \widehat{\psi}_r^{\max}$. Hence, it is easy to see that, given the initial estimates $\widehat{\mu}_r^{\max,0}$ we have $\widehat{\mu}_r^{\max,1} = \widehat{\mu}_r^{\max,0} \left(L_r^1/L_r^0\right)^{-k\xi}$. Thus, we can integrate agglomeration economies in a straightforward way into our framework by replacing $\widehat{\mu}_r^{\max}$ by $\widehat{\mu}_r^{\max} \left(L_r/L_r^0\right)^{-k\xi}$ in the market equilibrium conditions (51) and (52) when running the counterfactuals:

$$\widehat{\mu}_r^{\max} \left(\frac{L_r^1}{L_r^0}\right)^{-k\xi} = \sum_s L_s^1 \tau_{rs} \left(m_s^d \frac{\tau_{ss}}{\tau_{rs}} \frac{w_s}{w_r}\right)^{k+1} \tag{58}$$

$$\frac{S_r^1}{L_r^1} \frac{1}{\left(m_r^d\right)^{k+1}} = \sum_s S_s^1 \tau_{rr} \left(\frac{\tau_{sr}}{\tau_{rr}} \frac{w_s}{w_r}\right)^{-k} \frac{1}{\widehat{\mu}_s^{\max} \left(\frac{L_s^1}{L_s^0}\right)^{-k\xi}}, \tag{59}$$

We run both counterfactuals ('no urban frictions' and 'no trade frictions') with the agglomeration economies specification. The long-run impacts are summarized in the bottom panel of Table 4

[32] Formally, the right-shift in the *ex ante* productivity distribution implies that the distribution in a denser MSA first-order stochastically dominates that in a less dense MSA. Observe that firm selection afterwards acts as a truncation, so that the *ex post* distribution is both right-shifted and truncated.

(labeled CF3 and CF4, respectively). As can be seen, results change little compared to our previous specification without agglomeration economies (reported in the top panel). If anything, the implied aggregate changes become a bit larger, but the overall difference is small. Observe that this finding does not mean that agglomeration economies are unimportant. The reason why they do not matter much in our counterfactuals is that not that many people move between the initial and the counterfactual equilibria. Yet, given the measured elasticities of agglomeration economies, substantial population movements would be required for them to become quantitatively really important.

6.3 How important are spatial frictions?

Our paper is, to the best of our knowledge, the first to investigate the impact of both urban and trade frictions on the size distribution of cities.[33] A key novel insight of our analysis is that spatial frictions have a quite limited impact on that distribution. Although there would be small changes in the coefficient on log size, the rank-size rule would still hold with a statistically identical coefficient in a world without urban or trade frictions (both with and without the prevalence of agglomeration economies).[34] This result has important implications for future spatial modeling. As far as the city-size distribution is concerned, we can apparently abstract from either urban or trade frictions without much loss of generality. Hence, the modeling strategies taken by recent studies such as Gabaix (1999), Eeckhout (2004), Duranton (2007) and Rossi-Hansberg and Wright (2007), where trade frictions are assumed away, indeed appear to be good approximations.

Although spatial frictions hardly affect the city-size distribution, they do matter for the sizes of individual cities within that stable distribution. Indeed, eliminating spatial frictions leads to aggregate (gross) inter-MSA reallocations of about 4–10 million people. Whether or not large or small cities gain population crucially depends on which type of spatial frictions is eliminated. Actually, our numbers for the aggregate population movements might appear quite small at first glance, given that we contemplate major exogenous shocks in our counterfactual exercises. Yet, one has to keep in mind that we have considered *simultaneous* reductions in spatial frictions for all cities. We can

[33]The most closely related paper in that respect is Desmet and Rossi-Hansberg (2010). Their framework, however, abstracts from trade frictions, so that it is not suited to investigate their impact on the city-size distribution. Our result on urban frictions also contrasts with that of Desmet and Rossi-Hansberg (2010), who find that the size distribution tilts substantially when urban frictions are reduced. The difference in results can be understood as follows. In their analysis, the commuting friction parameter is common to all MSAs, whereas it is city specific in our model. In our setting, big cities like New York or Los Angeles tend to have the best commuting technologies in the initial equilibrium, so that the impacts of setting $\widehat{\theta}_r = 0$ are relatively small there. By contrast, in Desmet and Rossi-Hansberg (2010), the commuting technology improves equally across all MSAs so that big cities get very large due to larger efficiency gains in commuting than in our case. Another key difference is that in Desmet and Rossi-Hansberg (2010), all consumers react in the same way to changes in utility and amenities, whereas those reactions are idiosyncratic in our model and, therefore, less extreme.

[34]This insight is also consistent with the relative stability of the US city-size distribution over the 20th century as documented by Black and Henderson (2003). Note that although urban and trade frictions changed a lot over that century, such aspects are not explicitly incorporated into their stochastic modeling framework.

also look at a *unilateral* reduction for a single city only. Specifically, let us briefly consider two additional counterfactuals. In the first one, we only change, with respect to the initial equilibrium, urban costs for New York where they fall to zero. In that case, New York grows by about 19.73% (i.e., by about 3.7 million people) in the specification without, and by 20.61% in the specification with agglomeration economies. In the second one, we set $\tau_{sr} = \tau_{rr}$ for all s only when r is New York. That is, we improve the market access to New York for all firms that are located elsewhere, while holding the market access of firms located in New York to other MSAs constant. In that case, New York shrinks by a remarkable 15.57% (i.e., about 3 million people), and if we additionally allow for agglomeration economies it even shrinks by 15.95%. Hence, a unilateral change in spatial frictions for a single city has a much larger impact on the size of that city. More generally, these results show that the *relative levels across cities* of both types of frictions matter a lot to understand the sizes of individual cities.

Finally, our experiments show that urban and trade frictions matter, though to a different extent, for the distributions of productivity and markups – and ultimately welfare – across MSAs. Eliminating trade frictions would lead to significant productivity gains and substantially reduced markups. These changes are highly heterogeneous across space and tend to reduce differences in productivity and city sizes across MSAs. Concerning urban frictions, their elimination would not give rise to such significant productivity gains, but would still considerably intensify competition and generate lower markups.

7 Conclusions

We have developed a new NEG-cum-'urban systems' model and analyzed how city sizes, on the one hand, and productivity and competition, on the other hand, simultaneously respond to shocks in spatial frictions. Using 2007 US data at the state and at the metropolitan statistical area (MSA) levels, we have quantified our model using all of its market and spatial equilibrium conditions, a gravity equation for trade flows, and a logit model for consumers' location choice probabilities. The quantified model performs well empirically and is able to reproduce – both at the MSA and the firm levels – a number of empirical features that are linked to urban and trade frictions

To assess the importance of spatial frictions, we have used our model to study two counterfactual scenarios. Those allow us to trace out the impacts of both trade and urban frictions on the city-size distribution, the sizes of individual cities, as well as on productivity and competition across space. A first key insight is that the city-size distribution is little sensitive to the presence of either trade or urban frictions. A second key insight is that, within the stable distribution, the sizes of individual cities can be affected substantially by changes in spatial frictions. Last, our third key insight is that their presence imposes quite significant welfare costs. The reasons are too high price-cost margins and, depending on the type of spatial frictions we consider, foregone productivity or reduced product diversity.

We believe that our framework provides a useful starting point for further general equilibrium counterfactual analysis in spatial models. In particular, our model: (i) endogenizes productivity, markups, and product diversity at the firm level, three aspects that loom large in the recent trade literature; (ii) encompasses many key elements identified as being relevant by the NEG and urban economics literature;; (iii) allows to deal with heterogeneity along several dimensions (across space, across firms, across consumers); (iv) can be readily brought to data in very a self-contained way; (v) fits quite nicely features of the data not used in the quantification stage, including spatial shipping patterns and aggregate land rents; and (vi) provides a more spatially oriented approach to the classical Rosen-Roback type of analysis widely used in the urban economics literature.

There are many additional relevant questions that could be investigated within our framework, and we here suggest two of them. First, our model delivers a MSA-specific measure of underlying productivity, our technological possibilities $\widehat{\mu}_r^{\max}$. This measure is, by construction, filtered for agglomeration effects that stem from either local market size or accessibility. The correlation with an observed measure of productivity, such as GDP per employee (m_r^d), is far from perfect (0.6512) thus providing substantial additional information on the determinants of an MSA's productivity. Analyzing the economic fundamentals behind the spatial and temporal variation in the $\widehat{\mu}_r^{\max}$ certainly represents an interesting avenue of further research. Second, it would be desirable to replicate our results for countries other than the US. The features of the spatial distribution of economic activity in the European Union are, for example, quite different from those of the US.

Acknowledgements. We thank Klaus Desmet, Gilles Duranton, Yannis Ioannides, Henry Overman, Esteban Rossi-Hansberg, Daniel Sturm, as well as seminar and conference participants at the Sauder Business School of UBC, the LSE, the 2008 Annual Meeting of the NARSC in Brooklyn, New York, the 2010 CURE Conference in Princeton, and the 2011 Göttingen Trade Workshop for helpful comments and suggestions. We also thank Daniel Arribas-Bel and Fernando Sanz Gracia for sharing their data on polycentric cities with us, and Gilles Duranton for discussion on the CFS distance elasticities. Behrens is holder of the *Canada Research Chair in Regional Impacts of Globalization*. Financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada is gratefully acknowledged. He also gratefully acknowledges financial support from FQRSC Québec (Grant NP-127178). Murata gratefully acknowledges financial support from the Japan Society for the Promotion of Science (17730165). The usual disclaimer applies.

References

[1] Albouy, D. (2008) Are big cities bad places to live? Estimating quality of life across metropolitan areas, NBER *Working Paper* #14472, National Bureau of Economic Research: Boston, MA.

[2] Alonso, W.A. (1964) *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.

[3] Anderson, J.E. and E. van Wincoop (2003) Gravity with gravitas: A solution to the border puzzle, *American Economic Review* 93, 170–192.

[4] Arribas-Bel, D. and F. Sanz Gracia (2010) The validity of the monocentric city model in a polycentric age: US Metropolitan areas in 1900 and 2000. Mimeographed, Arizona State University.

[5] Behrens, K. and Y. Murata (2007) General equilibrium models of monopolistic competition: a new approach, *Journal of Economic Theory* 136, 776–787.

[6] Behrens, K., G. Mion, Y. Murata and J. Südekum (2009) Trade, wages and productivity, CEPR *Discussion Paper* #7369, Center for Economic Policy Research: London, UK.

[7] Behrens, K., G. Duranton and F.L. Robert-Nicoud (2010) Productive cities: Agglomeration, selection and sorting, CEPR Discussion Paper #7922, Center for Economic Policy Research: London, UK.

[8] Bernard, A.B., J. Eaton, J.B. Jensen and S. Kortum (2003) Plants and productivity in international trade, *American Economic Review* 93, 1268–1290.

[9] Black, D. and J.V. Henderson (2003) Urban evolution in the USA, *Journal of Economic Geography* 3, 343–372.

[10] Chen, Y. and S.S. Rosenthal (2008) Local amenities and life-cycle migration: Do people move for jobs or fun?, *Journal of Urban Economics* 64, 519–537.

[11] Combes, P.-Ph., G. Duranton and L. Gobillon (2008) Spatial wage disparities: Sorting matters!, *Journal of Urban Economics* 63, 723–742.

[12] Combes, P.-Ph., G. Duranton, L. Gobillon, D. Puga and S. Roux (2010) The productivity advantages of large cities: Distinguishing agglomeration from firm selection, CEPR *Discussion Paper* #7191, Center for Economic Policy Research: London, UK.

[13] Combes, P.-Ph. and M. Lafourcade (2011) Competition, market access and economic geography: Structural estimation and predictions for France, *Regional Science and Urban Economics*, forthcoming.

[14] Corless, R.M., G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey and D.E. Knuth (1996) On the Lambert W function, *Advances in Computational Mathematics* 5, 329–359.

[15] Desmet, K. and E. Rossi-Hansberg (2010) Urban accounting and welfare, CEPR *Discussion Paper* #8168, Center for Economic Policy Research: London, UK.

[16] Duranton, G. and D. Puga (2004) Micro-foundations of urban agglomeration economies. In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: North-Holland, pp.2063–2117.

[17] Duranton, G. (2007) Urban evolutions: the fast, the slow, and the still, *American Economic Review* 98, 197–221.

[18] Duranton, G., P.M. Morrow and M.A. Turner (2011) Roads and trade: Evidence from the US. Mimeographed, University of Toronto.

[19] Eaton, J. and S. Kortum (2002) Technology, geography and trade, *Econometrica* 70, 1741–1779.

[20] Eeckhout, J. (2004) Gibrat's law for (all) cities, *American Economic Review* 94, 1429–1451.

[21] Feenstra, R.C. and D.E. Weinstein (2010) Globalization, markups, and the U.S. price level, NBER *Working Paper* #15749, National Bureau of Economic Research: Boston, MA.

[22] Fujita, M. (1989) *Urban Economic Theory: Land Use and City Size*. Cambridge, MA: Cambridge Univ. Press.

[23] Fujita, M., P.R. Krugman and A.J. Venables (1999) *The Spatial Economy. Cities, Regions and International Trade.* Cambridge, MA: MIT Press.

[24] Fujita, M. and J.-F. Thisse (2002) *Economics of Agglomeration – Cities, Industrial Location, and Regional Growth*. Cambridge, MA: Cambridge University Press.

[25] Gabaix, X. (1999) Zipf's law for cities: an explanation, *Quarterly Journal of Economics* 114, 739–767.

[26] Gabaix, X. and R. Ibragimov (2011) Rank-1/2: A simple way to improve the OLS estimation of tail exponents. *Journal of Business Economics and Statistics* 29, 24–39.

[27] Glaeser, E.L. (2008) *Cities, Agglomeration, and Spatial Equilibrium*. Oxford, UK: Oxford University Press.

[28] Handbury, J. and D. Weinstein (2011) Is new economic geography right? Evidence from price data, NBER *Working Paper* #17067, National Bureau of Economic Research: Boston, MA.

[29] Hanson, G. (2005) Market potential, increasing returns, and geographic concentration, *Journal of International Economics* 67, 1–24.

[30] Helpman, E. (1998) The size of regions. In: D. Pines, E. Sadka, I. Zilcha (eds.), *Topics in Public Economics. Theoretical and Empirical Analysis*, Cambridge University Press, pp.33–54.

[31] Henderson, V.J. (1974) The sizes and types of cities, *American Economic Review* 64, 640–656.

[32] Hillberry, R. and D. Hummels (2008) Trade responses to geographic frictions: A decomposition using micro-data, *European Economic Review* 52, 527–550.

[33] Holmes, T.J., W.-T. Hsu and S. Lee (2010) Plants and productivity in regional agglomeration. Mimeographed, University of Minnesota *and* Chinese University Hong Kong *and* University of British Columbia.

[34] Holmes, T.J. and J.J. Stevens (2010) Exports, borders, distance, and plant size, NBER *Working Paper* #16046, National Bureau for Economic Research: Boston, MA.

[35] Krugman, P.R. (1979) Increasing returns, monopolistic competition, and international trade *Journal of International Economics* 9, 469–479.

[36] Krugman, P.R. (1980) Scale economies, product differentiation and the pattern of trade, *American Economic Review* 70, 950–959.

[37] Krugman, P.R. (1991) Increasing returns and economic geography, *Journal of Political Economy* 99, 483–499.

[38] Lucas, R.E. and E. Rossi-Hansberg (2002) On the internal structure of cities, *Econometrica* 70, 1445–1476.

[39] McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (ed.), *Frontiers in Econometrics*. Academic Press, New York, pp.105–142.

[40] Melitz, M.J. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* 71, 1695–1725.

[41] Melitz, M.J. and G.I.P. Ottaviano (2008) Market size, trade, and productivity, *Review of Economic Studies* 75, 295–316.

[42] Melo, P.C., D.J. Graham and R.B. Noland (2009) A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics* 39, 332–342.

[43] Murata, Y. (2003) Product diversity, taste heterogeneity, and geographic distribution of economic activities: market vs. non-market interactions, *Journal of Urban Economics* 53, 126–144.

[44] Murata, Y. and J.-F. Thisse (2005) A simple model of economic geography à la Helpman-Tabuchi, *Journal of Urban Economics* 58, 137–155.

[45] Ottaviano, G.I.P., T. Tabuchi and J.-F. Thisse (2002) Agglomeration and trade revisited, *International Economic Review* 43, 409–435.

[46] Redding, S.J. and A.J. Venables (2004) Economic geography and international inequality, *Journal of International Economics* 62, 53–82.

[47] Redding, S.J. and D. Sturm (2008) The costs of remoteness: Evidence from German division and reunification, *American Economic Review* 98, 1766–1797.

[48] Redding, S.J. (2010) The empirics of New Economic Geography, *Journal of Regional Science* 50, 297–311.

[49] Rosenthal, S.S. and W.C. Strange (2004) Evidence on the nature and sources of agglomeration economies. In: Henderson, J.V. and J.-F. Thisse (eds.), *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: North-Holland, pp.2119–2171.

[50] Rossi-Hansberg, E. and M. Wright (2007) Urban structure and growth, *Review of Economic Studies* 74, 597–624.

[51] Reza, F.M. (1994) *An Introduction to Information Theory*. Mineola, NY: Dover Publications Inc.

[52] Roback, J. (1982) Wages, rents, and the quality of life, *Journal of Political Economy* 90, 1257–1278.

[53] Syverson, C. (2004) Market structure and productivity: a concrete example, *Journal of Political Economy* 112, 1181-1222.

[54] Tabuchi, T. (1998) Urban agglomeration and dispersion: A synthesis of Alonso and Krugman, *Journal of Urban Economics* 44, 333–351.

[55] Tabuchi, T. and J.-F. Thisse (2002) Taste heterogeneity, labor mobility and economic geography, *Journal of Development Economics* 69, 155–177.

Appendix A: Proofs

A.1. Existence and uniqueness of the equilibrium cutoff m^d. To see that there exists a unique equilibrium cutoff m^d, we apply the Leibniz integral rule to the left-hand side of (14) and use $W(\mathrm{e}) = 1$ to obtain

$$\frac{\mathrm{e}L}{\alpha(m^d)^2}\int_0^{m^d} m^2\left(W^{-2}-1\right)W'\mathrm{d}G(m) > 0,$$

where the sign comes from $W' > 0$ and $W^{-2} \geq 1$ for $0 \leq m \leq m^d$. Hence, the left-hand side of (14) is strictly increasing. This uniquely determines the equilibrium cutoff m^d, because

$$\lim_{m^d\to 0}\int_0^{m^d} m\left(W^{-1}+W-2\right)\mathrm{d}G(m) = 0 \quad \text{and} \quad \lim_{m^d\to\infty}\int_0^{m^d} m\left(W^{-1}+W-2\right)\mathrm{d}G(m) = \infty.$$

A.2. Indirect utility in the single city. To derive the indirect utility, we first compute the (unweighted) average price across all varieties. Multiplying both sides of (10) by $p(i)$, integrating over Ω, and using (7), we obtain:

$$\overline{p} = \overline{m}w + \frac{\alpha E}{N}$$

where $\overline{m} \equiv (1/N)\int_\Omega m(j)\mathrm{d}j$ denotes the average marginal labor requirement of the surviving firms. Using $\overline{p}$, expression (8) can be rewritten as

$$U = \frac{N}{k+1} - \frac{S}{L}\frac{\alpha}{m^d}, \tag{60}$$

where we use $E = (S/L)w$, $p^d = m^d w$ and $\overline{m} = [k/(k+1)]m^d$. When combined with (18) and (19), we obtain the expression for U as given in (20).

We now show that U is single-peaked with respect to L. To this end, we rewrite the indirect utility (21) as $U = b(S/L)L^{1/(k+1)}$, where b is a positive constant capturing k, α, and $\mu^{\max}$, and then consider a log-transformation, $\ln U = \ln b + \ln S - [k/(k+1)]\ln L$. It then follows that

$$\frac{\partial \ln U}{\partial \ln L} = \frac{LS'}{S} - \frac{k}{k+1}.$$

To establish single-peakedness, we need to show that

$$\frac{LS'}{S} = \frac{\theta^2(L/\pi)}{2\left(\mathrm{e}^{\theta\sqrt{L/\pi}} - 1 - \theta\sqrt{L/\pi}\right)}$$

cuts the horizontal line $k/(k+1) \in (0,1)$ only once from above. Notice that $LS'/S \to 1$ as $L \to 0$, whereas $LS'/S \to 0$ as $L \to \infty$. Single-peakedness therefore follows if

$$\frac{\mathrm{d}}{\mathrm{d}L}\left(\frac{LS'}{S}\right) = -\frac{2+\theta\sqrt{L/\pi}+\mathrm{e}^{\theta\sqrt{L/\pi}}\left(\theta\sqrt{L/\pi}-2\right)}{(4/\theta^2)\left[\sqrt{\pi}\left(\mathrm{e}^{\theta\sqrt{L/\pi}}-1\right)-\theta\sqrt{L}\right]^2} < 0, \quad \forall L.$$

For this to be the case, the numerator must be positive. Let $y \equiv \theta\sqrt{L/\pi} > 0$. Then we can show that $H(y) \equiv 2 + y + \mathrm{e}^y(y-2) > 0$ for all $y > 0$. Obviously, $H(0) = 0$. So, if $H' > 0$ for all $y > 0$, the proof is complete. It is readily verified that $H' = 1 + y\mathrm{e}^y - \mathrm{e}^y > 0$ is equivalent to $\mathrm{e}^{-y} > 1 - y$, which is true for all y by convexity of e^{-y} (observe that $1 - y$ is the tangent to e^{-y} at $y = 0$ and that a convex function is everywhere above its tangent).

A.3. The mass of varieties consumed in the urban system. Using N_r^c as defined in (28), and the external cutoff and the mass of entrants as given by (27) and (33), and making use of the Pareto distribution, we obtain:

$$N_r^c = \frac{\kappa_2}{\kappa_1+\kappa_2}\left(m_r^d\right)^k \sum_s \frac{S_s}{F(m_s^{\max})^k}\left(\frac{\tau_{rr}}{\tau_{sr}}\frac{w_r}{w_s}\right)^k = \frac{\alpha}{\kappa_1+\kappa_2}\frac{\left(m_r^d\right)^k}{\tau_{rr}}\sum_s S_s\tau_{rr}\left(\frac{\tau_{rr}}{\tau_{sr}}\frac{w_r}{w_s}\right)^k \frac{\kappa_2}{\alpha F(m_s^{\max})^k}.$$

Using the definition of $\mu_s^{\max}$, and noting that the summation in the foregoing expression appears in the equilibrium relationship (34), we can then express the mass of varieties consumed in city r as given in (35).

A.4. The weighted average of markups in the urban system. Plugging (29) into the definition (36), the weighted average of markups in the open economy can be rewritten as

$$\overline{\Lambda}_r = \frac{1}{\alpha E_r \sum_s N_s^E G_s(m_{sr}^x)}\sum_s N_s^E \tau_{sr} w_s \int_0^{m_{sr}^x} m\left(W^{-2} - W^{-1}\right)\mathrm{d}G_s(m),$$

where the argument $\mathrm{e}m/m_{sr}^x$ of the Lambert W function is suppressed to alleviate notation. As shown in Appendix B.1, the integral term is given by $\kappa_3(m_s^{\max})^{-k}(m_{sr}^x)^{k+1} = \kappa_3 G_s(m_{sr}^x)m_{sr}^x$. Using this, together with (27) and $E_r = (S_r/L_r)w_r$, yields the expression in (36).

A.5. Indirect utility in the urban system. To derive the indirect utility, we first compute the (unweighted) average price across all varieties sold in each market. Multiplying both sides of (26) by $p_{rs}(i)$, integrating over Ω_{rs}, and summing the resulting expressions across r, we obtain:

$$\overline{p}_s \equiv \frac{1}{N_s^c}\sum_r \int_{\Omega_{rs}} p_{rs}(j)\mathrm{d}j = \frac{1}{N_s^c}\sum_r \tau_{rs}w_r\int_{\Omega_{rs}} m_r(j)\mathrm{d}j + \frac{\alpha E_s}{N_s^c},$$

where the first term is the average of marginal delivered costs. Under the Pareto distribution, $\int_{\Omega_{sr}} m_s(j)\mathrm{d}j = N_s^E \int_0^{m_{sr}^x} m\mathrm{d}G_s(m) = [k/(k+1)]m_{sr}^x N_s^E G_s(m_{sr}^x)$. Hence, the (unweighted) average price can be rewritten for city r as follows

$$\overline{p}_r = \frac{1}{N_r^c}\sum_s \tau_{sr}w_s\left(\frac{k}{k+1}\right)m_{sr}^x N_s^E G_s(m_{sr}^x) + \frac{\alpha E_r}{N_r^c} = \left(\frac{k}{k+1}\right)p_r^d + \frac{\alpha E_r}{N_r^c}, \tag{61}$$

where we have used (28) and $p_r^d = \tau_{sr}w_s m_{sr}^x$. Plugging (61) into (24) and using (27), the indirect utility is then given by

$$U_r = \frac{N_r^c}{k+1} - \frac{\alpha}{\tau_{rr}}\frac{S_r}{L_r m_r^d},$$

which together with (35) and (36) yields (37).

A.6. Some analytical results in the two-city case.

(i) *A reduction in θ from any given positive value to zero raises S_1/S_2.* In a world with urban frictions (where $\theta > 0$), and given that $\overline{h}_1 = \overline{h}_2 = \overline{h}$ and $\theta_1 = \theta_2 = \theta$, the term S_1/S_2 is given by

$$\frac{S_1}{S_2} = \frac{1 - \left(1 + \theta\sqrt{L_1/\pi}\right) \mathrm{e}^{-\theta\sqrt{L_1/\pi}}}{1 - \left(1 + \theta\sqrt{L_2/\pi}\right) \mathrm{e}^{-\theta\sqrt{L_2/\pi}}}. \tag{62}$$

In a world without urban frictions (where $\theta = 0$), we have $\widetilde{S}_1 = L_1\overline{h}$ and $\widetilde{S}_2 = L_2\overline{h}$, so that $\widetilde{S}_1/\widetilde{S}_2 = L_1/L_2$. Our aim is thus to prove that L_1/L_2 is larger than the term S_1/S_2 given in (62). Letting $y_r \equiv \theta\sqrt{L_r/\pi} > 0$, this is equivalent to proving that $y_1^2/\left(1 - \mathrm{e}^{-y_1} - y_1\mathrm{e}^{-y_1}\right) > y_2^2/\left(1 - \mathrm{e}^{-y_2} - y_2\mathrm{e}^{-y_2}\right)$. We thus need to show that $y^2/(1 - \mathrm{e}^{-y} - y\mathrm{e}^{-y})$ is increasing because $y_1 > y_2$. By differentiating, we have the derivative

$$\frac{y\mathrm{e}^{-y}}{\left(1 - \mathrm{e}^{-y} - y\mathrm{e}^{-y}\right)^2}Y, \quad \text{where} \quad Y \equiv 2\mathrm{e}^{y} - [(y+1)^2 + 1].$$

Noting that $Y = 0$ at $y = 0$ and $Y' = 2[\mathrm{e}^{y} - (y+1)] > 0$ for all $y > 0$, we know that the derivative is positive for all $y > 0$. Hence, $\widetilde{S}_1/\widetilde{S}_2 = L_1/L_2 > S_1/S_2$.

(ii) *$\widetilde{m}_1^d < m_1^d$ and $\widetilde{m}_2^d < m_2^d$ in the case without trade frictions.* Setting $\tau = t$, the market equilibrium conditions can be rewritten as

$$\frac{\mu_1^{\max}}{t} = L_1X_1 + L_2\frac{X_2}{\Omega} \tag{63}$$

$$\frac{\mu_2^{\max}}{t} = L_2X_2 + L_1\Omega X_1 \tag{64}$$

$$\Omega = \left(\frac{\rho}{\sigma}\frac{X_2}{X_1}\right)^{\frac{k+1}{2k+1}}, \tag{65}$$

where $X_1 \equiv (m_1^d)^{k+1}$, $X_2 \equiv (m_2^d)^{k+1}$, and $\Omega \equiv \omega^{k+1}$. From (63) and (64), we thus have $\Omega\frac{\mu_1^{\max}}{t} = \frac{\mu_2^{\max}}{t} = L_1\Omega X_1 + L_2X_2$. Hence, $\Omega = \rho$ must hold when $\tau = t$. We know by (65) that $X_2 = (\sigma/\rho)\Omega^{\frac{2k+1}{k+1}}X_1 = \sigma\rho^{\frac{k}{k+1}}X_1$. Plugging this expression into (63) yields the counterfactual cutoffs

$$\widetilde{X}_1 = (\widetilde{m}_1^d)^{k+1} = \frac{\mu_1^{\max}}{L_1t}\frac{1}{1 + \sigma\rho^{-\frac{1}{k+1}}(L_2/L_1)} \quad \text{and} \quad \widetilde{X}_2 = (\widetilde{m}_2^d)^{k+1} = \frac{\mu_2^{\max}}{L_2t}\frac{1}{1 + \sigma^{-1}\rho^{\frac{1}{k+1}}(L_1/L_2)}. \tag{66}$$

Establishing that $\widetilde{X}_1 < X_1$, i.e., that $\widetilde{m}_1^d < m_1^d$ requires

$$\begin{aligned}
&\frac{1 - \rho(t/\tau)^k\omega^{-(k+1)}}{1 - (t/\tau)^{2k}} > \frac{1}{1 + \sigma\rho^{-\frac{1}{k+1}}(L_2/L_1)} \\
\Rightarrow\quad & \sigma\rho^{-\frac{1}{k+1}}\left(\frac{L_2}{L_1}\right)\left[1 - \rho\left(\frac{t}{\tau}\right)^k\omega^{-(k+1)}\right] > \left(\frac{t}{\tau}\right)^k\left[\rho\omega^{-(k+1)} - \left(\frac{t}{\tau}\right)^k\right] \\
\Rightarrow\quad & \rho^{-\frac{1}{k+1}}\left(\frac{S_2}{S_1}\right)\omega^{-(k+1)}\left[\omega^{k+1} - \rho\left(\frac{t}{\tau}\right)^k\right] > \left(\frac{t}{\tau}\right)^k\omega^{-(k+1)}\left[\rho - \left(\frac{t}{\tau}\right)^k\omega^{k+1}\right] \\
\Rightarrow\quad & \rho\rho^{-\frac{1}{k+1}}\left(\frac{\tau}{t}\right)^k > \rho\left(\frac{S_1}{S_2}\right)\frac{\rho - (t/\tau)^k\omega^{k+1}}{\omega^{k+1} - \rho(t/\tau)^k} = \omega^k,
\end{aligned}$$

where the last equality holds by (44). We thus need to prove $\rho^{k/(k+1)}(\tau/t)^k > \omega^k$ or $\rho^{1/(k+1)}(\tau/t) > \omega$, which is straightforward since $\rho^{1/(k+1)}(\tau/t) > \rho^{1/(k+1)}(\tau/t)^{k/(k+1)} \equiv \overline{\omega} > \omega$. Hence, $\widetilde{m}_1^d < m_1^d$ must be true. Using a similar approach, it can be shown that $\widetilde{m}_2^d < m_2^d$ is also satisfied. The elimination of trade frictions thus leads to lower cutoffs in *both* cities.

(iii) $\widetilde{\Upsilon} < \Upsilon$ *for* $\rho^{1/(k+1)} \leq \sigma$. Let $\Delta m_r^d \equiv m_r^d - \widetilde{m}_r^d > 0$. Then, proving $h_1(1/\widetilde{m}_1^d - 1/m_1^d) < h_2(1/\widetilde{m}_2^d - 1/m_2^d)$ is equivalent to proving that

$$\frac{h_1 \Delta m_1^d}{m_1^d \widetilde{m}_1^d} < \frac{h_2 \Delta m_2^d}{m_2^d \widetilde{m}_2^d} \quad \Leftrightarrow \quad \frac{m_1^d \widetilde{m}_1^d \Delta m_2^d}{m_2^d \widetilde{m}_2^d \Delta m_1^d} \frac{h_2}{h_1} > 1. \tag{67}$$

This can be done by the following steps. First, we prove cutoff convergence, i.e., $\widetilde{m}_2^d/\widetilde{m}_1^d < m_2^d/m_1^d$. Using (66), the counterfactual cutoff ratio is given by $(\widetilde{m}_2^d/\widetilde{m}_1^d)^{k+1} = \sigma\rho^{k/(k+1)}$, whereas using (43), the cutoff ratio with trade frictions is

$$\left(\frac{m_2^d}{m_1^d}\right)^{k+1} = \frac{L_1}{L_2}\frac{1}{\omega^{-(k+1)}}\frac{\rho - (t/\tau)^k \omega^{k+1}}{\omega^{k+1} - \rho(t/\tau)^k} = \frac{L_1}{L_2}\frac{1}{\omega^{-(k+1)}}\frac{\omega^k}{\rho}\frac{S_2}{S_1} = \frac{\sigma}{\rho}\omega^{2k+1},$$

where we use (44) to obtain the second equality. Taking their difference, showing that $\widetilde{m}_2^d/\widetilde{m}_1^d < m_2^d/m_1^d$ boils down to showing that $\rho^{1/(k+1)} < \omega$ at the market equilibrium. This can be done by evaluating (44) at $\omega = \rho^{1/(k+1)}$. The LHS is equal to $\rho^{k/(k+1)}$, which falls short of the RHS given by $\rho S_1/S_2$ (because $\rho \geq 1$, $k \geq 1$, and $S_1/S_2 > 1$). Since the LHS is increasing and the RHS is decreasing, it must be that $\rho^{1/(k+1)} < \omega^*$. Thus, we have proved $\widetilde{m}_2^d/\widetilde{m}_1^d < m_2^d/m_1^d$. Turning to the second step, this cutoff convergence then implies

$$\frac{m_2^d}{m_1^d} > \frac{\widetilde{m}_2^d}{\widetilde{m}_1^d} \quad \Rightarrow \quad \frac{m_1^d}{m_2^d}\frac{\Delta m_2^d}{\Delta m_1^d} > 1 \quad \Rightarrow \quad \left(\frac{m_1^d}{m_2^d}\frac{\widetilde{m}_1^d}{\widetilde{m}_2^d}\frac{\Delta m_2^d}{\Delta m_1^d}\frac{h_2}{h_1}\right)\frac{\widetilde{m}_2^d}{\widetilde{m}_1^d}\frac{h_1}{h_2} > 1. \tag{68}$$

Recall from (67) that we ultimately we want to prove that $\left(\frac{m_1^d}{m_2^d}\frac{\widetilde{m}_1^d}{\widetilde{m}_2^d}\frac{\Delta m_2^d}{\Delta m_1^d}\frac{h_2}{h_1}\right) > 1$. A sufficient condition for this to be satisfied, given (68), is that $(\widetilde{m}_2^d/\widetilde{m}_1^d)(h_1/h_2) \leq 1$, i.e., that $[\sigma\rho^{k/(k+1)}]^{1/(k+1)}(1/\sigma) = [\rho^{1/(k+1)}/\sigma]^{k/(k+1)} \leq 1$. This is the case if $\rho^{1/(k+1)} \leq \sigma$.

Appendix B: Integrals involving the Lambert W function

To derive closed-form solutions for various expressions throughout the paper we need to compute integrals involving the Lambert W function. This can be done by using the change in variables suggested by Corless *et al.* (1996, p.341). Let

$$z \equiv W\left(\mathrm{e}\,\frac{m}{I}\right), \quad \text{so that} \quad \mathrm{e}\,\frac{m}{I} = z\mathrm{e}^z, \quad \text{where} \quad I = m_r^d, m_{rs}^x,$$

where subscript r can be dropped in the closed economy. The change in variables then yields $\mathrm{d}m = (1+z)\mathrm{e}^{z-1} I \mathrm{d}z$, with the new integration bounds given by 0 and 1. Under our assumption of a Pareto distribution for productivity draws, the change in variables allows to rewrite integrals in simplified form.

B.1. First, consider the following expression, which appears when integrating firms' outputs:

$$\int_0^I m\left[1 - W\left(\mathrm{e}\frac{m}{I}\right)\right]\mathrm{d}G_r(m) = \kappa_1\left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_1 \equiv k\mathrm{e}^{-(k+1)}\int_0^1 (1-z^2)\left(z\mathrm{e}^z\right)^k \mathrm{e}^z\mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k.

B.2. Second, the following expression appears when integrating firms' operating profits:

$$\int_0^I m\left[W\left(\mathrm{e}\frac{m}{I}\right)^{-1} + W\left(\mathrm{e}\frac{m}{I}\right) - 2\right]\mathrm{d}G_r(m) = \kappa_2\left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_2 \equiv k\mathrm{e}^{-(k+1)}\int_0^1 (1+z)\left(z^{-1}+z-2\right)\left(z\mathrm{e}^z\right)^k \mathrm{e}^z\mathrm{d}z > 0$ is also a constant term which solely depends on the shape parameter k.

B.3. Third, the following expression appears when deriving the weighted average of firm-level markups:

$$\int_0^I m\left[W\left(\mathrm{e}\,\frac{m}{I}\right)^{-2} - W\left(\mathrm{e}\,\frac{m}{I}\right)^{-1}\right]\mathrm{d}G_r(m) = \kappa_3\left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_3 \equiv k\mathrm{e}^{-(k+1)}\int_0^1 (z^{-2}-z^{-1})(1+z)(z\mathrm{e}^z)^k\mathrm{e}^z\mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k.

B.4. Finally, the following expression appears when integrating firms' revenues:

$$\int_0^I m\left[W\left(\mathrm{e}\,\frac{m}{I}\right)^{-1} - 1\right]\mathrm{d}G_r(m) = \kappa_4\left(m_r^{\max}\right)^{-k} I^{k+1},$$

where $\kappa_4 \equiv k\mathrm{e}^{-(1+k)}\int_0^1 (z^{-1}-z)\left(z\mathrm{e}^z\right)^k \mathrm{e}^z\mathrm{d}z > 0$ is a constant term which solely depends on the shape parameter k. Using the expressions for κ_1 and κ_2, one can verify that $\kappa_4 = \kappa_1 + \kappa_2$.

Appendix C: Equilibrium conditions in the urban system using the Lambert W function

By definition, the zero expected profit condition for each firm in city r is given by

$$\sum_s L_s \int_0^{m_{rs}^x} \left[p_{rs}(m) - \tau_{rs} m w_r\right] q_{rs}(m)\mathrm{d}G_r(m) = F w_r. \tag{69}$$

Furthermore, each labor market clears in equilibrium, which requires that

$$N_r^E\left[\sum_s L_s \tau_{rs}\int_0^{m_{rs}^x} m q_{rs}(m)\mathrm{d}G_r(m) + F\right] = S_r. \tag{70}$$

Last, in equilibrium trade must be balanced for each city

$$N_r^E \sum_{s\neq r} L_s \int_0^{m_{rs}^x} p_{rs}(m) q_{rs}(m)\mathrm{d}G_r(m) = L_r \sum_{s\neq r} N_s^E \int_0^{m_{sr}^x} p_{sr}(m) q_{sr}(m)\mathrm{d}G_s(m). \tag{71}$$

We now restate the foregoing conditions (69)–(71) in terms of the Lambert W function.

First, using (29), the labor market clearing condition can be rewritten as follows:

$$N_r^E \left\{ \frac{1}{\alpha} \sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[1 - W\left(\mathrm{e}\frac{m}{m_{rs}^x} \right) \right] \mathrm{d}G_r(m) + F \right\} = S_r. \tag{72}$$

Second, plugging (29) into (69), zero expected profits require that

$$\frac{1}{\alpha} \sum_s L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[W\left(\mathrm{e}\frac{m}{m_{rs}^x} \right)^{-1} + W\left(\mathrm{e}\frac{m}{m_{rs}^x} \right) - 2 \right] \mathrm{d}G_r(m) = F. \tag{73}$$

Last, the trade balance condition is given by

$$\begin{aligned} N_r^E w_r \sum_{s \neq r} L_s \tau_{rs} \int_0^{m_{rs}^x} m \left[W\left(\mathrm{e}\,\frac{m}{m_{rs}^x} \right)^{-1} - 1 \right] \mathrm{d}G_r(m) \\ = L_r \sum_{s \neq r} N_s^E \tau_{sr} w_s \int_0^{m_{sr}^x} m \left[W\left(\mathrm{e}\,\frac{m}{m_{sr}^x} \right)^{-1} - 1 \right] \mathrm{d}G_s(m). \end{aligned} \tag{74}$$

Applying the city-specific Pareto distribution $G_r(m) = (m/m_r^{\max})^k$ to (72)–(74) yields, using the results of Appendix B, expressions (30)–(32) given in the main text.

Appendix D: Data description

MSA-level data. We construct a dataset for 356 metropolitan statistical areas (see Table 1 for a full list of the MSAs). The bulk of our MSA-level data comes from the 2007 American Community Survey (ACS) of the US Census, from the Bureau of Economic Analysis (BEA) and from the Bureau of Labor Statistics (BLS). The geographical coordinates of each MSA are computed as the centroid of its constituent counties' geographical coordinates. The latter are obtained from the 2000 US Census Gazetteer county geography file, and the MSA-level aggregation is carried out using the county-to-MSA concordance tables for 2007. We then construct our measure of distance between two MSAs as $d_{rs} = \cos^{-1}\big(\sin(\mathrm{lat}_r)\sin(\mathrm{lat}_s) + \cos(|\mathrm{lon}_r - \mathrm{lon}_s|)\cos(\mathrm{lat}_r) \times \cos(\mathrm{lat}_s)\big) \times 6,378.137$ using the great circle formula, where lat_r and lon_r are the geographical coordinates of the MSA. The internal distance of an MSA is defined as $d_{rr} \equiv (2/3)\sqrt{\mathrm{surface}_r/\pi}$ as in Redding and Venables (2004). All MSA surface measures are given in square kilometers and include only land surface of the MSA's constitutent counties. That data is obtained from the 2000 US Census Gazetteer, and is aggregated from the county to the MSA level.

We further obtain total gross domestic product by MSA from the BEA metropolitan GDP files. Total employment at the MSA level is obtained from the 2007 BLS employment flat files (we use aggregate values for 'All occupations'). Using gross domestic product, total employment, and the average number of hours worked allows us to recover our measure of average MSA productivity (GDP per employee). Wages at the MSA level for 2007 are computed as total labor expenses (compensation of employees plus employer contributions for employee pension and insurance funds plus employer contributions for government social insurance) divided by total MSA employment. Data to compute total labor expenses is provided by the BEA.

Last, county-level data on natural amenities are from 1999 and provided by the US Department of Agriculture (USDA). The USDA data includes six measures of climate, topography, and water area that reflect environmental attributes usually valued by people. We use the standardized amenity score from that data as a proxy for our observed amenities. We aggregate the county-level amenities up to the MSA level by using the county-to-MSA concordance table and by weighting each county by its share in the total MSA land surface.

Urban frictions data. Total MSA population is taken from the 2007 ACS. The 2007 ACS further provides MSA-level data on average weekly hours worked and on average (one-way) commuting time in minutes. Both pieces of information are used to compute the internal cutoffs m_r^d, the aggregate labor supply $\overline{h}_r L_r$, and the effective labor supply S_r.

Trade frictions data. We estimate a gravity equation on state-to-state trade flows to obtain an estimate of the distance elasticity γ. To this end, we use aggregate bilateral trade flows X_{rs} from the 2007 Commodity Flow Survey (CFS) of the Bureau of Transportation Statistics (BTS) for the lower 48 contiguous US states, as these are the states containing the MSAs that will be used in our analysis. We work at the state level since MSA trade flows from the CFS public files can only be meaningfully exploited for a relatively small sample of large 'CFS regions'. Duranton *et al.* (2011, p.10), for example, work with that data to estimate the distance elasticity of trade flows. We ran several robustness checks using their estimate of γ instead of ours. Results are little sensitive to that choice. As to the specification of trade costs τ_{rs} we stick to standard practice and assume that $\tau_{rs} \equiv d_{rs}^{\gamma}$, where d_{rs} stands for the distance between r and s in kilometers computed using the great circle formula given above.[35] In that case, lat_r and lon_r denote the coordinates of the capital of state r, measured in radians, which are taken from Anderson and van Wincoop's (2003) dataset.

[35] Using CFS trade data, Duranton *et al.* (2011) show that the distance elasticity of trade within the US is basically insensitive to how distance is exactly measured (euclidian distance vs. various distance measures based on current or historical highway grids).

Table 1: MSA variables and descriptives for the initial equilibrium

FIPS	MSA name	State	$L_r/\overline{L}$	$\widehat{\mu}_r^{\max}$	$1/\overline{m}_r$	$\widehat{\theta}_r$	A_r^o	$\widehat{A}_r^u$
10180	Abilene	TX	0.2268	6.8852	0.8328	0.3925	1.3141	-0.6556
10420	Akron	OH	0.9956	17.4352	0.8212	0.2473	-2.2749	1.0062
10500	Albany	GA	0.2336	28.3000	0.7182	0.4608	-0.0435	-0.4451
10580	Albany-Schenectady-Troy	NY	1.2149	15.6558	0.8722	0.2015	-0.2432	1.1317
10740	Albuquerque	NM	1.1889	11.6475	0.8694	0.2232	3.7322	0.9275
10780	Alexandria	LA	0.2133	14.7747	0.7632	0.5445	-0.2067	-0.5842
10900	Allentown-Bethlehem-Easton	PA-NJ	1.1444	22.9469	0.8678	0.3088	0.3026	0.9760
11020	Altoona	PA	0.1787	28.9660	0.6877	0.5223	-0.8600	-0.7009
11100	Amarillo	TX	0.3449	7.1209	0.8305	0.3277	1.6304	-0.2289
11180	Ames	IA	0.1207	0.7978	0.9817	0.6556	-3.5400	-1.1175
11300	Anderson	IN	0.1869	6.1621	0.8247	0.8718	-3.4700	-0.6463
11340	Anderson	SC	0.2562	16.3593	0.7543	0.5571	0.7100	-0.4872
11460	Ann Arbor	MI	0.4983	2.9986	0.9738	0.2977	-2.1900	0.1721
11500	Anniston-Oxford	AL	0.1610	13.1516	0.7430	0.5613	0.2200	-0.9536
11540	Appleton	WI	0.3104	9.1579	0.7999	0.3684	-2.7304	-0.0904
11700	Asheville	NC	0.5756	31.3698	0.7609	0.3163	2.1012	0.2978
12020	Athens-Clarke County	GA	0.2668	15.4460	0.7858	0.4865	-1.0511	-0.3069
12060	Atlanta-Sandy Springs-Marietta	GA	7.5152	7.9312	1.0828	0.1174	0.2253	2.7880
12100	Atlantic City-Hammonton	NJ	0.3853	4.3460	0.9247	0.3301	-0.0400	-0.2364
12220	Auburn-Opelika	AL	0.1858	14.1079	0.7298	0.6358	-0.2400	-0.7240
12260	Augusta-Richmond County	GA-SC	0.7524	23.6409	0.8053	0.2920	-0.0192	0.6829
12420	Austin-Round Rock	TX	2.2752	5.6156	0.9979	0.1860	1.6141	1.5231
12540	Bakersfield	CA	1.1257	8.3291	0.9841	0.2453	4.8400	0.6741
12580	Baltimore-Towson	MD	3.7983	12.0935	0.9856	0.1519	-0.3557	2.1378
12620	Bangor	ME	0.2118	5.6207	0.8107	0.5506	-0.5200	-0.5302
12700	Barnstable Town	MA	0.3163	2.9345	0.8556	0.4759	1.5200	-0.4993
12940	Baton Rouge	LA	1.0962	3.7242	1.0012	0.2569	-0.6186	0.9311
12980	Battle Creek	MI	0.1945	7.2642	0.8301	0.4982	-2.7300	-0.6453
13020	Bay City	MI	0.1531	6.5755	0.7780	0.7995	-1.5300	-0.9167
13140	Beaumont-Port Arthur	TX	0.5356	8.3601	0.8672	0.2801	0.9407	0.1728
13380	Bellingham	WA	0.2748	1.1589	0.9747	0.4955	5.2600	-0.7955
13460	Bend	OR	0.2193	2.3869	0.8996	0.4620	6.1000	-1.0336
13740	Billings	MT	0.2131	7.1640	0.7761	0.3735	2.4532	-0.6830
13780	Binghamton	NY	0.3508	56.9535	0.6866	0.3785	-0.9289	0.0588
13820	Birmingham-Hoover	AL	1.5777	5.8973	1.0014	0.2055	0.5780	1.2351
13900	Bismarck	ND	0.1470	12.2467	0.7085	0.4403	-1.6258	-0.7564
13980	Blacksburg-Christiansburg-Radford	VA	0.2244	10.1677	0.8144	0.5208	0.5141	-0.5979
14020	Bloomington	IN	0.2616	14.7889	0.8140	0.5467	-0.4507	-0.3408
14060	Bloomington-Normal	IL	0.2338	2.4247	0.9891	0.3871	-3.5700	-0.4375
14260	Boise City-Nampa	ID	0.8367	10.6193	0.8491	0.2399	2.2919	0.6976
14460	Boston-Cambridge-Quincy	MA-NH	6.3819	2.7007	1.1870	0.1098	0.1444	2.4955
14500	Boulder	CO	0.4132	0.6188	1.1168	0.3373	5.8200	-0.6755
14540	Bowling Green	KY	0.1651	12.3177	0.7702	0.5611	-0.2160	-0.8510
14740	Bremerton-Silverdale	WA	0.3370	1.2068	1.0491	0.7249	2.6100	-0.6981
14860	Bridgeport-Stamford-Norwalk	CT	1.2742	0.0329	1.8325	0.2506	2.2500	-0.2081
15180	Brownsville-Harlingen	TX	0.5512	55.3719	0.5912	0.3178	2.4600	0.3482
15260	Brunswick	GA	0.1449	13.3594	0.7523	0.6313	1.3530	-1.0593
15380	Buffalo-Niagara Falls	NY	1.6061	15.4178	0.8225	0.1730	-0.6399	1.4505
15500	Burlington	NC	0.2069	16.5166	0.7377	0.6324	-0.9600	-0.6176
15540	Burlington-South Burlington	VT	0.2952	2.2778	0.9027	0.4271	-0.1238	-0.3845
15940	Canton-Massillon	OH	0.5797	27.4059	0.7541	0.3382	-1.4796	0.4955
15980	Cape Coral-Fort Myers	FL	0.8407	2.0378	0.9635	0.3210	5.2300	0.1676
16220	Casper	WY	0.1021	0.0797	1.3629	0.4917	2.4900	-1.9697
16300	Cedar Rapids	IA	0.3599	6.3374	0.8708	0.3126	-3.3035	0.0590
16580	Champaign-Urbana	IL	0.3145	14.7922	0.8363	0.3848	-4.3383	0.0884
16620	Charleston	WV	0.4327	6.2623	0.9251	0.3322	-0.7294	0.0286
16700	Charleston-North Charleston-Summerville	SC	0.8970	8.8536	0.8690	0.2777	0.5686	0.7409
16740	Charlotte-Gastonia-Concord	NC-SC	2.3512	0.6377	1.3186	0.1561	0.1000	1.3196
16820	Charlottesville	VA	0.2744	7.2636	0.9001	0.4341	-0.0364	-0.4526
16860	Chattanooga	TN-GA	0.7326	8.8814	0.8897	0.2830	0.2832	0.5342
16940	Cheyenne	WY	0.1229	2.1311	0.9176	0.5112	3.0500	-1.4960
16980	Chicago-Naperville-Joliet	IL-IN-WI	13.5596	7.6522	1.1400	0.0867	-2.1021	3.4958
17020	Chico	CA	0.3115	5.1269	0.8541	0.5341	5.1100	-0.5608
17140	Cincinnati-Middletown	OH-KY-IN	3.0376	14.2620	0.9455	0.1438	-0.7916	2.0448
17300	Clarksville	TN-KY	0.3727	1.4179	1.0663	0.5319	0.0733	-0.3729
17420	Cleveland	TN	0.1582	3.0055	0.9115	0.7279	0.8781	-1.1302
17460	Cleveland-Elyria-Mentor	OH	2.9846	7.3233	0.9836	0.1352	-1.4310	1.9676
17660	Coeur d'Alene	ID	0.1914	8.3418	0.7161	0.6066	3.5000	-0.9011
17780	College Station-Bryan	TX	0.2895	47.5407	0.7123	0.4095	0.8622	-0.2296
17820	Colorado Springs	CO	0.8671	7.0613	0.8860	0.2838	5.3867	0.3780
17860	Columbia	MO	0.2311	16.7125	0.7364	0.4196	0.1054	-0.4706
17900	Columbia	SC	1.0194	22.2288	0.8323	0.2385	0.5017	0.9371
17980	Columbus	GA-AL	0.4025	8.7851	0.8541	0.3100	-0.2353	-0.0490
18020	Columbus	IN	0.1064	2.9595	0.8788	0.4856	-2.3800	-1.3775
18140	Columbus	OH	2.4975	11.5892	0.9535	0.1398	-1.9162	1.8984
18580	Corpus Christi	TX	0.5899	5.0627	0.8543	0.2746	2.8551	0.1577

Table 1: MSA variables and descriptives for the initial equilibrium

FIPS	MSA name	State	$L_r/\overline{L}$	$\widehat{\mu}_r^{\max}$	$1/\overline{m}_r$	$\widehat{\theta}_r$	A_r^o	$\widehat{A}_r^u$
18700	Corvallis	OR	0.1159	0.1014	1.2152	0.7211	3.1000	-1.8133
19060	Cumberland	MD-WV	0.1414	56.7425	0.6576	0.7389	1.0076	-0.9889
19100	Dallas-Fort Worth-Arlington	TX	8.7483	3.2987	1.2029	0.0923	0.6857	2.8079
19140	Dalton	GA	0.1908	15.8567	0.7386	0.3339	0.4652	-0.8035
19180	Danville	IL	0.1156	13.3585	0.7769	0.7748	-3.2100	-1.0515
19260	Danville	VA	0.1506	34.1566	0.7025	0.6804	-0.3000	-0.8908
19340	Davenport-Moline-Rock Island	IA-IL	0.5355	8.2798	0.8791	0.2759	-2.6893	0.4377
19380	Dayton	OH	1.1895	14.1872	0.8640	0.1988	-2.1260	1.1962
19460	Decatur	AL	0.2125	3.5335	0.9214	0.6612	0.7910	-0.8247
19500	Decatur	IL	0.1548	2.7975	0.8839	0.4092	-2.7900	-0.9344
19660	Deltona-Daytona Beach-Ormond Beach	FL	0.7124	22.2777	0.7462	0.3743	3.4500	0.3884
19740	Denver-Aurora	CO	3.4326	2.2957	1.1516	0.1477	4.1942	1.7018
19780	Des Moines-West Des Moines	IA	0.7782	2.2274	1.0158	0.2050	-2.0346	0.6429
19820	Detroit-Warren-Livonia	MI	6.3602	8.3299	1.0380	0.1089	-1.6704	2.7501
20020	Dothan	AL	0.1986	49.5100	0.6561	0.4212	-0.4149	-0.5370
20100	Dover	DE	0.2168	1.9540	1.0020	0.5895	-0.0700	-0.8842
20220	Dubuque	IA	0.1315	5.7814	0.7869	0.3977	-0.7900	-1.1171
20260	Duluth	MN-WI	0.3905	18.6402	0.7996	0.3678	-0.8127	0.1938
20500	Durham	NC	0.6828	0.8200	1.1939	0.2552	0.0966	0.1845
20740	Eau Claire	WI	0.2247	12.7566	0.7611	0.4796	-2.6695	-0.3365
20940	El Centro	CA	0.2304	19.7182	0.7872	0.4081	6.4500	-0.8598
21060	Elizabethtown	KY	0.1589	3.7636	0.8891	0.5914	-0.8465	-1.0560
21140	Elkhart-Goshen	IN	0.2818	9.4337	0.7923	0.2901	-2.7200	-0.2450
21300	Elmira	NY	0.1253	16.7836	0.7000	0.6243	-1.1300	-1.0690
21340	El Paso	TX	1.0459	2.2083	0.9271	0.2441	4.4600	0.5021
21500	Erie	PA	0.3973	18.7253	0.7395	0.3204	-0.5700	0.0764
21660	Eugene-Springfield	OR	0.4891	13.2218	0.7821	0.3197	4.2900	0.0543
21780	Evansville	IN-KY	0.4979	8.0962	0.8860	0.2898	-1.6375	0.2844
22020	Fargo	ND-MN	0.2739	4.1400	0.8364	0.3067	-4.5908	-0.0388
22140	Farmington	NM	0.1743	0.2874	1.2203	0.5778	2.8300	-1.3307
22180	Fayetteville	NC	0.4968	0.7242	1.1132	0.3601	-0.9161	-0.1293
22220	Fayetteville-Springdale-Rogers	AR-MO	0.6203	13.9314	0.8230	0.2715	0.8552	0.4160
22380	Flagstaff	AZ	0.1814	41.4362	0.7797	0.4704	4.9300	-0.8937
22420	Flint	MI	0.6189	11.2936	0.8235	0.4086	-1.9000	0.4963
22500	Florence	SC	0.2829	14.4850	0.7801	0.4358	-0.2137	-0.3219
22520	Florence-Muscle Shoals	AL	0.2038	22.0682	0.7281	0.6420	0.8059	-0.6681
22540	Fond du Lac	WI	0.1411	5.1570	0.8386	0.6231	-1.9200	-1.0104
22660	Fort Collins-Loveland	CO	0.4094	9.8391	0.8295	0.3890	5.6200	-0.3039
22900	Fort Smith	AR-OK	0.4124	21.2879	0.7892	0.3342	1.6228	-0.0124
23020	Fort Walton Beach-Crestview-Destin	FL	0.2584	0.3985	1.1155	0.4967	2.0100	-0.9455
23060	Fort Wayne	IN	0.5838	20.3049	0.7882	0.2692	-3.0754	0.5929
23420	Fresno	CA	1.2803	22.9506	0.8468	0.2171	6.0300	0.8406
23460	Gadsden	AL	0.1469	27.7629	0.6669	0.7121	0.9600	-1.0397
23540	Gainesville	FL	0.3660	7.8664	0.8210	0.3731	2.0892	-0.2095
23580	Gainesville	GA	0.2565	4.7162	0.8383	0.6287	0.9600	-0.6703
24020	Glens Falls	NY	0.1835	53.2073	0.6769	0.6495	-0.3136	-0.6305
24140	Goldsboro	NC	0.1617	4.7743	0.8234	0.6350	-1.4100	-0.9470
24220	Grand Forks	ND-MN	0.1391	7.5933	0.7678	0.4540	-4.2873	-0.6426
24300	Grand Junction	CO	0.1980	14.4225	0.7324	0.5205	2.2600	-0.7599
24340	Grand Rapids-Wyoming	MI	1.1058	14.8202	0.8746	0.2091	-2.1226	1.1623
24500	Great Falls	MT	0.1164	3.0799	0.7954	0.5633	2.2000	-1.3183
24540	Greeley	CO	0.3470	11.1165	0.8543	0.6195	1.7000	-0.2422
24580	Green Bay	WI	0.4287	7.7067	0.8387	0.2912	-1.3945	0.1489
24660	Greensboro-High Point	NC	0.9944	12.2863	0.8764	0.2038	-0.2512	0.8794
24780	Greenville	NC	0.2455	8.4053	0.8048	0.4570	-1.9108	-0.3848
24860	Greenville-Mauldin-Easley	SC	0.8739	29.0690	0.7805	0.2293	1.3467	0.7392
25060	Gulfport-Biloxi	MS	0.3296	3.7705	0.8944	0.4062	0.1310	-0.3076
25180	Hagerstown-Martinsburg	MD-WV	0.3718	29.3045	0.7547	0.6204	0.3042	-0.0839
25260	Hanford-Corcoran	CA	0.2119	4.4956	0.8817	0.5882	3.4800	-0.9992
25420	Harrisburg-Carlisle	PA	0.7529	15.7008	0.8614	0.2220	-0.0004	0.5819
25500	Harrisonburg	VA	0.1674	3.5773	0.9210	0.4938	1.2500	-1.0739
25540	Hartford-West Hartford-East Hartford	CT	1.6929	0.6312	1.3157	0.1934	1.4760	0.8809
25620	Hattiesburg	MS	0.1967	14.5668	0.7576	0.6026	-0.2014	-0.6437
25860	Hickory-Lenoir-Morganton	NC	0.5132	43.2249	0.7227	0.3150	1.5055	0.2302
25980	Hinesville-Fort Stewart	GA	0.1022	0.0097	1.7152	1.4824	0.8063	-2.4818
26100	Holland-Grand Haven	MI	0.3690	4.6934	0.8693	0.4246	-0.0400	-0.1742
26300	Hot Springs	AR	0.1372	11.9767	0.7219	0.7581	1.6400	-1.1335
26380	Houma-Bayou Cane-Thibodaux	LA	0.2863	2.3685	0.9718	0.4086	0.3192	-0.5579
26420	Houston-Sugar Land-Baytown	TX	8.0123	0.7875	1.4273	0.1036	0.8426	2.4951
26580	Huntington-Ashland	WV-KY-OH	0.4043	18.9859	0.7879	0.3638	-0.1699	0.0365
26620	Huntsville	AL	0.5504	4.8277	0.9105	0.2864	-0.9066	0.2760
26820	Idaho Falls	ID	0.1700	14.9270	0.6994	0.6242	1.7783	-0.8152
26900	Indianapolis-Carmel	IN	2.4131	6.4117	1.0203	0.1453	-2.5367	1.8239
26980	Iowa City	IA	0.2093	3.0028	0.9098	0.4185	-2.9476	-0.5311
27060	Ithaca	NY	0.1439	7.6229	0.7882	0.5491	-0.2800	-0.9925
27100	Jackson	MI	0.2321	5.6531	0.8683	0.6124	-2.4500	-0.4931
27140	Jackson	MS	0.7603	9.3264	0.8735	0.2701	-0.6024	0.6792

Table 1: MSA variables and descriptives for the initial equilibrium

FIPS	MSA name	State	$L_r/\overline{L}$	$\widehat{\mu}_r^{\max}$	$1/\overline{m}_r$	$\widehat{\theta}_r$	A_r^o	$\widehat{A}_r^u$
27180	Jackson	TN	0.1604	8.0248	0.7820	0.4913	-1.6345	-0.8225
27260	Jacksonville	FL	1.8519	6.0828	0.9489	0.1930	2.0244	1.3020
27340	Jacksonville	NC	0.2317	0.1526	1.2201	0.6158	0.7400	-1.3510
27500	Janesville	WI	0.2272	17.1165	0.7514	0.5567	-2.6200	-0.3910
27620	Jefferson City	MO	0.2074	21.2752	0.7585	0.4518	0.3296	-0.5943
27740	Johnson City	TN	0.2755	15.4626	0.7613	0.4448	1.5055	-0.4559
27780	Johnstown	PA	0.2064	47.5556	0.6679	0.5599	-0.2300	-0.5483
27860	Jonesboro	AR	0.1657	19.0537	0.7332	0.4910	-2.2503	-0.6718
27900	Joplin	MO	0.2438	33.7469	0.6737	0.4025	-1.3200	-0.2872
28020	Kalamazoo-Portage	MI	0.4602	10.9030	0.8445	0.3422	-1.3239	0.2034
28100	Kankakee-Bradley	IL	0.1576	66.9572	0.6773	0.7130	-3.3000	-0.6326
28140	Kansas City	MO-KS	2.8265	9.2978	0.9719	0.1388	-1.3222	2.0201
28420	Kennewick-Pasco-Richland	WA	0.3260	1.7999	0.9386	0.4454	0.7491	-0.3261
28660	Killeen-Temple-Fort Hood	TX	0.5268	2.1655	1.0220	0.3488	1.5578	-0.0822
28700	Kingsport-Bristol-Bristol	TN-VA	0.4323	20.7011	0.7895	0.3835	0.3622	0.0800
28740	Kingston	NY	0.2589	38.4944	0.7621	0.7757	0.7000	-0.4394
28940	Knoxville	TN	0.9702	10.7076	0.8633	0.2284	1.0960	0.7774
29020	Kokomo	IN	0.1421	4.4454	0.8611	0.4794	-4.4522	-0.9032
29100	La Crosse	WI-MN	0.1864	15.4794	0.7197	0.4276	-1.1484	-0.6119
29140	Lafayette	IN	0.2736	6.6786	0.8963	0.4269	-3.4119	-0.2047
29180	Lafayette	LA	0.3652	0.3936	1.1340	0.3333	-0.9092	-0.4845
29340	Lake Charles	LA	0.2732	0.2160	1.2988	0.4158	0.1230	-0.8452
29460	Lakeland-Winter Haven	FL	0.8182	41.3451	0.7338	0.3320	3.9800	0.5254
29540	Lancaster	PA	0.7096	23.6630	0.8138	0.2773	0.4500	0.4974
29620	Lansing-East Lansing	MI	0.6498	8.5097	0.9034	0.3102	-3.3358	0.6664
29700	Laredo	TX	0.3319	40.7539	0.6586	0.3942	1.1200	-0.0710
29740	Las Cruces	NM	0.2830	14.1950	0.7658	0.4945	4.7700	-0.5204
29820	Las Vegas-Paradise	NV	2.6143	5.7538	0.9982	0.1449	4.8600	1.4990
29940	Lawrence	KS	0.1616	9.0883	0.7461	0.6893	0.3600	-0.9008
30020	Lawton	OK	0.1620	1.7247	0.9186	0.4717	2.2900	-1.2620
30140	Lebanon	PA	0.1821	21.6701	0.7301	0.6784	-0.6600	-0.7918
30340	Lewiston-Auburn	ME	0.1521	6.7201	0.7348	0.6650	-0.3200	-0.9631
30460	Lexington-Fayette	KY	0.6366	7.4339	0.8874	0.2408	-2.0342	0.5128
30620	Lima	OH	0.1498	6.3170	0.7978	0.4620	-2.3700	-0.9154
30700	Lincoln	NE	0.4160	6.3780	0.8194	0.2917	-2.8183	0.2242
30780	Little Rock-North Little Rock-Conway	AR	0.9487	8.6504	0.8992	0.2235	-0.0673	0.8521
30860	Logan	UT-ID	0.1724	17.5016	0.6920	0.6184	2.2845	-0.8079
30980	Longview	TX	0.2899	3.1890	0.9405	0.4235	1.0970	-0.5565
31020	Longview	WA	0.1430	5.9983	0.8127	0.8130	4.5400	-1.3338
31100	Los Angeles-Long Beach-Santa Ana	CA	18.3301	4.3306	1.2309	0.0708	10.0712	2.8862
31140	Louisville/Jefferson County	KY-IN	1.7564	14.2754	0.9145	0.1752	-0.7687	1.5113
31180	Lubbock	TX	0.3804	12.8002	0.7377	0.3094	1.7950	-0.0905
31340	Lynchburg	VA	0.3468	21.0406	0.7998	0.4312	0.4764	-0.1345
31420	Macon	GA	0.3272	31.5646	0.7452	0.3784	0.9051	-0.1751
31460	Madera	CA	0.2086	6.7275	0.8891	0.8123	6.0000	-1.0943
31540	Madison	WI	0.7910	4.1702	0.9806	0.2343	-0.4945	0.6170
31700	Manchester-Nashua	NH	0.5727	0.1167	1.4554	0.5151	0.0700	-0.3611
31900	Mansfield	OH	0.1789	33.4517	0.6730	0.4979	-2.8800	-0.5658
32580	McAllen-Edinburg-Mission	TX	1.0115	78.4494	0.6015	0.2479	0.4600	1.0886
32780	Medford	OR	0.2837	7.3664	0.7742	0.3762	4.5000	-0.5412
32820	Memphis	TN-MS-AR	1.8230	5.5326	0.9880	0.1653	-0.7140	1.4824
32900	Merced	CA	0.3495	3.4046	0.9806	0.6661	4.5100	-0.5673
33100	Miami-Fort Lauderdale-Pompano Beach	FL	7.7064	5.1829	1.0756	0.1063	5.2315	2.4562
33140	Michigan City-La Porte	IN	0.1563	21.9162	0.7391	0.6279	-1.8700	-0.8200
33260	Midland	TX	0.1800	0.0677	1.2915	0.3498	1.4200	-1.5392
33340	Milwaukee-Waukesha-West Allis	WI	2.1987	5.9256	0.9583	0.1410	-1.7072	1.6745
33460	Minneapolis-St. Paul-Bloomington	MN-WI	4.5673	4.2763	1.0673	0.1133	-2.1830	2.4717
33540	Missoula	MT	0.1504	2.8725	0.8180	0.4512	1.7400	-1.0344
33660	Mobile	AL	0.5757	9.1311	0.8016	0.3067	1.5200	0.2423
33700	Modesto	CA	0.7278	6.4113	0.9156	0.4128	7.2100	0.0268
33740	Monroe	LA	0.2453	9.2380	0.7899	0.4184	0.3390	-0.5074
33780	Monroe	MI	0.2187	2.0031	0.9750	0.9408	-1.4300	-0.7490
33860	Montgomery	AL	0.5210	12.6484	0.8354	0.3087	0.4625	0.2498
34060	Morgantown	WV	0.1677	4.0622	0.9172	0.6007	-0.5645	-0.9222
34100	Morristown	TN	0.1916	17.5432	0.7285	0.6252	1.4428	-0.8147
34580	Mount Vernon-Anacortes	WA	0.1657	0.7668	1.0340	0.7719	4.9400	-1.4000
34620	Muncie	IN	0.1643	21.3999	0.7009	0.5363	-2.6000	-0.6699
34740	Muskegon-Norton Shores	MI	0.2483	10.5424	0.7619	0.4962	-0.4000	-0.4569
34820	Myrtle Beach-North Myrtle Beach-Conway	SC	0.3558	14.1273	0.7514	0.3492	0.8800	-0.1685
34900	Napa	CA	0.1887	0.7977	1.1158	0.6025	7.5300	-1.5827
34940	Naples-Marco Island	FL	0.4496	0.8553	1.0987	0.3608	5.0000	-0.4961
34980	Nashville-Davidson–Murfreesboro–Franklin	TN	2.1660	8.8103	0.9775	0.1761	-0.8913	1.6814
35300	New Haven-Milford	CT	1.2037	0.3565	1.3393	0.3373	2.5200	0.3149
35380	New Orleans-Metairie-Kenner	LA	1.4669	0.3827	1.3139	0.1997	0.3337	0.8483
35620	New York-Northern New Jersey-Long Island	NY-NJ-PA	26.7870	2.3289	1.4318	0.0708	0.7740	3.7219
35660	Niles-Benton Harbor	MI	0.2272	4.2225	0.8899	0.4910	-0.3000	-0.7112
35980	Norwich-New London	CT	0.3806	2.5282	0.9939	0.3834	2.4300	-0.4626

Table 1: MSA variables and descriptives for the initial equilibrium

FIPS	MSA name	State	$L_r/\overline{L}$	$\widehat{\mu}_r^{\max}$	$1/\overline{m}_r$	$\widehat{\theta}_r$	A_r^o	$\widehat{A}_r^u$
36100	Ocala	FL	0.4625	26.5691	0.7385	0.4508	2.5900	0.0392
36140	Ocean City	NJ	0.1373	1.0674	0.9729	0.6085	0.0700	-1.4334
36220	Odessa	TX	0.1845	1.7012	0.8694	0.4434	2.5000	-1.1410
36260	Ogden-Clearfield	UT	0.7379	7.3733	0.8296	0.3433	4.0883	0.3479
36420	Oklahoma City	OK	1.6984	8.9525	0.9256	0.1702	0.1199	1.4212
36500	Olympia	WA	0.3396	2.6762	0.8761	0.5266	3.3200	-0.5078
36540	Omaha-Council Bluffs	NE-IA	1.1815	4.6939	0.9594	0.1726	-1.6836	1.1351
36740	Orlando-Kissimmee	FL	2.8935	9.3348	0.9478	0.1484	3.6792	1.6530
36780	Oshkosh-Neenah	WI	0.2308	3.4099	0.8448	0.3631	-1.3700	-0.5731
36980	Owensboro	KY	0.1596	5.0431	0.8563	0.4904	-0.9396	-0.9497
37100	Oxnard-Thousand Oaks-Ventura	CA	1.1366	1.0892	1.1665	0.3101	11.1700	-0.0195
37340	Palm Bay-Melbourne-Titusville	FL	0.7633	7.0268	0.8433	0.3242	3.9300	0.3194
37460	Panama City-Lynn Haven	FL	0.2335	3.9684	0.8128	0.4859	2.1500	-0.7925
37620	Parkersburg-Marietta-Vienna	WV-OH	0.2287	20.4051	0.7635	0.4824	-0.0229	-0.5302
37700	Pascagoula	MS	0.2164	3.3176	0.8870	0.6623	0.1912	-0.7469
37860	Pensacola-Ferry Pass-Brent	FL	0.6455	10.5757	0.8059	0.3574	2.0978	0.3456
37900	Peoria	IL	0.5285	6.0365	0.9428	0.2890	-2.5036	0.3764
37980	Philadelphia-Camden-Wilmington	PA-NJ-DE-MD	8.2969	5.0519	1.1876	0.1023	-0.6748	2.8345
38060	Phoenix-Mesa-Scottsdale	AZ	5.9500	13.0025	0.9713	0.1114	4.3136	2.4388
38220	Pine Bluff	AR	0.1445	18.4953	0.7485	0.5508	-1.2731	-0.8725
38300	Pittsburgh	PA	3.3537	10.5364	0.9970	0.1425	0.4012	2.0415
38340	Pittsfield	MA	0.1848	0.0590	1.5480	0.7997	0.8100	-1.5454
38540	Pocatello	ID	0.1247	18.4792	0.6806	0.5365	1.9030	-1.1149
38860	Portland-South Portland-Biddeford	ME	0.7305	0.3729	1.2367	0.3868	0.9595	0.1744
38900	Portland-Vancouver-Beaverton	OR-WA	3.0966	2.5795	1.0900	0.1534	2.8130	1.7475
38940	Port St. Lucie	FL	0.5696	4.4925	0.8792	0.4656	5.1827	-0.0890
39100	Poughkeepsie-Newburgh-Middletown	NY	0.9537	57.5790	0.7869	0.3958	0.0107	0.8914
39140	Prescott	AZ	0.3027	55.8791	0.7200	0.5665	5.2100	-0.4084
39300	Providence-New Bedford-Fall River	RI-MA	2.2790	1.8282	1.1372	0.2242	1.2849	1.3694
39340	Provo-Orem	UT	0.7023	15.6423	0.8210	0.3378	3.0296	0.5132
39380	Pueblo	CO	0.2200	33.0571	0.6806	0.5804	2.1100	-0.5738
39460	Punta Gorda	FL	0.2176	4.7904	0.8279	0.6776	5.1000	-1.0319
39540	Racine	WI	0.2777	2.6053	0.9046	0.5556	-0.5100	-0.5717
39580	Raleigh-Cary	NC	1.4914	4.1913	0.9997	0.2143	-0.6762	1.1883
39660	Rapid City	SD	0.1712	10.5487	0.7744	0.4558	-0.3579	-0.7024
39740	Reading	PA	0.5722	12.9659	0.8697	0.3670	-0.7300	0.2974
39820	Redding	CA	0.2554	5.9179	0.8368	0.4672	5.6900	-0.7588
39900	Reno-Sparks	NV	0.5841	6.1702	0.9153	0.2685	6.7038	-0.0559
40060	Richmond	VA	1.7268	11.1761	0.9742	0.1846	-0.9568	1.4730
40140	Riverside-San Bernardino-Ontario	CA	5.8104	104.4265	0.8632	0.1695	4.3817	2.5456
40220	Roanoke	VA	0.4222	22.5390	0.7805	0.3012	0.9380	0.0199
40340	Rochester	MN	0.2578	7.1786	0.8243	0.3375	-3.3458	-0.2406
40380	Rochester	NY	1.4670	9.7948	0.9057	0.1746	-0.6948	1.3292
40420	Rockford	IL	0.5015	16.7848	0.7779	0.3553	-2.7901	0.3797
40580	Rocky Mount	NC	0.2073	6.0239	0.8554	0.4688	-1.7475	-0.6464
40660	Rome	GA	0.1361	17.3345	0.7232	0.6475	0.3300	-1.0785
40900	Sacramento–Arden-Arcade–Roseville	CA	2.9770	4.8303	1.0444	0.1708	5.4091	1.5526
40980	Saginaw-Saginaw Township North	MI	0.2880	16.5948	0.7583	0.3910	-3.3300	-0.0839
41060	St. Cloud	MN	0.2642	12.5971	0.7626	0.4347	-3.0004	-0.1386
41100	St. George	UT	0.1905	23.2639	0.6948	0.4957	2.5700	-0.7385
41140	St. Joseph	MO-KS	0.1756	10.6024	0.7922	0.5409	-1.4641	-0.7059
41180	St. Louis	MO-IL	3.9914	19.9079	0.9226	0.1312	-0.4277	2.3707
41420	Salem	OR	0.5505	9.5532	0.8053	0.3850	3.4215	0.1330
41500	Salinas	CA	0.5803	1.2221	1.1497	0.3426	9.2400	-0.5045
41540	Salisbury	MD	0.1703	13.6356	0.7665	0.6063	-0.3934	-0.8133
41620	Salt Lake City	UT	1.5660	5.5353	0.9849	0.1645	3.3545	1.1401
41660	San Angelo	TX	0.1539	11.3999	0.7550	0.5001	1.5945	-0.9984
41700	San Antonio	TX	2.8340	12.2914	0.9238	0.1656	2.1287	1.8188
41740	San Diego-Carlsbad-San Marcos	CA	4.2351	1.5943	1.2222	0.1332	9.7800	1.4266
41780	Sandusky	OH	0.1101	4.8876	0.7919	0.5651	-0.9100	-1.3725
41860	San Francisco-Oakland-Fremont	CA	5.9848	0.3531	1.4952	0.1203	7.3604	1.6192
41940	San Jose-Sunnyvale-Santa Clara	CA	2.5677	0.1447	1.5878	0.1526	5.5612	0.8121
42020	San Luis Obispo-Paso Robles	CA	0.3736	2.4081	1.0086	0.3809	7.8700	-0.6538
42060	Santa Barbara-Santa Maria-Goleta	CA	0.5754	0.8643	1.1438	0.2810	10.9700	-0.5659
42100	Santa Cruz-Watsonville	CA	0.3584	0.6286	1.1396	0.6419	8.4900	-1.0716
42140	Santa Fe	NM	0.2035	0.1706	1.2396	0.6477	3.0200	-1.2264
42220	Santa Rosa-Petaluma	CA	0.6612	1.8173	1.0370	0.3670	7.9300	-0.2054
42260	Bradenton-Sarasota-Venice	FL	0.9783	8.0869	0.8481	0.2326	4.7123	0.5228
42340	Savannah	GA	0.4688	9.2001	0.8077	0.3385	0.7595	0.0822
42540	Scranton–Wilkes-Barre	PA	0.7822	62.6807	0.7348	0.2540	0.3497	0.7451
42660	Seattle-Tacoma-Bellevue	WA	4.7113	1.1719	1.2432	0.1332	4.6088	1.8885
42680	Sebastian-Vero Beach	FL	0.1877	1.2555	0.9359	0.6381	4.7200	-1.2862
43100	Sheboygan	WI	0.1630	3.2650	0.8625	0.4794	-0.3700	-1.0073
43300	Sherman-Denison	TX	0.1689	20.5729	0.7343	0.7441	0.7800	-0.9061
43340	Shreveport-Bossier City	LA	0.5518	0.5061	1.2082	0.2672	0.4263	-0.0654
43580	Sioux City	IA-NE-SD	0.2033	6.7056	0.8078	0.3518	-1.6477	-0.5531
43620	Sioux Falls	SD	0.3234	0.9176	1.0383	0.3194	-3.1981	-0.1810

Table 1: MSA variables and descriptives for the initial equilibrium

FIPS	MSA name	State	$L_r/\overline{L}$	$\widehat{\mu}_r^{\max}$	$1/\overline{m}_r$	$\widehat{\theta}_r$	A_r^o	$\widehat{A}_r^u$
43780	South Bend-Mishawaka	IN-MI	0.4508	5.9962	0.9017	0.3487	-2.3182	0.1576
43900	Spartanburg	SC	0.3923	11.2840	0.7992	0.3525	0.5200	-0.1066
44060	Spokane	WA	0.6494	3.8173	0.8466	0.2893	1.3300	0.3953
44100	Springfield	IL	0.2941	14.5944	0.7757	0.3680	-2.6215	-0.1150
44140	Springfield	MA	0.9719	48.7269	0.7653	0.2673	-0.0296	0.9868
44180	Springfield	MO	0.5980	42.4428	0.7162	0.3118	-0.1019	0.5377
44220	Springfield	OH	0.2000	20.6803	0.7124	0.6353	-2.0300	-0.5560
44300	State College	PA	0.2059	5.6983	0.8980	0.4912	-0.4000	-0.6733
44700	Stockton	CA	0.9552	9.1216	0.8869	0.3999	4.7700	0.4709
44940	Sumter	SC	0.1480	5.4151	0.8191	0.6486	0.4500	-1.1196
45060	Syracuse	NY	0.9187	11.6878	0.8621	0.2285	-1.0878	0.9094
45220	Tallahassee	FL	0.5016	15.0466	0.7887	0.3650	1.8418	0.1910
45300	Tampa-St. Petersburg-Clearwater	FL	3.8779	17.9295	0.8662	0.1303	4.0087	1.9781
45460	Terre Haute	IN	0.2411	20.4346	0.7766	0.5363	-2.2437	-0.3093
45500	Texarkana	TX	0.1911	11.9339	0.7701	0.4806	0.3401	-0.7535
45780	Toledo	OH	0.9267	18.0928	0.8282	0.2156	-2.2985	0.9937
45820	Topeka	KS	0.3256	22.9574	0.7672	0.3978	-1.2054	-0.0417
45940	Trenton-Ewing	NJ	0.5203	1.6191	1.0467	0.3137	-0.8000	-0.1181
46060	Tucson	AZ	1.3768	24.1671	0.8204	0.2328	4.0400	1.0965
46140	Tulsa	OK	1.2895	5.5205	0.9845	0.1913	0.4138	1.0760
46220	Tuscaloosa	AL	0.2922	7.7286	0.8737	0.3964	0.5956	-0.3554
46340	Tyler	TX	0.2829	3.5960	0.8892	0.4075	0.7200	-0.5192
46540	Utica-Rome	NY	0.4198	76.1905	0.6887	0.3637	-1.6177	0.3300
46660	Valdosta	GA	0.1853	33.3007	0.6831	0.4890	0.4906	-0.6906
46700	Vallejo-Fairfield	CA	0.5817	2.3184	1.0196	0.5800	5.8800	-0.2641
47020	Victoria	TX	0.1620	1.9775	0.9658	0.5431	0.7132	-1.1395
47220	Vineland-Millville-Bridgeton	NJ	0.2214	18.9165	0.7773	0.5472	0.3800	-0.6868
47260	Virginia Beach-Norfolk-Newport News	VA-NC	2.3615	6.6554	0.9682	0.1646	0.7721	1.5923
47300	Visalia-Porterville	CA	0.6001	20.2186	0.8264	0.3309	5.6500	0.1024
47380	Waco	TX	0.3248	14.4336	0.7623	0.3399	0.7600	-0.2405
47580	Warner Robins	GA	0.1865	2.0361	0.8817	0.5774	-0.0400	-0.9647
47900	Washington-Arlington-Alexandria	DC-VA-MD-WV	7.5546	2.1874	1.2875	0.1175	-0.5658	2.6267
47940	Waterloo-Cedar Falls	IA	0.2325	4.0817	0.8784	0.3123	-3.6928	-0.3363
48140	Wausau	WI	0.1850	8.5505	0.7840	0.4457	-3.3000	-0.5433
48260	Weirton-Steubenville	WV-OH	0.1745	12.5561	0.7784	0.6507	-0.4289	-0.8395
48300	Wenatchee	WA	0.1526	2.5064	0.9367	0.6415	1.1223	-1.0532
48540	Wheeling	WV-OH	0.2071	27.1680	0.7306	0.5045	-0.0508	-0.6087
48620	Wichita	KS	0.8491	7.0330	0.8959	0.2070	-0.5189	0.7748
48660	Wichita Falls	TX	0.2109	3.6100	0.9231	0.4866	-0.0733	-0.7295
48700	Williamsport	PA	0.1663	37.1189	0.7212	0.5359	0.3300	-0.8261
48900	Wilmington	NC	0.4833	4.2397	0.9124	0.3689	0.8620	0.0454
49020	Winchester	VA-WV	0.1725	8.0065	0.8765	0.8358	0.2643	-0.9449
49180	Winston-Salem	NC	0.6594	3.7013	0.9707	0.2738	-0.3283	0.3418
49340	Worcester	MA	1.1124	1.7596	1.1348	0.4121	0.2400	0.7079
49420	Yakima	WA	0.3318	3.8343	0.9066	0.4012	1.4800	-0.2958
49620	York-Hanover	PA	0.5994	20.5103	0.8111	0.4145	-0.5800	0.3817
49660	Youngstown-Warren-Boardman	OH-PA	0.8125	37.2035	0.7640	0.2679	-2.2828	0.9348
49700	Yuba City	CA	0.2337	1.2193	1.0373	0.9995	3.3821	-1.0057
49740	Yuma	AZ	0.2713	45.4247	0.6962	0.3985	4.2400	-0.5236

Notes: See Sections 4 and 5 for computational details on how to obtain the upper bounds, other amenities, and commuting friction parameters.

Table 2: Cross-MSA distribution of establishment numbers and average size – summary for observed and simulated data

Variable	Mean Model	Mean Observed	St.dev. Model	St.dev. Observed	Min Model	Min Observed	Max Model	Max Observed	Correlation Model-Observed
# of establishments total	18067.10	18067.09	16878.09	43138.45	1738	911	109210	541255	0.7253
# of establishments size 1-19	15444.74	15461.97	12066.43	37449.79	1550	804	79181	478618	0.3824
# of establishments size 20-99	2121.56	2162.09	6320.64	4728.28	49	93	52178	51310	0.9412
# of establishments size 100-499	429.83	397.50	1729.44	922.34	14	13	24365	9951	0.8890
# of establishments size 500+	70.94	45.52	132.67	113.75	2	1	1509	1376	0.9320
Avg establishment size	11.73	15.40	11.63	2.60	0.90	6.40	131.88	23.70	0.1716

Notes: Model values are computed from a representative sample of 6,431,886 establishments. The small difference (of 2 units) with respect the observed number of establishments in the 2007 County Business Patterns is due to rounding in the sampling procedure. Establishment sizes in the model are scaled to match the total employment figure for the 356 MSAs from the 2007 County Business Patterns. The number of observations is $N = 356$ MSAs in all cases.

Table 3: Shipment shares and shipping distances – summary for observed and simulated data

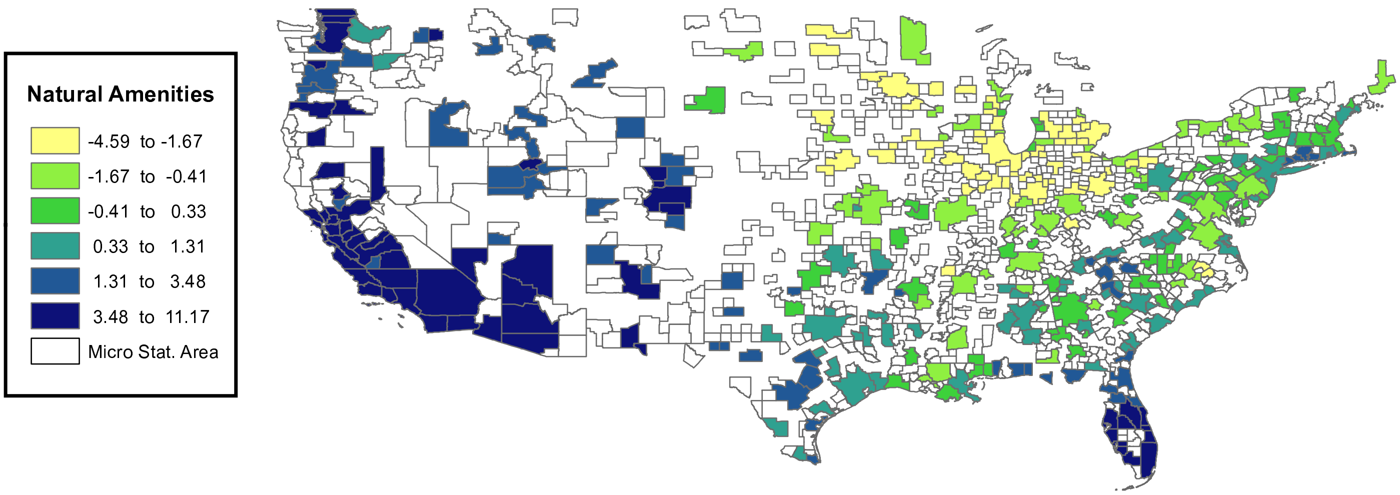
Employment	Number of establishments		Shipment shares by distance shipped to destination						Mean distance shipped		
			< 100 miles		100–500 miles		> 500 miles				
	Observed	Model	Observed	Model	Observed	Model	Observed	Model	Observed	Model	Model (wgt)
All	6,431,884	6,431,886	0.261	0.506	0.288	0.277	0.348	0.217	529.6	71.98	739.8
1–19	5,504,463	5,498,328	0.561	0.984	0.204	0.016	0.194	0.000	327.2	38.5	61.2
20–99	769,705	755,275	0.382	0.835	0.288	0.162	0.276	0.004	423.8	157.9	194.4
100–499	141,510	153,021	0.254	0.420	0.318	0.440	0.342	0.139	520.4	556.0	740.3
500+	16,206	25,255	0.203	0.079	0.272	0.332	0.388	0.590	588.6	1450.6	1519.1

Notes: Shipping distance and shipping share columns are adapted from calculations by Holmes and Stevens (2010, Table 1) who use confidential Census microdata from the 1997 Commodity Flow Survey. The small difference (of 2 units) between the observed and model total number of establishments is due to rounding in our sampling procedure. The last column reports distances shipped weighted by establishments' sales shares in total sales.

Table 4: Summary of the counterfactuals

Baseline counterfactuals (no agglomeration economies)						
	No urban frictions (CF1)			No trade frictions (CF2)		
	Mean	Std. dev.	Weighted mean	Mean	Std. dev.	Weighted mean
% change $1/\overline{m}_r$	-0.06	0.26	0.04	78.50	14.26	67.59
% change L_r	-2.15	3.60	0	4.30	15.28	0
% change $\overline{\Lambda}_r$	-8.79	1.82	-9.85	-43.55	4.27	-39.90
% change V_r	9.69	2.24	10.98	78.17	13.79	67.62
RS coefficient	-0.9178			-0.9392		
Baseline counterfactuals (short-run, no labor mobility)						
	No urban frictions (CF1)			No trade frictions (CF2)		
	Mean	Std. dev.	Weighted mean	Mean	Std. dev.	Weighted mean
% change $1/\overline{m}_r$	0.07	0.20	0.01	77.93	14.15	67.10
% change L_r	0	0	0	0	0	0
% change $\overline{\Lambda}_r$	-8.91	1.67	-9.83	-43.45	4.39	-39.68
% change V_r	9.83	2.05	10.93	77.93	14.15	67.10
RS coefficient	-0.9249			-0.9249		
Robustness checks (with agglomeration economies)						
	No urban frictions (CF3)			No trade frictions (CF4)		
	Mean	Std. dev.	Weighted mean	Mean	Std. dev.	Weighted mean
% change $1/\overline{m}_r$	-0.12	0.31	0.04	78.71	14.03	67.63
% change L_r	-2.21	3.74	0	4.50	16.15	0
% change $\overline{\Lambda}_r$	-8.74	1.89	-9.85	-43.60	4.33	-39.90
% change V_r	9.62	2.33	10.98	78.36	14.03	67.66
RS coefficient	-0.9176			-0.9394		

Notes: Weighted mean refers to the mean percentage change where the weights are given by the MSAs' initial population shares. The counterfactual scenarios CF3 and CF4 include the agglomeration economies specification developed in Section 6.2. RS coefficient refers to the slope of the estimated rank-size relationship.



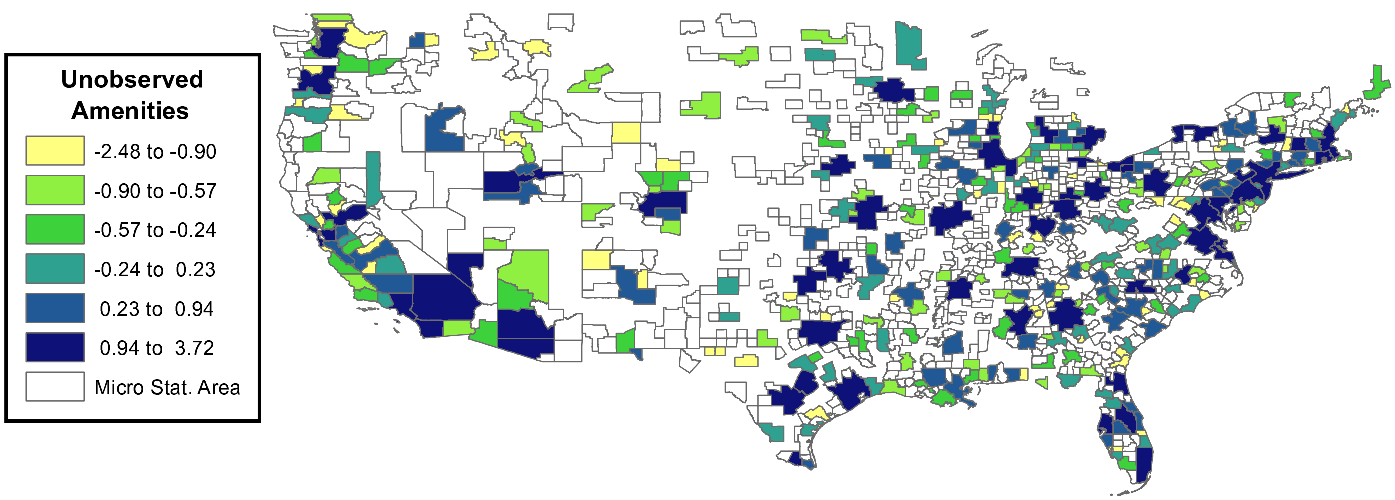


Figure 1: Distribution of natural A_r^o (top) and unobserved $\widehat{A}_r^u \equiv \widehat{\varepsilon}_r$ (bottom) amenities

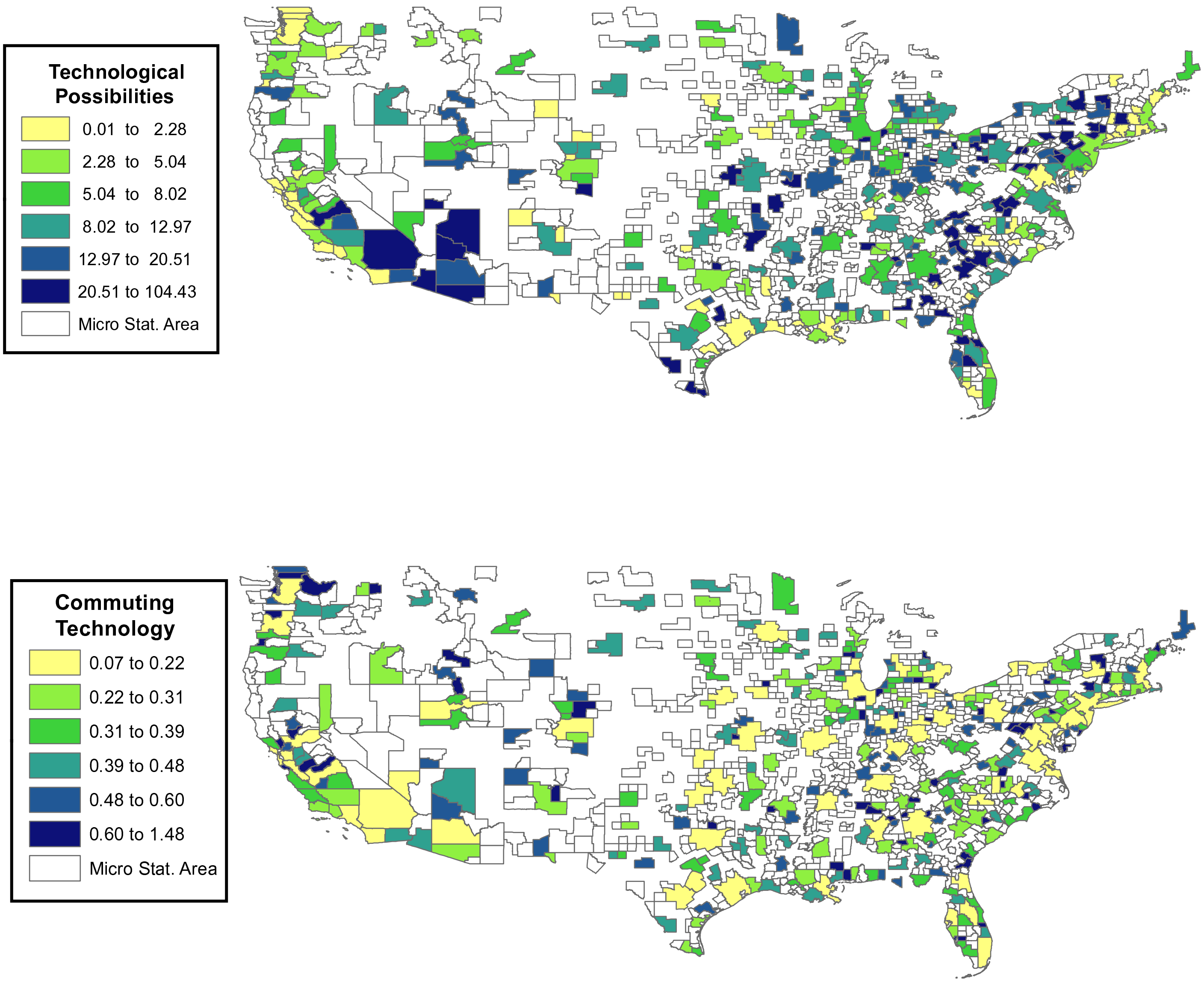


Figure 2: Distribution of technological possibilities $\widehat{\mu}_r^{\max}$ (top) and commuting technology $\widehat{\theta}_r$ (bottom)

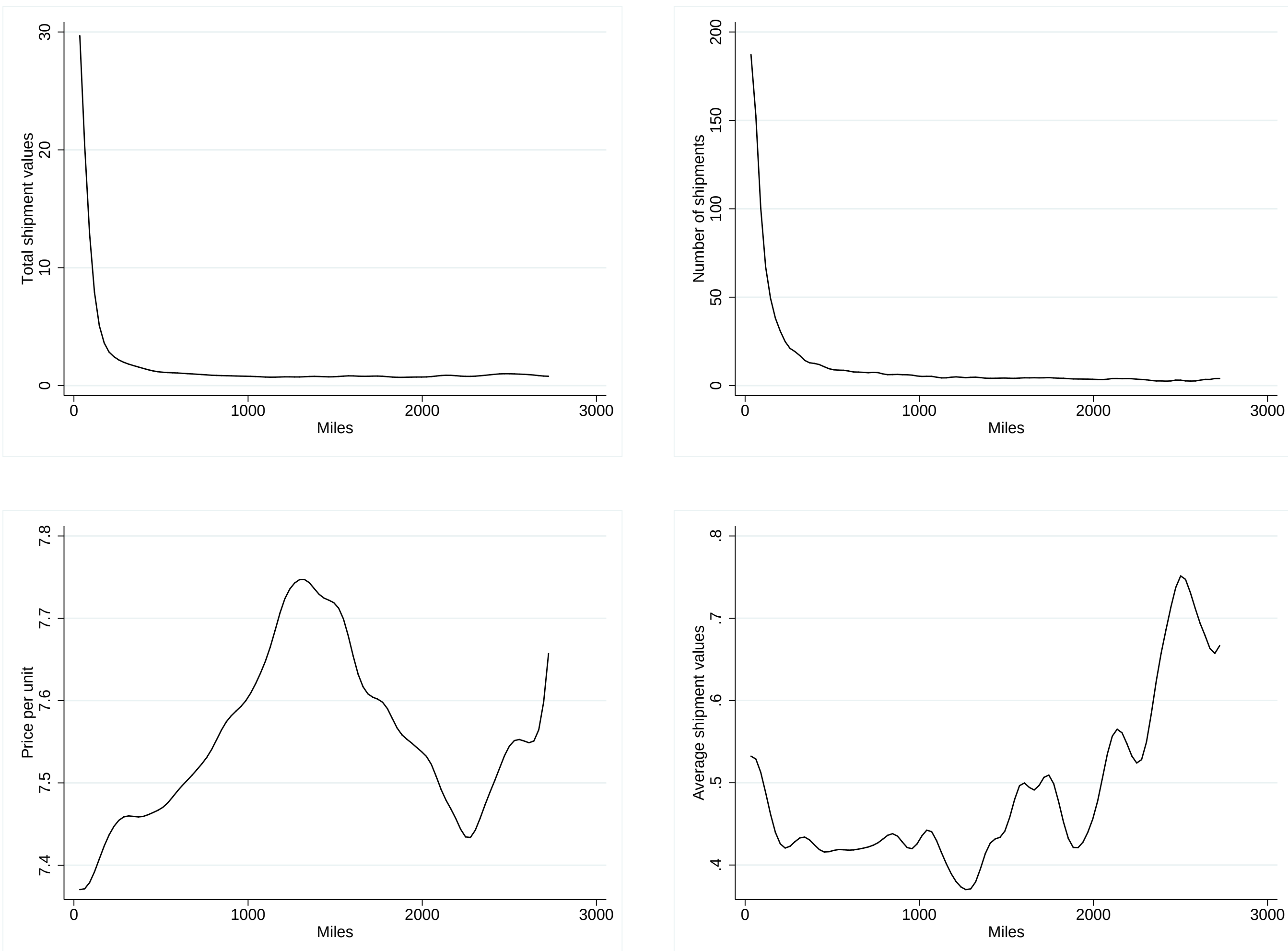


Figure 3: Micro-fit for establishment-level shipments across MSAs (kernel regressions on distance) (To be compared with Figures 1–3 in Hillberry and Hummels, 2008)

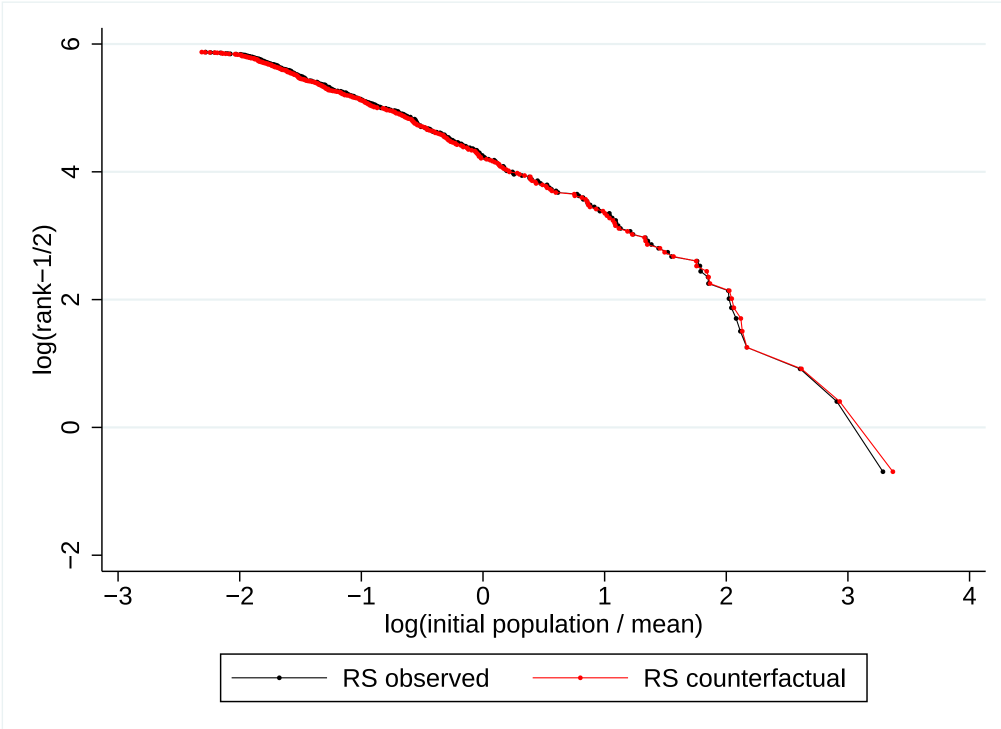


Figure 4: Rank-size rule, observed and counterfactual (CF1)

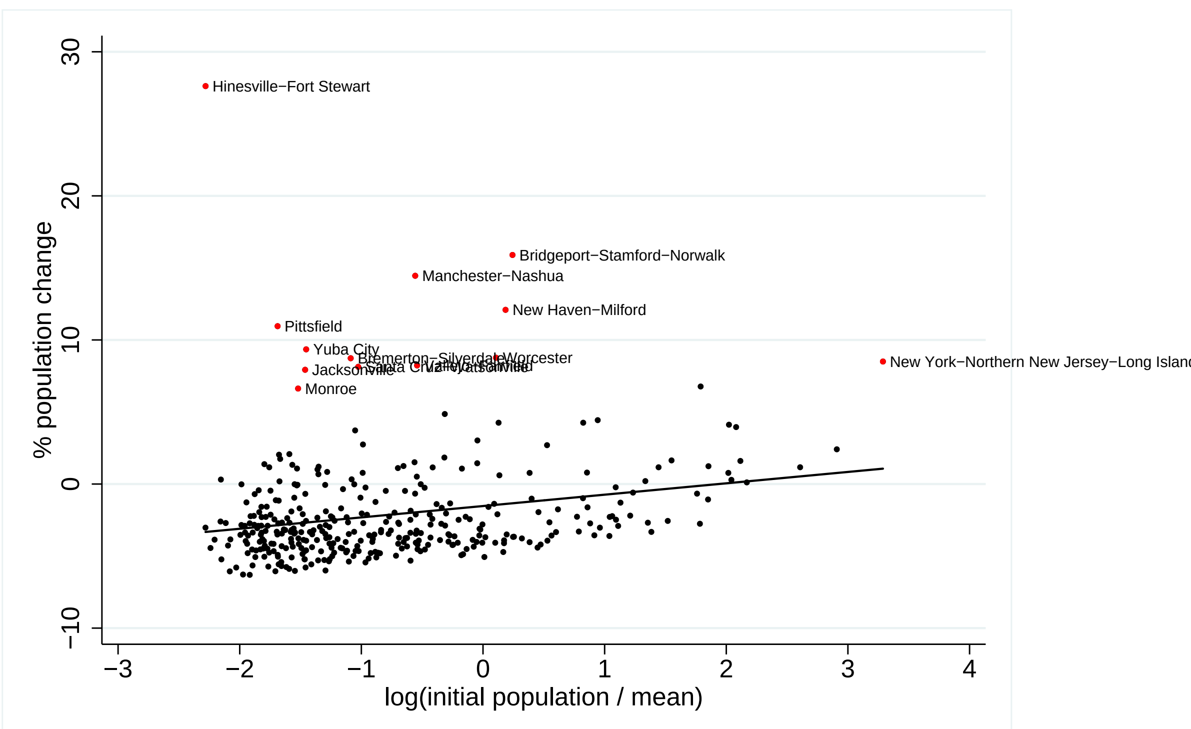


Figure 5: Changes in MSA populations and initial size (CF1)

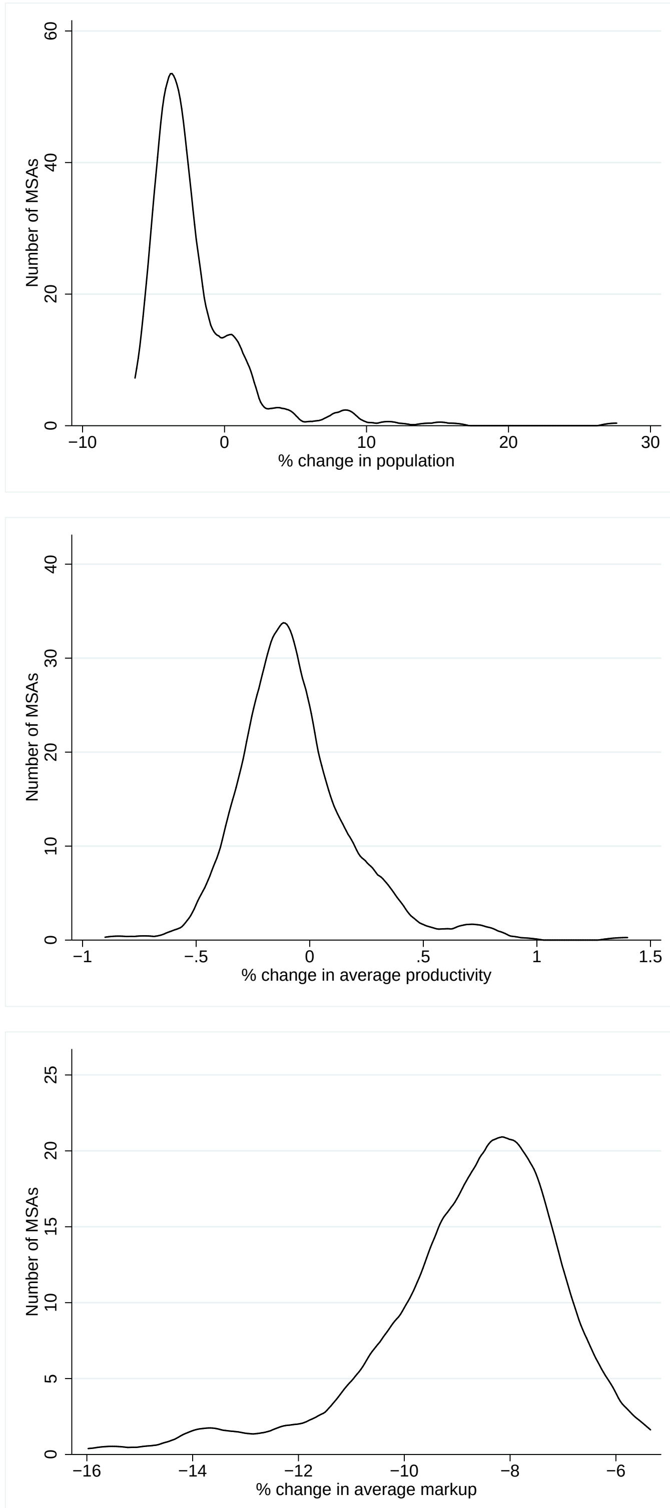


Figure 6: Distribution of counterfactual changes in L_r, $1/\overline{m}_r$ and $\overline{\Lambda}_r$ (CF1)

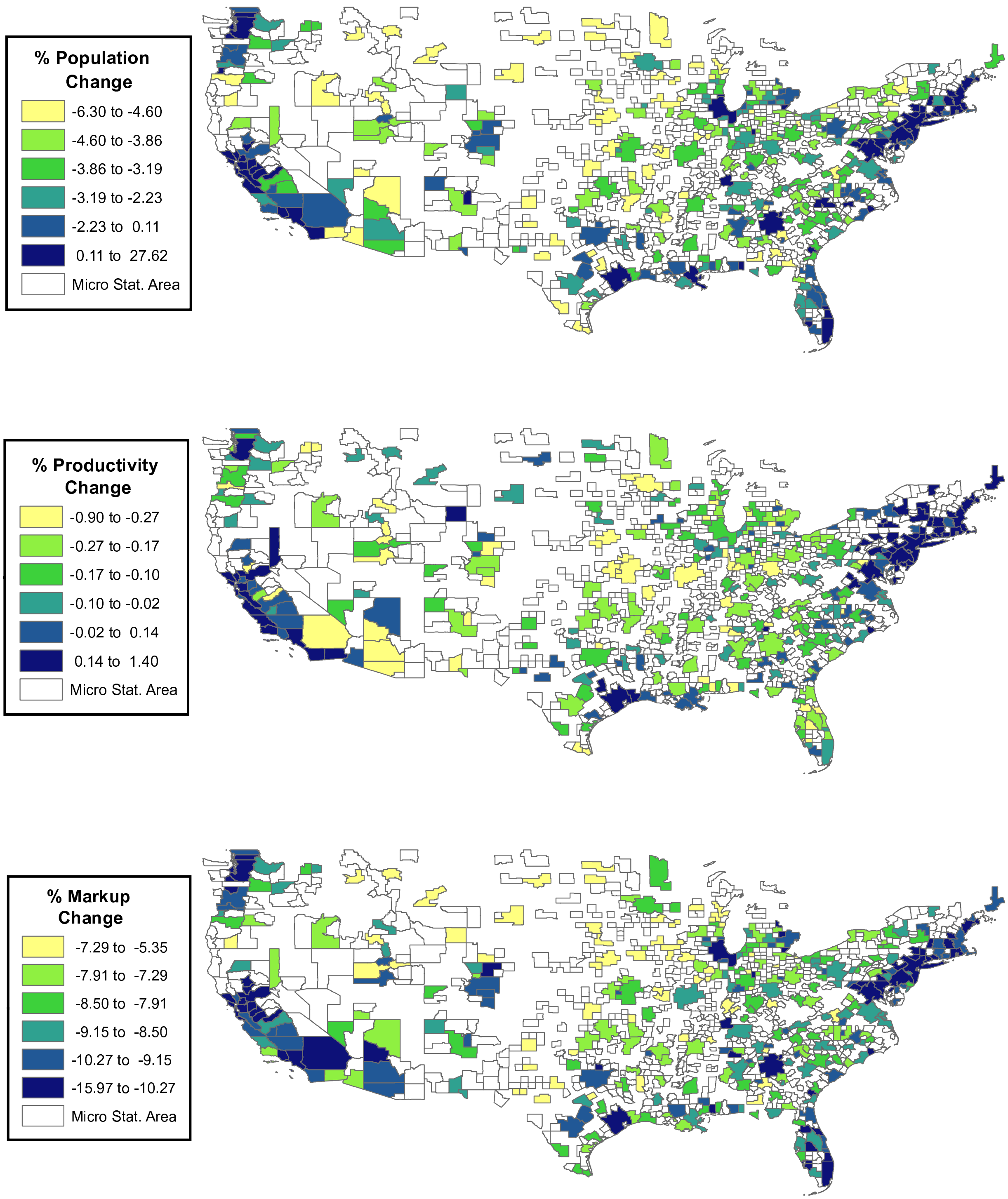


Figure 7: Spatial pattern of counterfactual changes in L_r, $1/m_r^d$ and $\overline{\Lambda}_r$ (CF1)

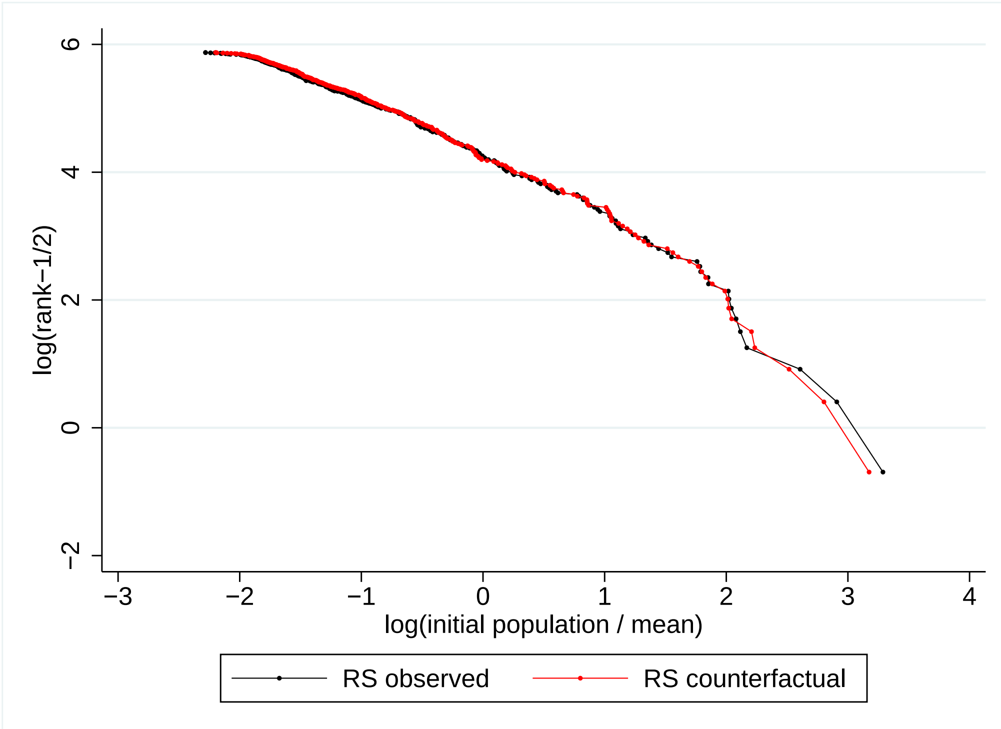


Figure 8: Rank-size rule, observed and counterfactual (CF2)

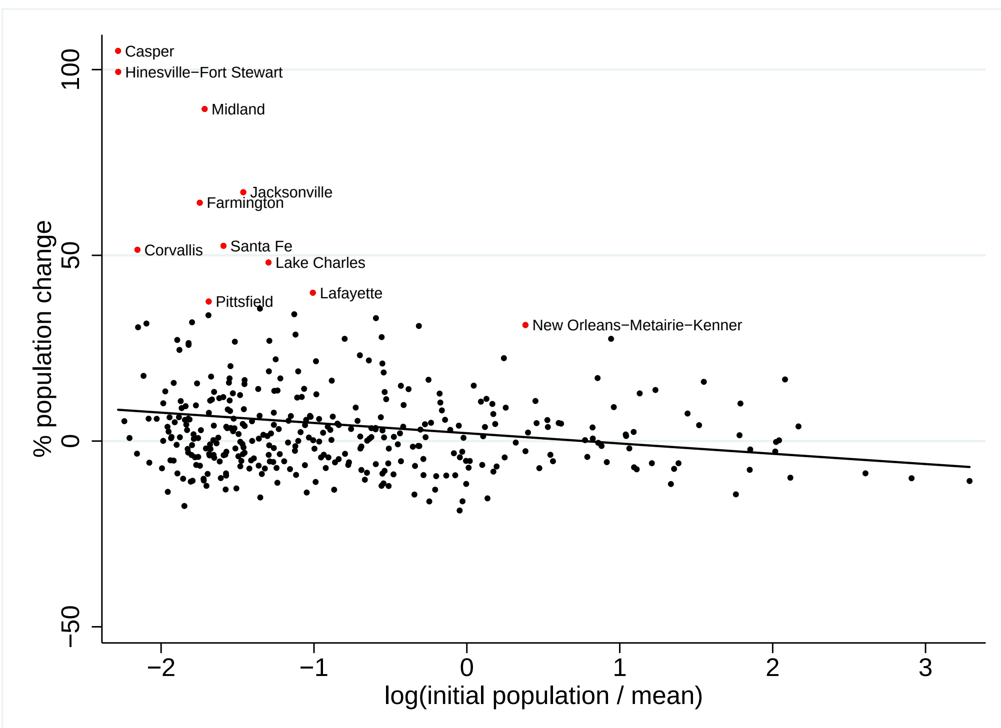


Figure 9: Changes in MSA populations and initial size (CF2)

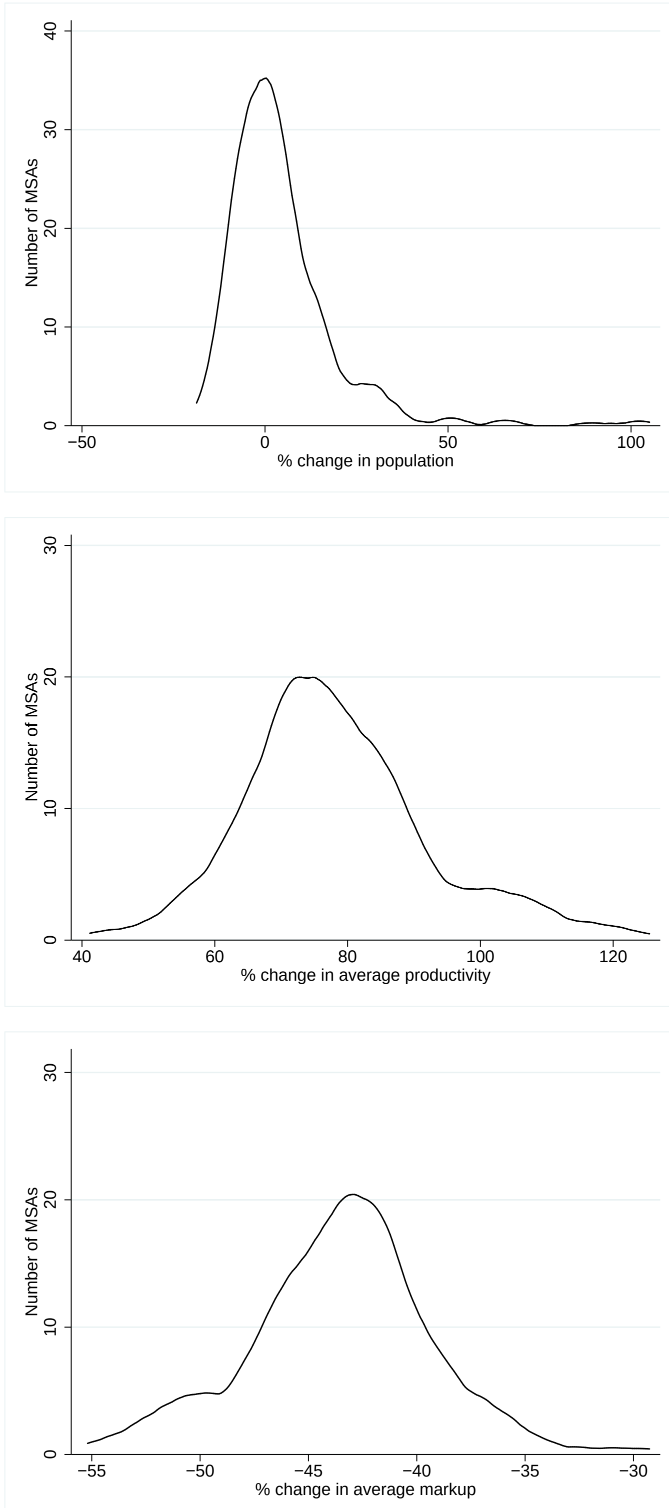


Figure 10: Distribution of counterfactual changes in L_r, $1/\overline{m}_r$ and $\overline{\Lambda}_r$ (CF2)

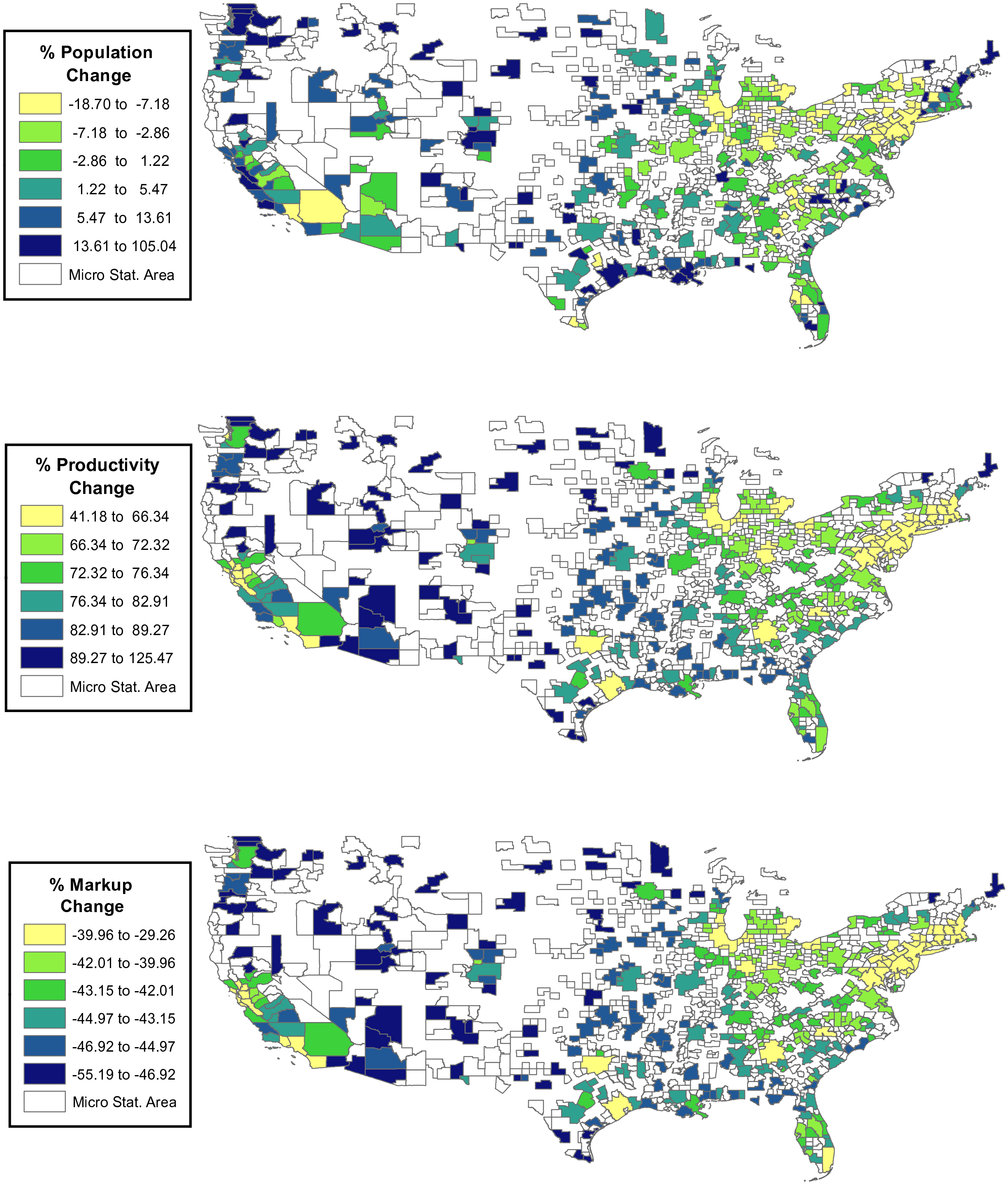


Figure 11: Spatial pattern of counterfactual changes in L_r, $1/m_r^d$ and $\overline{\Lambda}_r$ (CF2)

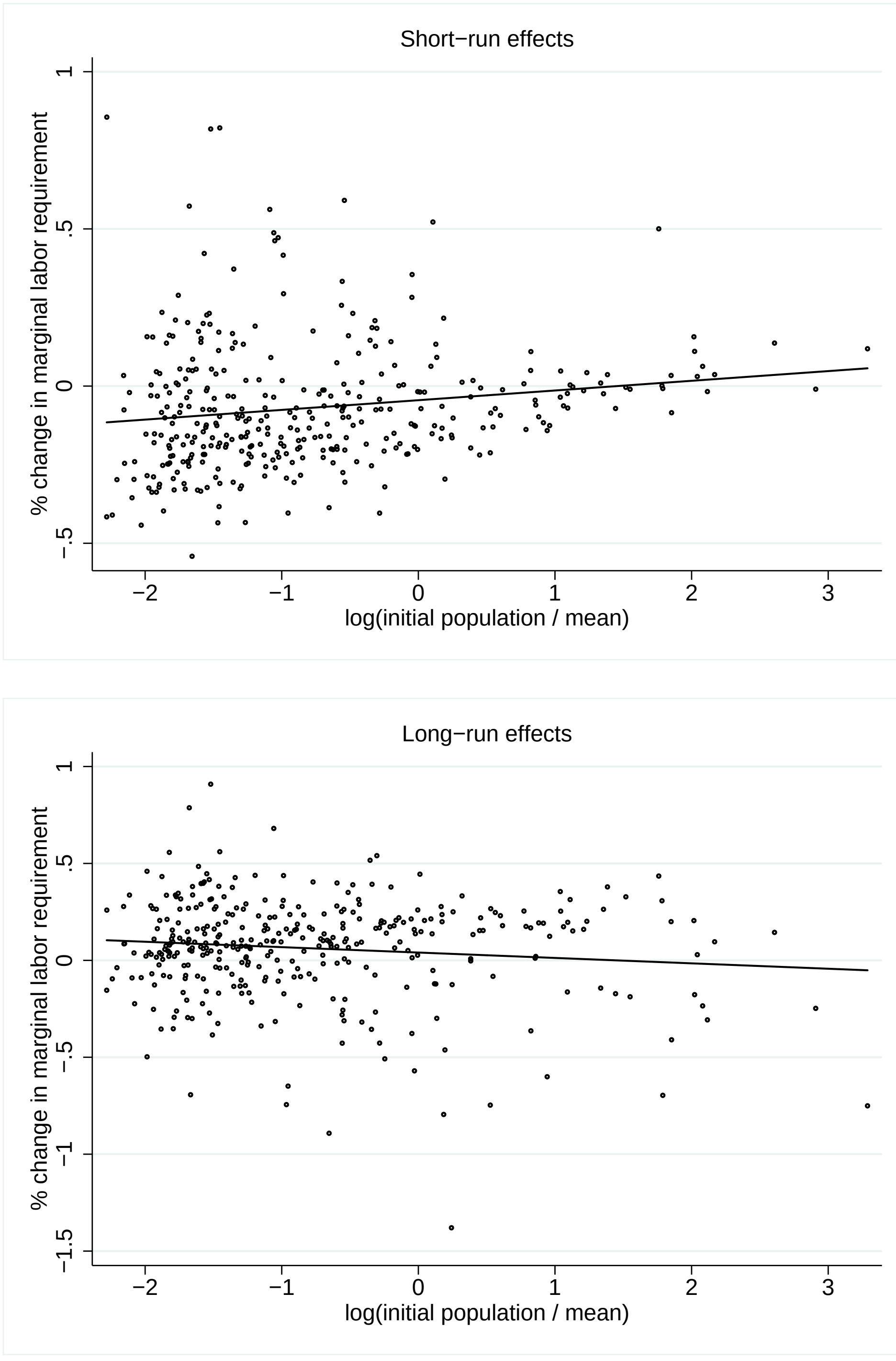


Figure 12: Difference in short- and long-run relationships between Δm_r^d and L_r (CF1)