Online Appendix for

Donald R. Davis and Jonathan I. Dingel "A Spatial Knowledge Economy"

May 2018

A Theory

A.1 Internal urban structure

To introduce congestion costs, we follow Behrens, Duranton and Robert-Nicoud (2014) and adopt a standard, highly stylized model of cities' internal structure.[1] City residences of unit size are located on a line and center around a single point where economic activities occur, called the central business district (CBD). Individuals commute to the CBD at a cost that is denoted in units of the numeraire. The cost of commuting from a distance x is τx^γ and independent of the resident's income and occupation.

Individuals choose a residential location x to minimize the sum of land rent and commuting cost, $r(x) + \tau x^\gamma$. In equilibrium, individuals are indifferent across residential locations. In a city with population mass L, the rents fulfilling this indifference condition are $r(x) = r\left(\frac{L}{2}\right) + \tau\left(\frac{L}{2}\right)^\gamma - \tau x^\gamma$ for $0 \leq x \leq \frac{L}{2}$. Normalizing rents at the edge to zero yields $r(x) = \tau\left(\frac{L}{2}\right)^\gamma - \tau x^\gamma$.

The city's total land rent is

$$TLR = \int_{\frac{-L}{2}}^{\frac{L}{2}} r(x)dx = 2\int_0^{\frac{L}{2}} r(x)dx = 2\tau\left(\left(\frac{L}{2}\right)^{\gamma+1} - \frac{1}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1}\right) = \frac{2\tau\gamma}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1}$$

The city's total commuting cost is

$$TCC = 2\int_0^{\frac{L}{2}} \tau x^\gamma dx = \frac{2\tau}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1} \equiv \theta L^{\gamma+1}$$

The city's total land rents are lump-sum redistributed equally to all city residents. Since they each receive $\frac{TLR}{L}$, every resident pays the average commuting cost, $\frac{TCC}{L} = \theta L^\gamma$, as her net urban cost. Since this urban cost is proportional to the average land rent, we say the "consumer price of housing" in city c is $p_{h,c} = \theta L_c^\gamma$.

[1] There is nothing original in this urban structure. We use notation identical to, and taken from, Behrens, Duranton and Robert-Nicoud (2014).

A.2 The number of cities

In section I.B, we define equilibrium for a finite set of locations $\{c\}$ in which each member of the set is populated, $L_c > 0$. This section discusses properties of this set. In our model, the equilibrium number of cities is not uniquely determined by exogenous parameters. This is a standard result in models with symmetric fundamentals, and our predictions about the cross section of cities do not depend upon the number of populated locations.

While the equilibrium number of cities is not uniquely determined, the equilibrium number of cities where idea exchange occurs is bounded by the exogenous parameters governing agglomeration and congestion. In the case of equations (5) and (6), there is an upper bound on the equilibrium number of cities with positive idea exchange for a given population L because the matching process in equation (6) features scale economies and the production function in equation (5) requires a minimum value of AZ_cz for positive participation. There is a lower bound on the equilibrium number of cities because congestion costs are unboundedly increasing in L_c while Z_c has a finite upper bound. Between these bounds, there may exist multiple equilibria that have distinct numbers of heterogeneous cities.

The equilibrium number of heterogeneous cities will tend to increase with population. The upper bound increases with population because a larger population makes it feasible to achieve the minimum scale for idea exchange in a larger number of cities. Holding other parameters fixed, a higher value of L can be accommodated by the same number of larger cities or an increase in the number of cities. The intensive margin cannot entirely absorb population increases of arbitrary size, since congestion costs must eventually exceed agglomeration benefits. Increases along the extensive margin – the number of cities – could result in a greater number of distinct city sizes or a greater number of instances of a given population size. The latter possibility is constrained by the fact that locally stable equilibria can have equal-sized cities only if the agglomeration force is weak relative to the congestion force at the margin, as we prove in Proposition 3 below.

Recent related research with heterogeneous agents and symmetric fundamentals has taken distinct approaches to thinking about the inter-related problems of city formation, the number (mass) of cities, and uniqueness of equilibrium. With heterogeneous firms, Gaubert (2015) assumes that there is a uniquely optimal city size distinct to each productivity level and that cities are created by developers who make zero profits. With a continuum of cities, this yields a one-to-one mapping between firm productivities and city sizes, and so the distribution of firm productivity determines the distribution of city sizes. With heterogeneous individuals, Behrens, Duranton and Robert-Nicoud (2014) assume a continuum of cities and characterize equilibria in which each city is talent-homogeneous, which yields a differential equation that maps between individual talents and city sizes.[2] Combined with the assumption of a boundary condition, this yields the distribution of city sizes as a function of the distribution of talent.[3]

We take a different path by assuming that the number of cities is an integer. This matches the empirical fact that cities are discrete. The top ten metropolitan areas account for one-quarter of the United States population. With a continuum, any countable set would be measure zero. Similarly, our model implies that the population size of the largest city is less than the economy's total population. In Behrens, Duranton and Robert-Nicoud (2014) and Gaubert (2015), the population size of the largest city is a function only of the talent/productivity distribution, so the fact that New York is larger than Zurich is attributable to differences in the US and Swiss talent/productivity distributions, not the fact that New York City has more residents than the entirety of Switzerland.

This greater realism comes at a cost. The equilibrium number and sizes of cities are not necessarily unique. In our numerical work, we take as given the number of cities and identify equilibria consistent with this number. For example, while we present a 275-city equilibrium in section II.D, the same parameter values are also consistent with a 270-city equilibrium. This multiplicity may simply be a feature of the world rather than something that needs to be refined away. Treating cities as discrete allows us to explain spatial variation in skill premia, whereas this form of within-city heterogeneity is absent in models with a one-to-one mapping between agents' heterogeneous characteristic and city size. We focus on results that are cross-sectional properties that do not rely on the number of cities or the uniqueness of equilibrium.

A.3 Existence of equilibrium with two heterogeneous cities

Here we characterize three sufficient conditions for $\{L, \mu(z), \bar{n}, B(\cdot), Z(\cdot), \theta, \gamma\}$ such that there exists a two-city equilibrium in which $L_1 < L_2$. The first is that idea exchange creates potential gains from agglomeration. The second is that congestion costs prevent the entire population from living in a single city. The third is that it is feasible for the entire population to live in two cities.

[2]While these authors focus on the properties of equilibria with talent-homogeneous cities, these are not the only equilibria in their model. It also yields equilibria with discrete number of cities, but in that case analytical results cannot be obtained in general.

[3] To obtain their city-size distribution that approximates Zipf's law, Behrens, Duranton and Robert-Nicoud (2014) impose the boundary condition that individuals of zero talent live in cities of zero population where they produce zero output.

To help define the three conditions, let $Z_c(x, y)$ denote the maximum value of Z_c satisfying equation (4) with $\beta_{z,c} = \beta(z, Z_c)$ when the population of tradables producers in city c is all individuals with abilities in the $[x, y]$ interval. Formally, the maximum value of Z_c satisfying that equation when $\mu(z, c) = \mu(z)\ \forall z \in [x, y]$ and $\mu(z, c) = 0\ \forall z \in [z_m, x) \cup (y, \infty)$ where z_m is given by $\bar{n} = \int_0^{z_m} \mu(z)dz$.

The *agglomeration condition* is that $\tilde{z}\left(\underline{z}, Z_c(\underline{z}, \infty)\right) > \tilde{z}\left(\underline{z}, Z_c(z_m, \underline{z})\right)$ where $\underline{z}$ is the median tradables producer, identified by $\frac{1-\bar{n}}{2} = \int_{z_m}^{\underline{z}} \mu(z)dz$. This condition says that technology $(Z(\cdot, \cdot), \bar{n})$ and population $(L, \mu(z))$ are such that the median tradables producer and every individual of greater ability would find idea exchange with one another profitable if they all colocated. In other words, there are potential gains from agglomeration via idea exchange. The *congestion condition* is that the congestion costs of locating the economy's entire population in a single city exceed the gains from idea exchange for the lowest-ability tradables producer, $\frac{\theta}{1-\bar{n}} L^{\gamma} > \tilde{z}(z_m, Z_c(z_m, \infty)) - \tilde{z}(z_m, 0)$. The *feasibility condition* is that the least-able tradables producer generates enough output to cover the congestion costs associated with two cities, $\tilde{z}(z_m, 0) \geq \frac{\theta}{1-\bar{n}} \left(\frac{L}{2}\right)^{\gamma}$.

We now characterize the economy in terms of L_1 and define a function $\Omega(L_1)$ that equals zero when the economy is in equilibrium. Choose a value $L_1 \leq \frac{1}{2}L$, which implies $L_2 = L - L_1$. Define values z_b and $z_{b,n}$ that respectively denote the highest-ability tradables and non-tradables producers in city 1 by

$$(1 - \bar{n})L_1 = L \int_{z_m}^{z_b} \mu(z)dz \qquad \bar{n}L_1 = L \int_0^{z_{b,n}} \mu(z)dz.$$

Because the support of $\mu(z)$ is connected, z_b is continuous in L_1. The locational assignments

$$\mu(z,1) = \begin{cases} \mu(z) & 0 \leq z < z_{b,n} \\ 0 & z_{b,n} \leq z < z_m \\ \mu(z) & z_m \leq z < z_b \\ 0 & z_b \leq z \end{cases} \qquad \mu(z,2) = \begin{cases} 0 & 0 \leq z < z_{b,n} \\ \mu(z) & z_{b,n} \leq z < z_m \\ 0 & z_m \leq z < z_b \\ \mu(z) & z_b \leq z \end{cases}$$

satisfy equations (8), (9), and (10). These assignments imply values for $p_{h,1}, p_{h,2}, p_{n,1}, p_{n,2}, Z_1, Z_2$, and $\beta_{z,c}$ via equations (4), (7), (12), and (13), where we select the maximal values of Z_1 and Z_2 satisfying those equations. The feasibility condition ensures these assignments are possible for all L_1.

This is a spatial equilibrium if z_b is indifferent between the two cities. Utility in the smaller city minus utility in the larger city for the marginal tradables producer, z_b, is

$$\tilde{z}(z_b, Z_1(z_m, z_b)) - \bar{n}p_{n,1} - p_{h,1} - (\tilde{z}(z_b, Z_2(z_b, \infty)) - \bar{n}p_{n,2} - p_{h,2})$$

Using equations (7) and (13) and rearranging terms, we call this difference $\Omega(L_1)$.

$$\Omega(L_1) \equiv \frac{\theta}{1-\bar{n}}\left(L_2^\gamma - L_1^\gamma\right) - \tilde{z}(z_b, Z_2(z_b, \infty)) + \tilde{z}(z_b, Z_1(z_m, z_b))$$

Ω can be written solely as a function of L_1 because all the other variables are given by L_1 via $z_{b,n}$ and z_b through the locational assignments and other equilibrium conditions.

$\Omega(L_1) = 0$ is an equilibrium. $\lim_{L_1 \to 0} \Omega(L_1) > 0$ due to the congestion condition. $\Omega\left(\frac{L}{2}\right) < 0$ since equal-sized cities have equal prices and the agglomeration condition ensures that $Z_2 > Z_1$ at $L_1 = \frac{1}{2}L$. If $\Omega(L_1)$ is appropriately continuous, then there is an intermediate value $L_1 \in (0, \frac{L}{2})$ satisfying $\Omega(L_1) = 0$. We now show that any discontinuity in $\Omega(L_1)$ is a discontinuous increase, so that such an intermediate value must exist.

The first term, $\frac{\theta}{1-\bar{n}}\left(L_2^\gamma - L_1^\gamma\right)$, is obviously continuous in L_1.

The second term is continuous in L_1 if the agglomeration condition holds. Since $\beta_{z,c}$ is a function of Z_c, the equilibrium value of Z_c satisfying equation (4) is a fixed point. The agglomeration condition means that such an intersection $Z_2 = Z(\{1-\beta(z, Z_2)\}, \{L \cdot \mu(z, 2)\},)$ exists for all values $L_1 \in (0, \frac{1}{2}L)$. Since $Z(\cdot, \cdot)$ is continuous by Assumption 3, $\beta(z, Z)$ is continuous by Assumption 1 and the maximum theorem, and our chosen $\mu(z, 2)$ is continuous in L_1, $Z_2(z_b, \infty)$ is a continuous function of L_1. Since $\tilde{z}(z, Z_c)$ is continuous by Assumption 1, $\tilde{z}(z_b, Z_2(z_b, \infty))$ is continuous in L_1.

The third term is increasing in L_1. By Assumption 3 and our chosen $\mu(z, 1)$, $Z(\{1 - \beta(z, Z_1)\}, \{L \cdot \mu(z, 1)\})$ is increasing in L_1 for any value of Z_1. By Assumption 3, for any L_1 the value of $Z(\{1 - \beta_{z,1}\}, \{L \cdot \mu(z, 1)\})$ is bounded above by z_b. Thus, if $Z_1(z_m, z_b) > 0$, for $\epsilon > 0$ $Z_1(z_m, z_b + \epsilon) > Z_1(z_m, z_b)$. Therefore, $Z_1(z_m, z_b)$ is increasing in L_1. By Assumption 1, $\tilde{z}(z_b, Z_1(z_m, z_b))$ is increasing in L_1. Therefore the third term in $\Omega(L_1)$ is increasing, and any discontinuity in $\Omega(L_1)$ is a discontinuous increase.

Since $\lim_{L_1 \to 0} \Omega(L_1) > 0$, $\Omega\left(\frac{L}{2}\right) < 0$, and Ω increases at any point at which Ω is not continuous in L_1, there exists a value of L_1 such that $\Omega(L_1) = 0$. This is an equilibrium with heterogeneous cities. Since $\Omega(L_1)$ crosses zero from above, it is a stable equilibrium, as will be defined in Appendix A.4.

A.4 Stability of equilibria

This section concerns the stability of equilibria. First, we adapt the notion of stability standard in the spatial-equilibrium literature to our setting. Second, we use this definition of local stability to show that stable equilibria can have equal-sized cities only if the agglomeration force is weak relative to the congestion force at the margin. This is the standard result.

The standard definition of stability in spatial-equilibrium models considers perturbations that reallocate a small mass of individuals away from their equilibrium locations (Henderson, 1974; Krugman, 1991; Behrens, Duranton and Robert-Nicoud, 2014; Allen and Arkolakis, 2014). If individuals would obtain greater utility in their initial equilibrium locations than in their arbitrarily assigned locations, then the equilibrium is stable.

Comparing equilibrium utilities to utilities under the perturbation requires calculating each individual's utility in a location given an arbitrary population allocation. This calculation is straightforward in models in which goods and labor markets clear city-by-city, so that an individual's utility in a location can be written solely as a function of the population in that location, as in Henderson (1974) and Behrens, Duranton and Robert-Nicoud (2014). It is also feasible in models in which the goods and labor markets clear for any arbitrary population allocation through inter-city trade, as in Krugman (1991) and Allen and Arkolakis (2014). In all these models, the spatial-equilibrium outcomes are identical to the economic outcomes that arise if individuals do not choose locations and are exogenously assigned to locations with assignments that coincide with the spatial-equilibrium population allocations.

In our model, spatial-equilibrium outcomes depend on the potential movement of individuals, so we cannot compute utility under an arbitrary population allocation without introducing additional assumptions. Our theory differs from the prior literature because nontradables prices are linked across cities in equilibrium by a no-arbitrage condition, equation (13). If we were to solve for an equilibrium with arbitrary population assignments rather than locational choice, clearing the goods and labor markets would require $p_{n,c} = \tilde{z}(z_{m,c}, Z_c)$ in each city, where $z_{m,c}$ is defined by $\int_0^{z_{m,c}} \mu(z,c)dz = \bar{n} \int_0^{\infty} \mu(z,c)dz$ for the arbitrary $\mu(z,c)$. Therefore, the prices and utilities obtained when clearing markets conditional on an arbitrary population allocation would not equal the equilibrium prices and utilities even when evaluated at the equilibrium population allocation. The inseparability of labor-market outcomes and labor mobility through this no-arbitrage condition distinguishes our model from prior work and require us to adapt the standard definition of stability to our setting.

We define a class of perturbations that maintains spatial equilibrium amongst non-tradables producers so that stability can be assessed in terms of tradables producers' incentives. Starting from an equilibrium allocation $\mu^*(z, c)$, we consider perturbations in which a small mass of tradables producers and a mass of non-tradables producers whose net supply equals the tradables producers' demand for non-tradables move from one city to another. The equilibrium allocation is stable if the tradables producers who moved would obtain higher utility in their equilibrium city than in their new location.

Definition 1 (Perturbation)**.** *A perturbation of size ϵ is a measure $d\mu(z, c)$ satisfying*

- $\{c : d\mu(z, c) > 0\}$ *is a singleton and* $\{c : d\mu(z, c) < 0\}$ *is a singleton, location changes are in one direction from a single city to another;*
- $L \sum_c \int |d\mu(z, c)| dz = 2\epsilon$*, individuals changing location have mass* ϵ*;*
- $(1 - \bar{n}) \int_0^{z_m} |d\mu(z, c)| dz = \bar{n} \int_{z_m}^{\infty} |d\mu(z, c)| dz$*, the movement of non-tradables producers satisfies demand from the movement of tradables producers; and*
- $\sum_c d\mu(z, c) = 0 \ \forall z$*, the aggregate population of any* z *is unchanged.*

Definition 2 (Local stability)**.** *An equilibrium with prices* $\{p^*_{h,c}, p^*_{n,c}\}$ *and populations* $\mu^*(z, c)$ *is locally stable if there exists an* $\bar{\epsilon} > 0$ *such that*

$$\tilde{z}(z, Z'_{c_1}) - \frac{\theta}{1 - \bar{n}} L'^{\gamma}_{c_1} \geq \tilde{z}(z, Z'_{c_2}) - \frac{\theta}{1 - \bar{n}} L'^{\gamma}_{c_2} \ \forall z, c_1, c_2 : z > z_m \ \& \ d\mu(z, c_1) < 0 \ \& \ d\mu(z, c_2) > 0$$

for all population allocations $\mu'(z, c) = \mu^*(z, c) + d\mu(z, c)$ *in which* $d\mu$ *is a perturbation of size* $\epsilon \leq \bar{\epsilon}$*, where* Z'_c*, and* L'_c *denote the values of these variables when the population allocation is* μ'*, individuals maximize (1) by their choices of* σ *and* β*, markets clear, and prices satisfy equations (12) and (13).*

Using this definition of local stability, we obtain the standard result that locally stable equilibria can have equal-sized cities only if the marginal gains from idea exchange are small relative to marginal congestion costs.

Proposition 3 (Instability of symmetric cities)**.** *Suppose Assumptions 1 and 2 hold.*

(a) If the population elasticity of congestion costs γ *is sufficiently small, two cities of equal population size with positive idea exchange cannot coexist in a locally stable equilibrium.*

(b) If the production function is equation (5) and A *is sufficiently large, two cities of equal population size with positive idea exchange cannot coexist in a locally stable equilibrium.*

(c) *If the production function is equation (5) and* $\sup\{z : \mu(z,c) > 0 \text{ or } \mu(z,c') > 0\}$ *is sufficiently large, then cities* c *and* c' *cannot coexist with* $L_c = L_{c'}$ *and* $Z_c = Z_{c'} > 0$ *in a locally stable equilibrium.*

In a canonical model with symmetric fundamentals and homogeneous agents (i.e., Henderson 1974), net agglomeration benefits are strictly concave in city size. Stability is therefore closely connected to whether the city is smaller or larger than the utility-maximizing population size. In our model, net agglomeration benefits are not necessarily concave, but stability is still defined in terms of the relative strength of agglomeration and congestion forces at the margin. As an example, with a uniform ability distribution and the functional forms in equations (5) and (6), agglomeration benefits are bounded from above for a given value of A, while the congestion costs in equation (7) are a convex function of population size. Thus, two identical cities of sufficient population size could be a stable equilibrium because their size generates a sufficiently large marginal congestion cost. As parts (a) and (b) of Proposition 3 report, the comparison of marginal congestion costs and marginal agglomeration benefits is governed by γ and A.

In distinction from the canonical model, our model's heterogeneity of abilities can make symmetric equilibria unstable. While all tradables producers face the same congestion costs, their benefits from idea exchange are heterogeneous, as higher-ability individuals benefit more from better opportunities. When these differences in benefits are unbounded from above, as in part (c) of Proposition 3, a symmetric equilibrium cannot be stable. Individuals of arbitrarily high ability have arbitrarily high willingness to pay for a better idea-exchange environment, so any perturbation generating a difference in idea-exchange benefits breaks the symmetric arrangement.

Finally, our sufficient conditions for existence of a two-city equilibrium with heterogeneous cities are also sufficient for it to be locally stable.

Proposition 4 (Stability of two heterogeneous cities)**.** *Suppose Assumptions 1, 2, and 3 hold. If the agglomeration, congestion, and feasibility conditions defined in Appendix A.3 hold, there exists a locally stable equilibrium with two heterogeneous cities.*

A.5 Properties of production functions

A.5.1 Assumptions on $B(1-\beta, z, Z_c)$ that imply Assumption 2

Assumption 2 is written in terms of the function $\tilde{z}(z, Z_c)$, which depends on an optimizing individual's choice of $1-\beta$. Assumption 2′ states conditions on the production function for tradables $B(1-\beta, z, Z_c)$ sufficient for Assumption 2 to be true.

Assumption 2′. *$B(1-\beta, z, Z)$ is supermodular in $(1-\beta, z, Z)$ and if $\beta < 1$, $B(1-\beta, z, Z_c)$ is strictly supermodular in (z, Z_c).*

Assumption 2′ implies Assumption 2 by Theorems 2.7.6 and 2.7.7 of Topkis (1998).

While sufficient, Assumption 2′ is not necessary. The production function in equation (5) satisfies Assumption 2 but not Assumption 2′.

A.5.2 Examples of production functions satisfying Condition 1

What are examples of production functions satisfying Condition 1? The main text focuses on the case of equation (5). Here, we define a class of $B(1-\beta, z, Z_c)$ functions that satisfy Condition 1 because their ability elasticity of tradable output is constant and therefore trivially non-decreasing in z and Z_c.

Suppose that $B(1-\beta, z, Z_c)$ can be written as $B(1-\beta, z, Z_c) = z\mathbb{B}(1-\beta, Z_c)$. Note that the output-maximizing choice of β is independent of ability z for this class of production functions. Denoting this optimal choice $\beta(Z_c)$, tradables output is $\tilde{z}(z, Z_c) = z\mathbb{B}(1-\beta(Z_c), Z_c)$. Thus, the ability elasticity of tradable output is $\frac{\partial \ln \tilde{z}(z, Z_c)}{\partial \ln z} = 1$ and the production function satisfies Condition 1.

We now identify conditions such that $B(1-\beta, z, Z_c) = z\mathbb{B}(1-\beta, Z_c)$ satisfies Assumptions 1 and 2. If $\mathbb{B}(1-\beta, Z_c)$ is continuous, strictly positive, strictly concave in $1-\beta$, and increasing in Z_c, if $\mathbb{B}(0, Z_c) = 1\ \forall Z_c$, and if $\mathbb{B}(1-\beta, 0) = \beta$, then the production function satisfies Assumption 1. If $\mathbb{B}(1-\beta, Z_c)$ is continuously differentiable, increasing in Z_c, and strictly increasing in Z_c when $\beta < 1$, then the production function satisfies Assumption 2. If $\mathbb{B}(1-\beta, Z_c)$ is supermodular, then the production function satisfies Assumption 2′.

A.6 Proofs

This appendix contains proofs of our main results.

A.6.1 Special case

The special case described in equations (5) and (6) satisfies Assumptions 1-3. The $B(1-\beta, z, Z_c)$ specified in equation (5) satisfies Assumption 1. To confirm that it satisfies Assumption 2, we explicitly derive $\tilde{z}(z, Z_c)$ and $\frac{\partial^2}{\partial z \partial Z_c}\tilde{z}(z, Z_c)$ by solving for $1 - \beta(z, Z_c)$:

$$1 - \beta(z, Z_c) = \begin{cases} 1 - \frac{1}{2}\frac{AZ_c z+1}{AZ_c z} & \text{if } AZ_c z \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

The resulting $\tilde{z}(z, Z_c)$ and $\frac{\partial^2 \tilde{z}(z,Z_c)}{\partial z \partial Z_c}$ are

$$\tilde{z}(z, Z_c) = \begin{cases} \frac{1}{4AZ_c}\left(AZ_c z + 1\right)^2 & \text{if } AZ_c z \geq 1 \\ z & \text{otherwise} \end{cases} \qquad \frac{\partial^2}{\partial z \partial Z_c}\tilde{z}(z, Z_c) = \begin{cases} \frac{Az}{2} & \text{if } AZ_c z \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

The twice-differentiable function $\tilde{z}(z, Z_c)$ is supermodular if and only if $\frac{\partial^2}{\partial z \partial Z_c}\tilde{z}(z, Z_c) \geq 0$ (Topkis, 1998). It is thus evident that this $\tilde{z}(z, Z_c)$ satisfies Assumption 2.

The function $Z(\{(1-\beta_{z,c})\}, \{L \cdot \mu(z, c)\})$ specified in equation (6) satisfies Assumption 3. $Z(\cdot, \cdot) = 0$ if $M_c = 0$. It is continuous. It is bounded above because $\bar{z}_c \leq \sup\{z : 1 - \beta_{z,c} > 0, \mu(z, c) > 0\}$. If $\{(1 - \beta_{z,c})\mu(z, c)\}$ stochastically dominates $\{(1 - \beta_{z,c'})\mu(z, c')\}$, then $\bar{z}_c \geq \bar{z}_{c'}$. If $\bar{z}_c \geq \bar{z}_{c'}$ and $M_c > M_{c'}$, then $Z(\{1 - \beta_{z,c}\}, \{L \cdot \mu(z, c)\}) > Z(\{1 - \beta_{z,c'}\}, \{L \cdot \mu(z, c')\})$, satisfying Assumption 3.

The $B(1 - \beta, z, Z_c)$ specified in equation (5) satisfies Condition 1. Using the expression for $\tilde{z}(z, Z_c)$ above, we obtain the following ability elasticity of tradable output:

$$\frac{\partial \ln \tilde{z}\left(z, Z_c\right)}{\partial \ln z} = \begin{cases} \frac{2AZ_c z}{AZ_c z+1} & \text{if } AZ_c z \geq 1 \\ 0 & \text{otherwise} \end{cases} .$$

This is non-decreasing in Z_c and z.

A.6.2 Lemma 1: Comparative advantage

Lemma 1: Suppose that Assumption 1 holds. There is an ability level z_m such that individuals of greater ability produce tradables and individuals of lesser ability produce non-tradables.

$$\sigma(z) = \begin{cases} t & \text{if } z > z_m \\ n & \text{if } z < z_m \end{cases}$$

Proof. First, we can identify an ability level dividing tradables and non-tradables producers in each city. Denote it $z_{m,c}$. Consider city c with price $p_{n,c} \geq 0$ and idea-exchange opportunities Z_c. If $p_{n,c} > \tilde{z}(\sup(z), Z_c)$, then $z_{m,c} = \sup(z)$ and all individuals in c produce non-tradables. If $p_{n,c} < \tilde{z}(\inf(z), Z_c)$, then $z_{m,c} = \inf(z)$ and all individuals in c produce tradables. Otherwise, since tradables output $\tilde{z}(z, Z_c)$ is strictly increasing and continuous in z by Assumption 1, there is a unique value $z_{m,c}$ such that $p_{n,c} = \tilde{z}(z_{m,c}, Z_c)$. Individuals of ability $z < z_{m,c}$ produce non-tradables and individuals of ability $z > z_{m,c}$ produce tradables in city c.

Second, there is an ability level dividing tradables and non-tradables producers across all locations, which we denote z_m. Individuals of ability $z \leq z_m$ produce non-tradables and individuals of ability $z \geq z_m$ produce tradables. Suppose not. If there is not an ability level dividing tradables and non-tradables production across all locations, there are abilities z', z'' such that, without loss of generality, $z' < z''$ and z' produces tradables in city c' and z'' produces non-tradables in city c''. The former's choice means $\tilde{z}(z', Z_{c'}) - p_{n,c'}\bar{n} - p_{h,c'} \geq (1-\bar{n})p_{n,c''} - p_{h,c''}$. The latter's choice means $(1-\bar{n})p_{n,c''} - p_{h,c''} \geq \tilde{z}(z'', Z_{c'}) - p_{n,c'}\bar{n} - p_{h,c'}$. Together, these imply $\tilde{z}(z', Z_{c'}) \geq \tilde{z}(z'', Z_{c'})$, contrary to the fact that $\tilde{z}(z, Z_c)$ is strictly increasing in z by Assumption 1. $\square$

A.6.3 Lemma 2: Spatial sorting

Lemma 2: Suppose that Assumptions 1 and 2 hold. For $z > z' > z_m$, if $\mu(z, c) > 0$, $\mu(z', c') > 0$, $\beta(z, Z_c) < 1$, and $\beta(z', Z_{c'}) < 1$, then $Z_c \geq Z_{c'}$.

Proof. By Assumption 1, $\beta(z, Z_c) < 1$ implies $(z, Z_c) \in \otimes$ and $\beta(z', Z_{c'}) < 1$ implies $(z', Z_{c'}) \in \otimes$.
$\mu(z, c) > 0 \Rightarrow \tilde{z}(z, Z_c) - \bar{n}p_{n,c} - p_{h,c} \geq \tilde{z}(z, Z_{c'}) - \bar{n}p_{n,c'} - p_{h,c'}$
$\mu(z', c') > 0 \Rightarrow \tilde{z}(z', Z_{c'}) - \bar{n}p_{n,c'} - p_{h,c'} \geq \tilde{z}(z', Z_c) - \bar{n}p_{n,c} - p_{h,c}$
Therefore $\tilde{z}(z, Z_c) + \tilde{z}(z', Z_{c'}) \geq \tilde{z}(z, Z_{c'}) + \tilde{z}(z', Z_c)$. By Assumption 2, $\tilde{z}$ is strictly supermodular on $\otimes$, so this inequality requires $Z_c \geq Z_{c'}$. $\square$

A.6.4 Proposition 1: Heterogeneous cities' characteristics

Proposition 1: Suppose that Assumptions 1 and 2 hold. In any equilibrium, a larger city has higher housing prices, higher non-tradables prices, a better idea-exchange environment, and higher-ability tradables producers. If $L_c > L_{c'}$ in equilibrium, then $p_{h,c} > p_{h,c'}$, $p_{n,c} > p_{n,c'}$, $Z_c > Z_{c'}$, and $z > z' > z_m \Rightarrow \mu(z, c)\mu(z', c') \geq \mu(z, c')\mu(z', c) = 0$.

Proof.

- Equation (7) says that $L_c > L_{c'} \iff p_{h,c} > p_{h,c'}$.
- Equation (13) says that $p_{h,c} > p_{h,c'} \iff p_{n,c} > p_{n,c'}$

- If $p_{h,c} > p_{h,c'}$ and $p_{n,c} > p_{n,c'}$, then $Z_c > Z_{c'}$. Suppose not. Then, since $\tilde{z}(z, Z_c)$ is increasing in Z_c by Assumption 1, $\tilde{z}(z, Z_c) - \bar{n}p_{n,c} - p_{h,c} < \tilde{z}(z, Z_{c'}) - \bar{n}p_{n,c'} - p_{h,c'}$ $\forall z > z_m$ and $\mu(z, c) = 0$ $\forall z > z_m$. Then $L_c = 0$ by equations (9) and (10), contrary to the premise that $L_c > L_{c'}$.
- If $z > z' > z_m$ and $L_c > L_{c'}$, then $\mu(z, c')\mu(z', c) = 0$. Suppose not, such that $\mu(z, c')\mu(z', c) > 0$. By equations (1) and (2) and Assumption 2, if z' strictly prefers c to c' then z strictly prefers c to c'. Since $Z_c > Z_{c'}$, $\mu(z, c')\mu(z', c) > 0$ is possible only if z and z' are both indifferent between c and c'. Since $p_{n,c} > p_{n,c'}$ and $p_{h,c} > p_{h,c'}$, such indifference implies that $(z, Z_c) \in \otimes$, $(z', Z_c) \in \otimes$ and $\tilde{z}(z, Z_c) - \tilde{z}(z', Z_c) = \tilde{z}(z, Z_{c'}) - \tilde{z}(z', Z_{c'})$. By continuity of $\tilde{z}(z, Z_c)$, there exists a $Z'' \in (Z_{c'}, Z_c)$ such that $(z, Z'') \in \otimes$ and $(z', Z'') \in \otimes$. By Assumption 2, the strict supermodularity of $\tilde{z}$ on $\otimes$, $\tilde{z}(z, Z_c) - \tilde{z}(z', Z_c) > \tilde{z}(z, Z'') - \tilde{z}(z', Z'')$. By Assumption 2, the supermodularity of $\tilde{z}$, $\tilde{z}(z, Z'') - \tilde{z}(z', Z'') \geq \tilde{z}(z, Z_{c'}) - \tilde{z}(z', Z_{c'})$. Thus $\tilde{z}(z, Z_c) - \tilde{z}(z', Z_c) > \tilde{z}(z, Z_{c'}) - \tilde{z}(z', Z_{c'})$, so z and z' cannot both be indifferent between c and c'. $\mu(z, c')\mu(z', c) = 0$.

□

A.6.5 Proposition 2: Skill premia

Proposition 2: Suppose that Assumptions 1 and 2 hold. In an equilibrium in which the smallest city has population L_1 and the second-smallest city has population $L_2 > L_1$,

1. if the ability distribution is decreasing, $\mu'(z) \leq 0$, $\tilde{z}(z, Z_c)$ is log-convex in z, and $\tilde{z}(z, Z_c)$ is log-supermodular, then $\frac{\bar{w}_2}{p_{n,2}} > \frac{\bar{w}_1}{p_{n,1}}$;
2. if the ability distribution is Pareto, $\mu(z) \propto z^{-k-1}$ for $z \geq z_{\min}$ and $k > 0$, and the production function satisfies Condition 1, then $\frac{\bar{w}_2}{p_{n,2}} > \frac{\bar{w}_1}{p_{n,1}}$;
3. if the ability distribution is uniform, $z \sim U(z_{\min}, z_{\max})$, the production function satisfies Condition 1, and $\frac{L_2 - L_1}{L_1^2} > \frac{1}{L} \frac{(1-\bar{n})(z_{\max} - z_{\min})}{z_{\min} + \bar{n}(z_{\max} - z_{\min})}$, then $\frac{\bar{w}_2}{p_{n,2}} > \frac{\bar{w}_1}{p_{n,1}}$.

Proof. By $L_2 > L_1$ and Proposition 1, the abilities of tradables producers in the two cities are intervals, which we can denote by (z_m, z_b) and $(z_b, \hat{z})$. The skill premium is higher in city 2 when $\frac{\bar{w}_2}{p_{n,2}} > \frac{\bar{w}_1}{p_{n,1}}$, which can rewritten as

$$\frac{1}{L_2 p_{n,2}} \int_{z_b}^{\hat{z}} \tilde{z}(z, Z_2)\mu(z)dz > \frac{1}{L_1 p_{n,1}} \int_{z_m}^{z_b} \tilde{z}(z, Z_1)\mu(z)dz.$$

We now obtain this condition in four steps.

1. Implicitly define the function $f(z)$ by a differential equation, $f'(z) = \frac{L_2}{L_1}\frac{\mu(z)}{\mu(f(z))}$, with the endpoint $f(z_m) = z_b$.

2. $\tilde{z}\left(f\left(z_m\right), Z_2\right) > \frac{p_{n,2}}{p_{n,1}}\tilde{z}\left(z_m, Z_1\right)$ because $\tilde{z}\left(z_m, Z_1\right) = p_{n,1}$ and $\tilde{z}\left(z_b, Z_2\right) > p_{n,2}$.

3. If $\frac{\partial \ln(\tilde{z}(x,Z_2))}{\partial x}|_{x=f(z)}f'(z) \geq \frac{\partial \ln(\tilde{z}(z,Z_1))}{\partial z}$ $\forall z \in (z_m, z_b)$, then $\tilde{z}\left(f\left(z\right), Z_2\right) > \frac{p_{n,2}}{p_{n,1}}\tilde{z}\left(z, Z_1\right)$ $\forall z \in (z_m, z_b)$.

4. Multiplying each side of $\tilde{z}\left(f(z), Z_2\right) > \frac{p_{n,2}}{p_{n,1}}\tilde{z}\left(z, Z_1\right)$ by $\mu(z)$ and integrating from z_m to z_b yields the desired result after a change of variables:

$$\begin{aligned}
&\int_{z_m}^{z_b} \tilde{z}\left(f(z), Z_2\right)\mu(z)dz > \frac{p_{n,2}}{p_{n,1}}\int_{z_m}^{z_b} \tilde{z}\left(z, Z_1\right)\mu(z)dz \\
\iff &\int_{z_m}^{z_b} \tilde{z}\left(f(z), Z_2\right) f'(z)\mu(f(z))\frac{L_1}{L_2}dz > \frac{p_{n,2}}{p_{n,1}}\int_{z_m}^{z_b} \tilde{z}\left(z, Z_1\right)\mu(z)dz \\
\iff &\frac{1}{L_2 p_{n,2}}\int_{z_b}^{\hat{z}} \tilde{z}(z, Z_2)\mu(z)dz > \frac{1}{L_1 p_{n,1}}\int_{z_m}^{z_b} \tilde{z}\left(z, Z_1\right)\mu(z)dz
\end{aligned}$$

The sufficient condition in step three, $\frac{\partial \ln(\tilde{z}(x,Z_2))}{\partial x}|_{x=f(z)}f'(z) \geq \frac{\partial \ln(\tilde{z}(z,Z_1))}{\partial z}$ $\forall z \in (z_m, z_b)$, depends jointly on the production function $B(1-\beta, z, Z_c)$, the ability distribution $\mu(z)$, and endogenous equilibrium outcomes. Knowing only $Z_2 > Z_1$, $L_2 > L_1$, the following joint assumptions on $\tilde{z}(z, Z)$ and $\mu(z)$ are sufficient to yield the result:

1. Suppose the ability distribution is decreasing, $\mu'(z) \leq 0$, $\tilde{z}(z, Z_c)$ is log-supermodular, and $\tilde{z}(z, Z_c)$ is log-convex in z. If $\mu'(z) \leq 0$, then $f'(z) \geq 1$. If $\tilde{z}(z, Z_c)$ is log-supermodular in (z, Z_c) and log-convex in z, then $\frac{\partial \ln(\tilde{z}(x,Z_2))}{\partial x} \geq \frac{\partial \ln(\tilde{z}(z,Z_1))}{\partial z}$ for any $x \geq z$, including $x = f(z)$. Thus, $\frac{\partial \ln(\tilde{z}(x,Z_2))}{\partial x}|_{x=f(z)}f'(z) \geq \frac{\partial \ln(\tilde{z}(z,Z_1))}{\partial z}$ $\forall z \in (z_m, z_b)$.

2. The condition in step three, $\frac{\partial \ln(\tilde{z}(x,Z_2))}{\partial x}|_{x=f(z)}f'(z) \geq \frac{\partial \ln(\tilde{z}(z,Z_1))}{\partial z}$, can be written as

$$\frac{\partial \ln(\tilde{z}(x, Z_2))}{\partial \ln x}\Big|_{x=f(z)}\frac{L_2}{L_1}\frac{\mu(z)}{\mu(f(z))}\frac{z}{f(z)} \geq \frac{\partial \ln(\tilde{z}(z, Z_1))}{\partial \ln z}.$$

   If the ability distribution is Pareto, $\mu(z) \propto z^{-k-1}$ for $z \geq z_{\min}$ and $k > 0$, then $\frac{\mu(z)}{\mu(f(z))}\frac{z}{f(z)} = \left(\frac{f(z)}{z}\right)^k$. The inequality is true because $Z_2 > Z_1$, $L_2 > L_1$, $f(z) > z$, and by Condition 1 the ability elasticity of tradable output is non-decreasing in z and Z_c.

3. Suppose the ability distribution is uniform, $z \sim U\left(z_{\min}, z_{\max}\right)$, the production function satisfies Condition 1, and $\frac{L_2 - L_1}{L_1^2} > \frac{1}{L}\frac{(1-\bar{n})(z_{\max}-z_{\min})}{z_{\min}+\bar{n}(z_{\max}-z_{\min})}$. In this case, the condition in step

three can be written as

$$\frac{\partial \ln(\tilde{z}(x, Z_2))}{\partial \ln x}\Big|_{x=f(z)} \frac{L_2}{L_1}\frac{z}{f(z)} \geq \frac{\partial \ln(\tilde{z}(z, Z_1))}{\partial \ln z}.$$

Since $f(z) = z_b + \frac{L_2}{L_1}(z - z_m)$, $\frac{L_2}{L_1}\frac{z}{f(z)} \geq 1 \iff \frac{L_2}{L_1} z_m \geq z_b$. $\frac{L_2 - L_1}{L_1^2} > \frac{1}{L}\frac{(1-\bar{n})(z_{\max} - z_{\min})}{z_{\min} + \bar{n}(z_{\max} - z_{\min})} = \frac{1}{L}\frac{z_{\max} - z_m}{z_m}$ implies that

$$\frac{L_2}{L_1} z_m - z_b = \frac{L_2 - L_1}{L_1} z_m - \frac{L_1}{L}(z_{\max} - z_m) > 0.$$

This inequality and the fact that the ability elasticity of tradable output is non-decreasing in z and Z_c are sufficient for the inequality in step three to be true.

□

A.6.6 Proposition 3: Instability of symmetric equilibria

Proof. Suppose $L_1 = L_2$ and $Z_1 = Z_2 > 0$. Without loss of generality, consider perturbations of size $\epsilon \leq \bar{\epsilon}$ moving individuals from city 1 to city 2. By Assumption 2, the highest-ability producers have the most to gain from a move and it is sufficient to consider perturbations of size ϵ in which all tradables producers in the range $[z^*(\epsilon), \infty]$ move from city 1 to city 2; these are perturbations $d\mu$ that satisfy $L\int_{z^*(\epsilon)}^{\infty} \mu(z,1)dz = (1-\bar{n})\epsilon$ and $d\mu(z,2) = -d\mu(z,1) = \mu(z,1)\ \forall z \geq z^*(\epsilon)$. Since an interval of the highest-ability tradables producers, accompanied by the appropriate mass of non-tradables producers, moves from city 1 to city 2, $Z_2' > Z_1'$ and $L_2' > L_1'$ with $L_2' = L_1 + \epsilon$ and $L_1' = L_1 - \epsilon$. Denote $\hat{z} = \sup\{z : \mu(z,1) > 0\}$. The equilibrium is stable with respect to this perturbation only if

$$\tilde{z}(\hat{z}, Z_2') - \tilde{z}(\hat{z}, Z_1') \leq \frac{\theta}{1-\bar{n}}\left((L_1 + \epsilon)^{\gamma} - (L_1 - \epsilon)^{\gamma}\right)$$

By Assumptions 1 and 2, $Z_2' > Z_1'$, and $Z_2' > 0$, the left side is strictly greater than zero. The right side is arbitrarily small if γ is arbitrarily small. This proves part (a). If the production function is that of equation (5), the left side is increasing without bound in A and z. This proves parts (b) and (c). This inequality is violated if A or $\hat{z}$ is sufficiently high relative to γ. □

A.6.7 Proposition 4: Stability of two heterogeneous cities

Proof. Appendix section A.3 shows that these three conditions are sufficient for the existence of an equilibrium with two cities in which $L_1 < L_2$ and $\Omega(L_1)$ crosses zero from above. Amongst tradables producers in city 1, those with the most to gain by moving to city 2 are those of the highest ability. Amongst tradables producers in city 2, those with the most to gain by moving to city 1 are those of the lowest ability. It is therefore sufficient to consider perturbations that are changes in z_b and consummate changes in $z_{b,n}$ as defined in appendix section A.3. Since $\Omega(L_1)$ crosses zero from above, this equilibrium is stable. □

B Numerical results

This appendix reports numerical results that complement the analytical results in Proposition 2. In all our numerical work, we use the functional forms for $B(\cdot)$ and $Z(\cdot)$ given by equations (5) and (6). Section B.1 shows that the sufficient condition on the equilibrium size of the smallest city in the uniform-ability case of Proposition 2 is typically true in two-city equilibria and that larger cities exhibit lower skill premia only when the skill premia are unrealistically large. Sections B.2 and B.3 extend our results for uniform and Pareto ability distributions, respectively, to greater numbers of cities. The overwhelming pattern is that larger cities have higher skill premia.

B.1 Uniform ability distribution and two cities

In the case of the uniform ability distribution, the sufficient condition in Proposition 2 is written in terms of exogenous parameters and the two cities' equilibrium population sizes, L_1 and L_2. For the larger city's skill premium to be lower, it must be the case that $\frac{L_2-L_1}{L_1^2} < \frac{1}{L}\frac{z_{\max}-z_m}{z_m}$. This will occur when L_1 is sufficiently large. However, we also know that the relative compensation effect on the right-hand side of inequality (14) approaches zero as $L_1 \to L_2$, so it is clear that this sufficient condition is not necessary for the larger city to have a higher skill premium. An equilibrium in which the larger city has a lower skill premium must exhibit some intermediate value of L_1.

To examine whether such an equilibrium exists and to more generally characterize the properties of two-city equilibria when the ability distribution is uniform, we compute equilibria for a range of parameter values. Our choice of the parameter values is admittedly arbitrary, but the results are sufficiently stark that they are suggestive of broader patterns.

We examine vectors of the form $[A, \bar{n}, \theta, \gamma, L, \nu, z_{\min}, z_{\max}]$ obtained by combining the following possible parameter values: $A \in [1, 2, 3, 4, 5, 10], \bar{n} \in [.1, .2, .3, .4, .5], \theta \in [.1, .5, 1, 2, 5], \gamma \in [.01, .1, .5, 1, 5], L \in [2, 6, 10, 15, 20, 40], \nu \in [1, 5, 10, 25, 50], z_{\min} \in [0, 1, 2.5, 5, 10, 25, 50], z_{\max} - z_{\min} \in [1, 5, 10, 25, 50]$. The Cartesian product of these sets has 787,500 elements. For each parameter vector, we seek values of L_1 and L_2, with $L_1 < L_2 = L - L_1$, constituting an equilibrium as defined in section I.B.

A large number of these 787,500 parameter combinations are inconsistent with the existence of any two-city equilibrium. We nonetheless explore these parts of the parameter space in order to identify exceptions to the pattern predicted by Proposition 2. For example, we find that a two-city equilibrium often does not exist when $z_{\max} - z_{\min}$ is large, but these parameter combinations also are more likely to violate the sufficient condition in Proposition 2 and yield an equilibrium in which the larger city has a smaller skill premium. The cost of exploring the extremes of the parameter space is that sometimes no equilibrium is feasible and sometimes the entire population lives in a single city.

In some cases, existence of an equilibrium can be ruled out prior to computing potential solutions. Consider two conditions that are necessary for a two-city equilibrium to exist. A modest feasibility condition is $\tilde{z}(\underline{z}, z_{\max}) \geq \theta\,(L/2)^{\gamma}$, which requires that the greatest conceivable tradables output for the median tradables producer be greater than the lowest conceivable housing cost in the larger city. If this failed, every conceivable two-city population allocation would be infeasible. A modest agglomeration condition is $A \cdot z_{\max} \cdot \underline{z} > 1$, which requires that the median tradables producer would find idea exchange with the most able producer profitable. If this failed, there would be no benefits to agglomeration. Of the 787,500 parameter combinations, 94,903 fail the former, 985 fail the latter, and 3,015 fail both modest necessary conditions.

An equilibrium with two heterogeneous cities exists for 58,005 of the parameter combinations. For the parameter vectors that do not yield a two-city equilibrium, this is overwhelmingly due to the entire population agglomerating in a single city (598,197 combinations). Since the necessary conditions described in the previous paragraph are very modest, there are also a number of parameter combinations for which agglomeration is not realized in equilibrium (644) or is insufficient to cover housing costs (31,226).

Of the 58,005 parameter combinations yielding two-city equilibria, only 159 (0.3%) yield equilibria in which the larger city has a lower skill premium. 46,055 of the equilibria satisfy the sufficient condition of Proposition 2 case 3, and 11,791 of the equilibria not satisfying that sufficient condition nonetheless have a higher skill premium in the more populous city. Table B.1 reports the fraction of equilibria in which the larger city has a lower skill premium for each parameter value.

The equilibria in which the larger city has a lower skill premium exhibit implausibly large skill premia. Large skill premia make the relative compensation effect large. Across the 159 equilibria in which the larger city has a lower skill premium, the mean skill premium in the smaller city is 563%. By contrast, the 95$^{\text{th}}$ percentile of $\frac{\bar{w}_1}{p_{n,1}}$ for equilibria with increasing skill premia is only 195%. Recall that, in the data, the college wage premium varies across metropolitan areas in the range of 47% to 71%.

Table B.1: Share of equilibria with decreasing skill premia, by parameter

A		$\bar{n}$		θ		γ		L		ν		$z_{\min}$		$z_{\max} - z_{\min}$	
1	.0017	0.1	0.0134	0.1	0.0038	0.01	0	2	0.0013	1	0.0036	0	0.0096	1	0
2	.0016	0.2	0	0.5	0.0028	0.1	0	6	0.005	5	0.003	1	0.0057	5	0.0003
3	.0035	0.3	0	1	0.0026	0.5	0	10	0.0043	10	0.0027	2.5	0.0019	10	0.001
4	.0034	0.4	0	2	0.0021	1	0	15	0.002	25	0.0025	5	0	25	0.0064
5	.0037	0.5	0	5	0.0024	5	0.0046	20	0	50	0.0023	10	0	50	0.0106
10	.0028							40	0			25	0		
												50	0		

NOTES: This table summarizes the parameter values yielding two-city equilibria in which the larger city has a lower skill premium. For each pair of columns, the first column lists the value of the parameter and the second column lists the share of the 58,005 equilibria in which the premium-size relationship is negative. Since the latter occurs in only 159 cases, these shares are typically less than 1% and often zero.

The parameter values yielding equilibria in which the larger city has a lower skill premium can be understand in terms of facilitating large equilibrium values of $\frac{\bar{w}_1}{p_{n,1}}$. When $z_{\max} - z_{\min}$ is larger and $z_{\min}$ and $\bar{n}$ are smaller, there is greater heterogeneity of ability within tradables producers, raising the value of $\frac{\bar{w}_1}{p_{n,1}}$. Since greater heterogeneity in these abilities generates larger differences in idea-exchange environments, two-city equilibria only exist when these greater agglomeration benefits are offset by higher congestion costs, governed by γ. We obtain a lower skill premium in the larger city only when γ is 5. This is a very large population elasticity of congestion costs. Empirical work typically estimates a value of γ below 0.1; Combes, Duranton and Gobillon (2012) report an estimate of 0.041. Empirically plausible values of the congestion-cost elasticity yield zero cases of non-increasing skill premia.

Thus, our examination of a large set of parameter vectors suggests that the larger city typically has a higher skill premium in two-city equilibria with a uniform ability distribution. The sufficient condition in Proposition 2 holds for most two-city equilibria, and the larger city almost always has a higher skill premium. Deviations from the predicted pattern are produced only by assuming empirically implausible values of γ that generate skill premia much higher than those observed in the data.

B.2 Uniform ability distribution and more than two cities

We now extend the uniform-ability-distribution results to more than two heterogeneous cities. We examine the same values of $[A, \bar{n}, \theta, \gamma, \nu, z_{\min}, z_{\max}]$ examined in the previous section. The population L is proportional to the number of cities under consideration (so as to facilitate existence of these equilibria). That is, $L \in C \times [1, 3, 5, 7.5, 10, 20]$, where C is the number of cities and the previous section considered $C = 2$. There are therefore, again, 787,500 parameter combinations for each C. We solve for equilibria in which $L_1 < L_2 < \cdots < L_C$.

The results for equibrilia with three to seven cities, summarized in Table B.2, are consistent with those found for two cities. First, in the vast majority (more than 99.5%) of equilibria, larger cities have higher skill premia. Second, the correlation between population size and skill premia is frequently positive even when the relationship isn't monotone. As in the two-city case, the exceptions to these patterns occur when equilibria exhibit very large values of $\frac{\bar{w}_1}{p_{n,1}}$. The 95$^{\text{th}}$ percentile of $\frac{\bar{w}_1}{p_{n,1}}$ in equilibria with monotonically increasing premia lies below the 25$^{\text{th}}$ percentile for equilibria with non-monotone premia. These non-monotone equilibria with very high skill premia can arise only when $z_{\max} - z_{\min}$ and γ are large and $z_{\min}$ and $\bar{n}$ are small. For this set of parameter combinations, the value of $\frac{\bar{w}_1}{p_{n,1}}$ is lower in equilibria with larger numbers of cities. Thus, all equilibria with six or seven cities exhibit monotonically increasing skill premia, and all equilibria with four or more cities exhibit positive premia-population correlations.

Table B.2: Uniform-ability equilibria, 2 to 7 cities

Number of cities	Number of equilibria	Share of non-monotone premia	Share of $corr(L_c, \frac{\bar{w}_c}{p_{n,c}}) < 0$	Percentiles of $\bar{w}_1/p_{n,1}$: monotone premia, 95$^{\text{th}}$	Percentiles of $\bar{w}_1/p_{n,1}$: non-monotone premia, 25$^{\text{th}}$	Maximal $\bar{n}$	Maximal $z_{\min}$	Minimal $z_{\max} - z_{\min}$	Minimal γ
2	58005	.0027	.0027	1.95	5.02	0.1	2.5	5	5
3	43367	.0157	.0006	1.46	1.53	0.2	2.5	1	5
4	38300	.0056	0	1.32	1.79	0.1	2.5	5	5
5	33305	.0005	0	1.21	1.41	0.1	0	5	5
6	30903	0	0	1.15					
7	26213	0	0	1.09					

NOTES: This table summarizes the existence and properties of equilibria with the number of cities listed in the first column. The second column lists the number of equilibria that exist for uniformly distributed abilities and the 787,500 parameter combinations described in the text. The third column lists the share of those equilibria that exhibit skill premia that are not monotone in city population size. The fourth column lists the share of equilibria that exhibit negative premia-size correlations. The fifth and sixth columns list the 95$^{\text{th}}$ and 25$^{\text{th}}$ percentiles of $\bar{w}_1/p_{n,1}$ for equilibria with monotonically increasing and non-monotone skill premia, respectively. The seventh through tenth columns list the maximal values of $\bar{n}$ and $z_{\min}$ and minimal values of $z_{\max} - z_{\min}$ and γ that yield equilibria with non-monotone skill premia.

Since the computational burden increases with the number of cities, we have also examined tens of thousands of parameter combinations for the cases of 10-, 20-, and 30-city equilibria, rather than hundreds of thousands. All these equilibria exhibit monotonically increasing skill premia.

In short, for uniformly distributed abilities and over a wide range of parameter values, equilibria typically exhibit monotonically increasing skill premia. In fact, equilibria with larger numbers of heterogeneous cities yield more consistently monotone premia-size relationships than those obtained for the two-city case. The exceptions involve very large relative compensation effects due to very large skill premia.

B.3 Pareto ability distribution and more than two cities

This section extends the analytical result of Proposition 2 for Pareto-distributed abilities to greater numbers of cities. We compute equilibria for a wide range of parameter values to examine their properties. We find that the skill premium is monotonically increasing with city size in every case.

We compute equilibria for vectors of the form $[A, \bar{n}, \theta, \gamma, L, \nu, z_{\min}, k]$, where k is the shape parameter of the Pareto distribution. We examine parameter vectors obtained by combining the following possible values: $A \in [1, 3, 5, 10]$, $\bar{n} \in [.1, .2, .3, .4, .5]$, $\theta \in [.1, .5, 1, 2, 5]$, $\gamma \in [.01, .1, .5, 1, 5]$, $L \in C \times [1, 3, 5, 10]$, $\nu \in [1, 5, 10, 25, 50]$, $k\left(z_{\min}\right)^k \in [1, 5, 10, 50]$, $k \in [2.1, 3, 5, 10, 50]$. The Cartesian product yields 200,000 parameter combinations. We solve for equilibria in which $L_1 < L_2 < \cdots < L_C$.

Once again, a large number of these 200,000 parameter combinations are inconsistent with the existence of equilibrium. For example, in the two-city case, 16,925 do not satisfy the modest feasibility condition that $\tilde{z}(\underline{z}, z_{\max}) \geq \theta\left(L/2\right)^{\gamma}$. For more than half the parameter values (primarily those with high θ and L), there is no pair of L_1 and L_2 that is feasible in the sense that tradables output is less than congestion costs. Nonetheless, we examine a wide range of parameter values in an effort to find a counterexample. As Table B.3 reports, we find none. Among hundreds of thousands of parameter combinations, zero yield a case in which a larger city has a lower skill premium. This suggests that the result proved in Proposition 2 for two cities extends to all cities in equilibrium when ability is Pareto distributed.[4]

[4] While we have found that skill premia are increasing in city size for every parameter vector examined in the case of the Pareto ability distribution, the technique employed to analytically prove the two-city result in Proposition 2 cannot be extended to apply to an arbitrary number of cities. Step 2 of our proof employs the fact that $\tilde{z}\left(f\left(z_m\right), Z_2\right) > \frac{p_{n,2}}{p_{n,1}}\tilde{z}\left(z_m, Z_1\right)$, where z_m is the least talented tradables producer in city 1. The analogous condition that, for example, $\tilde{z}\left(f\left(z_{b,1}\right), Z_3\right) > \frac{p_{n,3}}{p_{n,2}}\tilde{z}\left(z_{b,1}, Z_2\right)$, where $z_{b,1}$ is the lowest-ability tradables producer in city 2, is not necessarily true in equilibrium; in fact, some of the equilibria reported in Table B.3 fail to exhibit this property.

Table B.3: Pareto-ability equilibria, 2 to 7 cities

Number of cities	Number of equilibria	Number of non-monotone premia
2	31806	0
3	18871	0
4	17461	0
5	15214	0
6	14388	0
7	12643	0

NOTES: This table summarizes the existence and properties of equilibria with the number of cities listed in the first column. The second column lists the number of equilibria that exist for Pareto-distributed abilities and 200,000 parameter combinations described in the text. The third column lists the number of those equilibria that exhibit skill premia that are not monotone in city population size.

C Data and estimates

C.1 Data description

Data sources: Our population data are from the US Census website (1990, 2000, 2007). Our data on individuals' wages, education, demographics, and housing costs come from public-use samples of the decennial US Census and the annual American Community Survey made available by IPUMS-USA (Ruggles et al., 2010). We use the 1990 5% and 2000 5% Census samples and the 2005-2007 American Community Survey 3-year sample. We use the 2005-2007 ACS data because ACS data from 2008 onwards only report weeks worked in intervals.

Wages: We exclude observations missing the age, education, or wage income variables. We study individuals who report their highest educational attainment as a high-school diploma or GED or a bachelor's degree and are between ages 25 and 55. We study full-time, full-year employees, defined as individuals who work at least 40 weeks during the year and usually work at least 35 hours per week. We obtain weekly and hourly wages by dividing salary and wage income by weeks worked during the year and weeks worked times usual hours per week. Following Acemoglu and Autor (2011), we exclude observations reporting an hourly wage below $1.675 per hour in 1982 dollars, using the GDP PCE deflator. We define potential work experience as age minus 18 for high-school graduates and age minus 22 for individuals with a bachelor's degree. We weight observations by the "person weight" variable provided by IPUMS.

Housing: To calculate the average housing price in a metropolitan statistical area, we use all observations in which the household pays rent for their dwelling that has two or three bedrooms. We do not restrict the sample by any labor-market outcomes. We drop observations that lack a kitchen or phone. We calculate the average gross monthly rent for each metropolitan area using the "household weight" variable provided by IPUMS.

Note that both income and rent observations are top-coded in IPUMS data.

College ratio: Following Beaudry, Doms and Lewis (2010), we define the "college ratio" as the number of employed individuals in the MSA possessing a bachelor's degree or higher educational attainment plus one half the number of individuals with some college relative to the number of employed individuals in the MSA with educational attainment less than college plus one half the number of individuals with some college. We weight observations by the "person weight" variable provided by IPUMS.

Geography: We map the public-use microdata areas (PUMAs) to metropolitan statistical areas (MSAs) using the "'MABLE Geocorr90, Geocorr2K, and Geocorr2010" geographic correspondence engines from the Missouri Census Data Center. For 1990 and 2000, we consider both primary metropolitan statistical areas (PMSAs) and consolidated metropolitan statistical areas (CMSAs). The 2005-2007 geographies are MSAs. In some sparsely populated areas, only a fraction of a PUMA's population belongs to a MSA. We include PUMAs that have more than 50% of their population in a metropolitan area. Table 1 describes PMSAs in 2000.

C.2 Empirical estimates

Our empirical approach is to estimate cities' college wage premia and then study spatial variation in those premia. Our first-stage estimates of cities' skill premia are obtained by comparing the average log hourly wages of full-time, full-year employees whose highest educational attainment is a bachelor's degree to those whose highest educational attainment is a high school degree.

Our first specification uses the difference in average log hourly wages y in city c without any individual controls as the first-stage estimator. The dummy variable college_i indicates that individual i is a college graduate. Expectations are estimated by their sample analogues.

$$\text{premium}_c = \mathbb{E}(y_{ic}|\text{college}_i = 1) - \mathbb{E}(y_{ic}|\text{college}_i = 0)$$

Our second approach uses a first-stage Mincer regression to estimate cities' college wage premia after controlling for experience, sex, and race. The first-stage equation describing variation in the log hourly wage y of individual i in city c is

$$y_i = \gamma X_i + \alpha_c + \rho_c \text{college}_i + \epsilon_i$$

X_i is a vector containing years of potential work experience, potential experience squared, a dummy variable for males, dummies for white, Hispanic, and black demographics, and the college dummy interacted with the male and demographic dummies. The estimated skill premium in each city, $\hat{\rho}_c$, is the dependent variable used in the second-stage regression. We refer to these estimates as "composition-adjusted skill premia."

One may be inclined to think that the estimators that control for individual characteristics are more informative. But if differences in demographics or experience are correlated with differences in ability, controlling for spatial variation in skill premia attributable to spatial variation in these factors removes a dimension of the data potentially explained by our model. To the degree that individuals' observable characteristics reflect differences in their abilities, the unadjusted estimates of cities' skill premia are more informative for comparing our model's predictions to empirical outcomes.

Table 1 describes variation in skill premia using the first skill-premium measure that lacks individual controls. Table C.1 reports analogous regressions for composition-adjusted skill premia that yield very similar results. In the lower panel, we use a quality-adjusted annual rent from Chen and Rosenthal (2008) that includes both owner-occupied housing and rental properties. This reduces the number of observations because Chen and Rosenthal do not report quality-adjusted rent values for every PMSA in 2000, but the results are very similar.

Table C.2 shows the correlation between estimated skill premia and population sizes for various years and geographies using log weekly wages. These specifications are akin to those appearing in the first column of Table 1 and the first column of the upper panel of C.1.

Table C.1: Skill premia and metropolitan characteristics, 2000

Composition-adjusted skill premia				
log population	0.026	0.029	0.029	0.028
	(0.0031)	(0.0047)	(0.0036)	(0.0045)
log rent		-0.027		0.0051
		(0.032)		(0.034)
log college ratio			-0.027	-0.029
			(0.016)	(0.016)
Observations	325	325	325	325
R^2	0.146	0.151	0.162	0.162
Composition-adjusted skill premia and quality-adjusted rent				
log population	0.025	0.029	0.026	0.029
	(0.0033)	(0.0044)	(0.0038)	(0.0044)
log quality-adjusted rent		-0.030		-0.030
		(0.023)		(0.023)
log college ratio			-0.014	0.000086
			(0.016)	(0.015)
Observations	297	297	297	297
R^2	0.130	0.144	0.134	0.144

Notes: Robust standard errors in parentheses. In both panels, the dependent variable is a metropolitan area's skill premium, measured as the difference in average log hourly wages between college and high school graduates after controlling for for experience, sex, and race. The upper panel uses average gross monthly rent; the lower panel uses quality-adjusted annual rent from Chen and Rosenthal (2008). Details in text of appendix C.

Table C.2: Skill premia and metropolitan populations

Dependent variable	1990 PMSA	1990 CMSA	2000 PMSA	2000 CMSA	2005-7 MSA
Skill premia	0.015	0.014	0.033	0.029	0.040
	(0.0038)	(0.0039)	(0.0038)	(0.0036)	(0.0038)
Composition-adjusted	0.013	0.013	0.029	0.025	0.028
skill premia	(0.0030)	(0.0031)	(0.0032)	(0.0030)	(0.0033)
Observations	322	271	325	270	353

Notes: Robust standard errors in parentheses. Each cell reports the coefficient and standard error for log population from an OLS regression of the estimated college premia for weekly wages on log population (and a constant). The sample is full-time, full-year employees whose highest educational attainment is a bachelor's degree or a high-school degree.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics.* Vol. 4, , ed. O. Ashenfelter and D. Card, 1043–1171. Elsevier.

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

Beaudry, Paul, Mark Doms, and Ethan Lewis. 2010. "Should the Personal Computer Be Considered a Technological Revolution? Evidence from U.S. Metropolitan Areas." *Journal of Political Economy*, 118(5): 988 – 1036.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive Cities: Sorting, Selection, and Agglomeration." *Journal of Political Economy*, 122(3): 507 – 553.

Chen, Yong, and Stuart S. Rosenthal. 2008. "Local amenities and life-cycle migration: Do people move for jobs or fun?" *Journal of Urban Economics*, 64(3): 519–537.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2012. "The Costs of Agglomeration: Land Prices in French Cities." C.E.P.R. Discussion Papers CEPR Discussion Papers 9240.

Gaubert, Cecile. 2015. "Firm Sorting and Agglomeration." mimeo.

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

Krugman, Paul. 1991. "Increasing Returns and Economic Geography." *Journal of Political Economy*, 99(3): 483–99.

Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. "Integrated Public Use Microdata Series: Version 5.0 [Machine-readable database]." Minneapolis, MN: Minnesota Population Center.

Topkis, David M. 1998. *Supermodularity and Complementarity.* Princeton University Press.