

# Productive cities: Sorting, selection, and agglomeration

Kristian Behrens\* Gilles Duranton<sup>†</sup>

Frédéric Robert-Nicoud<sup>‡</sup>

Revised: June 11, 2012

## Abstract

Large cities produce more output per capita than small cities. This may occur because more talented individuals sort into large cities, because large cities select more productive entrepreneurs and firms, or because of agglomeration economies. We develop a model of systems of cities that combines all three elements and suggests interesting complementarities between them. The model can replicate stylised facts about sorting, agglomeration, and selection in cities. It also generates Zipf's law for cities under empirically plausible parameter values. Finally, it provides a useful framework within which to reinterpret extant empirical evidence.

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

**JEL Classification:** J24; R10; R23

---

\*Canada Research Chair, Département des Sciences Économiques, Université du Québec à Montréal (UQAM), Canada; CIRPÉE; and CEPR. E-mail: [behrens.kristian@uqam.ca](mailto:behrens.kristian@uqam.ca)

<sup>†</sup>Department of Economics, University of Toronto, Canada; Spatial Economics Research Centre, London School of Economics; and CEPR. E-mail: [gilles.duranton@utoronto.ca](mailto:gilles.duranton@utoronto.ca)

<sup>‡</sup>Département des Sciences Économiques, Université de Genève, Switzerland and CEPR. E-mail: [frederic.robert-nicoud@unige.ch](mailto:frederic.robert-nicoud@unige.ch)

# 1 Introduction

Output per capita is higher in larger cities. For instance, across 276 US metropolitan areas in 2000, the measured elasticity of average city earnings with respect to city population is 8.2%. This paper proposes a model that integrates three main reasons for this fact. The first is agglomeration economies: economies external to firms taking place within cities lead to citywide increasing returns. The second is sorting: more talented individuals may *ex ante* choose to locate in larger cities. The third is selection: larger cities make for larger markets where selection is tougher so that only the most productive firms may *ex post* profitably operate there.<sup>1</sup>

Integrating these three explanations of the urban premium into a theoretical framework where cities are determined endogenously is important for three reasons. First, it yields a better theoretical understanding of how sorting, selection, and agglomeration interact. Our results emphasise some interesting complementarities between these three forces. Tougher selection in larger cities implies that only more talented individuals will locate there in the first place: selection induces sorting. Conversely, the presence of more talented individuals reinforces selection. Cities with more talented individuals, where selection is tougher, also end up with more productive firms paying higher wages. In turn, this attracts more individuals and makes these cities larger, thereby strengthening agglomeration economies.

Second, our model matches a number of key stylised facts about cities. The literature strongly suggests the existence of a causal effect of city population on productivity, even after controlling for sorting and selection. There is also evidence that the returns to talent increase with city population which leads to the sorting of more talented individuals into larger cities. At the same time, there is a non-degenerate distribution of firm productivities in any city. There are fewer less productive firms in larger cities but there is no evidence of stronger selection after conditioning out agglomeration and sorting. Finally, the population distribution of cities is well described by a Pareto distribution with a unitary shape parameter. We discuss these facts in greater detail after deriving our results.

Third, our model provides a useful framework within which to interpret extant quantitative evidence. As already mentioned, in a city earnings regression for the US the coefficient on log city population is 8.2%. In our model, and because of sorting, this coefficient actually reflects the intensity of urban costs. Our estimate drops to 5.1% when conditioning out sorting, using the log share of city college graduates as a proxy for talent. In that case, the coefficient on population measures our key agglomeration parameter. In our model, the small difference between 8.2% and 5.1% should also be equal to the elasticity of city talent with respect to city population. Our data for the US are consistent with this result. Finally, our model also predicts that cities are appreciably

---

<sup>1</sup>We ignore a fourth possible reason: natural advantage. While fundamental for early urban development, the role of natural advantage in mature urban systems may be more limited. Ellison and Glaeser (1999) conclude that it only accounts for a small fraction of industrial concentration in the US. Combes, Duranton, and Gobillon (2008) find that sorting and agglomeration account for the bulk of spatial wage disparities in France.

oversized but the economic costs of this are tiny.

Formally, we extend the monopolistic competition framework of Dixit and Stiglitz (1977) to a two-stage production process, as in Ethier (1982), with heterogeneous entrepreneurs, borrowing from Lucas (1978) and Melitz (2003), to generate local increasing returns.<sup>2</sup> Following Henderson (1974), we then embed this production structure in a system of cities where urban costs increase with city population. The key to our model is that firms are operated by entrepreneurs whose productivity is revealed in two stages. Each individual initially knows about her talent and chooses a location. Upon moving, she gets a draw, which we call luck. Productivity is a combination of talent and luck and more productive individuals have a comparative advantage in entrepreneurship. In equilibrium, individuals sort across cities *ex ante* depending on their talent and they select *ex post* into entrepreneurship or become workers depending on their productivity.<sup>3</sup>

Cities result from a tradeoff between agglomeration economies and urban costs. Entrepreneurial profit increases with productivity and city population. Hence, more talented individuals, who stand a higher chance of becoming highly productive entrepreneurs, have more to gain from locating in larger cities. This complementarity between talent and city population leads to the sorting of more talented individuals into larger cities. Then, tougher selection in more talented cities increases observed average firm productivity. A higher productivity, in turn, complements the agglomeration benefits of cities and this justifies why more talented cities are larger in equilibrium.

Integrating sorting, selection, and agglomeration economies in a model with endogenous cities is the main innovation of our paper. Our model builds on and expands the large theoretical literature in urban economics on agglomeration economies (see Duranton and Puga, 2004, for a review). There is also a large literature about sorting on income and preferences within cities and its fiscal implications (see Epple and Nechyba, 2004, for a review). The theoretical literature about ability sorting across cities is more limited. In an important paper, Nocke (2006), like us, assumes that entrepreneurs are heterogeneous in talent but, unlike us, he ignores talent and maps talent directly into productivity in a partial equilibrium setting. He shows that perfect productivity sorting across exogenously determined cities occurs under weak conditions, a strong but counterfactual result.<sup>4</sup> The literature on selection in cities is even smaller. Behrens and Robert-

---

<sup>2</sup>We work with a single sector. For the issues at stake here, we believe this simplification is appropriate. Hendricks (2011) shows that about 80% of cross-city variations in skills are accounted for by variations within sectors and only 20% by sectoral composition effects.

<sup>3</sup>While the choice of city by individuals can be conceived as an assignment problem, the difficulty with regards to standard assignment theory (e.g., Sattinger, 1993) is that cities are endogenous and their characteristics depend on the location choices of everyone in a general equilibrium framework. Monte (2011) also takes a general equilibrium assignment approach. He considers the assignment of heterogeneous managers to heterogeneous firms to explore the relationship between trade integration and skill-biased technological change.

<sup>4</sup>In earlier work, Abdel-Rahman and Wang (1997) consider the sorting of skilled workers in core cities and that of unskilled workers in peripheral satellite cities. Sorting by talent also occurs in Mori and Turrini (2005) in a two-region setting. In a two-region setting with immobile workers and *ex ante* identical firms, Baldwin and Okubo (2006) assume that firms can relocate at a cost after receiving their productivity draw. This leads to the relocation

Nicoud (2009) propose a multi-region framework that builds on Melitz and Ottaviano (2008) where *ex ante* identical individuals can move from a rural hinterland to cities. In cities, they benefit from agglomeration but may get a poor entrepreneurial draw so that urbanisation also generates inequalities. In work also related to ours, Davis (2009) develops an original model of learning and matching inspired by Antràs, Garicano, and Rossi-Hansberg (2006). Individuals with heterogeneous abilities simultaneously choose to locate in one of two cities and their occupation. In equilibrium, the most talented individuals become managers, those of intermediate ability become workers, and the least talented end up producing a local good. Complementarities in production lead to positive assortative matching and the pairing of the best managers with the best workers. Learning among managers then leads the best manager-and-workers teams to cluster in the same city. Finally, Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) use a simple framework featuring agglomeration, sorting, and selection to assess empirically the effect of human capital on regional development. We return to their findings later in the paper.

A second innovation of our model is to generate Zipf’s law within a static setting. Zipf’s law and the size distribution of cities have attracted much attention recently. In random growth models, the population of a city reflects its balance of past shocks (see Gabaix and Ioannides, 2004, for a review). Our approach is radically different. It builds on a static model of cities. In equilibrium, the population of a city depends on the productivity of its entrepreneurs, magnified by the tradeoff between agglomeration economies and urban costs. More specifically, city population is a power function of the talent of its residents where the power is inversely related to the difference between the intensity of agglomeration economies and that of urban costs. When this difference is small, as is the case in the data, small productivity differences caused by sorting lead to large differences in city population sizes. The resulting size distribution of cities is thus close to degeneracy and approximately Zipf.<sup>5</sup>

In what follows, Section 2 presents the model. Section 3 solves for its equilibrium taking the distribution of population as given. Section 4 solves for location choices and Section 5 derives our results about the size distribution of cities. Section 6 proposes two extensions of our model, and Section 7 discusses its quantitative implications. Finally, Section 8 concludes.

## 2 The model

There is a continuum of individuals of mass  $\Lambda$  in the economy, each choosing a location and an occupation. Individuals are identical except for their ‘talent’,  $t$ , and their ‘luck’,  $s$ . There is also a continuum of homogeneous sites that can be used as cities. The number of cities, their population

---

of the most productive firms from the small market to the large one and incomplete productivity sorting.

<sup>5</sup>We know of two other papers that generate Zipf’s law from a static model. The argument of Lee and Li (2009) is the static equivalent of random growth models where population is determined by the multiplicative aggregation of many randomly distributed local characteristics. Hsu (2012) relies on central place theory.

size, and their composition are endogenous.

Each individual initially knows her talent and chooses where to locate. Upon moving to a city, her luck is revealed. The product of an individual's talent and luck determines her productivity:  $\varphi \equiv t \times s$ . 'Luck' subsumes many local interactions that are uncertain and affect productivity such as being acquainted with the right people at the right time, etc. Knowing her productivity, each individual then selects into an occupation, entrepreneur or worker. An entrepreneur sets up a firm that produces with productivity  $\varphi$  a variety of a differentiated intermediate good using labour. A worker supplies  $\varphi^a$  efficiency units of labour, with  $a \geq 0$ . For more productive individuals to retain a comparative advantage in entrepreneurship, we impose an upper bound on  $a$  below. After occupations are chosen, firms maximise profit, markets clear, and production and consumption take place.

Empirically, there are frictions to mobility. In our static model we formalise these frictions in a parsimonious and tractable way by assuming free mobility before luck is realised and prohibitive mobility costs afterwards.<sup>6</sup> The knowledge of their talent allows individuals to sort across cities. The revelation of luck after choosing a city leads to their selection into occupations. That is, our two-step revelation process enables us to consider both the spatial sorting of individuals and the productivity selection of firms. Selection without sorting would lead all cities to be symmetric in equilibrium. Sorting without selection would imply that all firms in any one city have the same productivity. Both predictions are counterfactual.

To avoid the introduction of arbitrary productivity differences across cities, the cumulative distribution of luck is assumed to be the same in all cities. The distributions of talent and luck are summarised by the continuously differentiable cumulative probability distribution functions  $G_t$  over  $[t, \bar{t}] \subset \mathbb{R}^{+*}$  and  $G_s$  over  $\mathbb{R}^+$ , respectively. We also note  $F(\varphi) = F(t \times s)$  the joint distribution of the product  $t \times s$ .

Individuals consume two goods: a final good and land. For simplicity, they require one unit of land for accommodation and do not increase their utility by consuming more land. They are also risk-neutral so that their utility can be taken to be linear in final good consumption. To produce the final good, competitive final producers in each city use locally produced differentiated intermediate inputs, which enter into their technology with constant elasticity of substitution  $1 + 1/\varepsilon$ , with  $\varepsilon > 0$ . Aggregate output in city  $c$  is given by

$$Y_c = \left[ \int_{\Omega_c} x_c(i)^{\frac{1}{1+\varepsilon}} di \right]^{1+\varepsilon}, \quad (1)$$

where  $x_c(i)$  is the amount of variety  $i$  used, and  $\Omega_c$  is the endogenously determined set of varieties of intermediate inputs produced in city  $c$ . Without loss of generality and for simplicity, we make the final good  $Y$  freely tradeable to use it as numéraire.

---

<sup>6</sup>In Section 6, we develop an extension where individuals can obtain additional draws of luck at a cost and show that this does not affect the main properties of our model.

As in Ethier (1982), intermediate inputs are produced by monopolistically competitive firms à la Dixit and Stiglitz (1977). Each entrepreneur sets up a firm which employs labour to produce a different variety. Hence  $\Omega_c$ , the set of varieties, also denotes the set of entrepreneurs and  $i$  refers equivalently to an entrepreneur, her firm, or the variety she produces. Entrepreneurs differ in their productivity as in Lucas (1978) and Melitz (2003).<sup>7</sup> Output of variety  $i$  is

$$x_c(i) = \varphi_c(i) l_c(i), \quad (2)$$

where  $l_c(i)$  is labour demand (in efficiency units) for the production of variety  $i$  and  $\varphi_c(i)$  is entrepreneur  $i$ 's productivity, which, in turn, depends on her 'talent',  $t$ , and her 'luck',  $s$ .

Turning to the urban structure of our model, we assume that each resident of a city of population  $L$  pays  $\theta L^\gamma$  as urban cost to reside in that city. In a separate web appendix (Appendix F) we develop micro-economic foundations that justify this functional form. These foundations rely on a standard monocentric urban framework, where an increase in population leads to greater commuting costs. For cities to remain of finite size in equilibrium we require  $\gamma$ , the elasticity of urban costs, to be larger than  $\varepsilon$ , which turns out to be the equilibrium elasticity of agglomeration economies.

The set of cities,  $C \equiv [0, \bar{c}]$ , is endogenous. To use the terminology of Henderson and Becker (2000), cities arise under 'self-organisation', i.e., they are the outcome of the mutually compatible optimal choices of a continuum of individuals. The endogenous population of each city,  $L_c$ , is given by

$$L_c \equiv \int_{\underline{t}}^{\bar{t}} L_c(t) dt, \quad (3)$$

where  $L_c(t)$  is the population with talent  $t$  in city  $c$ . In equilibrium all individuals must live in some city. With a slight abuse of notation, the population condition for each type of talent thus requires that

$$\Lambda g_t(t) = \int_0^{\bar{c}} L_c(t) dc, \quad \forall t \in [\underline{t}, \bar{t}], \quad (4)$$

where  $g_t$  is the probability distribution function of talent. Equation (4) states that the mass of individuals of talent  $t$  across all cities must be equal to the mass of individuals of talent  $t$  in the population. Summing across all talents requires the full population condition in the model to be satisfied.

---

<sup>7</sup>As in Lucas (1978), firms only differ in the productivity of their entrepreneur. Because differentiated varieties are imperfect substitutes we do not need to impose limits on entrepreneurial span of control for firms to remain of finite size. We nonetheless consider this extension in Section 7 below. As in Melitz (2003), we embed heterogeneous firms in a constant-elasticity-of-substitution demand system. Unlike in Melitz (2003), there is no sunk cost to create a firm and receive a productivity draw. This investment decision is potentially complex in our case because would-be entrepreneurs differ by talent.

### 3 Selection and agglomeration

In equilibrium, each individual optimally chooses a city based on her talent. After learning her luck, she optimally chooses an occupation, worker or entrepreneur. Entrepreneurs then maximise their profit with respect to the price of their variety. Markets for intermediate goods, final goods, and labour clear, and the population condition is satisfied.

For expositional purposes, it is convenient to solve for the equilibrium in steps. In this section, we study each city in isolation and take its population and its productivity distribution as given. Thus, individuals know their own productivity, the cumulative distribution of productivity in their city  $F_c(\cdot)$ , which we assume for now to be continuously differentiable over a closed support, and their city population size,  $L_c$ . The focus in this section is on selection (the occupational choice between worker and entrepreneur) and agglomeration (the increase in productive efficiency caused by an increase in city population and city talent). In the next section, we solve for the sorting of individuals across endogenously determined cities based on their talent.

To ease notation, we drop the city subscript  $c$  wherever possible. Minimising production costs in the final goods sector subject to the technology described by equation (1) yields the demand for intermediates inputs:

$$x(i) = \left[ \frac{p(i)}{\mathbb{P}} \right]^{-\frac{1+\varepsilon}{\varepsilon}} \frac{Y}{\mathbb{P}}, \quad \text{where} \quad \mathbb{P} \equiv \left[ \int_{\Omega} p(j)^{-\frac{1}{\varepsilon}} dj \right]^{-\varepsilon} \quad (5)$$

is the appropriate price index. It is immediate from equation (5) that the own-price elasticity of demand is  $-(1+\varepsilon)/\varepsilon$ . Hence, the profit-maximising price for each intermediate displays a constant markup over marginal cost:

$$p(i) = (1 + \varepsilon) \frac{w}{\varphi(i)}, \quad (6)$$

where  $w$  is the wage per efficiency unit of labour. This allows us to re-write the demand (5) as:

$$x(i) = \left[ \frac{\varphi(i)}{\Phi} \right]^{1+\frac{1}{\varepsilon}} \frac{Y}{\mathbb{P}}, \quad \text{where} \quad \Phi \equiv \left[ \int_{\Omega} \varphi(j)^{\frac{1}{\varepsilon}} dj \right]^{\varepsilon} \quad (7)$$

is the appropriate measure of aggregate productivity in the city. More entrepreneurs in a city (i.e., a larger measure of  $\Omega$ ) and/or better entrepreneurs (i.e., on average larger  $\varphi$ 's) imply a larger aggregate productivity,  $\Phi$ . In turn, individual sales are negatively affected by aggregate productivity through a market crowding effect. Using expressions (6) and (7), we rewrite the price index  $\mathbb{P}$  in (5) as a function of aggregate productivity,  $\Phi$ , and obtain

$$\mathbb{P} = (1 + \varepsilon) \frac{w}{\Phi}. \quad (8)$$

After combining this equation with (6) and (7), operating profit becomes

$$\pi(i) = \frac{\varepsilon}{1 + \varepsilon} p(i) x(i) = \frac{\varepsilon}{1 + \varepsilon} Y \left[ \frac{\varphi(i)}{\Phi} \right]^{\frac{1}{\varepsilon}}. \quad (9)$$

This expression shows that the profit of entrepreneurs increases with the economic size of their city,  $Y$ , and with their own productivity relative to aggregate productivity,  $\varphi(i)/\Phi$ . Put differently, holding her own productivity constant an entrepreneur would like to be in an economically large city with low aggregate productivity. As this does not happen in equilibrium, equation (9) contains the germ of our main tradeoff which occurs below when individuals need to choose a location.

Individuals choose their occupation by comparing their prospective entrepreneurial profit, as given by (9), with their labour income  $w \times \varphi^a$ . We assume  $a < 1/\varepsilon$  for more productive individuals to have a comparative advantage in entrepreneurship. Then, there exists a productivity cutoff for selection into entrepreneurship  $\underline{\varphi}$  defined by  $\pi(\underline{\varphi}) = w \underline{\varphi}^a$  such that all individuals with productivity above  $\underline{\varphi}$  become entrepreneurs and all individuals with productivity below  $\underline{\varphi}$  become workers. Because the set of individual productivities in the city is convex (as luck is distributed over  $\mathbb{R}^+$ ), this selection cutoff is unique and, using equation (9), given by

$$\underline{\varphi} \equiv \left[ \Phi \left( \frac{1 + \varepsilon}{\varepsilon} \frac{w}{Y} \right)^\varepsilon \right]^{\frac{1}{1-a\varepsilon}}. \quad (10)$$

Selection is tougher when aggregate productivity is higher ( $\partial \underline{\varphi} / \partial \Phi > 0$ ), for it is more difficult to compete against more productive and numerous entrepreneurs. Selection is also tougher when demand is lower ( $\partial \underline{\varphi} / \partial Y < 0$ ) and when wages are higher ( $\partial \underline{\varphi} / \partial w > 0$ ).

Labour in a city is supplied by all individuals with productivity below  $\underline{\varphi}$ . In efficiency units, city labour supply is equal to:  $L^S \equiv L \int_0^{\underline{\varphi}} \varphi^a dF(\varphi)$ . From equation (2), labour demand for an entrepreneur with productivity  $\varphi$  is  $l(\varphi) = x(\varphi)/\varphi$ . Combining this expression with equations (7) and (8) and aggregating over all entrepreneurs, we obtain city labour demand:  $L^D = L \int_{\underline{\varphi}}^{+\infty} l(\varphi) dF(\varphi) = \frac{1}{1+\varepsilon} \frac{Y}{w}$ . Equating labour supply and demand implies that workers receive a share  $\frac{1}{1+\varepsilon}$  of city output:

$$w L \int_0^{\underline{\varphi}} \varphi^a dF(\varphi) = \frac{Y}{1 + \varepsilon}. \quad (11)$$

That workers (and thus entrepreneurs) receive a constant share of output plays a key role to facilitate the analysis below.

Competition and cost minimisation by final good producers yield  $\mathbb{P} = 1$ , for  $\mathbb{P}$  is the marginal cost of final producers and the price of the final good is normalised to unity. Then equation (8) yields

$$w = \frac{1}{1 + \varepsilon} \Phi, \quad (12)$$

where aggregate productivity,  $\Phi$ , as defined in equation (7) can be rewritten as

$$\Phi = \left[ L \int_{\underline{\varphi}}^{+\infty} \varphi^{\frac{1}{\varepsilon}} dF(\varphi) \right]^\varepsilon. \quad (13)$$

Expressions (10) to (13) fully characterise the equilibrium tuple  $\{\underline{\varphi}, \Phi, w, Y\}$ .



**Proposition 1 (Existence and selection)** *Given population,  $L$ , and its productivity distribution,  $F(\cdot)$ , the equilibrium in a city is characterised by equations (10) to (13), exists, is unique, and the productivity cutoff for selection does not depend on city population. In addition, in any two cities 1 and 2 with ‘scaled’ distributions of productivity such that  $F_1(\varphi) = F_2(\lambda\varphi)$  with  $\lambda > 0$ , the selection cutoffs are such that  $\underline{\varphi}_2 = \lambda\underline{\varphi}_1$*

**Proof.** Using equations (10), (11), and (13) to eliminate  $w$ ,  $Y$ , and  $\Phi$  yields an implicit solution for  $\underline{\varphi}$ :

$$\underline{\varphi}^{\frac{1}{\varepsilon}-a} \int_0^{\underline{\varphi}} \varphi^a dF(\varphi) = \frac{1}{\varepsilon} \int_{\underline{\varphi}}^{+\infty} \varphi^{\frac{1}{\varepsilon}} dF(\varphi). \quad (14)$$

Since  $a < 1/\varepsilon$ , the left-hand side of this expression is monotonically increasing in  $\underline{\varphi}$ , starting from 0 and strictly positive when  $\underline{\varphi} \rightarrow +\infty$ . The right-hand side is monotonically decreasing in  $\underline{\varphi}$  and equal to 0 when  $\underline{\varphi} \rightarrow +\infty$ . By continuity, this establishes the existence of a unique equilibrium. Next, by inspection of equation (14),  $\underline{\varphi}$  does not depend on city population.

To prove the last part of the proposition assume that equation (14) holds for city 1. Since  $F_1(\varphi) = F_2(\lambda\varphi)$  we have  $dF_1(\varphi) = dF_2(\lambda\varphi)$  and we can write the equilibrium condition for city 1 as:

$$\underline{\varphi}_1^{\frac{1}{\varepsilon}-a} \int_0^{\underline{\varphi}_1} \varphi^a dF_2(\lambda\varphi) = \frac{1}{\varepsilon} \int_{\underline{\varphi}_1}^{\infty} \varphi^{\frac{1}{\varepsilon}} dF_2(\lambda\varphi).$$

The change of variable  $\lambda\varphi = z$  and  $\lambda d\varphi = dz$  implies that the previous equation can be rewritten as:

$$(\lambda\underline{\varphi}_1)^{\frac{1}{\varepsilon}-a} \int_0^{\lambda\underline{\varphi}_1} z^a dF_2(z) = \frac{1}{\varepsilon} \int_{\lambda\underline{\varphi}_1}^{\infty} z^{\frac{1}{\varepsilon}} dF_2(z).$$

It is then immediate to verify that  $\underline{\varphi}_2 = \lambda\underline{\varphi}_1$ . ■

Aside from existence and uniqueness, Proposition 1 shows that the equilibrium selection cutoff does not depend on city population, conditional on the distribution of productivity. This result is the outcome of two offsetting forces. Larger cities have both a higher demand (which lowers the selection cutoff) and more entrepreneurs (which raises it). These two effects exactly offset each other in our framework.<sup>8</sup> The reason behind this exact offset can be found in equation (11) which

---

<sup>8</sup>There are at least two ways to make the productivity cutoff vary with city population conditionally on the distribution of productivity. The first is to impose a different demand structure for varieties. In the spirit of Melitz and Ottaviano (2008), Behrens and Robert-Nicoud (2009) use non-CES preferences to generate markups that decrease with the number of local varieties. This naturally leads to tougher selection in larger markets. The second possibility is to change the supply side and have the ratio of fixed to variable costs for firms depend on city population. On the one hand, a fixed cost (in addition to the entrepreneur’s foregone labour) paid in numéraire would be relatively less costly in larger cities where productivity is higher. This would imply a greater proportion of entrepreneurs in larger cities. On the other hand, a fixed cost paid with a factor that is in fixed supply locally (such as land) would increase faster than operating profit as cities get larger. In turn, this would mean a lower proportion of entrepreneurs in larger cities.

shows that labour market clearing implies a constant aggregate entrepreneurial income as a share of city output. Hence, keeping the distribution of individual productivity constant, a city hosts the same proportion of workers and entrepreneurs regardless of its size. As we show in a separate web appendix (Appendix G), the equilibrium selection is also optimal as a result of constant mark-ups.

Proposition 1 also demonstrates that scaling up the distribution of productivity by a factor  $\lambda$  scales up the selection cut-off by the same factor. Again, this occurs because aggregate entrepreneurial income is a constant share of city output. Simply put, a city whose residents are twice as productive has a selection cutoff twice as high. This case is of particular empirical relevance since Combes, Duranton, Gobillon, Puga and Roux (2012) find that the distribution of log productivity in larger cities in France is, to a first approximation, a shifted version of its counterpart in smaller cities.<sup>9</sup> In addition, the share of entrepreneurs should also be constant across cities. Empirically, the share of self-employed workers – a proxy for entrepreneurship – is uncorrelated with city population in the US. Regressing the employment share of self-employed workers on log city population in 276 US MSAs using 2000 Census data yields a coefficient of 0.0003 with a standard error is 0.001.

Proposition 1 has two further implications. First, sorting induces selection. If larger cities attract more talented individuals, they will be tougher markets. As emphasised by Sinatra in his 1979 *New York, New York* song: “If I can make it there, I’ll make it anywhere”. Second, conditional on sorting there are no differences in selection across cities. This result is also compatible with the findings of Combes, Duranton, Gobillon, Puga and Roux (2012) for the productivity of French firms. They find no differences in selection cutoffs across cities of different population sizes after accounting for common productivity differences that affect all firms. Put differently, there are no differences in selection between large and small French cities after controlling for sorting and agglomeration.

**Proposition 2 (Agglomeration)** *Given the productivity distribution,  $F(\cdot)$ , the elasticity of aggregate productivity, per capita income, and the wage rate with respect to city population is  $\varepsilon$ . Scaling up the distribution of talent by a factor  $\lambda$  scales up output per worker by a factor  $\lambda^{1+a}$  and the wage rate by a factor  $\lambda$ .*

**Proof.** By equations (12) and (13), the wage can be written as:

$$w = \frac{1}{1 + \varepsilon} \left( \int_{\underline{\varphi}}^{+\infty} \varphi^{\frac{1}{\varepsilon}} dF(\varphi) \right)^{\varepsilon} L^{\varepsilon}. \quad (15)$$

Since by equation (14),  $\underline{\varphi}$  does not depend on  $L$ ,  $w$  is proportional  $L^{\varepsilon}$ . In turn by equation (11), we find:

$$Y = \left( \int_{\underline{\varphi}}^{+\infty} \varphi^{\frac{1}{\varepsilon}} dF(\varphi) \right)^{\varepsilon} \left( \int_0^{\underline{\varphi}} \varphi^a dF(\varphi) \right) L^{1+\varepsilon}, \quad (16)$$

---

<sup>9</sup>Scaling up the distribution of productivity across cities implies a shift when comparing the distribution of *log* productivity across cities.

which shows that  $Y/L$  is also proportional to  $L^\varepsilon$ .

By the same change of variable as in Proposition 1, it is easy to show from equations (15) and (16) that scaling the talent of every individual by a factor  $\lambda$  multiplies wages by the same factor and multiplies total output by a factor  $\lambda^{1+a}$ . ■

Our model thus displays agglomeration economies. They first take the form of scale externalities since per capita income increases with city population keeping the distribution of talent constant. The reason is that an increase in population increases the number of entrepreneurs and thus the number of intermediate inputs. Final producers become more productive as they have access to a wider range of varieties. Sharing local differentiated inputs produced under increasing returns is a popular way to generate scale externalities in the literature (Duranton and Puga, 2004). Our innovation here is to enrich the standard framework by considering heterogeneous firms.

The empirical evidence in favour of scale externalities is very strong. According to Rosenthal and Strange (2004) and Melo, Graham, and Noland (2009), in many countries the estimates for the elasticity of wages or productivity with respect to city population are close to the 8% we report for US MSAs in the introduction. Recent research suggests that about half of this estimate actually reflects the causal effect of a greater population size of cities on their wages (see Combes, Gobillon, and Duranton, 2011, for a recent discussion of identification issues in the estimation of agglomeration economies). Consistent with our modelling strategy, recent evidence also points at input-output linkages as an important source of agglomeration economies (see Puga, 2010, for a discussion).

In our model, agglomeration economies also take the form of talent externalities since scaling up the talent of everyone in a city raises the selection cutoff, which leads to more productive firms and increases the wage rate. The literature often refers to these externalities as human capital externalities. While they are often conceived as a consequence of direct spillovers, we show here that they can also arise from the same mechanism used to model scale externalities.

Empirically, we can think of education as a noisy proxy for talent. Higher earnings in more educated cities is another salient feature of the data. This form of agglomeration economies has been documented in many countries and the best evidence suggests that average education in a city has a causal effect on earnings in this city (see Moretti, 2004, for a review).

Last, observe that we make the final good tradable to monetise the benefits of agglomeration easily. While this matters for the quantitative exercise we conduct below, it is unimportant for our theory. In a separate web appendix (Appendix H) we show that our model is isomorphic to one where individuals consume a continuum of a nontraded local final good. Hence, our approach also subsumes the ‘consumer city’ model of urban economics where the benefits from local diversity cannot be traded (as in, e.g., Lee, 2010).

Before turning to location choices, it is useful to show that talent and city population are complements. This is the main force pushing towards sorting in our model.

**Proposition 3 (Complementarity between talent and city population)** *More talented individuals benefit more from being located in larger cities. Expected indirect utility is such that:*

$$\frac{\partial^2 \mathbb{E}V(t)}{\partial t \partial L} \Big|_{F(\cdot)} \geq 0.$$

**Proof.** Using equations (9) and (10) and the selection cutoff condition  $\pi(\underline{\varphi}) = w \times (t \underline{\varphi})^a$ , the expected indirect utility of an individual with talent  $t$  in her city before she learns her luck is

$$\begin{aligned} \mathbb{E}V(t) &= \int_0^{+\infty} \max\{w \times (ts)^a, \pi(ts)\} dG_s(s) - \theta L^\gamma \\ &= w t^a \left[ \int_0^{\underline{\varphi}/t} s^a dG_s(s) + \left(\frac{t}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}-a} \int_{\underline{\varphi}/t}^{+\infty} s^{\frac{1}{\varepsilon}} dG_s(s) \right] - \theta L^\gamma, \end{aligned} \quad (17)$$

where  $\underline{\varphi}$  is the solution to equation (14). Using equation (15), it is easy to see that the wage in the first term of equation (17) is proportional to  $L^\varepsilon$  as a result of agglomeration economies. The wage also increases with  $\underline{\varphi}$  as a result of selection. In turn, by Proposition 1,  $\underline{\varphi}$  is independent of  $L$  conditional on  $F$ . The product of  $t^a$  and the term in square brackets in (17) is the expected premium associated with being of talent  $t$ . This premium increases with talent.

The cross-partial derivative of  $\mathbb{E}V(t)$  in  $L$  and  $t$  sums the cross partials of the first and second terms in (17). The first is positive since the wage increases with  $L$  and does not depend on  $t$ , whereas the rest of that term increases with  $t$  and does not depend on  $L$ . The second term of (17), urban costs, does not depend on  $t$  and thus vanishes. This proves our claim. ■

The earnings of both entrepreneurs and workers increase with their talent and with city population. For workers, earnings increase with population through the wage rate because of the scale externalities highlighted above. Earnings also increase with talent because a higher talent implies a larger effective supply of labour for an individual. For entrepreneurs, profits increase with individual productivity (and thus talent) and city income (and thus population) as highlighted in equation (9).

This complementarity between talent and city population is underscored by the empirical literature. Taking again education as a proxy for talent, Wheeler (2001) and Glaeser and Resseger (2010) find stronger agglomeration benefits for more educated workers relative to less educated workers. Taking cognitive and people skills as another proxy for talent, Bacolod, Blum, and Strange (2009) find a similar result for individuals with better cognitive and people skills.

For future reference, we also note that the cross-partial derivative in Proposition 3 resembles a single-crossing condition. Such condition pushes towards sorting along talent. In our case, however, this cross-partial holds only conditionally on the distribution of productivity  $F(\cdot)$ , which is itself endogenous. Hence the sign of this cross-partial derivative does not immediately ensure the existence of a perfect sorting equilibrium since different cities may face different distributions of talent and thus productivity. Contrary to standard assignment problems, cities are endogenous and their equilibrium characteristics depend on the location choices of everyone.

## 4 Sorting and cities

Until now, we have taken the distribution of talent across cities as given. We now turn to location choices and the sorting of individuals across cities depending on their talent. When choosing a city  $c$  an individual with talent  $t$  maximises expected utility,  $\mathbb{E}V_c(t)$ , since luck is still unknown at that time. We define the assignment function  $\mu : T \rightarrow C$  which maps talents into cities. An equilibrium choice of cities is such that

$$\mu(t) = \{c \in C : \mathbb{E}V_c(t) \geq \mathbb{E}V_{c'}(t), \forall c' \in C\} \quad (18)$$

for all  $t \in T$ . In equilibrium, no individual wants to deviate to another city given the location choices of all other individuals. Once in a city, individuals make their occupational choice as described in Section 3. Entrepreneurs choose employment in their firm to maximise profit, and all markets clear. Formally, an equilibrium satisfies the population condition (4), selection and agglomeration as described by equations (10)–(13), and optimal location choice as given by equation (18).

Proposition 3 suggests that more talented individuals benefit more from being located in larger cities. It does not, however, preclude the existence of a symmetric equilibrium where all types of talents are equally represented in all cities. A first natural question to ask is, therefore, under which conditions a symmetric equilibrium is stable. We show in Appendix A that such an equilibrium is stable only if the variation in talent across the population is small enough. In other words, ability sorting is a natural equilibrium outcome when individuals are sufficiently heterogeneous.

Symmetric equilibria are both empirically counterfactual and theoretically not very illuminating. From now on we thus focus on equilibria with sorting. Specifically, we construct an equilibrium with a single type of talent,  $t_c$ , in each city. We refer to cities in this equilibrium as talent-homogeneous cities. While we postpone our discussion of how equilibria are selected, we note that this equilibrium allows us to account for the key stylised facts mentioned above. This equilibrium also displays all the main tradeoffs in a tractable analytical setting.

Each individual of talent  $t$  chooses a talent-homogeneous city of talent  $t_c$  by solving  $\max_{c \in C} \mathbb{E}V_c(t)$ , where  $\mathbb{E}V_c(t)$  is given by equation (17). Let  $S$  denote the common luck threshold to become an entrepreneur  $S$  and  $\sigma$  the share of efficiency units of labour used in production. By Proposition 1, these two quantities are constant across talent-homogeneous cities:

$$S \equiv \underline{\varphi}_c / t_c \quad \text{and} \quad \sigma \equiv \int_0^S (s/S)^a dG_s(s) = \frac{1}{\varepsilon} \int_S^\infty (s/S)^{\frac{1}{\varepsilon}} dG_s(s),$$

where the last equality follows from (14). Observe that the support of  $t$  is convex by assumption and that  $\mathbb{E}V_c(t)$  is continuously differentiable in  $L_c$ ,  $t_c$ , and  $t$ . As a result, we can order cities so that  $t_c = t(c)$  and  $L_c = L(c)$  are continuous functions of  $c$ , where  $t(c)$  comes from equation (18) and  $L(c)$  is given by equation (3).<sup>10</sup> Hence, an equilibrium with talent-homogeneous cities is

<sup>10</sup>For  $\mu(\cdot)$  to be invertible we need either a strict inequality in equation (18) (which occurs in equilibrium) or that ties are always broken in the same way in case of equality.

characterised by a function  $L(t_c)$  that assigns one city population to each talent. Individuals choose their preferred city from a ‘menu’ of possible combinations of talent and population, *knowing* that the choice of a city talent  $t_c$  implies the choice of a population  $L(t_c)$ .

**Proposition 4 (Equilibrium population of talent-homogeneous cities)** *The talent-homogeneous equilibrium is unique and such that*

$$L^*(t_c) = \left( \frac{1 + \gamma}{1 + \varepsilon} \xi t_c^{1+a} \right)^{\frac{1}{\gamma - \varepsilon}} \quad \text{where} \quad \xi \equiv \frac{(\varepsilon \sigma)^{1+\varepsilon} S^{1+a}}{\gamma \theta}. \quad (19)$$

*Equilibrium city population is too large, increases with talent  $t$ , agglomeration economies  $\varepsilon$ , and worker heterogeneity  $a$ , and decreases with urban costs  $\theta$  and  $\gamma$ .*

**Proof.** We first solve for the optimal population of talent-homogeneous cities. Using, equations (12)-(14) and (17), the first-order condition evaluated at  $t = t_c$  implies

$$L^o(t_c) = (\xi t_c^{1+a})^{\frac{1}{\gamma - \varepsilon}} \quad (20)$$

after simplifications.

In equilibrium, each individual solves a constrained optimisation problem which consists in picking the city with talent  $t_c$  that maximises her expected indirect utility from the menu of possible cities. For an individual of talent  $t$  the first-order condition to the city selection problem (18) with talent-homogeneous cities can be written as:

$$\left. \frac{\partial \mathbb{E}V_c(t)}{\partial L_c} \right|_{t=t_c} dL_c + \left. \frac{\partial \mathbb{E}V_c(t)}{\partial t_c} \right|_{t=t_c} dt_c = 0. \quad (21)$$

Using equations (12)-(14) and expression (17) valued at  $t = t_c$ , equation (21) becomes:

$$\left[ (\varepsilon \sigma)^{1+\varepsilon} (S t_c)^{1+a} L_c^\varepsilon - \theta \gamma L_c^\gamma \right] \frac{dL_c}{L_c} + \frac{1+a}{1+\varepsilon} (\varepsilon \sigma)^{1+\varepsilon} (S t_c)^{1+a} L_c^\varepsilon \frac{dt_c}{t_c} = 0.$$

Since the talent-homogeneous equilibrium is unique, we can directly express  $L_c$  as a function of  $t_c$  and write  $L_c \equiv L(t_c)$ . Plugging this into the previous equation yields a differential equation that determines the menu of talents and populations that supports the talent-homogeneous equilibrium:

$$\gamma \theta L(t_c)^\varepsilon \left[ \frac{\xi t_c^{1+a} - L(t_c)^{\gamma - \varepsilon}}{L(t_c)} dL(t_c) + \frac{1+a}{1+\varepsilon} \xi t_c^a dt_c \right] = 0. \quad (22)$$

To solve this differential equation, we can verify that  $L(t_c)$  is of the form  $L(t_c) = (z \xi t_c^{1+a})^{\frac{1}{\gamma - \varepsilon}}$ , for some  $z$ . Plugging this into (22) yields an equation involving the parameters of the model that is linear in  $z$ . Solving for  $z$  then gives

$$z = \left[ \frac{(\gamma - \varepsilon)(1+a)}{1+\varepsilon} \xi + \xi \right]^{\frac{1}{\gamma - \varepsilon}},$$

which satisfies (22) and allows us to obtain (19) after simplifications. Comparing (19) and (20) shows immediately that cities are overpopulated in equilibrium since  $\gamma > \varepsilon$  and  $a > 0$ . The comparative static results also follow directly from  $\gamma > \varepsilon$ .

Last, Appendix B shows that a necessary second-order condition for the talent-homogeneous equilibrium to exist is given by

$$\frac{(1 - a\varepsilon)^2}{\varepsilon} + \left[ (1 + a) - \left( \frac{1}{\varepsilon} - a \right) \right] (1 + \gamma) > \frac{Sg(S)}{\sigma} (\gamma - \varepsilon) \left( \frac{1}{\varepsilon} - a \right), \quad (23)$$

which always holds if  $\gamma - \varepsilon$  is small. ■

As made clear by equation (19), four elements determine equilibrium city population. The first is the standard trade-off between agglomeration economies (as given by  $\varepsilon$ ) and urban costs (as given by  $\gamma$  and  $\theta$ ). Unsurprisingly, equilibrium population size increases with agglomeration economies and decreases with urban costs. The second determinant of city population is talent. More talented cities have a larger population. This is because more talented individuals are on average more productive as entrepreneurs and more efficient as workers. Both aspects increase productivity and are magnified by agglomeration economies. The third determinant of city population is the distribution of luck which affects both the luck threshold to become an entrepreneur ( $S$ ) and the share of efficiency units of labour used in production ( $\sigma$ ). Last, heterogeneity among workers ( $a$ ) also affects city population size. With a higher  $a$ , more talented individuals are relatively more productive as workers and this again gets magnified by agglomeration economies.

We may view the equilibrium function  $L(t_c)$  as describing an envelope of indifference curves in  $(t_c, L_c)$  space. This function is represented by the bold curve in Figure 1. It is convex when  $\gamma < 1 + \varepsilon$ , which is empirically the case as highlighted in Section 7. Consider an individual with talent  $t_0$  choosing from the menu of equilibrium cities described by  $L(t_c)$ . Assume that she picks city  $c_1$ , which offers  $(t_1, L_1)$ . In that case, this individual faces the indifference curve  $\mathbb{E}V_{c_1}(t_0)$ , which describes all the combinations of talent  $t_c$  and population  $L_c$  that offer her the same expected utility as city  $c_1$  conditional on her talent  $t_0$ . The lower indifference curve  $\mathbb{E}V_{c_0}(t_0)$  describes all the combinations of talent  $t_c$  and population  $L_c$  that offer the same expected utility as city  $c_0$  conditional on a talent  $t_0$ .<sup>11</sup> Since expected indirect utility is increasing as indifference curves move down and right (i.e., in the direction represented by the arrow),  $\mathbb{E}V_{c_0}(t_0)$  maximises the expected utility of an individual with talent  $t_0$  subject to the equilibrium menu of cities. Hence, for this individual with talent  $t_0$  utility is maximised in a city where all individuals have the same talent  $t_0$  as hers. More generally, the bold curve  $L(t_c)$  is the envelope of indifference curves for all levels of talent. As we move up this curve, we progressively read the optimal choices of individuals with higher talent. These are larger cities. The convexity of the relationship implies that small differences in talent may translate into large differences in city size.

---

<sup>11</sup>Observe that this curve yields higher utility as it has smaller cities (lower urban costs) and more talent (higher

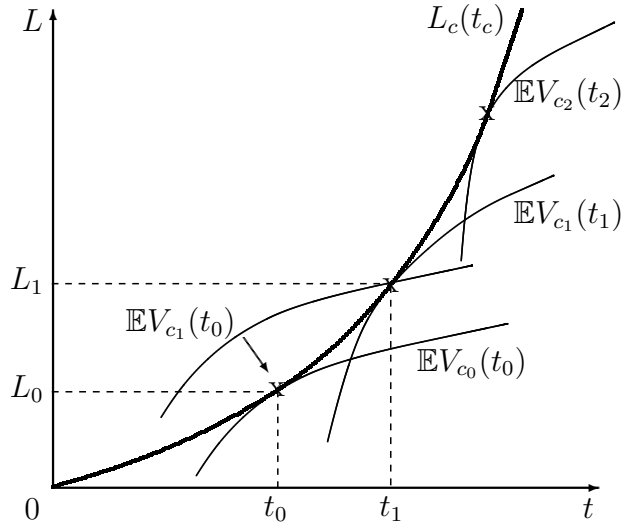


Figure 1: Equilibrium with talent-homogeneous cities

Cities are too large in equilibrium. To understand why this occurs and gain insight about sorting, consider the following hypothetical situation of a small isolated city and a mass of individuals outside. Provided the reservation level of the latter is low enough, they should gradually move to the city. Because  $\gamma > \varepsilon$ , there exists an optimal city size and the expected utility of all individuals in the city, as it grows, should first increase before decreasing after the city passes its optimal size. In standard models of urban systems (e.g., Henderson, 1974), an equilibrium with cities that are too large is reached when expected utility inside the city is equal to the reservation level.

Because individuals differ in talent, things are more complicated in our case. Heuristically, as shown by Proposition 3, more talented workers benefit more from the city being larger. Hence, as the city grows, we reach a situation where the expected utility of more talented individuals keeps growing while that of less talented individuals declines. This leads to further in-migration of more talented individuals and out-migration of less talented individuals. Eventually, the city ends up being too large and populated only by the most talented individuals. Taking these individuals out, we can repeat the same thought experiment for the city with the most talented individuals among those that remain, etc.<sup>12</sup> However, for talent-homogeneous cities to be an equilibrium outcome we

---

productivity): this is confirmed (locally at  $t = t_c$ ) by computing  $\partial \mathbb{E}V_c(t) / \partial L_c \big|_{t=t_c} > 0$  and  $\partial \mathbb{E}V_c(t) / \partial t_c \big|_{t=t_c} > 0$ , which yields the shape of the indifference curves.

<sup>12</sup>Interestingly, city sizes are uniquely determined in equilibrium. The trade-off between agglomeration economies and urban costs leads to net output per resident being a bell-shaped function of city population. With homogeneous individuals, there is a coordination failure in city formation so that any population size between optimal city size and grossly oversized cities – leaving their residents with zero consumption – can occur in equilibrium (Henderson and Becker, 2000). In our model, the sorting of heterogeneous individuals makes this indeterminacy disappear entirely. Formally, this follows from Proposition 3 and from the uniqueness of the solution to the differential equation. Intuitively, more talented cities must be larger in equilibrium to attract more talented individuals and



also need more talented individuals not to move to less talented cities. Individuals face a tradeoff between being in a larger city where population size has a positive effect on the wage rate and entrepreneurial profits and being in a less talented city where they face lower urban costs and where they expect to be relatively more productive. At the talent-homogeneous equilibrium this tradeoff is resolved with individuals optimally choosing cities of the same level of talent as theirs.

Empirically, our equilibrium matches well several of the key features of the data. That larger cities host more talented individuals is documented extensively in the literature (e.g., Berry and Glaeser, 2005; Bacolod, Blum, and Strange, 2009; Lee, 2010; Combes, Duranton, Gobillon, and Roux, 2012). For 276 US MSAs in 2000, the elasticity of the share of college graduates with respect to population is 6.8%. For more talented individuals to sort into larger cities where urban costs are higher, their rewards must be relatively higher there. This is exactly what the empirical literature finds (Wheeler, 2001; Bacolod, Blum, and Strange, 2009; Glaeser and Resseger, 2010). It is also the case that more talented individuals migrate to areas that offer them higher rewards (Dahl, 2002).

In our model, ability sorting does not imply perfect productivity sorting for firms or workers. Large cities host on average more productive firms but they also contain lots of firms with low productivity (Combes, Duranton, Gobillon, Puga, and Roux, 2012). The same, of course, holds for worker productivity (Combes, Duranton, Gobillon, and Roux, 2012). More specifically, these two papers find that the empirical distributions of log firm productivity, worker fixed effects, and log wages in denser employment areas in France are, to a good first approximation, shifted to the right relative to their corresponding distribution in less dense areas. Our equilibrium with talent-homogeneous cities also predicts these three shifts.

Finally, our results are consistent with a recurrent finding in the literature that the higher per capita output in larger cities is in part a reflection of the sorting of more productive individuals (Combes, Duranton, and Gobillon, 2008; Glaeser and Resseger, 2010; Baum-Snow and Pavan, 2012). We develop this point further in Section 7.

## 5 The size distribution of cities

Our next proposition establishes a number of properties about the ‘number’ (or mass) of cities and their population size distribution. In particular, we show that the latter converges to Zipf’s law as the difference between  $\gamma$  and  $\varepsilon$  goes to zero. This result is striking because it holds regardless of the underlying distribution of talent. In other words, provided that the gap between urban costs and agglomeration economies is small – a condition that finds strong empirical support, as highlighted in Section 7 – the size distribution of cities will be close to log-linear with slope  $-1$  no matter the distribution of talent in the population.

---

discourage less talented individuals. At the same time, they cannot be so much larger without discouraging more talented individuals as well. At the limit with a continuum of talents and talent-homogeneous cities, equilibrium city population sizes are uniquely determined.

To establish this result, we need to impose some mild technical restrictions. Namely, we assume that the support of the distribution of talent  $g_t(\cdot)$  is compact and includes  $\xi^{-1/(1+a)}$ . We assume further that the distribution of talent is finite valued and infinitely continuously differentiable around  $\xi^{-1/(1+a)}$ . One last technical condition is added in the proof below.

**Proposition 5 (Number and size distribution of cities)** *The equilibrium ‘number’ of cities is proportional to population size  $\Lambda$  and too small relative to the social optimum. The size distribution of cities converges to Zipf’s law as  $\eta \equiv (\gamma - \varepsilon)/(1 + a)$  goes to zero.*

**Proof.** Using the assignment function defined in (18), we can write the full-population condition (4) as:

$$\begin{aligned} \Lambda \int_{\underline{t}}^t g_t(\nu) d\nu &= \int_0^{\mu(t)} L(c) dc, \quad \forall t \in [\underline{t}, \bar{t}] \\ \mu(\underline{t}) &= 0. \end{aligned} \quad (24)$$

Differentiating the first expression in equation (24) and using the definition of the assignment function yields  $\Lambda g_t(t) = L(\mu(t))\mu'(t) = L(t)\mu'(t)$ , where  $\mu'(t)$  can naturally be interpreted as the density of cities hosting individuals with talent  $t$ . Solving this differential equation for  $\mu(t)$  implies

$$\mu(t) = \mu(\underline{t}) + \Lambda \int_{\underline{t}}^t \frac{g_t(\nu)}{L(\nu)} d\nu. \quad (25)$$

The measure of the set of cities  $C$  is such that  $\mu(\bar{t}) = \bar{c}$ . Inserting this into equation (25) and using the fact that  $\mu(\underline{t})$  is equal to zero by equation (24) leads to:

$$\bar{c} = \mu(\bar{t}) = \Lambda \int_{\underline{t}}^{\bar{t}} \frac{g_t(\nu)}{L(\nu)} d\nu. \quad (26)$$

This shows immediately that  $\bar{c}$ , the ‘number’ of cities, increases proportionately with  $\Lambda$  and establishes the first part of the proposition.

Let  $\tilde{\xi} \equiv \xi$  at the optimal solution and  $\tilde{\xi} \equiv \frac{1+\gamma}{1+\varepsilon}\xi$  at the equilibrium. Then, using equations (19) and (20), equations (25) and (26) may be rewritten as

$$\mu(t) = \frac{\Lambda}{\tilde{\xi}^{\frac{1}{\gamma-\varepsilon}}} \int_{\underline{t}}^t \frac{g_t(\nu)}{\nu^{\frac{1+a}{\gamma-\varepsilon}}} d\nu \quad \text{and} \quad \bar{c} = \mu(\bar{t}) = \frac{\Lambda}{\tilde{\xi}^{\frac{1}{\gamma-\varepsilon}}} \int_{\underline{t}}^{\bar{t}} \frac{g_t(\nu)}{\nu^{\frac{1+a}{\gamma-\varepsilon}}} d\nu. \quad (27)$$

Because,  $\tilde{\xi}$  is smaller at the social optimum than at the equilibrium, the second equality in expression (27) immediately implies the second part of the result. Since all cities are oversized there are too few of them.

Using the change in variables from talent to equilibrium city size in expression (27) allows us to derive the probability distribution function for the population size of cities:

$$G_L(L; \eta) \equiv \frac{\mu(L)}{\mu(\bar{L})} = \frac{\int_{\underline{L}}^L g_t \left( \xi^{-\frac{1}{1+a}} \ell^\eta \right) \ell^{\eta-2} d\ell}{\int_{\underline{L}}^{\bar{L}} g_t \left( \xi^{-\frac{1}{1+a}} \ell^\eta \right) \ell^{\eta-2} d\ell} \Rightarrow g_L(L; \eta) = \frac{g_t \left( \xi^{-\frac{1}{1+a}} L^\eta \right)}{\int_{\underline{L}}^{\bar{L}} g_t \left( \xi^{-\frac{1}{1+a}} \ell^\eta \right) \ell^{\eta-2} d\ell} L^{\eta-2}, \quad (28)$$

where  $\underline{L} \equiv \underline{t}^{\frac{1+a}{\gamma-\varepsilon}} \tilde{\xi}^{\frac{1}{\gamma-\varepsilon}}$  and  $\bar{L} \equiv \bar{t}^{\frac{1+a}{\gamma-\varepsilon}} \tilde{\xi}^{\frac{1}{\gamma-\varepsilon}}$ .

Using a Taylor expansion of the second expression of (28) around  $\eta = 0$  yields

$$g_L(L; \eta) = \sum_{i=0}^{\infty} \frac{g_L^{(i)}}{i!} \eta^i = \frac{\bar{L} \underline{L}}{\bar{L} - \underline{L}} L^{-2} + O(\eta).$$

The second-order remainder term above converges to zero as  $\eta$  tends to zero. This holds under the assumption that  $g_L^{(i)}$ , which defines the  $i^{\text{th}}$  derivative of  $g_L(\cdot)$  with respect to  $\eta$  evaluated at  $\eta = 0$ , satisfies  $|g_L^{(i)}| \leq K$  for some  $K < \infty$  and for all  $i$ . ■

That the number of cities should be proportional to total population is natural in a context where cities are endogenous and there are no increasing aggregate returns. It is also easy to understand that if cities are too large, as indicated by Proposition 4, there will be too few of them. The third part of Proposition 5 shows that the size distribution of cities converges to Zipf's law for any distribution of talent when  $\eta$  approaches zero. This is an important result since Zipf's law is a reasonable first-order approximation for observed distributions of city population sizes (Gabaix and Ioannides, 2004). As shown in Section 7, the difference  $\gamma - \varepsilon$  is empirically small, around 3%. We show in Appendix C that for such values of the parameters, the Zipf approximation works extremely well.

To understand the key intuitions behind this result, it is useful to proceed in steps. First, it is easy to see that if talent follows a Pareto distribution, the size distribution of cities is also Pareto. This occurs because both equilibrium and optimal city sizes in equations (19) and (20) and are power functions of talent in the city. Then, any power transformation of a Pareto distribution is also a Pareto distribution and the result obtains. Second, by equation (19), city size is proportional to  $t^{(1+a)/(\gamma-\varepsilon)}$ . Hence, the 'number' of cities of talent  $t$  is given by the 'number' of individuals with this level of talent divided by the size of those cities. That is, if talent is Pareto distributed with shape parameter  $m$ , the distribution of city population sizes by talent is Pareto with shape parameter  $m + \frac{1+a}{\gamma-\varepsilon}$ . In turn, the size distribution of cities, is obtained from a change of variable using the fact that  $L^{(\gamma-\varepsilon)/(1+a)}$  is proportional to  $t$ . This yields a shape parameter of  $1 + m(\gamma - \varepsilon)/(1 + a)$  for the size distribution of cities. Hence, when  $(\gamma - \varepsilon)/(1 + a)$  is small, the shape parameter of the size distribution of cities is close to one. Finally, as shown by the proposition, Zipf's law occurs for any distribution of talent. The reason is that one can always write the local Pareto shape of any distribution as  $m(t)$ . Then, provided that  $(\gamma - \varepsilon)/(1 + a)$  remains small, any deviation of the distribution of talent from a Pareto distribution  $m(t)$  can be neglected.

## 6 Two extensions

### 6.1 Discrete cities with heterogeneous talent and variable selection

The talent-homogeneous equilibrium we investigate above is consistent with key stylised facts about agglomeration, sorting, selection, and the size distribution of cities. In particular, if we take seriously the empirical results of Combes, Duranton, Gobillon, Puga, and Roux (2012) that the intensity of selection is constant across cities, one should look for equilibria with constant selection. The equilibrium with talent-homogeneous cities is a particular case within this class of candidate equilibria.

In Appendix D we investigate an example of equilibrium with discrete cities. In this case, no analytical results can be obtained in general. Numerical computations are needed. Because cities are in finite number while there is a continuum of talents, cities are no longer talent homogeneous in equilibrium. Then, this implies that selection also generally differs across cities. Despite these important differences, many of the key properties of the equilibrium with talent-homogeneous cities are also properties of this equilibrium with variable selection including the links between city size, productivity, and ‘city talent’. Interestingly, the selection cutoffs across cities differ only marginally. These similarities are a good reason to focus on the simpler and analytically tractable case of talent-homogeneous cities.

In a separate web appendix (Appendix G), we also provide some results about the optimal allocation of talent across cities. Optimal and equilibrium agglomeration and selection coincide. Turning to sorting, we show that a benevolent planner may also want to create talent-homogeneous cities. Although the conditions under which talent-homogeneous cities occur in equilibrium and at the social optimum do not perfectly overlap, talent-homogeneous cities can also be an optimal outcome.

### 6.2 A dynamic version of the model with relocations

So far, our model is static and individuals are ‘stuck’ in the city they initially chose. Though convenient and useful, this assumption is extreme. As we underscore in Section 2, allowing individuals to relocate at no cost once their productivity  $\varphi$  is fully revealed yields perfect productivity sorting as in Nocke (2006), which is counterfactual. In an extension of the model in which time runs indefinitely, we now allow agents to relocate at a cost.

The setting is as follows. We assume that time  $T$  runs discretely. Each individual has a probability of dying  $\delta \in (0, 1)$  at each period. A fraction  $\delta$  of newborns also appear every period so that aggregate population,  $\Lambda$ , is constant. Individuals are endowed with a talent  $t$  for life but they may draw their luck  $s$  multiple times over their lifetime. In all other respects the setting is the same as in the model described above.

Consider some arbitrary time  $T$ . A newborn observes her talent  $t$ , makes a location decision,

and receives a draw  $s$  of luck for this location. Upon observing her productivity  $\varphi = t \times s$ , an individual may either ‘stay’ in the chosen city and select an occupation (worker or entrepreneur) or ‘move’ to get another draw of luck. Getting another draw of luck is costly.<sup>13</sup> It involves exiting the current location at time  $T$  (hence the term ‘move’), wait in the hinterland for one period during which utility is normalised to zero, and pick a new location at time  $T + 1$  (possibly different from the one at time  $T$ ). Likewise, individuals already alive at time  $T - 1$  have the choice between staying where they are and stick with their current productivity, and moving to get a new draw of luck.

We define a talent-homogeneous steady state as an equilibrium in which the following two conditions hold:

1. Individuals optimally choose to live in talent-homogeneous city  $c$  such that  $t_c = t$ .
2. The lifetime value of staying with luck  $s$  is higher than the value of moving  $M(s, t_c)$  (defined as waiting one period, choosing a location and getting another draw):

$$\frac{V(s, t_c)}{\delta} \geq M(s, t_c),$$

where  $V(s, t_c)$  denotes the instantaneous utility of an individual with talent  $t = t_c$  and current luck  $s$  in talent-homogeneous city  $c$ ; and  $M(s, t_c)$  is the value of moving out of city  $c$  for that individual.

We omit time subscripts to ease notation since we are describing a steady state. After redefining  $\sigma$  the share of effective labour used in production to account for the change in the support of the distribution of luck, we can write the following proposition.

**Proposition 6 (Existence and characterisation)** *A steady state with talent-homogeneous cities with the following properties exists:*

1. All cities are talent-homogeneous with population  $L(t_c) = \left(\frac{1+\gamma}{1+\varepsilon} \xi t_c^{1+a}\right)^{\frac{1}{\gamma-\varepsilon}}$  as in the static model above.
2. There exists a unique threshold  $\hat{s} \in (0, \infty)$  such that individuals with  $s \geq \hat{s}$  stay in the same city with the same productivity for the remainder of their lifetime while those with  $s < \hat{s}$  move.
3.  $\hat{s}$  is a decreasing function of the death rate  $\delta$  with  $\lim_{\delta \rightarrow 1} \hat{s}(\delta) = 0$ .
4.  $\hat{s}$  is the same in all cities.
5. The fraction of movers in the economy is constant and equal to  $\frac{G_s(\hat{s})}{1-G_s(\hat{s})}\delta$ .

**Proof.** See Appendix E. ■

---

<sup>13</sup>If it was not, everybody would keep drawing a new  $s$  until getting the upper bound of  $s$ .

That is to say, a steady state with the same characteristics as the equilibrium with talent-homogeneous cities described in Propositions 4 and 5 exist.

## 7 Quantitative implications

We now use our framework to revisit several empirical results. Since the model is highly stylised, this exercise should be viewed as ‘theory with numbers’, not empirical analysis.

Equation (16) provides an expression for output in each location. In this expression, which holds for any allocation of talent, output depends on local population and a complex function of the distribution of talent. If we are willing to proxy this complex function of the distribution of talent with average years of education, we obtain the first key estimation of Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) who regress output per capita  $y_c \equiv Y_c/L_c$  in 1499 regions of 105 countries and find:

$$\ln(y_c) = \epsilon \ln L_c + f(G_{t,c}(\cdot), G_s(\cdot)) \approx 0.068 \log L_c + 0.257 \text{Educ}_c + \text{controls}_c + v_c, \quad (29)$$

where  $G_{t,c}(\cdot)$  is the distribution of talent in  $c$ ,  $G_s(\cdot)$  is the distribution of luck, and  $f(\cdot)$  is a function that we obtain from equation (16). This regression implies a value of 0.068 for  $\epsilon$ . It also points at the importance of human capital and education as determinant of output per capita. However, their coefficient of 0.257 on average years of education does not have a structural interpretation in our framework given the complexity of the function  $f(\cdot)$  and the unknown mapping between education and talent.

Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) also work with micro data for 6314 firms in 76 regions of 20 countries. More specifically, they regress firm revenue  $Z_i$  on its employment, the education of its ‘entrepreneur’, the average education of its workers, and, as above, local population and local average education. Our model does not generate this specification. However, we can extend it to allow for limited span of control of entrepreneurs in the wake of Lucas (1978) and obtain this exact specification while leaving its other properties unchanged.<sup>14</sup> We develop this extension in a separate web appendix (Appendix I). Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) find

$$\ln Z_i = 0.126 \log L_{c(i)} + 0.073 \text{Educ}_{c(i)} + 0.860 N_i + 0.017 \text{Educ}_i^W + 0.026 \text{Educ}_i^E + \text{controls}_{c(i)} + v_i,$$

where  $N_i$  is the employment count of firm  $i$ ,  $\text{Educ}_i^W$  the average years of education of its workers, and  $\text{Educ}_i^E$  the years of education of its entrepreneur. According to our extended model, the coefficient on local population of 0.126 is another estimate of  $\epsilon$ , our agglomeration parameter. The coefficient on employment of 0.860 should be equal to  $(1 - \alpha)/(1 + \epsilon)$  where  $\alpha$  is the span of control parameter

---

<sup>14</sup>By an artefact of the constant elasticity of demand, a linear production function (as we use above) implies that the productivity of entrepreneurs and workers cancel out when computing firm revenue. This is easily avoided by imposing decreasing returns to scale in production, using for instance a limited span of control argument.

(corresponding implicitly to the share of the entrepreneur in production). Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) take  $\alpha$  to be about 0.1, which implies in our extended model a coefficient on employment very close to the one they measure:  $(1 - \alpha)/(1 + \varepsilon) = (1 - 0.1)/(1 + 0.126) = 0.800$ , instead of 0.860. Even more interesting, the coefficient on entrepreneurial education is higher than that on worker education. Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) interpret this as evidence of extremely high returns to education for entrepreneurs.<sup>15</sup> Our model suggests an alternative explanation. Recall that our model indicates that only the most productive individuals become entrepreneurs while the others become workers. Put differently, given talent (or education in this empirical implementation), only individuals with particularly good draws of luck become entrepreneurs whereas the others (with bad draws) become workers. Put differently, the coefficient on entrepreneurial education is biased upwards while that on worker education is biased downwards. Whether returns to education are particularly high for entrepreneurs and managers after accounting for positive selection into these occupations is an open question.

Next we exploit the restrictions of our original model at the talent-homogeneous equilibrium. Observe that the expected indirect utility (17) can be written as  $\mathbb{E}V_c(t_c) = \sigma^{1+\varepsilon}(St_c)^{1+a}(\varepsilon L_c)^\varepsilon - \theta L_c^\gamma = y_c - \theta L_c^\gamma$ . Taking logs, we have

$$\ln y_c = \kappa_1 + (1 + a) \ln t_c + \varepsilon \ln L_c, \quad (30)$$

where  $\kappa_1$  is a constant term. Expression (30) shows that regressing average earnings on population, while controlling for talent, yields an estimate of agglomeration economies,  $\varepsilon$ . Now, using the equilibrium relationship linking city population sizes to the distribution of talent (19),  $L_c^{\gamma-\varepsilon} = \xi((1 + \gamma)/(1 + \varepsilon))t_c^{1+a}$ , to control for the shift we get:

$$\ln y_c = \kappa_2 + \gamma \ln L_c, \quad (31)$$

where  $\kappa_2$  is another constant term. Hence, regressing average earnings on population *without* controlling for talent yields an estimate of the urban costs parameter,  $\gamma$ .<sup>16</sup>

We estimate equations (30) and (31) using standard US Census data for 276 metropolitan statistical areas in 2000. We measure  $y_c$  with city average earnings and  $t_c$  with the share of the population older than 18 years with at least an associate degree, following standard practice in

---

<sup>15</sup>With  $\alpha = 0.1$ , the returns to education for entrepreneurs in the framework of Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) are  $0.026/0.1 = 26\%$ .

<sup>16</sup>The result that the elasticity of income with respect to city size equals the elasticity of urban costs seems a priori surprising since utility is not equalised across cities in our framework. Yet, this result holds because in our model at equilibrium  $\mathbb{E}V_c(t_c) = \sigma^{1+\varepsilon}(St_c)^{1+a}(\varepsilon L_c)^\varepsilon - \theta L_c^\gamma = \kappa_3 L_c^\gamma$  where  $\kappa_3$  is a positive bundle of parameters. As can be seen from this expression, the population elasticity of equilibrium utility is  $\gamma$ , which is also the population elasticity of urban costs. We show in a separate web appendix (Appendix J) that this result holds beyond our specific model although this will not be true in general.

labour economics. We obtain:<sup>17</sup>

$$\ln y_c = 8.59 + 0.082 \ln L_c, \quad (32)$$

$$\ln y_c = 9.60 + 0.051 \ln L_c + 0.46 \ln t_c. \quad (33)$$

These two regressions imply  $\hat{\gamma} = 0.082$  and  $\hat{\varepsilon} = 0.051$ . These coefficients on log-population are robust to alternative measures of  $y_c$  and  $t_c$ . For instance, if we take income per capita instead of average earnings, we obtain estimates of 0.067 for  $\gamma$  and 0.039 for  $\varepsilon$ . Using the share of population older than 18 years with a graduate or professional degree to measure  $t_c$  in regression (33) yields a coefficient of 0.058 on log population.<sup>18</sup> Note that we refrain from interpreting the coefficient of  $\ln t_c$  as providing an estimate of  $1 + a$  since we do not know how talent maps into education. The reason is that while population is measured accurately, education is only a rough proxy for ‘talent’.

Our preferred estimate of the elasticity of earnings,  $\hat{\varepsilon} = 0.051$ , is within the usual range in the literature. See Glaeser and Resseger (2010) for recent results on US data and Rosenthal and Strange (2004) or Melo, Graham, and Noland (2009) for broader reviews.<sup>19</sup> The sizable drop in the coefficient for log population after adding a measure of city education is also typical (Combes, Duranton, and Gobillon, 2008; Glaeser and Resseger, 2010). Our favourite estimate for the elasticity of urban costs is  $\hat{\gamma} = 0.082$ . A monocentric model with linear commuting costs implies much higher elasticities: between 0.66 (for a two dimensional city) and 1 (for a one dimensional city as we use here). However, recent work on US cities (Albouy, 2009; Baum-Snow and Pavan, 2012) or French land markets (Combes, Duranton and Gobillon, 2012) reports estimates of  $\gamma$  between 0.033 and 0.12 that are close to ours.

To corroborate our findings further, we also estimate the elasticity of urban costs with respect to population size using housing rents,  $r_c$ , to measure urban costs directly:

$$\ln r_c = 5.19 + 0.085 \ln L_c.$$

This coefficient of 0.085 is remarkably close to the coefficient of 0.082 estimated in regression (32). Arguably, renters differ from homeowners and their rents may not reflect typical urban costs. As a further robustness test, assume that the price index for housing in city  $c$  is given by  $h_c = v_c^{\alpha_c} r_c^{1-\alpha_c}$ , where  $v_c$  is the value of owner-occupied housing and  $r_c$  the rents paid for renter occupied housing. We measure  $\alpha_c$  by MSA  $c$ ’s share of owner-occupied housing. Regressing the log of this housing

---

<sup>17</sup>All coefficients, including the constant terms, are significant at the 1% confidence level in all estimations.

<sup>18</sup>We ran all regressions with combinations of four different measures for  $y_c$  and three different proxies for  $t_c$ . The estimates of  $\varepsilon$  are between 0.039 and 0.078, with mean 0.043. The estimates of  $\gamma$  are between 0.066 and 0.082, with mean 0.074. Note that the average estimated difference  $\gamma - \varepsilon$  is 0.031, which is almost identical to the value we obtain in our preferred case below.

<sup>19</sup>We use city aggregated data and few controls. Using micro-data and more controls typically results in slightly lower estimates for the coefficient on city size (Combes, Duranton, and Gobillon, 2008; Glaeser and Resseger, 2010). These small differences are not important for our purpose here.



price index,  $h_c$ , on the log of population yields:

$$\ln h_c = 8.93 + 0.068 \ln L_c. \quad (34)$$

This estimate of 0.068 for urban costs remains reasonably close to that in (32) despite relying on a different estimating equation.

Using equilibrium city size as given by expression (19), the elasticity of talent to city population size should be equal to  $(\gamma - \varepsilon)/(1 + a)$ . We obtain  $\hat{\gamma} - \hat{\varepsilon} = 0.031$  using (32) and (33). Regressing directly the log-share of college graduates on log-population yields

$$\ln t_c = -2.21 + 0.068 \ln L_c.$$

This elasticity 0.068 is larger than 0.031 in a statistical sense but economically close (keeping in mind that we do not compute  $a$ ). Using a weaker definition of talent, namely the share of people who attended college irrespective of the degree they earned, yields a lower elasticity of 0.024.

At first sight, small values for the elasticity of talent to city population size seem to argue against the importance of ability sorting across cities. Our model shows instead that a small value for the population elasticity of talent corresponds in equilibrium to the small difference between the population elasticity of urban costs and that of agglomeration economies. Then, the counterpart of a small population elasticity of talent is a large talent elasticity of population. Put differently, city population size is proportional to  $t_c^{(1+a)/(\gamma-\varepsilon)}$ . A small difference between  $\gamma$  and  $\varepsilon$  is then enough for small differences in talent to translate into large differences in city population size. For instance, if  $a = 0.1$ ,  $\underline{L} = 10,000$  inhabitants for the smallest city in the economy, and  $\bar{L} = 10$  million for the largest, then the largest city is ‘only’ about 14% more talented than the smallest given our estimates of  $\gamma$  and  $\varepsilon$ . This result that small differences in talent lead to large differences in city size is reminiscent of Gabaix and Landier (2008), who find that small differences in CEO talent may translate into large pay differences because the best CEOs are assigned to the largest firms at the competitive equilibrium.

Our model also predicts that the share of expenditure on housing is independent of city population. To see this, we note that total land rents are given by  $\theta\gamma L_c^{1+\gamma}$  as shown in a separate web appendix (Appendix F), whereas aggregate income is equal to  $Y_c$ . Taking the ratio of total land rent to aggregate income, using equations (16) and (19) for talent-homogeneous cities, and the definitions of  $\sigma$  and  $\xi$ , we obtain

$$\frac{TLR_c}{Y_c} = \frac{(1 + \gamma)\varepsilon}{1 + \varepsilon},$$

after simplifications. This quantity does not depend on city population size and is equal to 0.052 for our preferred estimates of  $\varepsilon$  and  $\gamma$ . This result is important for two reasons. First, it is in line with findings by Davis and Ortalo-Magné (2011). They show that expenditure shares on housing are constant over time and across US MSAs at around 24%. If we take a share of land in housing

of 18% as in Combes, Duranton, and Gobillon (2012), we find an empirical value of  $TLR/Y$  equal to  $0.24 \times 0.18 = 0.043$ , which is close to 0.052. Second, this result of a constant share of land is obtained with an additively separable utility function. Hence, Cobb-Douglas preferences are not required to generate constant expenditure shares on housing across cities.

Next, using again  $\hat{\varepsilon} = 0.051$  and  $\hat{\gamma} = 0.082$ , it is easy to compute how oversized cities are:

$$\frac{\widehat{L}_c^*}{\widehat{L}_c^o} = \left( \frac{1 + \hat{\gamma}}{1 + \hat{\varepsilon}} \right)^{\frac{1}{\hat{\gamma} - \hat{\varepsilon}}} = 2.55. \quad (35)$$

This suggests that US cities may be, on average, about 155% larger than their optimal size. To check the robustness of this finding, Figure 2 plots the oversize of cities as computed in equation (35) for varying values of  $\gamma$  and three values of  $\varepsilon$ . This plot indicates that an oversize of 145 to 165% is to be expected. Consistent with the comparative statics of equation (35), the figure also shows that city oversize decreases in  $\gamma$  and  $\varepsilon$  ( $< \gamma$ ). Using a first-order linear approximation of equation (35) when  $\gamma - \varepsilon$  is small, we obtain  $L_c^*/L_c^o \approx \exp(\frac{1}{1+\varepsilon})$  which tends to Euler's number when  $\varepsilon$  and  $\gamma$  go to zero. Given that  $\varepsilon$  and  $\gamma$  are empirically small, cities are 'naturally' oversized by a factor close to  $e \approx 2.72$ .

This oversize may seem like a considerable inefficiency. However, the associated welfare loss in consumption is tiny. To see this, we use equations (19), (20), and (35) to compute an estimate of the indirect utility (consumption) loss:

$$\widehat{\Delta EV} \equiv \frac{\mathbb{E}V(\widehat{L}^*) - \mathbb{E}V(\widehat{L}^o)}{\mathbb{E}V(\widehat{L}^o)} = \frac{1}{1 + \hat{\gamma}} \left( \frac{\widehat{L}^*}{\widehat{L}^o} \right)^{\hat{\gamma}} - 1 = -0.2\%.$$

This loss in consumption is economically small, about one-fifth of a percentage point. To confirm the robustness of this magnitude, Figure 3 plots the economic loss associated with this oversize for the same parameter values as Figure 2. It is less than half a percentage point.

The reason why losses from oversized cities are so small is the following. Recall first that cities are oversized by a factor close to  $e = 2.72$ . Imagine next that earnings ( $Y_c/L_c$ ) are of the same magnitude as urban costs ( $\theta L_c^\gamma$ ). Then, the maximum loss from oversize would be  $1 - e^{-(\gamma - \varepsilon)}$  or about 2.8% for our preferred value of  $\hat{\gamma} - \hat{\varepsilon} = 0.031$ . However, equilibrium urban costs are only about 5% of earnings as shown above. Hence, the actual loss is much smaller than 2.8%. These results are consistent with those of Au and Henderson (2006) for Chinese cities. Using the fact that Chinese migration policies have limited the growth of Chinese cities, they estimate the shape of net benefits from cities as a function of their size. Like us, they find a very flat curve past the optimum. This suggests that restricting the size and growth of cities is unlikely to deliver substantial welfare improvements.<sup>20</sup>

---

<sup>20</sup>Figure 3 also shows that the loss from oversize increases with agglomeration economies,  $\varepsilon$ . This is because a higher  $\varepsilon$  (given  $\gamma$ ) implies larger cities. In turn this magnifies the inefficiency from oversize (that the oversize becomes relatively smaller with  $\varepsilon$  only partially offsets this).

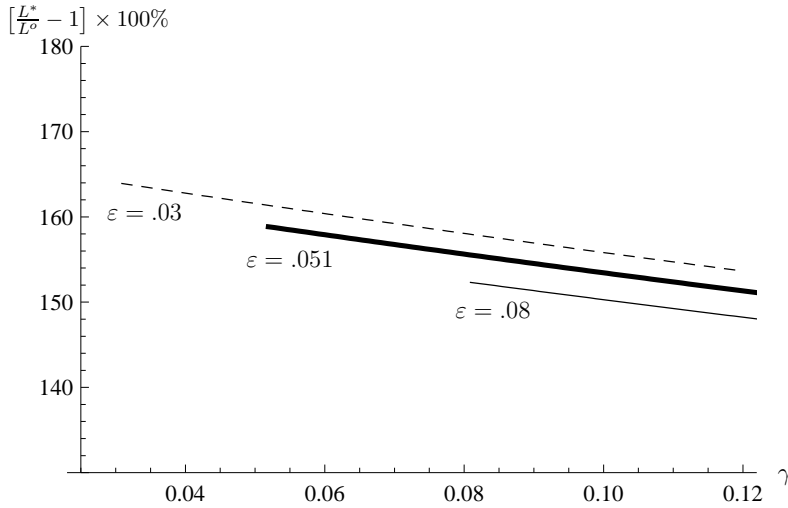


Figure 2: Oversize as a function of  $\gamma$  for different values of  $\epsilon < \gamma$

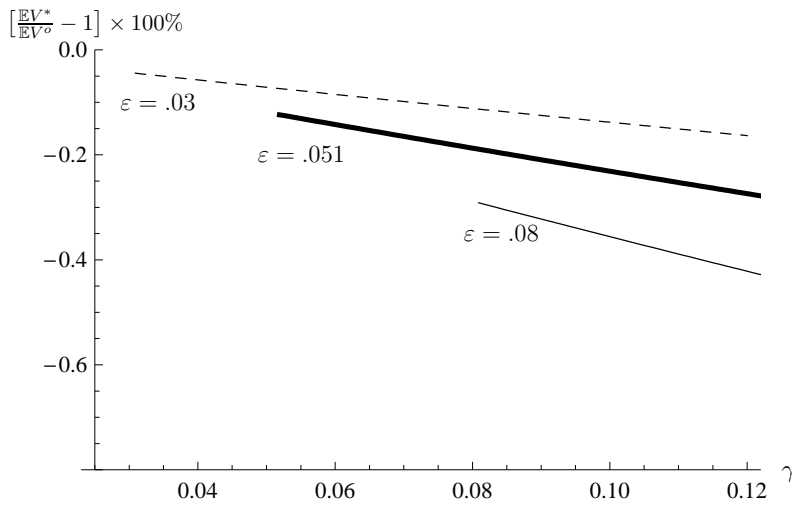


Figure 3: Efficiency loss as a function of  $\gamma$  for different values of  $\epsilon < \gamma$

## 8 Conclusion

Although abundant empirical research in urban economics has substantiated a significant positive correlation between skills and city population, theory has had much less to say about the spatial sorting of heterogeneous individuals across an urban hierarchy until now. Our paper is an attempt to make progress in this direction. We have shown that *ex ante* sorting along talent and *ex post* selection along productivity, when coupled with an otherwise standard model of agglomeration economies and monocentric cities, allow us to replicate key stylised facts: larger cities host more talented individuals, have more productive (but not a greater proportion of) entrepreneurs, pay

higher wages, and have higher urban costs. Importantly, even though firms in larger cities are on average more productive than in smaller cities, this hides considerable heterogeneity. Finally, the distribution of talent maps into the distribution of city population sizes and this provides a simple static explanation for Zipf’s law.

In addition to our theoretical contributions, our model also provides a unified setting within which to interpret quantitative evidence obtained from various standard regressions. It suggests, in particular, how regressions of measures of productivity, skills, and urban costs on log population can be consistently interpreted and how they relate quite naturally to each other. We believe that such an interpretative framework is useful for guiding future empirical analysis.

Second, our concept of spatial equilibrium is useful to shed new light on spatial arbitrage when static utility levels are not equalised across locations. Most of the literature in urban economics focuses on situations where all individuals are indifferent across all cities. But there may be no one at the margin between two very different cities when different types of individuals locate in different subsets of the urban hierarchy. This fact has numerous implications on how changes in various economic variables affect the urban landscape: when some individuals strictly prefer larger cities, local policy makers have an additional degree of freedom since their ‘tax base’ becomes less mobile. Our model may thus be especially useful for addressing local policy issues.

Finally, cities are essentially passive in our model. In reality, cities, especially the most talented ones, actively limit their population growth and this may foster sorting even further (Gyourko, Mayer, and Sinai, 2006). Allowing cities to play a more active role within our framework figures prominently on our research agenda.

**Acknowledgements.** We thank Arnaud Costinot, Xavier Gabaix, Ed Glaeser, Laurent Gobillon, Delfim Gomes Neto, Tom Holmes, Yannis Ioannides, Sam Kortum, Sanghoon Lee, Diego Puga, Jacques Thisse, Jonathan Vogel, many conference and seminar participants, and especially Vernon Henderson, Esteban Rossi-Hansberg, and two anonymous referees for very helpful discussions and comments. Behrens is holder of the *Canada Research Chair in Regional Impacts of Globalization*. Financial support from the CRC Program of the Social Sciences and Humanities Research Council (SSHRC) of Canada is gratefully acknowledged. Behrens further gratefully acknowledges financial support from FQRSC Québec (Grant NP-127178). Robert-Nicoud was a visiting scholar at Princeton University when the first draft of this paper was written; the hospitality of this institution is gratefully acknowledged. Any remaining errors are ours.

## References

- [1] Aban, Inmaculada B., Mark M. Meerschaert and Anna K. Panorska (2006) Parameter estimation for the truncated Pareto distribution, *Journal of the American Statistical Association* 101, No. 473, 270–277.

- [2] Abdel-Rahman, Hesham M. and Ping Wang (1997) Social welfare and income inequality in a system of cities, *Journal of Urban Economics* 41(3), 462–483.
- [3] Albouy, David (2009) Are big cities really bad places to live? Improving quality-of-life estimates across cities. *NBER Working Paper #14472*.
- [4] Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg (2006) Offshoring in a knowledge economy, *Quarterly Journal of Economics* 121(1), 31–77.
- [5] Au, Chun-Chung and J. Vernon Henderson (2006) Are Chinese cities too small?, *Review of Economic Studies* 73(3), 549–576.
- [6] Bacolod, Marigee, Bernardo S. Blum, and William C. Strange (2009) Skills in the city, *Journal of Urban Economics* 65(2), 136–153.
- [7] Baldwin, Richard E. and Toshihiro Okubo (2006) Heterogeneous firms, agglomeration and economic geography: spatial selection and sorting, *Journal of Economic Geography* 6(3), 323–346
- [8] Baum-Snow, Nathaniel and Ronni Pavan (2009) Inequality and city size. Processed, Brown University.
- [9] Baum-Snow, Nathaniel and Ronni Pavan (2012) Understanding the city size wage gap. *Review of Economic Studies* 79(1): 88–127.
- [10] Behrens, Kristian and Frédéric Robert-Nicoud (2009) Survival of the fittest in cities: Agglomeration, polarization, and income inequality. *CIRPÉE Discussion Paper #09-19*.
- [11] Berry, Christopher R. and Edward L. Glaeser (2005) The divergence of human capital levels across cities, *Papers in Regional Science* 84(3), 407–444.
- [12] Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008) Spatial wage disparities: Sorting matters!, *Journal of Urban Economics* 63(2), 723–742.
- [13] Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2011) The identification of agglomeration economies, *Journal of Economic Geography* 11(2), 253–266.
- [14] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux (2012) The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, forthcoming.
- [15] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2012) Sorting and local wage and skill distributions in France. *Regional Science and Urban Economics* 42(4), 663–680.

- [16] Dahl, Gordon B. (2002) Mobility and the return to education: Testing a Roy model with multiple markets, *Econometrica* 70(6), 2367–2420.
- [17] Davis, Donald R. (2009) A spatial knowledge economy. Processed, Columbia University.
- [18] Davis, Morris A. and Francois Ortalo-Magn. 2011. Household expenditures, wages, rents. *Review of Economic Dynamics* 14(2):248261.
- [19] Dixit, Avinash K. and Joseph E. Stiglitz (1977) Monopolistic competition and optimum product diversity *American Economic Review* 67(3), 297–308.
- [20] Duranton, Gilles and Diego Puga (2004) Micro-foundations of urban agglomeration economies. In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: North-Holland, 2063–2117.
- [21] Eeckhout, Jan (2004) Gibrat’s Law for (all) cities, *American Economic Review* 94(5), 1429–1451.
- [22] Ellison, Glenn D. and Edward L. Glaeser (1999) The geographic concentration of industry: Does natural advantage explain agglomeration?, *American Economic Review Papers and Proceedings* 89(2), 311–316.
- [23] Ethier, Wilfred J. (1982) National and international returns to scale in the modern theory of international trade, *American Economic Review* 72(3), 389–405.
- [24] Epple, Dennis and Thomas Nechyba (2004) Fiscal Decentralization. In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: North-Holland, 2423–2480.
- [25] Gabaix, Xavier and Rustam Ibragimov (2011) Rank - 1/2: A simple way to improve the OLS estimation of tail exponents, *Journal of Business & Economic Statistics*, 29(1), 24–39.
- [26] Gabaix, Xavier and Yannis M. Ioannides (2004) The evolution of city size distributions. In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: North-Holland, 2341–2378.
- [27] Gabaix, Xavier and Augustin Landier (2008) Why has CEO pay increased so much?, *Quarterly Journal of Economics* 123(1), 49–1000.
- [28] Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2012) Human capital and Regional Development. Processed, Harvard University.
- [29] Glaeser, Edward L. and Matthew G. Resseger (2010) The complementarity between cities and skills, *Journal of Regional Science* 50(1), 221–244.

- [30] Glen, Andrew G., Lawrence M. Leemis and John H. Drew (2004) Computing the distribution of the product of two continuous random variables, *Computational Statistics & Data Analysis* 44(3), 451–464.
- [31] Gyourko, Joseph, Christopher Mayer, and Todd Sinai (2006) Superstar cities. Processed, The Wharton School, University of Pennsylvania.
- [32] Henderson, J. Vernon (1974) The sizes and types of cities, *American Economic Review* 64(4), 640–656.
- [33] Henderson, J. Vernon and Randy Becker (2000) Political economy of city sizes and formation, *Journal of Urban Economics* 48(3), 453–484.
- [34] Hendricks, Lutz (2011) The skill composition of US cities, *International Economic Review* 52(1), 1–32.
- [35] Hsu, Wen-Tai (2012) Central place theory and city size distribution. *Economic Journal*, forthcoming.
- [36] Lee, Sanghoon (2010) Ability sorting and consumer city, *Journal of Urban Economics* 68(1), 20–33.
- [37] Lee, Sanghoon and Qiang Li (2009) Uneven landscapes and the city size distribution. Processed, University of British Columbia.
- [38] Lucas, Robert E., Jr. (1978) On the size distribution of business firms, *Bell Journal of Economics* 9(2), 508–523.
- [39] Melitz, Marc J. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* 71(6), 1695–1725.
- [40] Melitz, Marc J. and Gianmarco I.P. Ottaviano (2008) Market size, trade and productivity, *Review of Economic Studies* 75(1), 295–316.
- [41] Melo, Patricia C., Daniel J. Graham and Robert B. Noland (2009) A meta-analysis of estimates of urban agglomeration economies, *Regional Science and Urban Economics* 39(3), 332–342.
- [42] Monte, Ferdinando (2011) Skill bias, trade, and wage dispersion, *Journal of International Economics* 83(2), 202–218.
- [43] Nocke, Volker (2006) A gap for me: Entrepreneurs and entry, *Journal of the European Economic Association* 4(5), 929–956.
- [44] Puga, Diego (2010) The magnitude and causes of agglomeration economies, *Journal of Regional Science* 50(1), 203–219.

- [45] Rosenthal, Stuart S. and William C. Strange (2004) Evidence on the nature and sources of agglomeration economies. In: J. Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, Volume 4, Amsterdam: North-Holland, 2119–2171.
- [46] Sattinger, Michael (1993) Assignments models of the distribution of earnings, *Journal of Economic Literature* 31(2), 831–880.
- [47] Wheeler, Christopher H. (2001) Search, sorting, and urban agglomeration, *Journal of Labor Economics* 19(4), 879–899.

## Appendix A Symmetric equilibria of the model

In this appendix, we establish that symmetric equilibria exist provided population is homogeneous enough with respect to talent. We also show that is generally a continuum of symmetric equilibria, each one featuring a different ‘number’ of cities of different population size.

**Proposition A.1 (Symmetric equilibria)** *There exists a continuum of stable equilibria such that  $F_c(\cdot) = F(\cdot)$  and  $L_c = L$  for all cities  $c$  only if the variation in talent across the population is small enough.*

**Proof.** Assume for now that  $F_c(\cdot) = F(\cdot)$  for all  $c$ . By the uniqueness of the solution to equation(14), which does not depend on  $L_c$ , we then have  $\underline{\varphi}_c = \underline{\varphi}$  for all  $c$ . This implies that selection is constant across cities:  $F_c(\underline{\varphi}_c) = F(\underline{\varphi})$  for all  $c$ . Because all types of talent are located in all cities, it must be that  $\mathbb{E}V_c(t) = \mathbb{E}V(t)$  for all cities,  $c$ , and talents,  $t$ . With  $F_c(\cdot) = F(\cdot)$ , the condition in Proposition 3 is a true single-crossing condition: more talented individuals benefit more from larger cities. Hence, it must be that  $L_c = L$  for all  $c \in C$  for all talents to be indifferent across all cities.

Symmetry is a stable equilibrium only if  $\mathbb{E}V(t) \geq 0$  and  $\partial \mathbb{E}V(t)/\partial L < 0$  for all  $t \in [t, \bar{t}]$ . The first condition ensures that individuals want to stay in existing cities since the outside option of starting a new city yields zero utility. The second condition implies that no deviation of any small mass of representative individuals to another city is profitable. Using (17), and the fact that expected indirect utility is increasing in  $t$ , these two conditions will hold for all  $t \in [t, \bar{t}]$  if and only if

$$\begin{aligned} \frac{\varepsilon}{\gamma} \left[ \int_{\underline{\varphi}}^{+\infty} \varphi^{\frac{1}{\varepsilon}} dF(\varphi) \right]^{\varepsilon} \left[ \bar{t}^a \int_0^{\underline{\varphi}/\bar{t}} s^a dG_s(s) + \underline{\varphi}^a \left( \frac{\bar{t}}{\underline{\varphi}} \right)^{\frac{1}{\varepsilon}} \int_{\underline{\varphi}/\bar{t}}^{+\infty} s^{\frac{1}{\varepsilon}} dG_s(s) \right] &\leq \theta L^{\gamma-\varepsilon} (1 + \varepsilon) \\ &\leq \left[ \int_{\underline{\varphi}}^{+\infty} \varphi^{\frac{1}{\varepsilon}} dF(\varphi) \right]^{\varepsilon} \left[ \underline{t}^a \int_0^{\underline{\varphi}/\underline{t}} s^a dG_s(s) + \underline{\varphi}^a \left( \frac{\underline{t}}{\underline{\varphi}} \right)^{\frac{1}{\varepsilon}} \int_{\underline{\varphi}/\underline{t}}^{+\infty} s^{\frac{1}{\varepsilon}} dG_s(s) \right]. \quad (\text{A.1}) \end{aligned}$$



In addition, it implies that

$$\frac{\varepsilon}{\gamma} < \frac{\underline{t}^a \int_0^{\underline{\varphi}/\underline{t}} s^a dG_s(s) + \underline{\varphi}^a \left(\frac{\underline{t}}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}} \int_{\underline{\varphi}/\underline{t}}^{+\infty} s^{\frac{1}{\varepsilon}} dG_s(s)}{\bar{t}^a \int_0^{\underline{\varphi}/\bar{t}} s^a dG_s(s) + \underline{\varphi}^a \left(\frac{\bar{t}}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}} \int_{\underline{\varphi}/\bar{t}}^{+\infty} s^{\frac{1}{\varepsilon}} dG_s(s)} \quad (\text{A.2})$$

must hold at any stable symmetric equilibrium. Since  $\gamma > \varepsilon$ , the left-hand side of this expression is smaller than unity. The right-hand side is increasing with  $\underline{t}$  and decreasing with  $\bar{t}$ . Furthermore, it is also smaller than unity, but it tends to 1 as  $\underline{t} \rightarrow \bar{t}$ . Hence, this condition is fulfilled for a ‘sufficiently homogeneous’ population ( $\underline{t} \approx \bar{t}$ ). Additionally, when  $\underline{t} < \bar{t}$  symmetric equilibria are never stable when  $\gamma \approx \varepsilon$ , i.e., when ‘net urban costs’ are small.

Last, note that condition (A.1) bounds the population  $L$  of symmetric cities. It is then easy to verify that when (A.2) holds there exists in general a continuum of pairs  $(L, N)$  of city populations and ‘number of cities’ such that (A.1) and the full-population condition  $\Lambda g_t(t) = NL(t)$ ,  $\forall t \in [\underline{t}, \bar{t}]$  hold (the latter implying, of course, that  $\Lambda = NL$ ). ■

Proposition A.1 establishes that there generally exists a continuum of stable symmetric equilibria when the variation of talent across the population is small enough, or when agglomeration economies,  $\varepsilon$ , are small compared to urban costs,  $\gamma$ . Neither case seems empirically relevant. This suggests that ability sorting is a natural equilibrium outcome.

## Appendix B Second-order conditions for the equilibrium with talent-homogeneous cities

Rewrite expression (17), evaluated at the talent-homogeneous equilibrium (19), as follows:

$$\mathbb{E}V_c(t, t_c) = \theta^{-\frac{\varepsilon}{\gamma-\varepsilon}} \left[ \frac{1 + \gamma (\varepsilon\sigma)^{1+\varepsilon}}{1 + \varepsilon \gamma} (St_c)^{1+a} \right]^{\frac{\gamma}{\gamma-\varepsilon}} \left[ \frac{\gamma}{\varepsilon} \frac{1 + \varepsilon \sigma(t, t_c)}{1 + \gamma \sigma} - 1 \right], \quad (\text{B.1})$$

where

$$\sigma(t, t_c) \equiv \frac{1}{1 + \varepsilon} \left[ \left(\frac{t}{t_c}\right)^a \int_0^{St_c/t} \left(\frac{s}{S}\right)^a dG_s(s) + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}} \int_{St_c/t}^{\infty} \left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}} dG_s(s) \right].$$

A sufficient condition for the talent-homogeneous case to be an equilibrium is that  $\mathbb{E}V(t, t_c)$  is quasi-concave in  $t_c$  for all  $(t, t_c) \in T \times T$ . Imposing quasi-concavity on (B.1) yields an expression that is so unwieldy that it is impossible to assess how restrictive it is. Imposing concavity on (B.1) at the talent-homogeneous equilibrium, a more stringent sufficient condition than quasi-concavity also yields an expression that is still quite unwieldy in general. By contrast, the following necessary second order condition is not. That is to say, we require the second derivative of (B.1) with respect

to  $t_c$  to be negative when evaluated at  $t = t_c$ . Straightforward but tedious algebra yields

$$\begin{aligned} \left. \frac{d^2}{dt_c^2} \mathbb{E}V_c(t_c, t) \right|_{t=t_c} &\propto -t_c^{-2+\frac{\gamma}{\gamma-\varepsilon}(1+a)} \sigma \left\{ -\frac{Sg(S)}{\sigma} (\gamma - \varepsilon) (1 - a\varepsilon) \right. \\ &\quad \left. + (1 - a\varepsilon)^2 + [\varepsilon(1 + a) - (1 - a\varepsilon)](1 + \gamma) \right\}, \end{aligned}$$

where  $\sigma = \sigma(t_c, t_c)$ . Hence, if a talent-homogeneous equilibrium exists, then equation (23) in the main text holds. Note that condition (23) is not overly restrictive. Indeed, as shown in Section 7,  $\gamma - \varepsilon$  is empirically small so that the right-hand side of the foregoing equation is small. Furthermore, it always holds when  $a$  is close to  $1/\varepsilon$ , regardless of the values of  $\gamma$  and  $\varepsilon$ .

## Appendix C Zipf's law

As shown in Section 5, the size distribution of cities converges to Zipf's law when  $(\gamma - \varepsilon)/(1 + a)$  goes to zero, irrespective of the underlying distribution of talent. In this appendix, we quantify the quality of this approximation when  $\gamma - \varepsilon$  is within the range of empirically plausible estimates.<sup>21</sup>

Assume that talent,  $t$ , is distributed following  $g_t(\cdot)$  on  $[\underline{t}, \bar{t}]$ . As shown in Section 5, under perfect sorting the equilibrium city population sizes are a power function of talent:  $L \propto t^{1/\eta}$ , where  $1/\eta \equiv (1 + a)/(\gamma - \varepsilon) > 0$  is the power that magnifies talent  $t$  to derive city population size  $L$ . We are interested in the distribution  $g_L$  of city population sizes. If talent  $t$  occurs  $g_t(t)dt$  times in the population, and if city sizes are linked to talent by  $L \propto t^{1/\eta}$ , the mass  $g_L(\cdot)$  of cities of size  $L$  has to be  $g_t(t)dt/L$ . Since in equilibrium  $dt \propto \eta L^{\eta-1} dL$ , straightforward substitution yields  $g_L(L; \eta) \propto \eta L^{\eta-2} g_t(L^\eta)$ . This expression is (up to some scaling) the probability distribution function of  $L$  conditional on  $\eta$ . As shown in Section 5,  $\lim_{\eta \rightarrow 0} g_L(L; \eta) \propto L^{-2}$ . In words, as  $\eta$  gets small, the power  $1/\eta$  that magnifies talent gets large and the distribution of city population size converges to a truncated Pareto distribution over the support  $[\underline{L}, \bar{L}]$ . Writing the cumulative of the untruncated Pareto as  $P$ , we obtain:

$$\lim_{\eta \rightarrow 0} g_L(L; \eta) \rightarrow p(L; k) = \frac{1}{P(\bar{L})} \frac{k}{\underline{L}} \left( \frac{L}{\underline{L}} \right)^{-k-1}, \quad \text{with } k \rightarrow 1.$$

To evaluate the quality of the power law approximation of  $g_L(\cdot)$  for a given  $\eta$ , a first natural metric is given by the distance between the two distributions:

$$d(g_L, p; \eta) \equiv \int_{\underline{L}}^{\bar{L}} \left\{ \ln[g_L(L; \eta)] - \ln[p(L; 1)] \right\}^2 dL,$$

where  $\underline{L} \equiv \underline{t}^{\frac{1+a}{\gamma-\varepsilon}} \tilde{\xi}^{\frac{1}{\gamma-\varepsilon}}$  and  $\bar{L} \equiv \bar{t}^{\frac{1+a}{\gamma-\varepsilon}} \tilde{\xi}^{\frac{1}{\gamma-\varepsilon}}$ .

A second way to judge the quality of the approximation involves generating random samples from the approximation and to estimate the parameters of the truncated Pareto distribution from those samples. The better the approximation, the closer  $\hat{k}$  must be to 1.

---

<sup>21</sup>We assume  $a = 0$ . We cannot reliably estimate this quantity and taking a lower bound allows to remain conservative when assessing the quality of the Zipf approximation.

Table 1: Quality of the numerical approximation of the Pareto distributions

$\eta$	distance $d(g_L; \eta)$	distance $d(g_L; \eta)$	ML estimate of $\widehat{k}$	OLS estimate of $\widehat{k}$
	truncated Pareto	Pareto	truncated Pareto	Pareto
1	62.9127	62.9815	-0.2521	0.7175
1/2	21.0455	21.0858	0.3690	0.6840
1/4	4.4038	4.4222	0.7371	0.8362
1/8	0.9314	0.9399	0.8688	0.9701
1/16	0.2104	0.2144	0.9514	1.0362
<b>1/32</b>	<b>0.0498</b>	<b>0.0518</b>	<b>0.9843</b>	<b>1.0144</b>
1/64	0.0121	0.0131	0.9403	0.9901
1/128	0.0030	0.0035	0.9764	1.0802
1/256	0.0007	0.0010	1.0216	1.0392
1/512	0.0002	0.0003	0.9926	1.0328

*Notes:* Maximum likelihood (ML) estimates (column 3) for the truncated Pareto distribution follow Aban, Meerschaert, and Panorska (2006) and are computed from random samples of 1,000 observations that are generated using inverse transform sampling. Ordinary least squares (OLS) estimates of  $k$  (column 4) for the Pareto distribution are computed using the same random samples. Following Gabaix and Ibragimov (2011), we use  $\log(\text{rank} - 1/2)$  as the dependent variable in that case.

We illustrate how the approximation converges to both a truncated and a complete Pareto distribution by using for  $g_t(\cdot)$  a log-normal distribution with parameters  $(\mu_t, \sigma_t) = (1.4, 1.4)$  and with support  $[\underline{t}, \bar{t}] = [0.01, 5]$ .<sup>22</sup>

We evaluate the approximation using values of  $\eta = 1/2^n$ , for  $n = 1, 2, \dots, 10$ . Figure 4 shows that the approximation rapidly converges to a (truncated) Pareto distribution with shape parameter  $k$  close to 1. This result can be seen more formally from Table 1, which reports the maximum likelihood (ML) and the ordinary least squares (OLS) estimates of the shape parameter  $k$  of the truncated and the complete Pareto distribution. To estimate those parameters, we sample 1,000 points from the approximation using inverse transform sampling. The ML estimate of  $k$  is then obtained from that sample using the estimator developed by Aban, Meerschaert, and Panorska (2006). As can be seen from Table 1, the distance between the approximation and the truncated Pareto distribution vanishes quickly. The same occurs, to almost the same degree, for a complete Pareto distribution. Furthermore, as can be seen from the last column of Table 1, the estimates of the shape parameter converge to 1 relatively quickly. In particular, for a value of  $\eta = 1/32 \approx 0.031$ , which corresponds closely to our empirical value  $\widehat{\gamma} - \widehat{\varepsilon}$ , the approximation is already fairly close to 1. Hence, even if the underlying distribution of talent is lognormal, as in our example, the size

<sup>22</sup>Our results are robust to the underlying distribution and parametrization. As expected, the approximation is better and convergence is faster when the underlying distribution of talent is more skewed to the right.

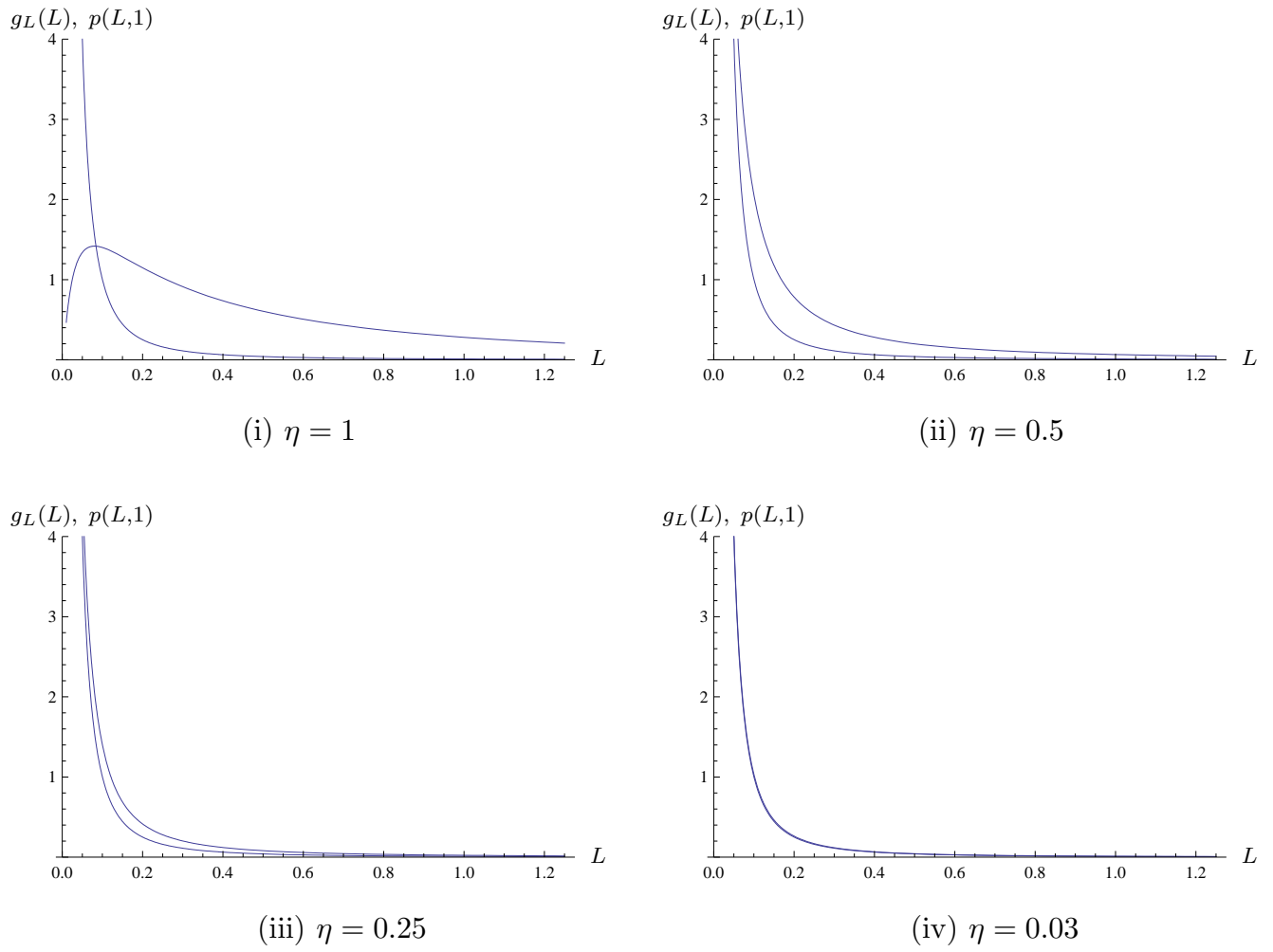


Figure 4: Quality of the numerical approximation of the Pareto distribution

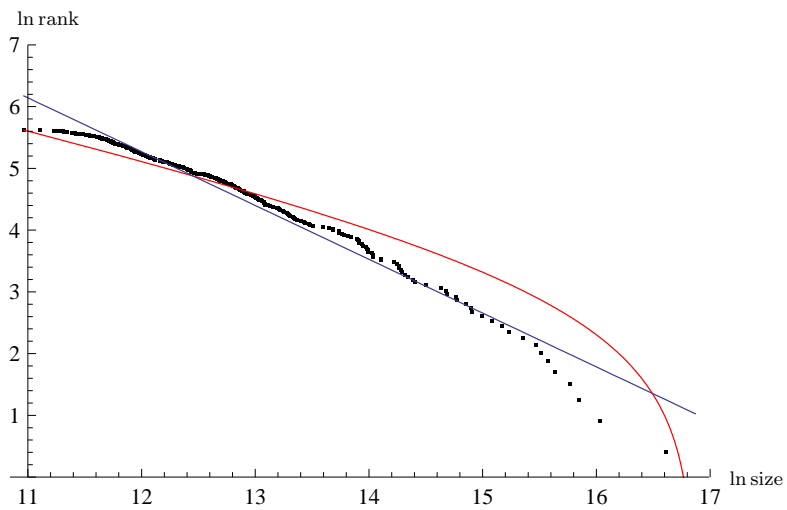


Figure 5: Pareto fit (blue line) and truncated Pareto fit (red line) for 276 MSAs in 2000

distribution of cities will be approximately Zipf with a shape parameter close to 1.

Estimating the Pareto parameter from the data on our 276 MSAs yields a coefficient of  $\widehat{k} = 0.8716$ , whereas the parameter for a truncated distribution is  $\widehat{k}_t = 0.4484$ . Figure 5 depicts the observed rank-size distribution for 276 MSAs in 2000 (black dots), as well as the fits for the truncated (red line) and complete (blue line) Pareto.

## Appendix D Example of equilibrium with discrete cities, heterogeneous talent, and varying selection

Equilibria with constant selection across cities seem empirically relevant. They are, however, special cases. While a complete analysis of all equilibria is beyond the scope of this appendix, we now provide examples of equilibria with varying selection across a discrete number of city-types. This situation is interesting because it shows that many of the properties of the equilibrium with talent-homogeneous cities remain true or approximately true in more general cases.

To keep things simple, we assume that  $a = 0$  so that all workers have the same productivity  $\varphi^a = 1$ . We also consider only three types of cities, type-1, type-2, and type-3 cities, and subscript variables accordingly. We also specify the distributions of talent and luck to be uniform over  $T = [\underline{t}, \bar{t}]$  and  $\Sigma = [\underline{s}, \bar{s}]$ , respectively. Total population is fixed to  $\Lambda$ , and we denote by  $n_i$  the mass (the number) of type- $i$  cities in the economy.

We first derive the distribution of the productivity variable  $\varphi \equiv t \times s$ . Using Theorem 1 of Glen, Leemis, and Drew (2004) and assuming, without loss of generality for our purpose, that  $\underline{t}\bar{s} < \bar{t}\underline{s}$ , the product of talent and luck is distributed as follows:

$$f(\varphi) = \begin{cases} \frac{1}{(\bar{s}-\underline{s})(\bar{t}-\underline{t})} \ln\left(\frac{\varphi}{\underline{st}}\right) & \text{if } \underline{ts} \leq \varphi \leq \underline{t}\bar{s} \\ \frac{1}{(\bar{s}-\underline{s})(\bar{t}-\underline{t})} \ln\left(\frac{\bar{s}}{\underline{s}}\right) & \text{if } \underline{t}\bar{s} \leq \varphi \leq \bar{t}\underline{s} \\ \frac{1}{(\bar{s}-\underline{s})(\bar{t}-\underline{t})} \ln\left(\frac{\bar{s}\bar{t}}{\varphi}\right) & \text{if } \bar{t}\underline{s} \leq \varphi \leq \bar{t}\bar{s} \end{cases} \quad (\text{D.1})$$

Using equation (D.1), we can easily derive the cumulative productivity distribution  $F(\cdot)$ .

In what follows, we focus on equilibria with two talent thresholds  $t_1$  and  $t_2$  such that all individuals with talent  $t \in [\underline{t}, t_1)$  choose to locate in type-1 cities; all individuals with talent  $t \in [t_1, t_2)$  choose to locate in type-2 cities; and all individuals with talent  $t \in [t_2, \bar{t}]$  choose to locate in type-3 cities. The thresholds  $t_1$  and  $t_2$ , the number of type- $i$  cities, their populations  $L_i$  and their selection cutoffs  $\underline{\varphi}_i$  for  $i = 1, 2, 3$  are all endogenously determined. Let  $\Delta\mathbb{E}V_i(t) = \mathbb{E}V_i(t) - \max_{j \neq i} \mathbb{E}V_j(t)$ . A spatial equilibrium is such that every individual with talent  $t$  picks the city that maximises her expected indirect utility. Formally,  $\Delta\mathbb{E}V_1(t) \geq 0$  for all  $t \in [\underline{t}, t_1)$  (and negative otherwise);  $\Delta\mathbb{E}V_2(t) \geq 0$  for all  $t \in [t_1, t_2)$  (and negative otherwise); and  $\Delta\mathbb{E}V_3(t) \geq 0$  for all  $t \in [t_2, \bar{t}]$  (and negative otherwise).

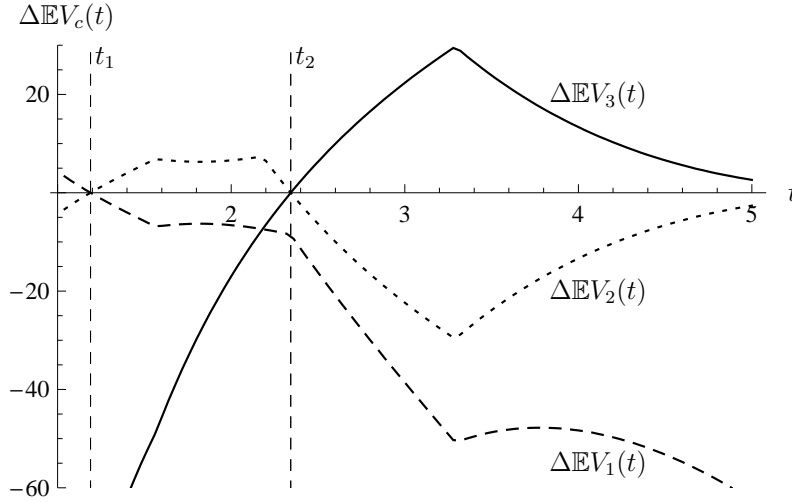


Figure 6: Example of spatial equilibrium with imperfect ability sorting

Figure 6 depicts the expected indirect utility differentials for the three types of cities, as well as the two talent cutoffs for:  $\varepsilon = 0.47$ ,  $\gamma = 0.5$ ,  $\theta = 0.5$ ,  $\Lambda = 5000$ ,  $\underline{t} = \underline{s} = 1$ ,  $\bar{t} = 5$  and  $\bar{s} = 2$ .<sup>23</sup> In Figure 6 we also set the numbers of cities to  $n_1 = 20$ ,  $n_2 = 6$  and  $n_3 = 2$ . We note that the mass of cities of each type is not uniquely determined in equilibrium, as was the case with talent-homogeneous cities. There exists instead a continuum of  $n_i$ ,  $i = 1, 2, 3$  that can be sustained as an equilibrium.

The allocation we have chosen is in equilibrium for  $t_1$  and  $t_2$  as determined on Figure 6 since all individuals located in type-1 cities (i.e., to the left of  $t_1$ ) get an expected utility no smaller than in type-2 or type-3 cities; all individuals in type-2 cities (i.e., between  $t_1$  and  $t_2$ ) get an expected utility no smaller than in type-1 or type-3 cities; and all individuals in type-3 cities (i.e., to the right of  $t_2$ ) get an expected utility no smaller than in type-1 or type-2 cities.

In line with the results derived in the case of talent-homogeneous cities, more talented cities are larger, more productive, and pay higher wages. We have  $L_3 = 1660.42$ ,  $\underline{\varphi}_3 = 6.58$  and  $w_3 = 89.46$ , whereas the corresponding figures for type-2 cities are  $L_2 = 240.38$ ,  $\underline{\varphi}_2 = 3.13$  and  $w_2 = 17.14$  and for type-1 cities are  $L_1 = 11.84$ ,  $\underline{\varphi}_1 = 1.88$  and  $w_1 = 2.45$ . In words, type-3 cities are about 7 times larger than type-2 cities, which are themselves about 20 times larger than type-1 cities. Furthermore, type-3 wages exceed type-2 wages by a factor of about 5, and type-2 wages exceed type-1 wages by a factor of 7. The selection cutoffs reflect a similar ranking. Importantly, the strong right-skew in the size distribution of cities does not stem from the right-skew in the distribution of talent. The latter is uniform. Instead, sorting, agglomeration economies, and the population-talent complementarity generate these strong asymmetries.

With talent-homogeneous cities, the degree of selection  $F_c(\underline{\varphi}_c)$  is the same for all cities, a knife-

<sup>23</sup>Our values for agglomeration economies,  $\varepsilon$ , and urban costs,  $\gamma$  are much larger than empirically reasonable to accentuate differences across cities in the figure.

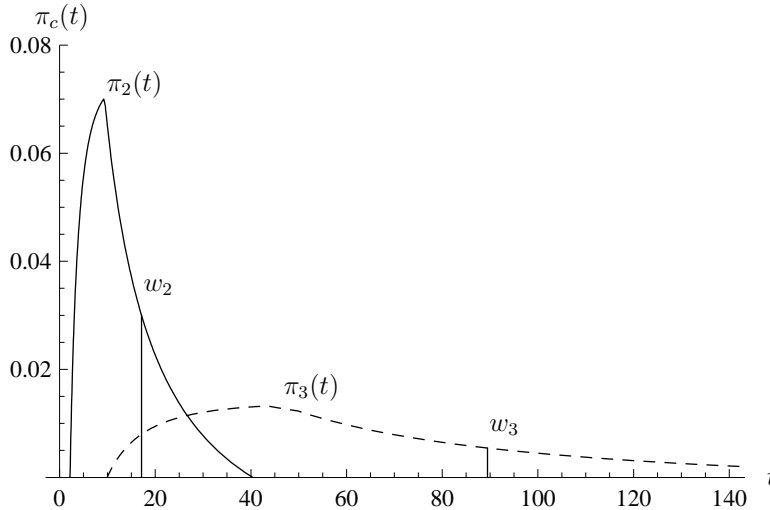


Figure 7: Distribution of profits in type-2 and type-3 cities

edge result. However, and quite remarkably, although larger cities may have tougher selection, the differences in the degree of selection are small in our example. We find that  $F_1(\underline{\varphi}_1) = 0.720$ , whereas  $F_2(\underline{\varphi}_2) = 0.746$  and  $F_3(\underline{\varphi}_3) = 0.749$ . Put differently, although the selection cutoff in type-3 cities is about 110% higher than in type-2 cities, itself 66% higher than in type-1 cities, selection differs by barely 4% between the two extremes. Larger cities provide entrepreneurs with access to more and richer consumers which almost fully offsets the tougher environment.

Figure 7 depicts the distribution of entrepreneurial profits for type-2 and type-3 cities (a similar figure can be drawn for type-1 and type-2 cities). The solid curve is for type-2 cities (i.e., medium-sized cities) whereas the dashed curve is for type-3 cities (i.e., large cities). All individuals with profit below the thresholds  $w_2$  and  $w_3$  choose to become workers instead of entrepreneurs. Hence, entrepreneurs are to the right of  $w_2$  for type-2 cities and to the right of  $w_3$  for type-3 cities. Comparing the two curves, two features are immediately apparent. First, entrepreneurial profits in the larger cities are significantly right-shifted relative to the ones in smaller cities. This is due to both agglomeration and sorting. Second, there is substantial dilation of profits in large cities relative to small cities. Large cities host, on average, more productive individuals but the most productive of them are benefiting disproportionately from being there. Large cities are thus more unequal than small cities by most conventional measures of inequality. This is consistent with the findings of the literature on inequalities in cities (Baum-Snow and Pavan, 2009; Behrens and Robert-Nicoud, 2009; Glaeser, Tobio, and Resseger, 2009). Interestingly, to map the distribution of profits in medium-sized cities into that of large cities, we need to apply a tiny truncation (small differences in selection), a large right-shift (for agglomeration and sorting) and a significant dilation (the interaction between sorting and agglomeration). This is clearly reminiscent of the findings of Combes, Duranton, Gobillon, Puga, and Roux (2012) regarding the distributions of firms' productivities in small and large French cities.

## Appendix E Proof of Proposition 6

**Proof.** Claim (1.) holds by construction. We thus prove claims (2.)-(4.) imposing claim (1.). Let

$$\mathbb{E}_{\hat{s}}V(t, c) \equiv \frac{1}{1 - G_s(\hat{s})} \int_{\hat{s}}^{\infty} V(s, t_c) dG_s(s) \quad (\text{E.1})$$

denote the expected utility conditional on drawing  $s \geq \hat{s}$ , and let  $M(s, t, t_c)$  denote the value of moving from talent-homogeneous city  $c$  for an individual with talent  $t$  and current luck  $s$ , with  $M(s, t_c) \equiv M(s, t_c, t_c)$ . Then the value of moving is equal to

$$\begin{aligned} M(s, t_c) &= 0 + (1 - \delta) \left\{ [1 - G_s(\hat{s})] \frac{\mathbb{E}_{\hat{s}}V(t, c)}{\delta} + G_s(\hat{s}) \max_{t_c} M(s, t, t_c) \right\} \\ &= (1 - \delta) \left\{ [1 - G_s(\hat{s})] \frac{\mathbb{E}_{\hat{s}}V(t, c)}{\delta} + G_s(\hat{s}) M(s, t_c) \right\}, \end{aligned}$$

where the second line follows from the optimal location choice (in steady state, choosing  $c$  remains optimal at the current period if it was the optimal city to chose at the previous period). Rearranging and using (E.1) yields the following expression for the value of moving:

$$M(s, t_c) = M(t_c) \equiv \frac{1}{1 - G_s(\hat{s}) + \delta G_s(\hat{s})} \frac{1 - \delta}{\delta} \int_{\hat{s}}^{\infty} V(\tilde{s}, t_c) dG_s(\tilde{s}),$$

that is,  $M(s, t_c)$  does not depend on  $s$ . Using this expression and (E.1), we may rewrite the condition requiring that the net current value of staying is (weakly) positive for all  $s$  greater than  $\hat{s}$  as follows:

$$\begin{aligned} 0 &\leq V(s, t_c) - \delta M(t_c) \\ &= V(s, t_c) - \frac{1 - \delta}{1 - (1 - \delta)G_s(\hat{s})} \int_{\hat{s}}^{\infty} V(\tilde{s}, t_c) dG_s(\tilde{s}), \quad \forall s \geq \hat{s} \end{aligned} \quad (\text{E.2})$$

which is increasing in  $s$  by inspection and where the cutoff for luck  $\hat{s}$  is implicitly defined as

$$V(\hat{s}, t_c) = \frac{1 - \delta}{1 - (1 - \delta)G_s(\hat{s})} \int_{\hat{s}}^{\infty} V(\tilde{s}, t_c) dG_s(\tilde{s}), \quad (\text{E.3})$$

that is, the opportunity cost of moving is equal to the expected value of moving for some  $\hat{s}$ . The left-hand side (LHS) of equation (E.3) is strictly increasing in  $\hat{s}$  from 0 to  $+\infty$  over  $(0, \infty)$  by inspection. The right-hand side (RHS) of equation (E.3) evaluated at  $\hat{s} = 0$  is equal to  $(1 - \delta)\mathbb{E}_0V(t_c) = (1 - \delta)\mathbb{E}V(t_c) > 0$  and, at the limit  $\hat{s} \rightarrow +\infty$ , it is equal to zero. Thus, by continuity and the intermediate value theorem, there exists  $\hat{s} \in (0, \infty)$  that satisfies equation (E.3). To show uniqueness of  $\hat{s}$ , differentiate the RHS of (E.3) to obtain

$$\frac{\partial}{\partial \hat{s}} \text{RHS} = \frac{(1 - \delta) g_s(\hat{s})}{1 - (1 - \delta)G_s(\hat{s})} [\text{RHS} - V(\hat{s}, t_c)] \leq 0,$$

where the last inequality follows from (E.2). This establishes uniqueness of  $\hat{s}$  and thus claim (2.).



From (E.3), the LHS does not depend on  $\delta$ , whereas the RHS is decreasing in  $\delta$  by inspection. Hence,  $\hat{s}$  must be a decreasing function of  $\delta$ , which proves the first part of claim (3.). The second part of this claim immediately follows from the fact that  $\lim_{\delta \rightarrow 1} M(s, t_c) = 0$  for all values of  $s > 0$ . When death is certain, waiting for better times is worthless.

To obtain claim (4.), we use the definition of  $M(t_c)$  to rewrite (E.3) as  $0 = V(\hat{s}, t_c) - \delta M(t_c)$ . We then use the definitions of  $V(\cdot)$  and  $M(\cdot)$  in order to get

$$\begin{aligned} 0 &= V(\hat{s}, t_c) - \delta M(t_c) \\ &\equiv w_c(St_c)^a \left[ \max \left\{ \left( \frac{\hat{s}}{S} \right)^a, \left( \frac{\hat{s}}{S} \right)^{\frac{1}{\varepsilon}} \right\} - \frac{\theta L^\gamma}{w_c(St_c)^a} \right] \\ &\quad - \frac{1 - \delta}{1 - (1 - \delta)G_s(\hat{s})} w_c(St_c)^a \left\{ \left[ \int_{\min\{\hat{s}, S\}}^S \left( \frac{s}{S} \right)^a dG_s(s) + \int_{\max\{\hat{s}, S\}}^\infty \left( \frac{s}{S} \right)^{\frac{1}{\varepsilon}} dG_s(s) \right] - \frac{\theta L^\gamma}{w_c(St_c)^a} \right\}. \end{aligned}$$

Recall that  $w_c(St_c)^a$  is proportional to  $L^\varepsilon t_c^{1+a}$ , where  $t_c^{1+a}$  is itself proportional to  $L_c^{\gamma-\varepsilon}$  at the equilibrium with talent-homogeneous cities (see equation (20) in the text). Hence, the final condition at the equilibrium with talent-homogeneous cities is given by

$$\begin{aligned} 0 &= \max \left\{ \left( \frac{\hat{s}}{S} \right)^a, \left( \frac{\hat{s}}{S} \right)^{\frac{1}{\varepsilon}} \right\} - \zeta \\ &\quad - \frac{1 - \delta}{1 - (1 - \delta)G_s(\hat{s})} \left\{ \left[ \int_{\min\{\hat{s}, S\}}^S \left( \frac{s}{S} \right)^a dG_s(s) + \int_{\max\{\hat{s}, S\}}^\infty \left( \frac{s}{S} \right)^{\frac{1}{\varepsilon}} dG_s(s) \right] - \zeta \right\}, \end{aligned}$$

where the term  $\zeta$  collects parameters and variables that are constant across cities (like  $\sigma$  and  $S$ ). Consequently, the whole condition is independent of city-specific variables. Thus,  $\hat{s}_c = \hat{s}$  for all  $c$ , which establishes claim (4.).

Last, define the fraction of movers at time  $T$  as  $m_T$ . Then the stock of movers varies according to  $m_{T+1} = G(\hat{s})(m_T + \delta)$ , where  $\delta$  is the exogenous fraction of newborns. At steady state, the fraction of movers is constant and equal to  $\frac{G_s(\hat{s})}{1 - G_s(\hat{s})}\delta$ , which establishes claim (5.) and completes the proof. ■

Separate Appendices with Supplemental Material for:

Productive cities:  
Sorting, selection, and agglomeration

Kristian Behrens\*    Gilles Duranton<sup>†</sup>

Frédéric Robert-Nicoud<sup>‡</sup>

June 11, 2012

**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 (2012) can be derived from a simple extension of our model. Finally, Appendix J shows that the interpretation of the estimation results of Section 7 rely 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

---

\*Canada Research Chair, Département des Sciences Économiques, Université du Québec à Montréal (UQAM), Canada; CIRPÉE; and CEPR. E-mail: [behrens.kristian@uqam.ca](mailto:behrens.kristian@uqam.ca)

<sup>†</sup>Department of Economics, University of Toronto, Canada; Spatial Economics Research Centre, London School of Economics; and CEPR. E-mail: [gilles.duranton@utoronto.ca](mailto:gilles.duranton@utoronto.ca)

<sup>‡</sup>Département des Sciences Économiques, Université de Genève, Switzerland and CEPR. E-mail: [frederic.robert-nicoud@unige.ch](mailto:frederic.robert-nicoud@unige.ch)

## Appendix 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 (2012) 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  $\tau$  and  $\gamma$  are positive parameters.<sup>1</sup> 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 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. The equilibrium land rent schedule is thus such that

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

Without loss of generality, the rent at the city edges can be normalised to zero. This 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:

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

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

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

---

<sup>1</sup>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, 2012) suggests that, empirically, the elasticity of urban costs to city population is well below unity. We confirm this in Section 7 of the paper and show that a small value of  $\gamma$  has important implications regarding the size distribution of cities.

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.

## Appendix 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. Thus we may write the utilitarian planner's objective as

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

such that

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

where  $T(c)$  is the set of talents allocated to  $c$ ,  $g_{t,s,c}(t, s)$  is the joint probability density function of talents and luck in  $c$ ,  $L(t, c)$  is the mass of workers of talent  $t$  allocated to  $c$ ,  $L(c) \equiv \int_{t \in T(c)} L(t, c) dt$  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 to the density distribution of productivity to the joint distribution function of talent  $t$  and luck  $s$ , where  $t$  is taken over  $T(c)$  with density  $g_t(t, c)$  and  $s$  is taken over  $[0, +\infty)$  with density  $g_s(s)$ . The third constraint above 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 ones to production work since  $a < 1/\varepsilon$ . Thus, there exists a productivity cutoff in each city,  $\underline{\varphi}(c)$ , such that  $\varphi^W(c) = [0, \underline{\varphi}(c)]$  and  $\varphi^E(c) = (\underline{\varphi}(c), +\infty)$ .

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

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

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

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

Note that this expression is identical to its equilibrium counterpart as given by expression (14) in the main text, despite entrepreneurs charging a (constant) markup. This occurs 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 up in 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  $c$  so as 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

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

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 (20) 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).** The problem of the optimal city composition cannot be characterised precisely without making specific assumptions about the distribution of talent  $G_t(\cdot)$  and the distribution of luck  $G_s(\cdot)$  in the economy. This is because (i) talent and luck interact to determine productivity  $\varphi$  and, especially, (ii) because luck is realised after the planner’s location decisions. Relaxing either would solve the problem entirely. Equivalently, we may instead assume that the distribution of luck 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  $t_1 > t_2$  (like in the paper, we abuse notations in using subscripts). 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,  $\sigma$ , 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:

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

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  $V(t_{12})$  denote the average indirect utility prevailing in such a heterogeneous city. It is equal to:

$$V(t_{12}) = D (t_1 t_2^a)^{\frac{\gamma}{\gamma-\varepsilon}}.$$

Forming talent-homogeneous cities is optimal if

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

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  $\varepsilon$ , 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  $f : t_1 \times t_2 \times a \rightarrow \varepsilon$  defined from  $T \times T \times \mathbb{R}^+$  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,

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

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 Appendix B) 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.

## Appendix H A 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

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

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  $\pi(\varphi_c) = w_c \varphi_c^a$ . 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. This 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  $\Phi$  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  $\theta$  denotes the same bundle of parameters as in Appendix F.<sup>2</sup>

---

<sup>2</sup>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.

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:

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

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.

## Appendix 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). This implies 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 (2012) but leads to different interpretations of their results.<sup>3</sup>

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

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

---

<sup>3</sup>In the version of our model that appears in the main text, firm revenue does not depend directly on entrepreneurial productivity. This is 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  $\varphi$  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.



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

$$p(\varphi)^{\frac{1}{\varepsilon}(\frac{1+\varepsilon}{1-\alpha}-1)} = \frac{1+\varepsilon}{1-\alpha} w Y^{\frac{1}{1-\alpha}-1} \varphi^{-\frac{1}{1-\alpha}}. \quad (\text{I.3})$$

At the limit  $\alpha \rightarrow 0$  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

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

Given that Gennaioli, La Porta, Lopez-de-Silanes, and Shleifer (2012) use log firm revenue as dependent variable, it is also useful to write this quantity:

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

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, 2012, 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:

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

By definition of the price index:

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

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

Next, by definition of  $Y$ :

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

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

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

which clearly shows that  $\varepsilon$  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  $L \int_0^{\underline{\varphi}} \varphi^a dF(\varphi)$  and the labour demand is equal to  $L \int_{\underline{\varphi}}^{\infty} l(\varphi) dF(\varphi)$ . These yield

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

where the last equality follows from equating production (I.1) with demand  $x(p(\varphi)) = Yp(\varphi)^{-\frac{1+\varepsilon}{\varepsilon}}$ , 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

$$\underline{\varphi}^{\frac{1}{\alpha+\varepsilon}-a} \int_0^{\underline{\varphi}} \varphi^a dF(\varphi) = \frac{1-\alpha}{\alpha+\varepsilon} \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} dF(\varphi). \quad (\text{I.10})$$

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

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

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

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

At the limit  $\alpha \rightarrow 0$  we get  $Y/L = \left[ \int_0^{\underline{\varphi}} \varphi^a dF(\varphi) \right] \left[ \int_{\underline{\varphi}}^{\infty} \varphi^{\frac{1}{\alpha+\varepsilon}} dF(\varphi) \right]^\varepsilon L^\varepsilon$ , 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:

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

where the wage is given by

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

with (using equation I.10),

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

With talent-homogeneous cities the first-order condition becomes

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

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

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

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:

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

## Appendix 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

$$\left( \frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c} \right) dL_c + \left( \frac{\partial y}{\partial t_c} \frac{\partial t_c}{\partial L_c} - \frac{\partial z}{\partial x_c} \frac{\partial x_c}{\partial L_c} \right) dL_c = 0. \quad (\text{J.1})$$

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

$$dy = \frac{\partial y}{\partial t_c} \frac{\partial t_c}{\partial L_c} dL_c + \frac{\partial y}{\partial L_c} dL_c \quad \text{and} \quad dz = \frac{\partial z}{\partial x_c} \frac{\partial x_c}{\partial L_c} dL_c + \frac{\partial z}{\partial L_c} dL_c. \quad (\text{J.2})$$

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

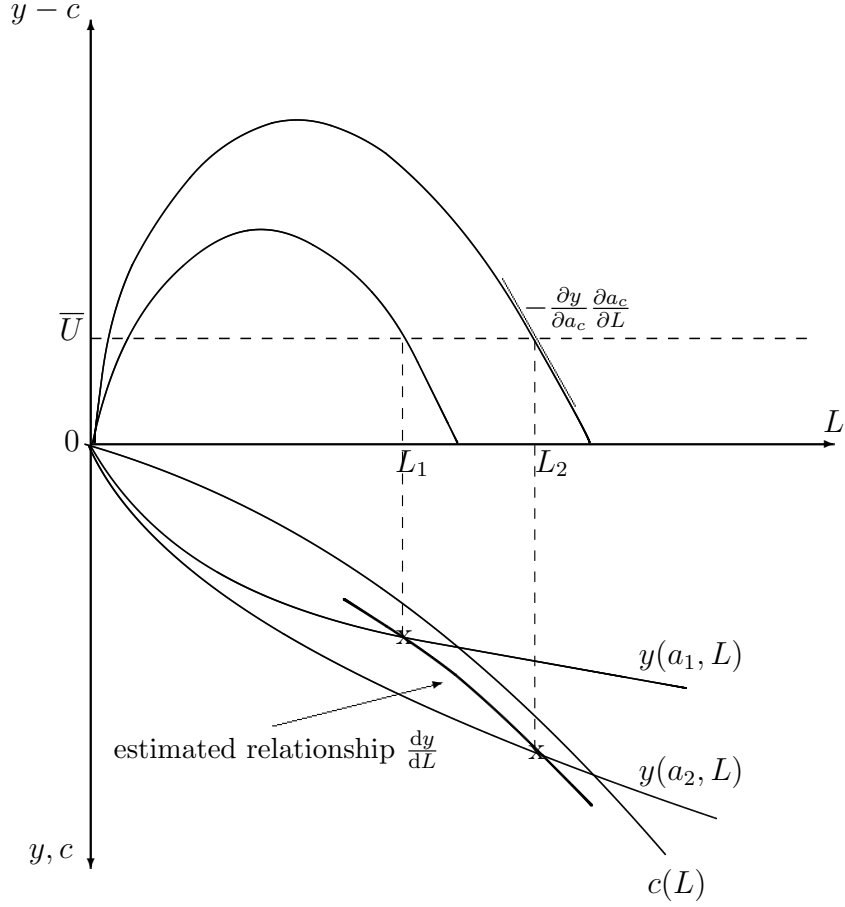


Figure 1: Estimating  $\partial y/\partial L$  with constant equilibrium utility.

to  $(\partial y/\partial t_c)/(\partial t_c/\partial L_c) = -(\partial y/\partial L_c - \partial z/\partial L_c)$ , which, together with equation (J.2) yields

$$dy = -\left(\frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c}\right)dL_c + \frac{\partial y}{\partial L_c}dL_c \Rightarrow \frac{dy}{dL_c} = \frac{\partial z}{\partial L_c}. \quad (\text{J.3})$$

In words, at equilibrium the impact of a change in population on income just equals the change in urban costs. To understand why this 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.

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 ( $dy/dL_c \geq \partial y/\partial L_c$ ). 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 1 illustrates these results for the case where there is no shift in urban costs.<sup>4</sup>

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 productivity measure on urban population yields the elasticity of urban costs,  $\gamma$ .

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. This 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

$$\left( \frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c} \right) dL_c + \left( \frac{\partial y}{\partial t_c} \frac{\partial t_c}{\partial L_c} - \frac{\partial z}{\partial x_c} \frac{\partial x_c}{\partial L_c} \right) dL_c = dU.$$

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  $(\partial y/\partial t_c)/(\partial t_c/\partial L_c) = -(\partial y/\partial L_c - \partial z/\partial L_c) + dU_c/dL_c$ . 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

$$dy = - \left( \frac{\partial y}{\partial L_c} - \frac{\partial z}{\partial L_c} \right) dL_c + dU_c + \frac{\partial y}{\partial L_c} dL_c = dU_c + \frac{\partial z}{\partial L_c} dL_c$$

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 2 illustrates this case.<sup>5</sup>

---

<sup>4</sup>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  $dz/dL_c = \partial y/\partial L_c$ . 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.

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

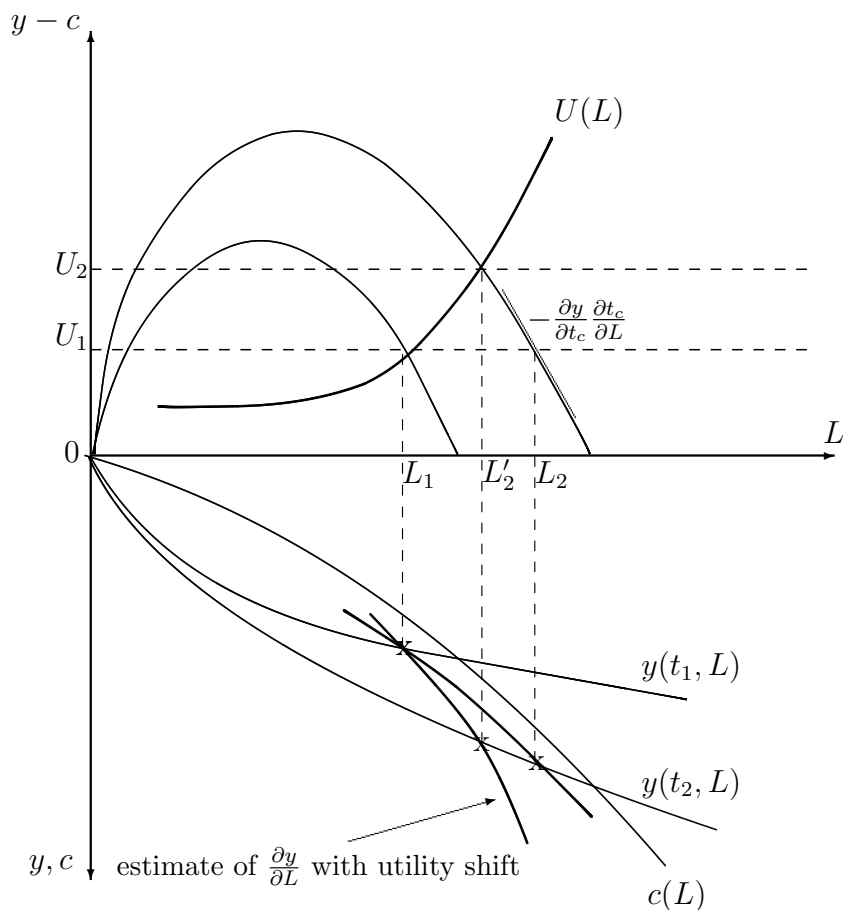


Figure 2: Estimating  $\partial y/\partial L$  with varying equilibrium utility.

## References

- [1] Albouy, David (2009) Are big cities really bad places to live? Improving quality-of-life estimates across cities. *NBER Working Paper #14472*.
- [2] Alonso, William (1964) *Location and Land Use; Toward a General Theory of Land Rent*. Cambridge, MA: Harvard University Press.
- [3] Baum-Snow, Nathaniel and Ronni Pavan (2012) Understanding the city size wage gap. *Review of Economic Studies* 79(1): 88–127.
- [4] Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2012). The costs of agglomeration: Land prices in French cities. Processed, University of Toronto.
- [5] Duranton, Gilles and Diego Puga. (2012) The growth of cities. Processed, University of Toronto.

- [6] Ethier, Wilfred J. (1982) National and international returns to scale in the modern theory of international trade, *American Economic Review* 72(3), 389–405.
- [7] Fujita, Masahisa (1989) *Urban Economic Theory: Land Use and City Size*. Cambridge, MA: Cambridge University Press.
- [8] Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer (2012) Human capital and Regional Development. Processed, Harvard University.
- [9] Mills, Edwin S. (1967) An aggregative model of resource allocation in a metropolitan area, *American Economic Review (Papers and Proceedings)* 57(2), 197–210.
- [10] Muth, Richard F. (1969) *Cities and Housing. The Spatial Pattern of Urban Residential Land Use*. Chicago, IL: University of Chicago Press.
- [11] Small, Kenneth A. and Erik T. Verhoef (2007) *The Economics of Urban Transportation*. London, UK: Routledge, Taylor& Francis Group.
- [12] Zenou, Yves. 2009. *Urban Labour Economics*. Cambridge: Cambridge University Press.