Separate Online Appendices with Supplemental Material for:

Productive cities: Sorting, selection, and agglomeration

Kristian Behrens

Université du Québec à Montréal (ESG–UQAM), Canada;

CIRPÉE, Canada; and CEPR, UK.

Gilles Duranton

Wharton School, University of Pennsylvania, USA;

SERC, London School of Economics, UK; and CEPR, UK.

Frédéric Robert-Nicoud

Geneva School of Economics and Management, Université de Genève, Switzerland;

SERC, London School of Economics, UK; and CEPR, UK.

Abstract

This document contains a set of appendices with supplemental material. Appendix F provides microeconomic foundations for our specification of urban costs. Appendix G presents results regarding the planner's problem. Appendix H establishes the equivalence between our model and a 'consumer-city' version of that model. Appendix I shows how the main estimating equations of Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2013) can be derived from a simple span-of-control extension of our model. Finally, Appendix J shows that the interpretation of the estimation results of Section 7 relies on principles more general than the assumptions made in the model.

Keywords: sorting; selection; agglomeration; urban premium; city size; Zipf's law.

JEL Classification: J24; R10; R23

F City structure and urban costs

We model cities in a simple way that builds on the pioneering work of Alonso (1964), Muth (1969), and Mills (1967). See Fujita (1989), Zenou (2011), or Duranton and Puga (2013) for more recent treatments. In each city, production takes place at a single point, defined as the central business district (CBD). Surrounding a city's CBD, there is a line with residences of unit length. Residents commute from their residence to the CBD and back at a cost. Commuting costs are paid in numéraire, and we assume that the cost of a resident's round-trip from a location at distance x to the CBD is $t(x) = \tau x^\gamma$ where τ and γ are positive parameters.[1] To keep matters simple and avoid making the differential land rent disappear to absentee landowners, we assume that this rent is taxed in each city and redistributed equally to current residents.

Each resident chooses her location of residence so as to maximise utility given her income and the land rent schedule in the city. Because of fixed lot size, this assumption is equivalent to choosing a location to minimise the sum of the differential land rent and commuting, $r(x) + t(x)$, with respect to x. At the residential equilibrium, the lack of arbitrage across residential locations ensures that this sum is the same for all residents. Lack of arbitrage also implies that the city is symmetric with respect to its edges at a distance $L/2$ from the CBD.

[1] In practice, commuting costs include both the direct monetary cost of travelling and the opportunity cost of the time spent on the journey (Small and Verhoef, 2007). Ignoring the time cost of commuting avoids having to deal with residential choices for individuals with heterogeneous values of time. The location of workers and entrepreneurs within cities is not a focus of this paper. Observe further that the literature often imposes $\gamma = 1$. Recent evidence (e.g., Albouy, 2009; Baum-Snow and Pavan, 2012; Combes, Duranton, and Gobillon, 2013) suggests that, empirically, the elasticity of urban costs to city population is well below unity. We confirm this finding in Section 5 of the paper and show that a small value of γ has important implications regarding the size distribution of cities.

The equilibrium land rent schedule is thus such that

$$\tau x^{\gamma} + r(x) = \tau \times \left(\frac{L}{2}\right)^{\gamma} + r\left(\frac{L}{2}\right), \quad \forall x \in [0, L/2]. \tag{F.1}$$

Without loss of generality, the rent at the city edges can be normalised to zero. This normalisation yields the land rent schedule

$$r(x) = \tau \left[\left(\frac{L}{2}\right)^{\gamma} - x^{\gamma}\right],$$

with $\partial r(x)/\partial x < 0$, that is land rents fall with distance to the CBD. Integrating land rent over the entire city after making use of its symmetry yields total land rent:

$$\mathrm{TLR} = 2\int_0^{L/2} r(x)\mathrm{d}x = \frac{2\tau\gamma}{\gamma+1}\left(\frac{L}{2}\right)^{\gamma+1}. \tag{F.2}$$

For a resident living at distance x from the CBD, urban costs are the sum of her differential rent and her commuting costs minus her share of total land rent. Using equations (F.1) and (F.2) and the normalisation $r(L/2) = 0$ we find, after simplifications, that urban cost for each resident is given by

$$UC(x) \equiv t(x) + r(x) - \frac{\mathrm{TLR}}{N} = \theta L^{\gamma}, \tag{F.3}$$

where $\theta \equiv 2^{-\gamma}\tau(\gamma+1)^{-1}$ is a bundle of parameters. The expression in (F.3) corresponds to the reduced-form proposed in the main text in Section 2.

G The planner's problem

We consider a planner seeking to maximise aggregate output net of urban costs. The planner's problem is to create cities and allocate individuals to them. We denote C the set of potential sites suitable for cities (we assume C to be sufficiently large relative to Λ for some sites to be left empty at the socially optimal allocation). We also let $\mathcal{T} \equiv [\underline{t}, \overline{t}]$. Thus we may write the utilitarian planner's objective as

$$\max_{\{\mathcal{T}(c),\varphi^W(c),\varphi^E(c),L(c)\}_{c\in C}} \int_{c\in C} \left[Y(c) - \theta L(c)^{\gamma+1}\right] \mathrm{d}c$$

such that

$$\begin{aligned} Y(c) &= \left[\int_{\varphi\in\varphi^W(c)} \varphi^a \mathrm{d}F(\varphi,c)\right]\left[\int_{\varphi\in\varphi^E(c)} \varphi^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi,c)\right]^{\varepsilon} L(c)^{1+\varepsilon} \\ f(\varphi,c) &= \int_{t\in\mathcal{T}(c)} \frac{1}{t} g_{t,s,c}\left(t,\frac{\varphi}{t}\right) \mathrm{d}t \\ \Lambda &= \int_{c\in C} L(c)\,\mathrm{d}c, \end{aligned}$$

where $\mathcal{T}(c) \subseteq \mathcal{T}$ is the set of talents allocated to c, $g_{t,s,c}(t,s)$ is the joint probability density function of talents and serendipity in c, $L(t,c)$ is the mass of workers of talent t allocated to c, $L(c) \equiv \int_{t\in\mathcal{T}(c)} L(t,c)\,\mathrm{d}t$ is the population of c, $f(\varphi,c)$ is the density distribution of productivity in c, $\varphi^W(c)$ is the set of productivities in c allocated to production work, and $\varphi^E(c)$ is the complement set of productivities allocated to entrepreneurship.

The first constraint in the maximisation programme above is city c's output, which depends on its productivity distribution. The second constraint relates the density distribution of productivity to the joint distribution function of talent t and serendipity s, where t is taken over $\mathcal{T}(c)$ with density $g_t(t,c)$ and s is taken over $[0,+\infty)$ with density $g_s(s)$. The third constraint is the full-population condition.

Characterising the solution(s) to the problem above is a complex task: the set and composition of cities and the productivity distribution of each city are all endogenous. We proceed in steps.

Optimal selection. We note that the planner will allocate the most productive individuals of each city to entrepreneurship and the least productive to production work since $a < 1/\varepsilon$. Thus, there exists a productivity cutoff in each city, $\underline{\varphi}(c)$, such that $\varphi^W(c) = \left[0, \underline{\varphi}(c)\right]$ and $\varphi^E(c) = \left(\underline{\varphi}(c), +\infty\right)$.

Maximising $Y(c)$ conditional on $L(c)$ and $F(\varphi, c)$ requires solving

$$\max_{\underline{\varphi}(c)} Y(c) = \left[\int_0^{\underline{\varphi}(c)} \varphi^a \mathrm{d}F(\varphi, c)\right] \left[\int_{\underline{\varphi}(c)}^{\infty} \varphi^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi, c)\right]^{\varepsilon} L(c)^{1+\varepsilon}.$$

The unique solution to this problem is the fixed point of:

$$\underline{\varphi}^0(c)^{\frac{1}{\varepsilon}-a} \int_0^{\underline{\varphi}^0(c)} \varphi^a \mathrm{d}F(\varphi, c) = \frac{1}{\varepsilon} \int_{\underline{\varphi}^0(c)}^{\infty} \varphi^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi, c).$$

Note that this expression is identical to its equilibrium counterpart (equation 14 in the main text) despite entrepreneurs charging a (constant) markup. This property arises because the market and the optimal share of profits and labour wages coincide under Dixit-Stiglitz monopolistic competition as a result of constant and identical markups.

Optimal city population size. We work with a continuum of cities. Thus, each city can be sliced into any arbitrary 'number' of smaller cities, and conversely. Since the location of c is immaterial by assumption (we are ruling out natural advantages), the planner may choose the population of city c to maximise its per capita output net of urban costs, conditional on its composition, namely:

$$\max_{L(c)} \frac{Y(c) - \theta L(c)^{\gamma+1}}{L(c)} = A(c)L(c)^{\varepsilon} - \theta L(c)^{\gamma},$$

where

$$A(c) \equiv \left[\int_0^{\underline{\varphi}^0(c)} \varphi^a \mathrm{d}F(\varphi, c)\right] \left[\int_{\underline{\varphi}^0(c)}^{\infty} \varphi^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi, c)\right]^{\varepsilon}$$

is a measure of city productivity that takes into account both workers' and entrepreneurs' productivity net of agglomeration economies. The term in the first bracket above adjusts for the effective units of labour held by the city's workforce. The term in the second bracket is the unconditional average productivity of entrepreneurs. The unique solution to this problem is

$$L^0(c) = \left[\frac{\varepsilon}{\gamma\theta} A(c)\right]^{\frac{1}{\gamma-\varepsilon}}.$$

Note again the similarity with the market solution. With talent-homogeneous cities, this expression boils down to expression (33) in the main text. In addition, optimal city size increases with the measure of city productivity defined above with an elasticity of $1/(\gamma - \varepsilon)$ and this measure of productivity is entirely driven by the talent composition of c. Thus, cities endowed with more talented individuals are larger at the optimal allocation, just as they are in equilibrium.

Optimal sorting (or city composition). Optimal city composition cannot be characterised precisely without making specific assumptions about the distribution of talent $G_t(\cdot)$ and the distribution of serendipity $G_s(\cdot)$ in the economy because (i) talent and serendipity interact to determine productivity φ and (ii) serendipity occurs after the planner's location decisions. Relaxing either would solve the problem entirely. Equivalently, we may instead assume that the distribution of serendipity is degenerate and can take only one value so that there is a one-to-one mapping between talent and productivity. Without further loss of generality, let us normalise $s = 1$.

Our first result here emphasises a key inefficiency at work in our model. Consider two possible talents t_1 and t_2, with $t_1 > t_2$. Does the planner prefer creating talent-homogeneous cities or mixed cities with entrepreneurs of talent t_1 and workers of talent t_2? Consider now two talent-homogeneous cities, $c = 1$ and $c = 2$ with sizes and talent (L_1, t_1) and (L_2, t_2), respectively. Here the share of efficiency units of labour used in production, σ, is simply equal to $\sigma = 1/(1+\varepsilon)$. Define $V(t_c)$ as the indirect utility enjoyed by an individual with talent t_c in talent-homogeneous cities c. Using the expressions above and simplifying, we get:

$$V(t_c) = D\left(t_c^{1+a}\right)^{\frac{\gamma}{\gamma-\varepsilon}},$$

for $c = 1, 2$ and where $D > 0$ collects parameters.

Next, consider forming cities with both types of individuals, selecting type-1 individuals as entrepreneurs and type-2 individuals as production workers. The optimal composition of mixed cities has a fraction $1/(1+\varepsilon)$ of workers and the complementary fraction $\varepsilon/(1+\varepsilon)$ of entrepreneurs. Let Vt_{12} denote the average indirect utility prevailing in such a heterogeneous city. It is equal to:

$$V_{12} = D\left(t_1 t_2^a\right)^{\frac{\gamma}{\gamma-\varepsilon}}.$$

Forming talent-homogeneous cities is optimal if

$$\begin{aligned} 0 &< \frac{\varepsilon}{1+\varepsilon}V(t_1) + \frac{1}{1+\varepsilon}V(t_2) - V_{12} \\ &= D\frac{1}{1+\varepsilon}\left(t_1 t_2^a\right)^{\frac{\gamma}{\gamma-\varepsilon}} S(t_1, t_2, a, \varepsilon), \end{aligned}$$

where

$$S(t_1, t_2, a, \varepsilon) \equiv \varepsilon\left[\left(\frac{t_1}{t_2}\right)^{a\frac{\gamma}{\gamma-\varepsilon}} - 1\right] + \left[\left(\frac{t_1}{t_2}\right)^{-\frac{\gamma}{\gamma-\varepsilon}} - 1\right].$$

Note that $S(t_1, t_2, a, \varepsilon)$ is increasing in a and in ε, and that $S(t_1, t_2, 0, \varepsilon) < 0$ and $S(t_1, t_2, a, 0) < 0$ hold by $t_1 > t_2$. Thus, there exists a function $f : t_1 \times t_2 \times a \to \varepsilon$ defined from $\mathcal{T} \times \mathcal{T} \times \mathbb{R}^+$ to $\mathbb{R}^+$ such that $S(t_1, t_2, a, f(t_1, t_2, a)) = 0$ and $\partial f(\cdot)/\partial a < 0$. Second,

$$\begin{aligned} \lim_{t_1 \to t_2} \frac{\partial}{\partial t_1}\left(t_1 t_2^a\right)^{\frac{\gamma}{\gamma-\varepsilon}} S(t_1, t_2, a, \varepsilon) &= -(1-a)\frac{\gamma}{\gamma-\varepsilon} t_2^{-1+(1+a)/(\gamma-\varepsilon)} \\ &< 0, \end{aligned}$$

where the inequality holds by $a \in (0,1)$. Thus, the planner prefers to create talent-homogeneous cities instead of mixed cities if and only if $\varepsilon > f(t_1, t_2, a)$.

We show in the text (and especially in) that an equilibrium with talent-homogeneous cities exists under some conditions. A necessary condition for talent-homogeneous cities to be optimal is that $t_1 - t_2 > \kappa$, for some $\kappa \in \mathbb{R}^{+*}$. This condition is violated if T is a compact subset of $\mathbb{R}^+$, as we assume in the text. Thus, talent-homogeneous cities may be suboptimal and yet arise in equilibrium.

H Consumer city version of the model

In the main text, we use a specification with local intermediates as in Ethier (1982). We now show that all our results continue to hold true in a 'consumer city' version of the model with non-tradable varieties of a differentiated consumption good.

Assume that residents consume inelastically one unit of land and a continuum of local varieties of a differentiated consumption good. The consumer problem in city c is given by

$$\max\ U_c \equiv \left[\int_{\Omega_c} x_c(i)^{\frac{1}{1+\varepsilon}} \mathrm{d}i\right]^{1+\varepsilon} \quad \text{s.t.} \quad y_c = \int_{\Omega_c} x_c(i) p_c(i) \mathrm{d}i\,,$$

where y_c denotes the disposable income of an individual after paying for land and commuting. Solving the consumer problem yields the same aggregate demand $x_c(i)$ for each variety and the same price index as in equation (5). To produce their variety of the final good, entrepreneurs operate using $1/\varphi(i)$ units of labour paid w_c to produce each unit of their variety. Maximising profits

$$\pi(i) = \left[p_c(i) - \frac{w_c}{\varphi_c(i)}\right] x(i)$$

yields the same constant markup pricing rule as in equation (6). Hence, expressions (7), (8), and (9) continue to hold true in the consumer city version of the model. As all varieties are local to the city where they are produced, and since consumers in city c all face the same price index irrespective of their productivity, the occupational choice is still determined by $\pi(\underline{\varphi}_c) = w_c \underline{\varphi}_c^a$. Consequently, labour supply and demand, as well as the wage bill, are unchanged.

Since consumption varieties are local to each city, and since there is no trade, we no longer have a single price index that can serve as a natural numéraire. Put differently, $\mathbb{P}_c \equiv 1$ for all c no longer holds. Instead, we now have city-specific price differences in the (final) consumption bundle as given by cross-city differences in the price index. It implies that the nominal wage relationship (12) no longer holds either. Instead, we now have

$$\frac{w_c}{\mathbb{P}_c} = \frac{\Phi_c}{1+\varepsilon},$$

which pins down real wages. Observe that more productive cities still have higher Φ and, therefore, higher real wages.

To close the model, we assume that commuting costs are incurred in terms of the local consumption aggregate, i.e., the price of commuting is equal to the local price index. The cost of a resident's round-trip from a location at distance x from the CBD is then equal to $t_c(x) = \mathbb{P}_c \tau x^\gamma$. Per capita commuting costs are $TCC_c/L_c = \mathbb{P}_c \theta L_c^\gamma$, where θ denotes the same bundle of parameters as in .[2]

Recalling that $\varphi \equiv t \times s$, utility net of urban costs can finally be expressed as

$$U_c = \frac{\max\{w_c \varphi^a - \mathbb{P}_c \theta L_c^\gamma, \pi_c - \mathbb{P}_c \theta L_c^\gamma\}}{\mathbb{P}_c} = \frac{\max\{w_c (ts)^a, \pi_c\}}{\mathbb{P}_c} - \theta L_c^\gamma.$$

We then successively get:

$$\begin{aligned}
\mathbb{E}V(t) &= \int_0^{+\infty} \frac{\max\{w_c \times (ts)^a, \pi(ts)\}}{\mathbb{P}_c} \mathrm{d}G_s(s) - \theta L_c^\gamma \\
&= t^a \frac{w_c}{\mathbb{P}_c} \int_0^{\underline{\varphi}_c/t} s^a \mathrm{d}G_{s,c}(s) + \frac{w_c}{\mathbb{P}} \underline{\varphi}_c^a \left(\frac{t}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \int_{\underline{\varphi}_c/t}^{+\infty} s^{\frac{1}{\varepsilon}} \mathrm{d}G_{s,c}(s) - \theta L_c^\gamma \\
&= \frac{\Phi_c}{1+\varepsilon} t^a \left[\int_0^{\underline{\varphi}_c/t} s^a \mathrm{d}G_{s,c}(s) + \left(\frac{t}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}-a} \int_{\underline{\varphi}_c/t}^{+\infty} s^{\frac{1}{\varepsilon}} \mathrm{d}G_{s,c}(s)\right] - \theta L^\gamma
\end{aligned}$$

which is the same as equation (17) in the main text because w_c and $w_c/\mathbb{P}_c$ are identical in the intermediate good and the final good version of the model. It then follows that all of our equilibrium results hold true in the consumer city version of the model.

[2]Observe that this formulation implies that per unit distance commuting costs are city specific. Indeed, per unit distance commuting costs will be lower (in nominal terms) in larger cities with lower price indices. An alternative interpretation is that commuting causes (pure) disutility.

I Entrepreneurial span of control

In this appendix, we show that our model can be extended to include limited span of control for entrepreneurs as in Lucas (1978)by adding an extra parameter to capture entrepreneurial span of control in the production function. As we show in this appendix, this extension does not change the nature of our results. However, it allows us to generate the same estimating equations as Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2013) but leads to different interpretations of their results.[3]

Let the output of an intermediate variety be given by

$$x(\varphi) = \varphi\, l(\varphi)^{1-\alpha}, \tag{I.1}$$

where $1-\alpha \in (0,1)$ is a measure of entrepreneurial span of control. When $1-\alpha$ is close to zero, firms operate under sharply decreasing returns (entrepreneurs have a low span of control) whereas when $1-\alpha$ is close to one, firms operate close to constant returns and can get much larger (entrepreneurs have a high span of control). We note that equation (I.1) replaces equation (2) in the main text. Then, profits are given by

$$\pi(\varphi) = Y\mathbb{P}^{\frac{1}{\varepsilon}} p^{-\frac{1}{\varepsilon}} - w(Y\mathbb{P}^{\frac{1}{\varepsilon}})^{\frac{1}{1-\alpha}} \varphi^{-\frac{1}{1-\alpha}} p^{-\frac{1+\varepsilon}{1-\alpha}\frac{1}{\varepsilon}}, \tag{I.2}$$

where $\mathbb{P}$ is equal to unity by marginal cost pricing in sector Y as in the main text. The first-order condition for profit maximisation implies that

$$p(\varphi)^{\frac{1}{\varepsilon}\left(\frac{1+\varepsilon}{1-\alpha}-1\right)} = \frac{1+\varepsilon}{1-\alpha} wY^{\,\frac{1}{1-\alpha}-1} \varphi^{-\frac{1}{1-\alpha}}\,. \tag{I.3}$$

[3] In the version of our model that appears in the main text, firm revenue does not depend directly on entrepreneurial productivity. This features arises because with a production function linear in labour, the equalisation of marginal products across firms implies that $\varphi\, p(\varphi)$ must be constant. Since the product of these two terms also appears to determine firm revenue, entrepreneurial productivity φ disappears from the expression giving firm revenue. A simple way to prevent this pathological simplification from occurring is to impose instead decreasing returns in production in the form of limits on entrepreneurial span of control.

At the limit $\alpha \to 0$ we verify that $p(\varphi) = (1+\varepsilon)\, w/\varphi$, as in the main text. Using equations (I.1), (I.2), and (I.3), profits can be rewritten as

$$\pi(\varphi) = Y^{\frac{\varepsilon}{1+\varepsilon}} \frac{\alpha+\varepsilon}{1+\varepsilon} \varphi^{\frac{1}{1+\varepsilon}}\, l(\varphi)^{\frac{1-\alpha}{1+\varepsilon}} \,. \tag{I.4}$$

Given that Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2013) use log firm revenue as dependent variable, it is also useful to write this quantity:

$$\ln Z(\varphi) = \frac{\varepsilon}{1+\varepsilon} \ln Y + \frac{1}{1+\varepsilon} \ln \varphi + \frac{1-\alpha}{1+\varepsilon} \ln l(\varphi)\,, \tag{I.5}$$

where we keep in mind that firm employment $l(\varphi)$ is measured in efficiency units of labour. Hence, log employment must be corrected for the productivity of employees (as Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer, 2013, do). We also note that the entrepreneur's productivity enters expression (I.5) both directly and indirectly since more entrepreneurs manage more productive firms and hire more workers.

As in the main model, the productivity cutoff $\underline{\varphi}$ solves $\pi(\varphi) = w\varphi^a$. Using equations (I.1), (I.3), and (I.4), we obtain:

$$\underline{\varphi}^{\frac{1}{\alpha+\varepsilon}-a} = A^{-\frac{\varepsilon}{\alpha+\varepsilon}} w^{\frac{1+\varepsilon}{\alpha+\varepsilon}} \frac{1+\varepsilon}{\alpha+\varepsilon} \left(\frac{1+\varepsilon}{1-\alpha}\right)^{\frac{1-\alpha}{\alpha+\varepsilon}} . \tag{I.6}$$

By definition of the price index:

$$1 = \mathbb{P}^{-\frac{1}{\varepsilon}} = \int_{\Omega_+} p(\varphi)^{-\frac{1}{\varepsilon}} \mathrm{d}\varphi = \left(\frac{1+\varepsilon}{1-\alpha} w\right)^{-\frac{1-\alpha}{\alpha+\varepsilon}} Y^{-\frac{\alpha}{\alpha+\varepsilon}} L \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi), \tag{I.7}$$

where the last equality follows from (I.3), the definition of $F(\cdot)$ as the cumulative distribution function of φ, and rearranging terms.

Next, by definition of Y:

$$\begin{aligned} Y &\equiv wL \int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi) + L \int_{\underline{\varphi}}^{\infty} \pi(\varphi) \mathrm{d}F(\varphi) \\ &= wL \left[\int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi) + \underline{\varphi}^{a-\frac{1}{\alpha+\varepsilon}} \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi) \right], \end{aligned} \tag{I.8}$$

where the last equality follows from rewriting $\pi(\varphi)$ as $\pi(\underline{\varphi}) \times \left(\varphi/\underline{\varphi}\right)^{\frac{1}{\alpha+\varepsilon}}$ and $\pi(\underline{\varphi}) = w\underline{\varphi}^a$. Plugging this expression back into (I.7) and simplifying, we obtain an expression for the equilibrium wage:

$$w = \left(\frac{1-\alpha}{1+\varepsilon}\right)^{1-\alpha} \left[\int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi) + \underline{\varphi}^{a-\frac{1}{\alpha+\varepsilon}} \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi)\right]^{-\alpha} \left[\int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi)\right]^{\alpha+\varepsilon} L^{\varepsilon}, \tag{I.9}$$

which clearly shows that ε still captures agglomeration economies in the model extended to allow for entrepreneurial span of control.

Turning to the full-employment condition, the labour supply (in effective labour units) is equal to $L \int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi)$ and the labour demand is equal to $L \int_{\underline{\varphi}}^{\infty} l(\varphi)\mathrm{d}F(\varphi)$. These yield

$$\int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi) = \int_{\underline{\varphi}}^{\infty} l(\varphi)\mathrm{d}F(\varphi) = \frac{1-\alpha}{\alpha+\varepsilon} \frac{1}{\underline{\varphi}^{\frac{1}{\alpha+\varepsilon}-a}} \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi),$$

where the last equality follows from equating production (I.1) with demand $x\left(p(\varphi)\right) = Yp(\varphi)^{-\frac{1+\varepsilon}{\varepsilon}}$, equations (I.3) and (I.6), and simplifying. Imposing the comparative advantage condition

$$\frac{1}{\alpha+\varepsilon} > a,$$

the cutoff is the unique fixed point of

$$\underline{\varphi}^{\frac{1}{\alpha+\varepsilon}-a} \int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi) = \frac{1-\alpha}{\alpha+\varepsilon} \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi). \tag{I.10}$$

Like with the corresponding expression (14) in the model, the selection cutoff does not depend on city size. Plugging this expression back into (I.9) yields

$$w = L^{\varepsilon} \left(\frac{1-\alpha}{1+\varepsilon}\right)^{1-\alpha} \underline{\varphi}^{-\alpha\left(a-\frac{1}{\alpha+\varepsilon}\right)} \left(\frac{1+\varepsilon}{\alpha+\varepsilon}\right)^{-\alpha} \left[\int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi)\right]^{\varepsilon}.$$

Using equation (I.8), we obtain the following expression for per-capita output:

$$\frac{Y}{L} = L^{\varepsilon} \left(\frac{1-\alpha}{\alpha+\varepsilon}\right)^{-\alpha} \underline{\varphi}^{-\alpha\left(a-\frac{1}{\alpha+\varepsilon}\right)} \left[\int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi)\right] \left[\int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi)\right]^{\varepsilon}.$$

At the limit $\alpha \to 0$ we get $Y/L = \left[\int_0^{\underline{\varphi}} \varphi^a \mathrm{d}F(\varphi)\right] \left[\int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}F(\varphi)\right]^{\varepsilon} L^{\varepsilon}$, which is expression (16) in the main text.

To characterise the equilibrium with talent-homogeneous cities, we note first that expected indirect utility is given by:

$$\begin{aligned} \mathbb{E}V_c(t) &= \int_0^{+\infty} \max\{w_c \times (ts)^a, \pi(ts)\} \mathrm{d}G_s(s) - \theta L_c^{\gamma} \\ &= w_c\,(St)^a \left[\int_0^{St_c/t} \left(\frac{s}{S}\right)^a \mathrm{d}G_s(s) + \left(\frac{t}{t_c}\right)^{\frac{1}{\alpha+\varepsilon}-a} \int_{St_c/t}^{+\infty} \left(\frac{s}{S}\right)^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}G_s(s)\right] - \theta L^{\gamma}, \end{aligned}$$

where the wage is given by

$$\begin{aligned} w_c &= L_c^{\varepsilon} \left(\frac{1-\alpha}{1+\varepsilon}\right)^{1-\alpha} (St_c)^{-\alpha\left(a-\frac{1}{\alpha+\varepsilon}\right)} \left(\frac{1+\varepsilon}{\alpha+\varepsilon}\right)^{-\alpha} \left[t_c^{\frac{1}{\alpha+\varepsilon}} \int_S^{\infty} s^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}G_s(s)\right]^{\varepsilon} \\ &= L_c^{\varepsilon}\,(St_c)^{1-\alpha a}\,(1+\varepsilon)^{-1}\,(1-\alpha) \left(\frac{\alpha+\varepsilon}{1-\alpha}\right)^{\varepsilon+\alpha} \sigma^{\varepsilon}\,, \end{aligned}$$

with (using equation I.10),

$$\sigma \equiv \int_0^S \left(\frac{s}{S}\right)^a \mathrm{d}G_s(s) = \frac{1-\alpha}{\alpha+\varepsilon} \int_S^{\infty} \left(\frac{s}{S}\right)^{\frac{1}{\alpha+\varepsilon}} \mathrm{d}G_s(s)\,.$$

With talent-homogeneous cities the first-order condition becomes

$$\begin{aligned} 0 &= \left.\frac{\partial \mathbb{E}V_c(t)}{\partial L_c}\right|_{t=t_c} \mathrm{d}L_c + \left.\frac{\partial \mathbb{E}V_c(t)}{\partial t_c}\right|_{t=t_c} \mathrm{d}t_c \\ &= \left[\varepsilon w_c\,(St_c)^a \frac{1+\varepsilon}{1-\alpha}\sigma - \gamma\theta L^{\gamma}\right] \frac{\mathrm{d}L_c}{L_c} + w_c\,(St_c)^a \left[(1-a\alpha)\frac{1+\varepsilon}{1-\alpha} - \left(\frac{1}{\alpha+\varepsilon} - a\right)\frac{\alpha+\varepsilon}{1-\alpha}\right] \sigma \frac{\mathrm{d}t_c}{t_c}. \end{aligned}$$

As a result, $L(t_c)$ is of the form:

$$L(t_c) = \left[\xi_2 t_c^{1+a(1-\alpha)}\right]^{\frac{1}{\gamma-\varepsilon}}\,,$$

for some $\xi_2 > 0$, and the size distribution of cities can be shown to have the same form as given by (28) in the main text:

$$g_L(L) = \frac{\overline{L}\underline{L}}{\overline{L} - \underline{L}} L^{-2} + O(\eta)\,, \quad \text{where} \quad \eta \equiv \frac{\gamma - \varepsilon}{1 + a\,(1-\alpha)}.$$

J Measuring agglomeration effects in a more general context

Consider the general case where agglomeration benefits, y, and urban costs, z, determine equilibrium utility: $U_c \equiv y(t_c, L_c) - z(x_c, L_c)$. Both y and z depend on population size, L_c, and specific shifters, t_c and x_c, that may differ across cities (and in turn depend on population). Most models with a spatial equilibrium assume that U_c is equalised across cities. Let us start by imposing that assumption. We also assume that all functions are continuously differentiable. In equilibrium we then have

$$\left(\frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c}\right) \mathrm{d}L_c + \left(\frac{\partial y}{\partial t_c}\frac{\partial t_c}{\partial L_c} - \frac{\partial z}{\partial x_c}\frac{\partial x_c}{\partial L_c}\right) \mathrm{d}L_c = 0. \tag{J.1}$$

The first term is the common net marginal benefit from agglomeration within each city, whereas the second term is the marginal change in the net benefits from agglomeration across cities. For cities of different population sizes to coexist, it must be that an increase in city size is offset by a corresponding shift in t_c and x_c that leaves individuals indifferent. Since urban costs dominate agglomeration economies at any stable equilibrium ($\partial y/\partial L_c < \partial z/\partial L_c$), larger cities must have higher net shifts. Differentiating y and z, we readily obtain

$$\mathrm{d}y = \frac{\partial y}{\partial t_c}\frac{\partial t_c}{\partial L_c}\mathrm{d}L_c + \frac{\partial y}{\partial L_c}\mathrm{d}L_c \quad \text{and} \quad \mathrm{d}z = \frac{\partial z}{\partial x_c}\frac{\partial x_c}{\partial L_c}\mathrm{d}L_c + \frac{\partial z}{\partial L_c}\mathrm{d}L_c. \tag{J.2}$$

Assume that urban costs differ across cities only because of differences in population but not because of systematic differences in the shift parameter ($\partial x_c/\partial L_c = 0$). In that case, equation (J.1) reduces to $(\partial y/\partial t_c)/(\partial t_c/\partial L_c) = -\left(\partial y/\partial L_c - \partial z/\partial L_c\right)$, which, together with equation (J.2) yields

$$\mathrm{d}y = -\left(\frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c}\right)\mathrm{d}L_c + \frac{\partial y}{\partial L_c}\mathrm{d}L_c \quad \Rightarrow \quad \frac{\mathrm{d}y}{\mathrm{d}L_c} = \frac{\partial z}{\partial L_c}. \tag{J.3}$$

In words, at equilibrium the impact of a change in population on income just equals the change in urban costs. To understand why it is so, recall that cities result from a tradeoff between agglomeration benefits and urban costs. Cities can be of different population sizes either because they differ in how agglomeration benefits are affected by their size or in how urban costs are affected by their size. Here, we assume that urban costs are the same across cities, whereas expected earnings depend on t_c. As a result, if we do not control for t_c, we are looking at a situation where all cities face the same urban cost function but differ in how they benefit from agglomeration. Regressing log average earnings against log population then estimates the population elasticity of urban costs.

Insert Figure 8 about here.

Since the first term in (J.3) is positive at equilibrium, regressing log income on log population leads to an upward biased estimate of agglomeration economies ($\mathrm{d}y/\mathrm{d}L_c \geq \partial y/\partial L_c$). However, it delivers the correct estimate for urban costs. Observe from equation (J.2) that regressing y on L will only give an estimate of the agglomeration economies, $\partial y/\partial L_c$, when the cross-city shift t_c is controlled for. Put differently, we have to take out the equilibrium shift that naturally arises in the presence of sorting along talent across cities, for instance. Figure 8 illustrates these results for the case where there is no shift in urban costs.[4]

Stronger results can be obtained if both agglomeration benefits and urban costs are log-linear: $y(t_c, L) = \ln t_c + \varepsilon \ln L_c$ and $z(x_c, L_c) = \ln x_c + \gamma \ln L_c$. In that case, when there is no shift in urban costs across cities we readily obtain that a log-linear regression of the

[4] In a symmetric way, assume that the gross benefits from agglomeration depends only on size but not on the shift parameter ($\partial t_c/\partial L_c = 0$). In that case, it is easy to see that $\mathrm{d}z/\mathrm{d}L_c = \partial y/\partial L_c$. Hence, at equilibrium the impact of a change in population on urban costs just equals the agglomeration economies, whereas regressing log urban costs on log population leads to an downward biased estimate. The in-between cases, where both agglomeration benefits and urban costs shift, do not generally deliver clear results. In that case, estimating either relationship in equation (J.2) will deliver biased estimates of agglomeration and dispersion forces, as a mix of both is captured. However, the direction of the bias can be signed.

productivity measure on urban population yields the elasticity of urban costs, γ.

The foregoing discussion builds on the assumption that there is a common equilibrium utility level across cities. Matters may be more complicated because expected indirect utility need not be equalised across cities. It is the case in the model at hand as shown by Proposition 4. In larger cities, where more talented individuals locate, expected indirect utility is higher. Hence, not only do cities differ in their production function but they also differ in how much they offer to individuals. Assume hence that, in equilibrium, $U(L_c) = y(t_c, L_c) - z(x_c, L_c)$. We then have

$$\left(\frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c}\right) \mathrm{d}L_c + \left(\frac{\partial y}{\partial t_c}\frac{\partial t_c}{\partial L_c} - \frac{\partial z}{\partial x_c}\frac{\partial x_c}{\partial L_c}\right) \mathrm{d}L_c = \mathrm{d}U.$$

The term on the left is the same as in (J.1). The term on the right captures how equilibrium utility changes with city population size. Assume that urban costs depend only on city size ($\partial x_c/\partial L_c = 0$) so that $(\partial y/\partial t_c)/(\partial t_c/\partial L_c) = -\left(\partial y/\partial L_c - \partial z/\partial L_c\right) + \mathrm{d}U_c/\mathrm{d}L_c$. In words, at equilibrium the change in the shift equals the opposite of the net agglomeration benefits plus the increase in utility across cities of different sizes. The former is negative at any stable equilibrium, whereas the latter is usually positive. Replacing into (J.2) then yields

$$\mathrm{d}y = -\left(\frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c}\right) \mathrm{d}L_c + \mathrm{d}U_c + \frac{\partial y}{\partial L_c}\mathrm{d}L_c = \mathrm{d}U_c + \frac{\partial z}{\partial L_c}\mathrm{d}L_c$$

Hence, the impact of a change in population on productivity equals the change in urban costs augmented by the utility shift across cities. Any positive shift thus yields an upward biased estimate of urban costs, while a negative shift biases the estimate downwards. Observe that, when compared to the equal utility case, the bias gets stronger when utility is not equalised. Figure 9 illustrates this case.[5]

Insert Figure 9 about here.

[5]The case with $\partial t_c/\partial L_c = 0$ yields analogous results and we do not discuss it in more detail.

References

Albouy, David. 2009. "Are big cities really bad places to live? Improving quality-of-life estimates across cities." Working Paper no. 14472, NBER, Cambridge.

Alonso, William. 1964. *Location and Land Use; Toward a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

Baum-Snow, Nathaniel and Ronni Pavan. 2012. "Understanding the city size wage gap." *Review of Economic Studies* 79(1): 88-127.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2013. "The costs of agglomeration: Land prices in French cities." Wharton School, University of Pennsylvania.

Duranton, Gilles and Diego Puga. 2013. "The growth of cities." In *Handbook of Economic Growth*, Volume 2, edited by Philippe Aghion and Steven Durlauf. Amsterdam: North-Holland, forthcoming.

Ethier, Wilfred J. 1982. "National and international returns to scale in the modern theory of international trade." *American Economic Review* 72(3): 389-405.

Fujita, Masahisa 1989. *Urban Economic Theory: Land Use and City Size.* Cambridge, MA: Cambridge University Press.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. "Human capital and Regional Development." *Quarterly Journal of Economics* 128(1): 105-64.

Mills, Edwin S. 1967. "An aggregative model of resource allocation in a metropolitan area." *American Economic Review (Papers and Proceedings)* 57(2): 197-210.

Muth, Richard F. 1969. *Cities and Housing. The Spatial Pattern of Urban Residential Land Use*. Chicago, IL: University of Chicago Press.

Small, Kenneth A. and Erik T. Verhoef. 2007. *The Economics of Urban Transportation*. London, UK: Routledge, Taylor& Francis Group.

Zenou, Yves. 2009. *Urban Labour Economics*. Cambridge: Cambridge University Press.

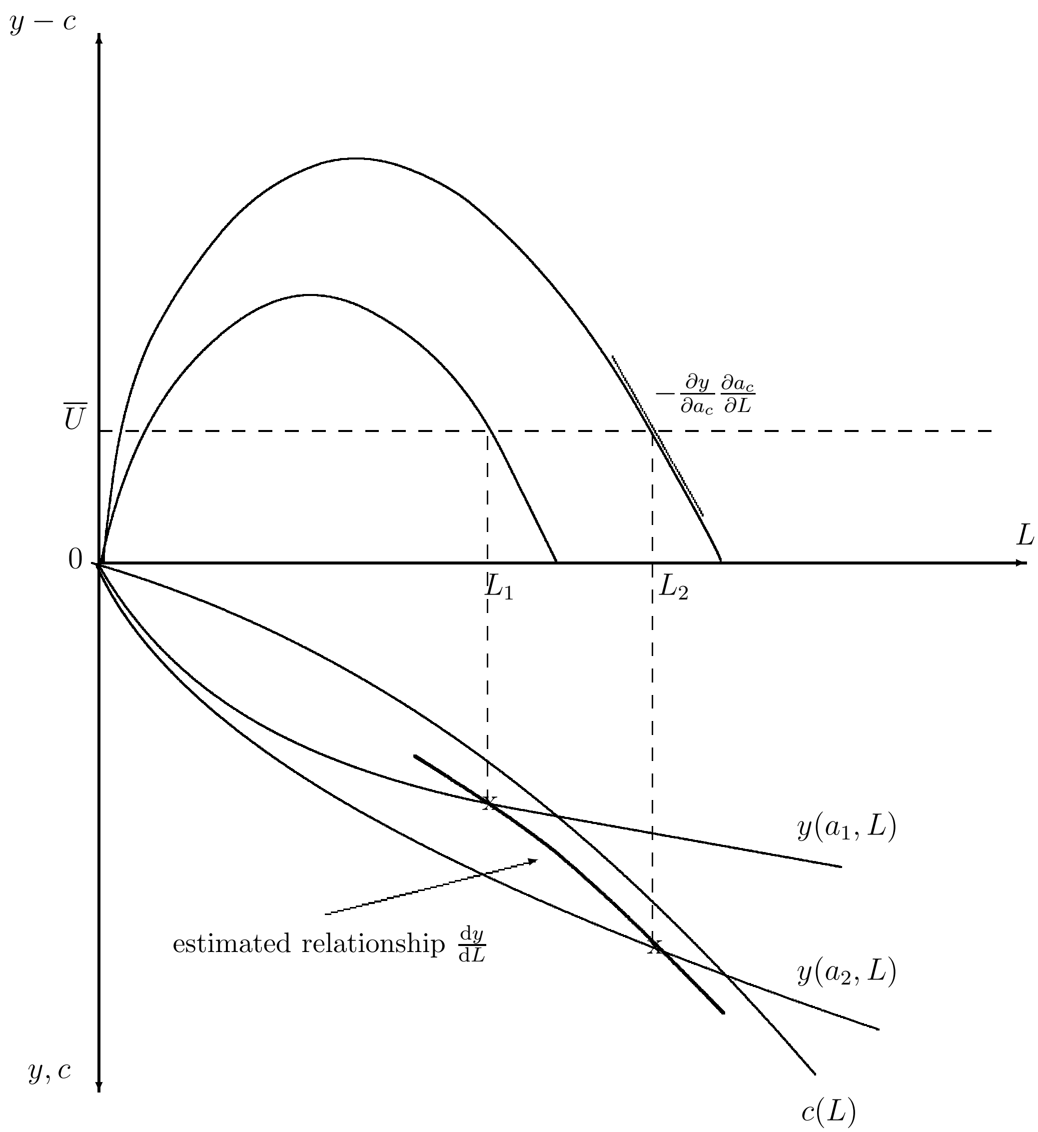

Figure 8: Estimating $\partial y/\partial L$ with constant equilibrium utility.

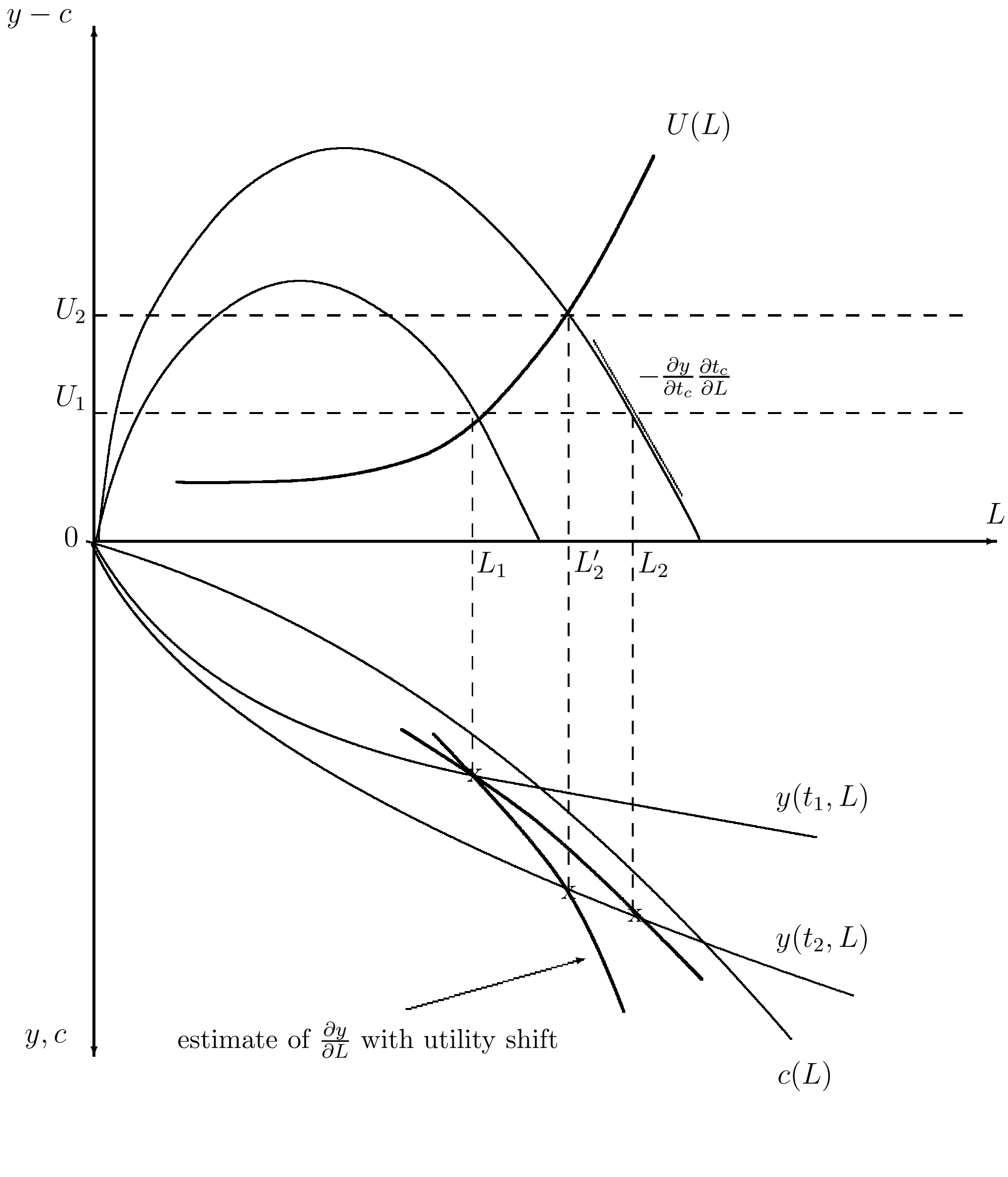

Figure 9: Estimating $\partial y/\partial L$ with varying equilibrium utility.