Productive cities: Sorting, selection, and agglomeration

Kristian Behrens* Gilles Duranton†
Frédéric Robert-Nicoud‡

August 11, 2010

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 can also generate Zipf's law for cities. Finally, it provides a useful framework within which to reinterpret existing 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

†Department of Economics, University of Toronto, Canada; Spatial Economics Research Centre *and* Centre for Economic Performance, London School of Economics; and CEPR. E-mail: gilles.duranton@utoronto.ca

‡Département d'économie politique, Université de Genève, Switzerland; Spatial Economics Research Centre *and* Centre for Economic Performance, London School of Economics; and CEPR. E-mail: 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 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.

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 suggest 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 size on productivity, even after controlling for sorting and selection. There is also evidence that returns to talent (or skills) increase with city size and that this 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. Less productive firms are less likely to operate in larger cities but there is no evidence of stronger selection after conditioning out agglomeration and sorting. Finally, the size distribution of cities is well described by a Pareto distribution with a unitary shape parameter. Section 2 below discusses these facts in greater detail.

Third, our model provides a useful framework within which to interpret existing quantitative evidence. As mentioned above, the coefficient on log city population is 8.2% in a city earnings regression for the US. This coefficient drops to 5.1% when conditioning out talent through the log share of city college graduates. Because of sorting, 8.2% is actually the elasticity of urban costs with respect to population size in our model. This number can be confirmed using direct measures of urban costs. When controlling for city talent, the coefficient of 5.1% reflects the elasticity of earnings with respect to population size. In our model, the (small) difference between these two numbers should also be equal to the elasticity of city talent with respect to city size. The data for the US are consistent with this result. Put differently, small differences in talent across cities are enough to generate very large differences in city sizes. Finally, our model also predicts that cities are 'naturally' oversized by a factor close to e (≈ 2.72, Euler's number) in equilibrium. Nevertheless, this significant oversize has trivial economic costs. These two results hold for a range of plausible values for the elasticities of earnings and urban costs with respect to population size.

Formally, we extend the monopolistic competition framework of Dixit and Stiglitz (1977) to a two-stage production process (as in Ethier, 1982) with heterogeneous firms (as in Melitz, 2003) to generate local increasing returns. We then embed this production structure in a system of cities in the tradition of Henderson (1974) where commuting is more costly in larger cities.[1] 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 draw of talent and chooses a location. Upon moving, she gets another draw, which we call luck. Productivity as an entrepreneur is a combination of talent and luck, whereas labour as a worker is homogeneous. Individuals sort across cities *ex ante* depending on their talent and they select *ex post* into entrepreneurship or become workers depending on their productivity. At its heart, our model is thus an assignment problem. We characterise the assignment function from individuals' talents to cities. 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 equilibrium, cities result from a tradeoff between agglomeration economies and urban costs. In each city, only the most productive individuals become entrepreneurs. The others become workers. Profits as an entrepreneur increase with productivity and city size. Hence, more talented individuals, who stand a higher chance of becoming highly productive entrepreneurs, have more to gain from larger cities. This complementarity between talent and city size, together with the fact that urban costs do not depend on talent, lead to the sorting of more talented individuals into larger cities. Then, tougher selection in more talented cities implies more productive firms. A higher productivity, in turn, complements the agglomeration benefits of cities. This justifies why more talented cities are larger in equilibrium.

The remainder of this paper is organised as follows. Section 2 discusses the stylised facts mentioned above in greater depth. Section 3 presents the model and solves for its 'short-run' equilibrium, which takes the distribution of population as given. Section 4 solves for the 'long-run' equilibrium with endogenous cities. Section 5 discusses the quantitative implications of our model. Section 6 provides some extensions. Finally, Section 7 concludes.

2 Stylised facts and related literature

Agglomeration economies. It would be hard to justify the existence of cities without some form of underlying increasing returns. The positive correlation between various measures of output per capita (or productivity) and city size is extensively documented in the literature. See Rosenthal and Strange (2004), Melo, Graham, and Noland (2009), and Puga (2010) for reviews. Estimates

[1] For tractability reasons, we ignore sectoral issues and work with a single sector. The distribution of sectors across and within cities is the focus of much of the literature on systems of cities. We leave it to future work to combine the insights developed here with a more sectoral perspective.

for the elasticity of wages or productivity with respect to city size in many countries are usually close to the $5-8\%$ we report above for US MSAs. There are strong indications that city size has a causal effect on local productivity. Instrumental variable approaches (Ciccone and Hall, 1996; Combes, Duranton, Gobillon, and Roux, 2010) and natural experiments (Greenstone, Hornbeck, and Moretti, 2010) suggest that the existence of agglomeration effects is robust to reverse causation and missing variables. Furthermore, positive agglomeration effects persist even after controlling for ability sorting (Combes, Duranton, and Gobillon, 2008) and firm selection (Combes, Duranton, Gobillon, Puga, and Roux, 2009).

Recent evidence also points at input-output linkages as the single most important source of agglomeration economies, though such effects potentially arise from a broad range of economic mechanisms. See Holmes (1999), Amiti and Cameron (2007), Overman and Puga (2009), and Ellison, Glaeser, and Kerr (2010).

Sorting. A higher per capita output in larger cities may also reflect the sorting of individuals (Combes, Duranton, and Gobillon, 2008; Baum-Snow and Pavan, 2009). 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). 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 larger, their rewards must be relatively higher there. This is exactly what the literature finds. Agglomeration economies are stronger for individuals with more education (Wheeler, 2001; Glaeser and Resseger, 2010) or individuals with better cognitive and people skills (Bacolod, Blum, and Strange, 2009). At the same time, more talented individuals migrate to areas that offer them higher rewards (Dahl, 2002).

Finally, ability sorting by individuals does not imply perfect productivity sorting for the firms they operate or work for. Large cities host on average more productive firms but they also contain lots of poorly productive firms (Combes, Duranton, Gobillon, Puga, and Roux, 2009). Heterogeneity in firm productivity within cities is a major feature of the data.

Selection. A third reason for the higher per capita output in larger cities is tougher selection.[2] Larger cities are larger markets. They may also 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". More rigourous evidence about selection shows that less productive firms exit more frequently (Bartelsman and Doms, 2000; Foster, Haltiwanger, and Syverson, 2008). It also highlights that the survival productivity cutoff is higher in larger markets (Syverson, 2004).

However, after conditioning out sorting and agglomeration, there is no evidence that selection is really tougher in larger cities (Combes, Duranton, Gobillon, Puga, and Roux, 2009). Consistent

[2] Our model below ignores 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.

with this, the share of self-employed – a proxy for entrepreneurship – is independent from city population in the US.[3] In short, selection occurs but it is not directly apparent in urban data.

Size distribution of cities. Last, it is well known that the size distribution of cities is right-skewed and reasonably well approximated by a Pareto distribution with unitary shape parameter (Gabaix and Ioannides, 2004; Soo 2005).

Related theoretical literature. There is a large literature about sorting on income and preferences within cities and its fiscal implications (see Epple and Nechyba, 2004, for a review). The literature about ability sorting across cities is more limited. 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. Like us, Nocke (2006) assumes a continuum of talents for entrepreneurs. Unlike us, he maps talent directly into productivity (i.e., he does not consider selection) in a partial equilibrium setting. Importantly, he shows that perfect productivity sorting across exogenously determined cities generally occurs. 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 of the most productive firms from the small market to the large one and incomplete productivity sorting.

The two models most closely related to ours are Behrens and Robert-Nicoud (2009) and Davis (2009). The former 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. To our knowledge, this is the only prior model to embed market selection in an urban framework. Unlike in our model, inter-city trade plays a role but sorting is ignored. Davis (2009) develops an original model of learning and matching inspired by Antràs, Garicano, and Rossi-Hansberg (2006). Individuals with *ex ante* 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 abilities 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, Zipf's law and the size distribution of cities have attracted much attention recently. In random growth models, the current size of a city reflects its balance of past shocks (amenity for Gabaix, 1999; productivity for Eeckhout, 2004, or Rossi-Hansberg and Wright, 2007; innovation

[3] We regress the employment share of self-employed on log city population in 276 US MSAs using 2000 Census data. The share of self-employed is computed as the share of civilian employees above 16 who are self-employed (both incorporated and non-incorporated) in sectors other than agriculture, forestry, fishing, and mining. Although entrepreneurship and self-employment are not identical, the latter is often used in the literature as a proxy for the former (e.g., Doms, Lewis, and Robb, 2010). The coefficient is 0.0003 and nowhere near statistical significance.

for Duranton, 2007). Our approach is radically different. It builds on a static model of cities. In equilibrium, the size of a city depends on the productivity of its entrepreneurs magnified by the tradeoff between agglomeration economies and urban costs. More specifically, city size is a function of the talent of its residents elevated to a power which is the inverse of the difference between the intensity of agglomeration economies and that of urban costs. When this difference is small, as seems to be the case in the data, small differences in productivity between cities lead to large differences in size. The resulting size distribution of cities is thus close to degeneracy and approximately Zipf.[4]

3 The model

There is a continuum of individuals in the economy. They are identical except for their 'talent', t, and their 'luck', s. Talent and luck determine an individual's entrepreneurial productivity, $\varphi \equiv t \times s$. There is also a continuum of sites that can be used as cities. The 'number' of cities, their population size, and their composition are endogenous.

3.1 Timing

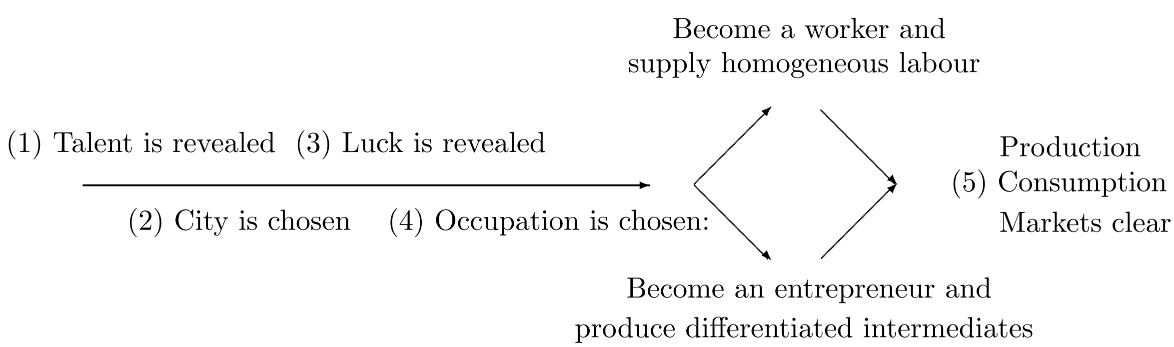

Figure 1. Timing

Each individual initially knows her talent and chooses where to locate. Upon moving to a city, her luck is revealed. Knowing her luck and thus her productivity, each individual selects into an occupation, worker or entrepreneur. A worker supplies a unit of homogeneous labour. Each entrepreneur sets up a firm that produces with productivity φ a variety of differentiated intermediate goods using labour. Finally, firms maximise profit, markets clear, and production and consumption take place. Figure 1 illustrates this timing.

[4] We know of two other papers that generate Zipf's law from a static model. Hsu (2008) relies on central place theory. The argument of Lee and Li (2009) is the static equivalent of random growth models where size is determined by the (multiplicative) aggregation of many randomly distributed local characteristics.

A key feature of our model is that productivity is revealed in two steps. 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. 'Luck' subsumes many local interactions that are uncertain and affect productivity such as being acquainted with the right people at the right time, etc. 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 $T \equiv [\underline{t}, \overline{t}]$ and G_s over $\Sigma \equiv [\underline{s}, \overline{s}]$, respectively, where $0 \leq \underline{s} \leq \overline{s} \leq \infty$ and $0 \leq \underline{t} \leq \overline{t} \leq \infty$. We do not allow for relocations. Empirically, there are frictions to mobility. In our static model we formalise these frictions in a parsimonious way by assuming free mobility before luck is realised and prohibitive mobility costs afterwards.

This two-step revelation process enables us to consider both the spatial sorting of individuals and the productivity selection of firms in a simple framework.[5] 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.

3.2 Preferences and technology

Individuals consume two goods: a final good and land. For simplicity, individuals 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

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

where $x_c(i)$ is the amount of variety i, and Ω_c^+ is the endogenously determined set of varieties of intermediate inputs produced in city c. Unlike intermediate inputs, the final good is freely tradable across cities. We use it as the numéraire.

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 Ω_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

[5] Melitz (2003) and subsequent literature model the selection of firms by imposing a sunk cost to create a firm and receive a productivity draw. Firms with poor draws then exit the market. In our case, this investment decision is potentially complex because would-be entrepreneurs differ by talent. We simplify it by giving one free draw to each individual. Formally, this allows individuals to select directly into occupations. This simplification does not change the nature of our results in the equilibrium we examine below.

productivity as in Melitz (2003). Output of variety i is

$$x_c(i) = \varphi_c(i) l_c(i), \tag{2}$$

where $l_c(i)$ is labour demand 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.

3.3 'Short-run' equilibrium

For expositional purposes, it is convenient to solve for the equilibrium in two steps. First, we study each city in isolation and take the set of individuals, Ω_c, in that city and the distribution of their productivity as given. Thus, individuals know their own productivity, its cumulative distribution $F_c(\cdot)$, which we assume for now to be continuously differentiable over a closed support, and the population size of their city, L_c. The focus is on selection, i.e., the occupational choice between being a worker and an entrepreneur, conditional on productivity. In section 4, we solve for the 'long-run' equilibrium where individuals sort 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:

$$x(i) = \left[\frac{p(i)}{\mathbb{P}}\right]^{-\frac{1+\varepsilon}{\varepsilon}} \frac{Y}{\mathbb{P}}, \qquad \text{where} \qquad \mathbb{P} \equiv \left[\int_{\Omega^+} p(j)^{-\frac{1}{\varepsilon}} \mathrm{d}j\right]^{-\varepsilon} \tag{3}$$

is the appropriate price index. It is immediate from (3) 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:

$$p(i) = (1+\varepsilon)\frac{w}{\varphi(i)}, \tag{4}$$

where w is the wage. This allows us to re-write the demand (3) as follows:

$$x(i) = \left[\frac{\varphi(i)}{\Phi}\right]^{1+\frac{1}{\varepsilon}} \frac{Y}{\mathbb{P}}, \quad \text{where} \quad \Phi \equiv \left[\int_{\Omega^+} \varphi(j)^{\frac{1}{\varepsilon}} \mathrm{d}j\right]^{\varepsilon} \tag{5}$$

is a measure of aggregate productivity in the city. More entrepreneurs in a city (i.e., a larger measure of Ω^+) and/or better entrepreneurs (i.e., 'on average' larger φ's) imply a larger aggregate productivity, Φ. In turn, individual sales are negatively affected by aggregate productivity through a crowding effect. Using expressions (4) and (5), we rewrite the price index $\mathbb{P}$ in (3) as a function of aggregate productivity, Φ, and obtain

$$\mathbb{P} = (1+\varepsilon)\frac{w}{\Phi}. \tag{6}$$

After combining this equation with (4) and (5), operating profit becomes

$$\pi(i) = \frac{\varepsilon}{1+\varepsilon}\, p(i)\, x(i) = \frac{\varepsilon}{1+\varepsilon}\, Y \left[\frac{\varphi(i)}{\Phi}\right]^{\frac{1}{\varepsilon}}. \tag{7}$$

As made clear by this expression, the profit of entrepreneurs increases with the economic size of their city, Y, and with their own productivity relative to aggregate productivity, φ/Φ.

Recall that labour is homogeneous.[6] Hence, individuals choose their occupation by comparing their prospective profit, as given by (7), with the wage w. An individual with productivity φ becomes an entrepreneur if the expected benefits of doing so exceed the forgone opportunity w, that is if $\pi(\varphi) > w$. This individual becomes a worker if the opposite inequality is true and is indifferent across occupations otherwise. Assuming that the set of individuals in the city, Ω, is convex (which is true in equilibrium), there exists a unique productivity cutoff $\underline{\varphi}$ defined by $\pi(\underline{\varphi}) = w$ such that all individuals with productivity above $\underline{\varphi}$ become entrepreneurs and all individuals with productivity below $\underline{\varphi}$ become workers. Using (7), the productivity cutoff is

$$\underline{\varphi} \equiv \Phi \left(\frac{1+\varepsilon}{\varepsilon} \frac{w}{Y} \right)^{\varepsilon} . \tag{8}$$

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 in smaller markets where demand is lower ($\partial\underline{\varphi}/\partial Y < 0$) and in markets where workers are paid higher wages ($\partial\underline{\varphi}/\partial w > 0$).

The city supply of labour is composed of all individuals with productivity below $\underline{\varphi}$:

$$L^S = F(\underline{\varphi})L \, .$$

From equation (2), labour demand for an entrepreneur with productivity φ is $l(\varphi) = x(\varphi)/\varphi$. After combining this expression with equations (5) and (6) and aggregating over all entrepreneurs, we obtain city labour demand:

$$L^D = L \int_{\underline{\varphi}}^{\sup \Omega^+} l(\varphi) \mathrm{d}F(\varphi) = \frac{1}{1+\varepsilon} \frac{Y}{w} \, .$$

Equating labour supply and demand then implies that workers receive a share $\frac{1}{1+\varepsilon}$ of city output:

$$Y = (1+\varepsilon)F(\underline{\varphi})Lw \, . \tag{9}$$

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 (6) yields

$$w = \frac{1}{1+\varepsilon}\Phi, \tag{10}$$

where aggregate productivity, Φ, as defined in (5) can be rewritten as

$$\Phi = \left[L \int_{\underline{\varphi}}^{\sup \Omega^+} \varphi^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi) \right]^{\varepsilon} . \tag{11}$$

[6] This assumption is for simplicity. The equilibrium occupational choice obtained here only requires more productive individuals to have a comparative advantage at being entrepreneurs.

Finally, aggregating (7) over all firms yields aggregate profit $\Pi = \frac{\varepsilon}{1+\varepsilon} Y$. Combining this with equation (9) implies

$$\Pi = \varepsilon F(\underline{\varphi}) L w. \tag{12}$$

Expressions (8) to (12) fully characterise the 'short-run' equilibrium for the tuple $\{\underline{\varphi}, \Phi, w, \Pi, Y\}$.

Proposition 1 (Existence and uniqueness of the 'short-run' equilibrium) *Given population, L, and its productivity distribution, $F(\cdot)$, the equilibrium in a city exists, is unique, and characterised by* (8) *to* (12)*.*

Proof. Using equation (9) to eliminate w and Y from equation (8) yields a positive relationship between productivity and selection: $\underline{\varphi}^{\frac{1}{\varepsilon}} F(\underline{\varphi}) = \Phi^{\frac{1}{\varepsilon}} / (\varepsilon L)$. Using equation (11), the implicit solution for $\underline{\varphi}$ may be written as:

$$F(\underline{\varphi}) = \frac{1}{\varepsilon} \int_{\underline{\varphi}}^{\sup \Omega^{+}} \left(\frac{\varphi}{\underline{\varphi}} \right)^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi). \tag{13}$$

The left-hand side of this expression is monotonically increasing from 0 to 1 in $\underline{\varphi}$, whereas the right-hand side is monotonically decreasing and equal to 0 when $\underline{\varphi} = \sup \Omega^{+}$. By continuity, this establishes the existence of a unique 'short-run' equilibrium. Knowing $\underline{\varphi}$, we can retrieve w, Y, Π and Φ from (9)–(12). ■

Proposition 2 ('Short-run' equilibrium properties) *Given the productivity distribution, $F(\cdot)$, larger cities have higher aggregate productivity, per-capita income, and wages than smaller cities. The productivity cutoff for selection does not depend on city size.*

Proof. By inspection of (13), $\underline{\varphi}$ does not depend on city size. By contrast, Φ is increasing in L by (11). Once $\underline{\varphi}$ and Φ are known, the equilibrium values for Y, w, and Π follow from (9), (10), and (12), respectively. These are all increasing in Φ (and thus in L) by inspection. ■

Our model displays agglomeration economies since per capita income increases with population size. To see them more clearly, we can use equations (9), (10), and (11) to write aggregate city income as

$$Y = F(\underline{\varphi}) \Phi L = F(\underline{\varphi}) \left[\int_{\underline{\varphi}}^{\sup \Omega^{+}} \varphi^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi) \right]^{\varepsilon} L^{1+\varepsilon}. \tag{14}$$

City production exhibits aggregate increasing returns to scale and ε measures their intensity. 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 agglomeration economies in the literature (Duranton and Puga, 2004). Our innovation here is to enrich the standard framework by considering heterogeneous firms in the spirit of Melitz (2003). In the absence of individual heterogeneity equation (14) would take a simpler form: $Y = AL^{1+\varepsilon}$ with A being a constant.

The empirical evidence in favour of agglomeration economies is very strong (for recent reviews, see Rosenthal and Strange, 2004; Melo, Graham and Noland, 2009; and Puga, 2010). Casual observation and formal evidence also suggest substantial heterogeneity in productivity across firms within cities (e.g., Syverson, 2004, and Combes, Duranton, Gobillon, Puga, and Roux, 2009). Our model accounts for both agglomeration economies and productivity dispersion in cities.

The result that the equilibrium productivity cutoff does not depend on city size conditional on the distribution of productivity is the outcome of two offsetting forces. Larger cities have at once a higher demand (which lowers the productivity cutoff) and more entrepreneurs (which raises it). These two effects exactly offset each other in our framework. There would be at least two ways to make the productivity cutoff vary with city size conditionally on the distribution of productivity. The first would be 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 would be to change the supply side of our model and have the ratio of fixed to variable costs for firms depend on city size. For instance, 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. However, the empirical results of Combes, Duranton, Gobillon, Puga, and Roux (2009) support our specification. They show that there are no differences in selection for French firms across cities after correcting for differences in productivity levels that affect all firms uniformly (i.e., agglomeration effects and sorting).

3.4 Urban structure

To close our model, we assume a standard internal spatial structure for cities. 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 $\tau, \gamma > 0$ are parameters.[7] For cities to be of finite size in equilibrium, we require that urban costs eventually

[7] 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., Baum-Snow and Pavan, 2009a) suggests that, empirically, the elasticity of urban costs to city size is well below

dominate agglomeration benefits: $\gamma > \varepsilon$.

Each resident chooses her place 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 residence so as 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 $r(x) + t(x)$ is the same for all residents and that the city is symmetric with its edges at a distance $L/2$ from the CBD. The equilibrium land rent schedule is such that $\tau x^\gamma + r(x) = \tau(L/2)^\gamma + r(L/2)$ for all $0 \leq x \leq L/2$. Without loss of generality, the rent at the city edges is normalised to zero, which yields the land rent schedule

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

Integrating land rent over the city after making use of its symmetry yields total land rent:

$$\text{TLR} = 2 \int_0^{L/2} r(x) \mathrm{d}x = \frac{2\tau\gamma}{\gamma + 1} \left(\frac{L}{2} \right)^{\gamma+1} . \tag{15}$$

By the same token, integrating commuting costs over the city yields total commuting cost:

$$\text{TCC} = 2 \int_0^{L/2} t(x) \mathrm{d}x = \frac{\text{TLR}}{\gamma} = \theta L^{\gamma+1}$$

where $\theta \equiv 2^{-\gamma}\tau(\gamma+1)^{-1}$. Finally, income from land rents is equally redistributed across all local residents, who thus receive TLR/L each. All this implies that each and every resident pays the average commuting cost, $\text{TCC}/L = \theta L^\gamma$, as urban cost.

3.5 The complementarity between talent and city size

Before looking at location choices, it is useful to write the expected indirect utility of an individual with talent t in her city before she learns her luck

$$\begin{aligned} \mathbb{E}V(t) &= \int_{\underline{s}}^{\overline{s}} \max\{w, \pi(ts)\} \mathrm{d}G_s(s) - \theta L^\gamma = \int_{\underline{s}}^{s'} w \mathrm{d}G_s(s) + \int_{s'}^{\overline{s}} \pi(ts) \mathrm{d}G_s(s) - \theta L^\gamma \\ &= w\, G_s(s') + w \left(\frac{t}{\underline{\varphi}} \right)^{\frac{1}{\varepsilon}} \int_{s'}^{\overline{s}} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) - \theta L^\gamma \\ &= \frac{1}{1+\varepsilon} \left[\varepsilon F(\underline{\varphi}) L \right]^\varepsilon \underline{\varphi} \left[G_s(s') + \left(\frac{t}{\underline{\varphi}} \right)^{\frac{1}{\varepsilon}} \int_{s'}^{\overline{s}} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - \theta L^\gamma \end{aligned} \tag{16}$$

where $s' \equiv \min\{\overline{s}, \max\{\underline{\varphi}/t, \underline{s}\}\}$.[8] The second and third equalities follow from (7) and (8) and from $\pi(\underline{\varphi}) = w$. The final equality follows from (10) and (11). In this expression, $F(\underline{\varphi})$ is the solution

one. We confirm this below and show that a small value of γ has important implications.

[8] This expression for s' accounts for the fact that some poorly talented individuals may never become entrepreneur regardless of their (good) luck, whereas some highly talented individuals may always become entrepreneurs regardless of their (bad) luck.

to (13), where F is the joint distribution of the product $s \times t$ within city c.[9]

In the last line of expression (16), the first series of terms (up to and including $\underline{\varphi}$) is the equilibrium wage w, which is proportional to L^ε (as a result of agglomeration economies) and to $\underline{\varphi}$ (as a result of selection). The middle square bracket is the expected premium of becoming an entrepreneur. It is increasing in talent t and decreasing in the productivity cutoff $\underline{\varphi}$. The last term in (16) is urban crowding: urban costs are larger in more populated cities.

Proposition 3 (Complementarity between talent and city size) *More talented individuals benefit disproportionately from being located in larger cities:*

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

Proof. Using (16), making use of the fact that $\partial s'/\partial t \leq 0$, and noting that $\overline{\varphi}$ is independent of L conditional on F (Proposition 2), we have:

$$\begin{aligned}\frac{\partial^2 \mathbb{E}V(t)}{\partial t \partial L}\Big|_{F(.)} &= \frac{\varepsilon}{1+\varepsilon}\frac{\underline{\varphi}}{L}\left[\varepsilon F(\underline{\varphi})L\right]^\varepsilon \left\{\frac{1}{\varepsilon t}\left(\frac{t}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}} s^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) + \left[1-\left(\frac{s't}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}}\right] g_s(s')\frac{\partial s'}{\partial t}\right\} \\ &= \frac{1}{(1+\varepsilon)L}\left[\varepsilon F(\underline{\varphi})L\right]^\varepsilon \left(\frac{t}{\underline{\varphi}}\right)^{-1+\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}} s^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) \geq 0\,,\end{aligned}$$

where g_s is the p.d.f. associated with G_s. The second equality in the above expression follows from the definition of s'. At an interior point we have $s' = \underline{\varphi}/t$, in which case the expression in square brackets is zero; and at a non-interior point we simply have $\partial s'/\partial t = 0$. ■

As apparent in equation (7), there is a complementarity between the economic size of a city, Y, and the productivity of an entrepreneur, φ, so that more productive entrepreneurs benefit more from being in larger markets. As shown by Propositions 1 and 2, the economic size of a city increases more than proportionately with its population (keeping the distribution of talent constant). We also know that entrepreneurial productivity is the product of talent and luck. Given this, the complementarity between Y and φ naturally translates into a complementarity between population, L, and talent, t. Urban costs play no role here because they are the same for everyone in a city.

This complementarity between population size and talent highlighted by Proposition 3 pushes the most talented individuals to locate in the largest cities and acts as a sorting force. Empirical work (e.g., Wheeler, 2001; Bacolod, Blum and Strange, 2009; and Glaeser and Resseger, 2010) often underscores such complementarity between cities and what we call talent (i.e., education and predetermined skills).

[9] Even for simple distributions like the Pareto distribution, the joint distribution of the product $s \times t$ is involved (see, e.g., Glen, Leemis and Drew, 2004). Nevertheless, we can derive our main results without imposing any specific form for G_t or G_s.

Returning to equation (7), it also highlights that the profit of an entrepreneur increases with her productivity relative to aggregate city productivity. Using again the link between talent and productivity, equation (16) takes this relation one step back and shows how expected profit in a city depends on talent relative to the selection threshold in this city. That is expected profit depends on the talent of an individual relative to that of the others in her city. All else equal, this pushes talented individuals to locate in cities where they are more talented than the others.

To illustrate the tradeoff between these two opposing forces, we use the aphorism attributed to Julius Caesar: "I had rather be the first in a village than second in Rome". Caesar knew that for a man of his talent there was much to be gained from being in Rome, the dominant city of a nascent empire. This 'Rome' effect is the same as the complementarity described above. At the same time, Caesar also knew there were benefits from being first, i.e., be more talented than the others in a village. This 'first-in-village' effect pushes towards the dispersion of talent across cities.[10]

More formally, the cross-partial derivative in Proposition 3 resembles a single-crossing condition. There is a key difference. It is conditional on the distribution of productivity $F(.)$ in a city. Hence, unlike more standard cases, the sign of this cross-partial derivative does not ensure the existence of a separating equilibrium since different cities may face different distributions of talent and thus productivity. More specifically, the first-in-village effect makes it harder to exhibit equilibria where the likes of Julius Caesar eventually cross the Rubicon and take their chances in Rome.

4 Location choices and cities

We now turn to the 'long-run' equilibrium of our model. Individuals make optimal location choices based on their talent. No individual wants to deviate to another city given the location choices of all other individuals. When choosing a city c an individual with talent t maximises expected utility, $\mathbb{E}V_c(t)$ (luck is still unknown). We define the assignment function $\mu : T \to C$ which maps talents into cities. An equilibrium is a correspondence $c = \mu(t)$ such that, for all $t \in T$ and for all $c, c' \in C$:

$$\mu(t) = \{c : \mathbb{E}V_c(t) \geq \mathbb{E}V_{c'}(t),\ \forall c' \in C\}\,. \tag{17}$$

Then, once in a city individuals also make an occupational choice based on their productivity, φ, the product of their talent, t, and their luck, s, to maximise utility. Entrepreneurs also choose employment in their firm to maximise profit and all markets clear. Hence, in addition to the 'short-run' equilibrium conditions, the population of each city, L_c, is endogenous and is given by

$$L_c \equiv \int_{\underline{t}}^{\bar{t}} L_c(t)\mathrm{d}t\,, \tag{18}$$

[10]There is also a symmetric force pushing towards the dispersion of less talented individuals. Larger cities pay higher nominal wages (by Proposition 2). This encourages less talented individuals who have little chance of becoming entrepreneurs anyway to try their luck in a large city. "*Panem et circenses*" (bread and circuses) at the bottom of the distribution in Rome may be preferable to relative mediocrity in a village.

where $L_c(t)$ is the population with talent t in city c. In addition, the set of cities, $C \equiv [0, \overline{c}]$, is also endogenous and cities arise under 'self-organisation' (to use the terminology of Henderson and Becker, 2000).

4.1 Symmetric equilibrium

Proposition 3 suggests that more talented individuals benefit disproportionately 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 natural question to ask is under which conditions a symmetric equilibrium may be stable.

Proposition 4 (Symmetric equilibrium) *Assume that $F_c(\cdot) = F(\cdot)$ for all cities c. Then the only equilibrium is a symmetric equilibrium with $L_c = L$, $\phi_c = \phi$ and $\underline{\varphi}_c = \underline{\varphi}$ for all c. This equilibrium is stable only if the variation in talent across the population is small enough.*

Proof. Assume that $F_c(\cdot) = F(\cdot)$ for all c. By the uniqueness of the solution to (13), which is independent of L_c, we then have $\underline{\varphi}_c = \underline{\varphi}$ for all c. This implies that selection is constant across cities: $\phi_c = F_c(\underline{\varphi}_c) = \phi$ 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 and talents. Since $F_c \equiv F$, 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 $t \in [\underline{t}, \overline{t}]$. The first condition ensures that individuals want to stay in existing cities (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 (16) and the fact that expected indirect utility is increasing in t, these two conditions will hold for all $t \in [\underline{t}, \overline{t}]$ if and only if

$$\frac{\varepsilon}{\gamma}\underline{\varphi}\left[G_s\left(\frac{\underline{\varphi}}{\overline{t}}\right) + \left(\frac{\overline{t}}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}}\int_{\frac{\underline{\varphi}}{\overline{t}}}^{\overline{s}} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s\right] < \theta L^{\gamma-\varepsilon}\frac{1+\varepsilon}{(\phi\varepsilon)^{\varepsilon}} < \underline{\varphi}\left[G_s\left(\frac{\underline{\varphi}}{\underline{t}}\right) + \left(\frac{\underline{t}}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}}\int_{\frac{\underline{\varphi}}{\underline{t}}}^{\overline{s}} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s\right].$$

This expression bounds the size L of symmetric cities. In addition, it implies:

$$\frac{\varepsilon}{\gamma} < \frac{G_s\left(\frac{\underline{\varphi}}{\underline{t}}\right) + \left(\frac{\underline{t}}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}}\int_{\frac{\underline{\varphi}}{\underline{t}}}^{\overline{s}} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s}{G_s\left(\frac{\underline{\varphi}}{\overline{t}}\right) + \left(\frac{\overline{t}}{\underline{\varphi}}\right)^{\frac{1}{\varepsilon}}\int_{\frac{\underline{\varphi}}{\overline{t}}}^{\overline{s}} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s},$$

which must hold at any stable symmetric equilibrium. Since $\gamma > \varepsilon$, the left-hand size of this expression is smaller than unity. The right-hand side is increasing with $\underline{t}$ and decreasing with $\overline{t}$. Furthermore, it is also smaller than unity, but it limits 1 as $\underline{t} \to \overline{t}$. Hence, this condition is fulfilled for a 'sufficiently homogeneous' population ($\underline{t} \approx \overline{t}$). ■

Proposition 4 shows that symmetry is stable only if the variation of talent across the population is small enough. This suggests that ability sorting is a natural equilibrium outcome when individuals are sufficiently heterogeneous.

4.2 Perfect sorting equilibrium

We now 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 equilibrium selection, we note that this equilibrium allows us to account for the stylised facts described above. It also displays all the main tradeoffs in a tractable analytic setting.

Lemma 5 (Selection under perfect sorting) *In talent-homogeneous cities, the productivity cutoff is proportional to talent: there exist $S \in (\underline{s}, \overline{s})$ and $\phi \in (0,1)$ such that $\underline{\varphi}_c = St_c$ and $G_s(S) = \phi$ for all $c \in C$, with*

$$\phi \equiv G_s(S) = \frac{1}{\varepsilon} \int_S^{\overline{s}} \left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s). \tag{19}$$

Proof. It follows from the definition of $F_c(\underline{\varphi}_c)$ that the fraction of individuals choosing to become entrepreneurs after learning their luck, s, is independent from c: $F_c(\underline{\varphi}_c) \equiv \Pr\{\varphi < \underline{\varphi}_c\} = G_s(\underline{\varphi}_c/t_c)$. Inserting this expression into (13) and replacing φ by $s\,t_c$ and $\underline{\varphi}_c$ by $S_c\,t_c$ yields (19) since it it immediate that S_c must be the same for all c. ■

Lemma 5 highlights two important results. First, sorting induces selection. Talent-homogeneous cities where talent is higher have a proportionately higher productivity cutoff, $\underline{\varphi}_c = St_c$. Second, conditional on sorting there are no differences in selection across cities. We also note that S is the implicit solution to equation (19). Obtaining an explicit solution would require the specification of a tractable functional form for the cumulative distribution of luck, G_s.[11]

To establish the conditions under which perfect sorting is part of an equilibrium, let us first use equations (16) and (19) to write the expected indirect utility of an individual endowed with talent t in talent-homogeneous city t_c of size L_c:

$$\begin{aligned} \mathbb{E}V_c(t) &= \frac{1}{1+\varepsilon} (\varepsilon\phi L_c)^{\varepsilon}\, St_c \left[G_s(s') + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}} \int_{s'}^{\overline{s}} \left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - \theta L_c^{\gamma} \\ &= \frac{1}{1+\varepsilon} (\varepsilon\phi L_c)^{\varepsilon}\, St_c \left[\int_{\underline{s}}^{S} \mathrm{d}G_s(s) + \int_{S}^{s'} \mathrm{d}G_s(s) \right. \\ &\quad \left. + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}} \int_{S}^{\overline{s}} \left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}} \int_{s'}^{S} \left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - \theta L_c^{\gamma}. \end{aligned} \tag{20}$$

[11] The implicit function theorem shows that $\partial S/\partial\varepsilon < 0$: selection is tougher when varieties become closer substitutes. This is because greater substitutability leads to lower price markups and, in turn, lower profits. This then leaves room for fewer entrepreneurs.

It is easy to see that expected indirect utility increases with talent, t. A higher level of talent increases the probability for an individual to make it into entrepreneurship and, in turn, increases her expected earnings (given city size L_c). As an aside and for future reference, we note that for $t = t_c$, equation (20) indicates the expected consumption of a representative individual in talent-homogeneous city c. Since $t = t_c$ implies $s' = S$, equation (20) then simplifies into:

$$\mathbb{E}V_c(t_c) = \phi^{1+\varepsilon} S t_c (\varepsilon L_c)^{\varepsilon} - \theta L_c^{\gamma}. \tag{21}$$

Each individual chooses a city by solving $\max_{c \in C} \mathbb{E}V_c(t)$, where $\mathbb{E}V_c(t)$ is given by (20). 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 (17) and $L(c)$ is given by (18).[12] Hence, an equilibrium with talent homogeneous cities is characterised by a function $L_c(t_c)$ that assigns one city size to each talent. Individuals choose their preferred city from a 'menu' of possible combinations of talent and size, *knowing* that the choice of a talent t_c implies the choice of a size $L_c(t_c)$. Formally, this implies that 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. Inserting the constraint $L_c \equiv L_c(t_c)$ into the objective function, for an individual of talent t the first-order condition to the city selection problem (17) with talent-homogeneous cities can be written as:

$$\left.\frac{\partial \mathbb{E}V_c(t)}{\partial L_c}\right|_{t=t_c} \mathrm{d}L_c + \left.\frac{\partial \mathbb{E}V_c(t)}{\partial t_c}\right|_{t=t_c} \mathrm{d}t_c = 0. \tag{22}$$

Hence, for talent-homogeneous cities to be an equilibrium, there must exist a relationship between talent and size such that each and every individual chooses a city where everyone has the same talent as hers, and this choice maximises her expected indirect utility. More formally, we must find a function $L_c(t_c)$ such that (22) holds for all t at $t = t_c$.

We may view such equilibrium function $L_c(t_c)$ as describing an envelope of indifference curves in the (t_c, L_c) space. This function is represented by the bold curve in figure 2. As we will show, it is convex when $\gamma - \varepsilon < 1$ (which is empirically the case, as highlighted by section 5). Consider an individual with talent t_0 choosing from the menu of equilibrium cities described by $L_c(t_c)$. 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 size 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 size L_c that offer the same expected utility as city c_0 conditional on a talent t_0.[13] Since expected indirect utility is increasing in the direction represented

[12]For $\mu(t)$ to be invertible we need either a strict inequality in (17) (which occurs in equilibrium) or that ties are always broken in the same way in case of equality. Also, we can label cities in such a way that t is a continuous function of c. It then follows from this labelling and from (22) that L is also a continuous function of c.

[13]Observe that this curve yields higher utility as it has smaller cities (less congestion) and more talent (higher productivity): this is easily confirmed (locally at $t = t_c$) by computing $\partial \mathbb{E}V_c(t)/\partial L_c \left.\right|_{t=t_c} > 0$ and $\partial \mathbb{E}V_c(t)/\partial t_c \left.\right|_{t=t_c} >$

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 $L_c(t_c)$ 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.

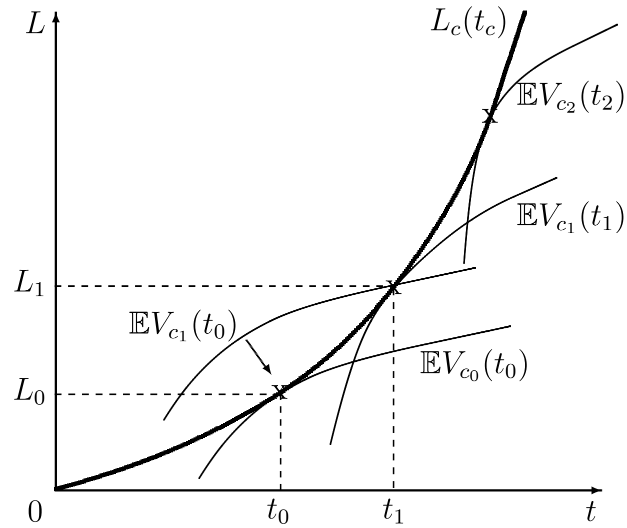

Figure 2. Equilibrium with talent-homogeneous cities

We now solve formally for the equilibrium. From expressions (20) and (22), we obtain:[14]

$$\left\{\frac{\varepsilon}{1+\varepsilon}\left(\varepsilon\phi L_c\right)^{\varepsilon} St_c\left[\phi\left[1+\varepsilon\left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\right]+\int_{s'}^{S}\left[\left(\frac{st}{St_c}\right)^{\frac{1}{\varepsilon}}-1\right]\mathrm{d}G_s(s)\right]-\gamma\theta L_c^{\gamma}\right\}\frac{\mathrm{d}L_c}{L_c}$$
$$+\frac{1}{1+\varepsilon}\left(\varepsilon\phi L_c\right)^{\varepsilon} S\left\{\phi\left[1+(\varepsilon-1)\left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\right]\right.$$
$$\left.-\int_{s'}^{S}\left[\left(\frac{st}{St_c}\right)^{\frac{1}{\varepsilon}}\frac{\varepsilon-1}{\varepsilon}+1\right]\mathrm{d}G_s(s)-t_c\left[\left(\frac{s't}{St_c}\right)^{\frac{1}{\varepsilon}}-1\right]g_s(s')\frac{\partial s'}{\partial t_c}\right\}\mathrm{d}t_c=0 \qquad (23)$$

Two properties of (23) are noteworthy. First, setting the term inside the first curly bracket to zero for $t=t_c$ admits an interior solution in L_c. Because this term corresponds to the first term in (22), this solution maximises expected utility with respect to size in a talent-homogeneous city. Put differently, for $t=t_c$, the solution to $\partial\mathbb{E}V_c(t)/\partial L_c=0$ in L_c is the (constrained) socially optimal size of talent-homogeneous city c.[15] Second, still for $t=t_c$ (and thus $s'=S$), the terms

0, which yields the shape of the indifference maps (the arrow in figure 2 points in the direction of increasing expected indirect utility).

[14] s' is not generally differentiable with respect to t. Hence, $\partial s'/\partial t_c$ should be viewed as a generalised derivative. This does not matter for our problem at hand since, at the talent-homogeneous equilibrium, the term involving $\partial s'/\partial t_c$ vanishes regardless of that (finite) derivative.

[15] Given labour supply, the equilibrium is optimal. However, the equilibrium choice of occupations is not in general socially efficient. The equilibrium equates the wage with the profit of the marginal entrepreneur whereas optimality

of the third line of (23) are equal to zero. In turn, this implies that the second curly bracket is positive. Since this term corresponds to the second term in (22), the first-order condition for a social optimum is violated at equilibrium city size for talent-homogeneous cities. Individuals have an incentive to move to a city less talented than socially optimal. This is a consequence of the first-in-village effect described above. The only reason why more talented individuals will accept to be in more talented cities is that they are larger and, as shown by Proposition 3 more talented individuals benefit relatively more from larger cities. These properties imply that cities are too large in equilibrium. Coming back to our earlier metaphor, Rome must grow sub-optimally large to offset the first-in-village effect and attract the likes of Julius Caesar.

Writing L and t as explicit functions of c, i.e. $L_c \equiv L(c)$ and $t_c \equiv t(c)$, plugging these into (23), and evaluating the resulting expression at $t = t(c)$ yields a differential equation that determines the menu (t_c, L_c) that supports the talent-homogeneous equilibrium. Formally, we have to solve:

$$\gamma\theta L(c)^{\varepsilon}\left[\xi t(c)\frac{L'(c)}{L(c)} - L(c)^{\gamma-\varepsilon}\frac{L'(c)}{L(c)} + \delta\right] = 0, \tag{24}$$

where

$$\xi \equiv \frac{S}{\gamma\theta}\left(\varepsilon\phi\right)^{1+\varepsilon} \quad \text{and} \quad \delta \equiv \frac{\xi}{1+\varepsilon} > 0,$$

and ξ is increasing in selection and decreasing in the intensity of urban costs. Observe that δ may be interpreted as the equilibrium market distortion that results from sorting since it arises from the second term in (22).

Equation (24) under $\delta = 0$ enables us to characterise the optimal size of talent-homogeneous cities. This size is such that total land rent is equal to output times the degree of increasing returns. This result is known as the Henry George Theorem and occurs in many urban models.[16]

Proposition 6 (Optimal city size under perfect sorting) *Talent-homogeneous cities of optimal size are such that:*

$$L^o(c) = [\xi t(c)]^{\frac{1}{\gamma-\varepsilon}}. \tag{25}$$

Optimal size increases with talent, t_c, and agglomeration economies, ε, decreases with urban costs, θ and γ, is such that the Henry George Theorem holds: $TLR(c) = \varepsilon Y(c)$, and leads to expected indirect utility: $\mathbb{E}V^0(c) = \frac{\theta}{\varepsilon}(\gamma - \varepsilon)[L^o(c)]^{\gamma}$.

Proof. By definition, talent-homogeneous cities of optimal size solve equation (24) for $\delta = 0$. Standard methods to solve such differential equations yield equation (25). The comparative statics results are immediate by $\gamma > \varepsilon$. Observe that $L^o(c)$ is actually a local maximum if $\mathbb{E}V_c(t)$ is concave in $L(c)$ at $t = t(c)$. Appendix A shows under which circumstances this necessary second-order condition is satisfied. Next, total land rent is equal to $TLR(c) = \theta\gamma L(c)^{1+\gamma}$ by (15) and city

requires equating the output of the marginal entrepreneur with foregone output from inframarginal ones.

[16]See Arnott (2004) for further discussions of the Henry George Theorem.

output is $Y(c) = \varepsilon^{\varepsilon} \left[\varphi\, L(c)\right]^{1+\varepsilon} St(c)$ by (21). Using (25), it is easy to verify that $TLR(c) - \varepsilon Y(c) \gtreqless 0$ if and only if $L(c) \gtreqless L^{o}(c)$. Finally, plugging (25) into (21) yields the expression for expected indirect utility. ■

As made clear by equation (25), optimal city size is determined by three elements. The first is the standard trade-off between agglomeration economies (as given by ε) and urban costs (as given by γ and θ). The second is talent in the city. The third is the proportion of entrepreneurs, which is endogenously determined (but constant across cities). Cities with more talented entrepreneurs have a larger optimal size. This is because more talent and better luck lead to higher productivity firms and, in turn, productivity is magnified by agglomeration economies as shown by equation (14).

We also note that this optimal size is equal to the one obtained when maximising expected utility with respect to size in equation (21). This is unsurprising. When cities are talent-homogeneous and in the absence of interactions between them, the optimal assignment of talent boils down to finding an optimal size for each city depending on the talent of its residents regardless of the overall distribution of talents. We now turn our attention to the decentralised solution, which solves equation (24) under $\delta \neq 0$:

Proposition 7 (Equilibrium size under perfect sorting) *The talent-homogeneous equilibrium is unique and such that*

$$L^{*}(c) = \left(\frac{1+\gamma}{1+\varepsilon}\right)^{\frac{1}{\gamma-\varepsilon}} L^{o}(c). \tag{26}$$

Equilibrium size is too large, increases with talent, t_c, and agglomeration economies, ε, decreases with urban costs, θ and γ, and leads to expected indirect utility: $\mathbb{E}V^{*}(c) = \frac{\theta}{\varepsilon}\frac{\gamma-\varepsilon}{1+\gamma} L^{*}(c)^{\gamma}$.

Proof. The solution to (24) for $\delta \neq 0$ is slightly more complicated than for $\delta = 0$. So we proceed in two steps. First, we know that it must be of the same form as (25) because δ only appears as a constant in (24). Hence, we impose $L(c) = zt(c)^{\frac{1}{\gamma-\varepsilon}}$, for some z, and plug this into (24). This yields an equation involving the parameters of the model that is linear in z. Solving for z gives

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

which establishes the result. For the second-order conditions, see Appendix A. ■

Cities are too large in equilibrium as in Henderson (1974). Because of the first-in-village effect, when cities are of optimal size, any individual would like to relocate to smaller and less talented cities to lessen selection. Less talented cities, when oversized, become less attractive in that respect. However, it should also be that individuals should not want to relocate to larger cities where wages are higher even if this comes at the cost of stronger selection. More talented cities, when oversized, have higher urban costs and thus become less attractive in that respect. In the end, the gap between optimal and equilibrium cities depends only on the intensity of urban costs, γ, and agglomeration economies, ε.

We also note that 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 size. With homogeneous individuals, there would be a coordination failure in city formation so that any 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 indetermination 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 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 sizes are uniquely determined.[17]

4.3 The size distribution of cities

To close the model, we return to the assignment function $\mu(t) = c$, with $\mu(\cdot) = \mu^o(\cdot)$ in the optimal allocation and $\mu(\cdot) = \mu^*(\cdot)$ in the equilibrium one. This assignment function assigns all individuals in the economy to a city where to live. Roughly speaking, the assignment function tells us to which share $\mu(t_c)$ of cities the share of individuals with talent less than t_c is assigned. After denoting Λ the population size and g_t the p.d.f. of G_t, the 'full-population condition' thus requires that

$$\begin{aligned} \Lambda \int_{\underline{t}}^{t} g_t(\nu)\mathrm{d}\nu &= \int_0^{\mu(t)} L(c)\mathrm{d}c, \quad \forall t \in [\underline{t}, \overline{t}] \\ \mu(\underline{t}) &= 0. \end{aligned} \tag{27}$$

Proposition 8 (Number of cities) *The equilibrium 'number' of cities is proportional to population size Λ and too small relative to the social optimum.*

Proof. Differentiating the first expression of (27) 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

$$\mu(t) = \mu(\underline{t}) + \Lambda \int_{\underline{t}}^{t} \frac{g_t(z)}{L(z)}\mathrm{d}z, \tag{28}$$

where the constant of integration $\mu(\underline{t})$ is equal to zero by (27). Inserting $\mu(\overline{t}) = \overline{c}$ into this expression pins down the measure of C:

$$\overline{c} = \mu(\overline{t}) = \Lambda \int_{\underline{t}}^{\overline{t}} \frac{g_t(z)}{L(z)}\mathrm{d}z, \tag{29}$$

[17] Henderson and Venables (2009) propose a dynamic fix for this coordination failure.

which shows immediately that $\bar{c}$ increases proportionately to Λ. Let $\widetilde{\xi} \equiv \xi$ at the optimal solution and $\widetilde{\xi} \equiv \frac{1+\gamma}{1+\varepsilon}\xi$ at the equilibrium. Then, using (25) and (26), equations (28) and (29) may be rewritten as

$$\mu(t) = \frac{\Lambda}{\widetilde{\xi}} \int_{\underline{t}}^{t} \frac{g_t(z)}{z^{\frac{1}{\gamma-\varepsilon}}} \mathrm{d}z \quad \text{and} \quad \bar{c} = \mu(\bar{t}) = \frac{\Lambda}{\widetilde{\xi}} \int_{\underline{t}}^{\bar{t}} \frac{g_t(z)}{z^{\frac{1}{\gamma-\varepsilon}}} \mathrm{d}z. \tag{30}$$

Because, $\widetilde{\xi}$ is smaller at the social optimum than at the market equilibrium, the second equality in (30) immediately implies the second part of the result. ■

To derive sharper results about the size distribution of cities, we make additional assumptions about the distribution of talent. Assume for now that it follows a truncated Pareto distribution with support $[\underline{t}, \bar{t}]$:

$$G_t(t) = \frac{1 - (\underline{t}/t)^m}{1 - \left(\underline{t}/\bar{t}\right)^m}. \tag{31}$$

We note that this allows us to consider distributions of talent that are right skewed ($m > -1$), uniform ($m = -1$), or even left skewed ($m < -1$).

Proposition 9 (Size distribution of cities) *When talent follows a truncated Pareto distribution with shape parameter m over $[\underline{t}, \bar{t}]$, the size distribution of talent-homogeneous cities is then a truncated Pareto with shape parameter $1 + (\gamma - \varepsilon)m$ at both equilibrium and optimum.*

Proof. Inserting the probability distribution function associated with (31) into both equalities in (30) and computing the integrals yields:

$$\mu(t) = \frac{\Lambda m \, \underline{t}^m}{\left(m + \frac{1}{\gamma-\varepsilon}\right) \widetilde{\xi}^{\frac{1}{\gamma-\varepsilon}} \left[1 - \left(\underline{t}/\bar{t}\right)^m\right]} \left[\underline{t}^{-(m+\frac{1}{\gamma-\varepsilon})} - t^{-(m+\frac{1}{\gamma-\varepsilon})}\right]$$

and

$$\bar{c} = \frac{\Lambda m}{\left(m + \frac{1}{\gamma-\varepsilon}\right) \left(\widetilde{\xi}\underline{t}\right)^{\frac{1}{\gamma-\varepsilon}} \left[1 - \left(\underline{t}/\bar{t}\right)^m\right]} \left[1 - \left(\underline{t}/\bar{t}\right)^{m+\frac{1}{\gamma-\varepsilon}}\right].$$

To get an expression for the size distribution of cities, $G_L(L) \equiv \frac{1}{\bar{c}} \int_0^{c(L)} \mathrm{d}c$, we use the assignment function $\mu(\cdot)$ and equation (26) (at the equilibrium) or (25) (at the optimum). This yields

$$G_L(t) = \frac{1 - (\underline{t}/t)^{m+\frac{1}{\gamma-\varepsilon}}}{1 - \left(\underline{t}/\bar{t}\right)^{m+\frac{1}{\gamma-\varepsilon}}} \quad \Rightarrow \quad G_L(L) = \frac{1 - \left(\underline{L}/L\right)^{1+(\gamma-\varepsilon)m}}{1 - \left(\underline{L}/\overline{L}\right)^{1+(\gamma-\varepsilon)m}} \tag{32}$$

which is a truncated Pareto distribution with shape parameter $1 + (\gamma - \varepsilon)m$ over $[\underline{L}, \overline{L}]$ where $\underline{L} \equiv L(\underline{t})$ and $\overline{L} \equiv L(\bar{t})$. ■

It is easy to understand that if talent follows a Pareto distribution, the size distribution of cities is also Pareto. This occurs because both optimal and equilibrium city sizes in (25) and (26) 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. Beyond this, note that the first part of equation (32) gives the distribution of city sizes by talent. Since city size is proportional to $t^{1/(\gamma-\varepsilon)}$, 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. This adds $\frac{1}{\gamma-\varepsilon}$ to the skew of the talent distribution and yields a Pareto distribution of sizes by talent with shape $m + \frac{1}{\gamma-\varepsilon}$. In turn, the size distribution of cities, is obtained from a change of variable using the fact that $L^{\gamma-\varepsilon}$ is proportional to t. This yields a shape parameter of $1 + m(\gamma - \varepsilon)$ for the size distribution of cities.

The fact that $\gamma - \varepsilon$ is empirically small (as highlighted by the next section) has two interesting implications. First, the equilibrium and optimum distributions for the size of cities are expected to be close to a Pareto distribution with unitary shape parameter, a.k.a. Zipf's law, which is a reasonable first-order approximation for observed city size distributions (Gabaix and Ioannides, 2004; Soo, 2005). Second, even if the distribution of talents does not follow a power law, the fact that $\gamma - \varepsilon$ is small still implies that the size distribution of cities remains approximately Zipf.

Proposition 9 completes the characterisation of the equilibrium with talent-homogeneous cities. This equilibrium is consistent with the stylised facts discussed above about agglomeration, sorting, selection, and the size distribution of cities. In particular, if we take seriously the empirical result of Combes, Duranton, Gobillon, Puga, and Roux (2009) 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 equilibria. In Appendix B, we explore more fully equilibria with constant selection but imperfect sorting. We show that the key qualitative properties of the equilibrium with talent-homogeneous cities continue to hold true.

There are two further reasons that justify our focus on the equilibrium with talent-homogeneous cities. In section 6, we investigate an example of equilibrium with variable selection across cities. We show that the key properties of the equilibrium with talent-homogeneous cities are also properties of this equilibrium with variable selection. Finally, in Appendix C, we show that the socially optimal solution in a simplified setting also allocates the most talented entrepreneurs to the largest cities under the exact same conditions that are necessary for the second-order condition in Proposition 7 to hold. Strikingly, this condition ensures that $\mathbb{E}V(t, L)$ is log supermodular in its arguments, in which case Proposition 3 holds *irrespective* of the composition of cities in terms of talent.

5 Quantitative implications

We now use our framework to revisit several well-known empirical results. Since the model is highly stylised, this exercise should be viewed as 'theory with numbers', not empirical analysis. Taking the log of average city earnings as given by the first term on the right hand side of equation (21), we obtain

$$\ln y_c = [\ln(\phi S) + \varepsilon \ln(\varepsilon\phi)] + \ln t_c + \varepsilon \ln L_c \,. \tag{33}$$

This expression indicates that regressing log average earnings on log population while controlling for talent yields an estimate of agglomeration economies, ε. Now, using equations (25) and (26) to obtain a relation between talent and equilibrium size and substituting for t_c into (33), we get:

$$\ln y_c = \ln\left(\frac{1+\varepsilon}{1+\gamma}\,\frac{\gamma\theta}{\varepsilon}\right) + \gamma \ln L_c\,. \tag{34}$$

Hence, regressing log average earnings on log population *without* controlling for talent yields an estimate of the urban costs parameter, γ. To understand why this is so, note first that cities result from a tradeoff between agglomeration benefits and urban costs. Cities can be of different sizes either because they differ in how they benefit from agglomeration depending on their size or in how urban costs are affected by size. In our model, urban costs are equal to θL^γ in all cities. As made clear by (33), expected earnings depend on the level of talent in the city. As a result, if we do not control for talent, we look 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 elasticity of urban costs with respect to size.[18]

Let us illustrate this with standard data. We estimate equations (33) and (34) using 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 labour economics. We obtain:[19]

$$\ln y_c = 8.59 + 0.082 \ln L_c\,, \tag{35}$$

$$\ln y_c = 9.60 + 0.051 \ln L_c + 0.46 \ln t_c\,. \tag{36}$$

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 γ and 0.050 for ε. Using the share of population older than 18 years with a graduate or professional degree to measure t_c in regression (36) yields a coefficient of 0.058 on log population.

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.[20] The sizable drop in

[18] Matters are actually more complicated than that because expected indirect utility is not equalised across cities as shown by Proposition 7. 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. However, in equilibrium, the elasticity of expected indirect utility with respect to city size is the same as that of urban costs, γ. Because of this, expected indirect utility varies across cities like urban costs and this creates no further problem.

[19] All coefficients, including the constant terms, are significant at the 1% confidence level in all estimations.

[20] 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.

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 reports estimates close to ours (Albouy, 2009; Baum-Snow and Pavan, 2009a). To corroborate our finding 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 (35). 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 α_c by MSA c's share of owner-occupied housing. Regressing the log of this housing price index, h_c, on the log of population yields:

$$\ln h_c = 9.72 + 0.11 \ln L_c. \tag{37}$$

This estimate of 0.11 for urban costs remains reasonably close to that in (35) despite relying on a different estimating equation.

Equation (26) also implies that the elasticity of talent (measured by the share of university graduates) to city size should be equal to $\gamma - \varepsilon$. We obtain $\hat{\gamma} - \hat{\varepsilon} = 0.031$ when using (35) and (36) and $\hat{\gamma} - \hat{\varepsilon} = 0.059$ when using (35) and (37). Regressing directly the log-share of graduates on log-population yields

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

This elasticity 0.068 is statistically undistinguishable from 0.059 (though it differs from 0.031). At first sight, small values for the elasticity of talent to city size appear to argue against the importance of ability sorting across cities. Our model shows instead that a small value for the size elasticity of talent corresponds in equilibrium to the small difference between the size elasticity of urban costs and that of agglomeration economies. Put differently, city size is proportional to $t_c^{1/(\gamma-\varepsilon)}$. A small difference between γ and ϵ, as seems to be the case empirically, is then enough for small differences in talent to translate into large differences in city size. For instance, if the smallest city in the economy has $\underline{L} = 10,000$ inhabitants and the largest has $\overline{L} = 10$ million, then this corresponds to the latter city being only about 24% more talented than the former given our estimates of γ and of ε.

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

$$\frac{\widehat{L_c^*}}{L_c^o} = \left(\frac{1+\widehat{\gamma}}{1+\widehat{\varepsilon}}\right)^{\frac{1}{\hat{\gamma}-\hat{\varepsilon}}} = 2.55. \tag{38}$$

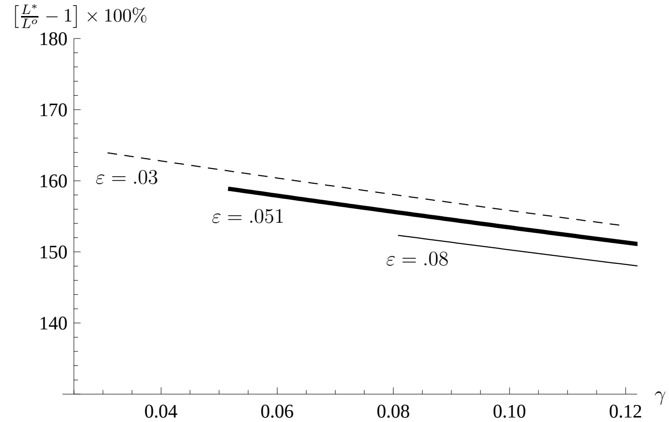

Figure 3. Oversize as a function of γ for different values of $\varepsilon < \gamma$

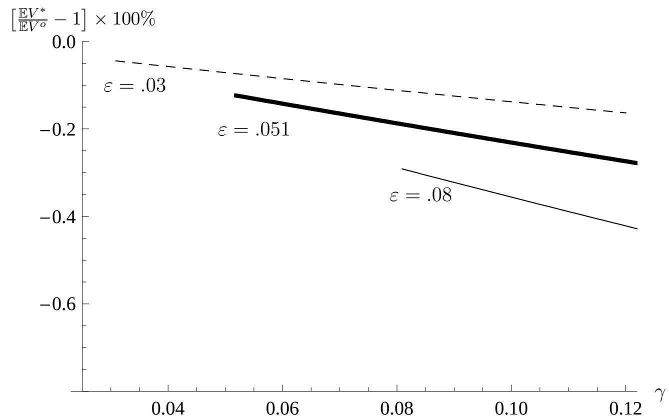

Figure 4. Efficiency loss as a function of γ for different values of $\varepsilon < \gamma$

This suggests that US cities may be, on average, about 155% larger than their optimal size.[21] To check the robustness of this finding, figure 3 plots the oversize of cities as computed in equation (38) for varying values of γ and three values of ε. Despite covering a broad range around our preferred estimates, this plot indicates that an oversize of 145 to 165% is to be expected. Consistently with the comparative statics of equation (38), the figure also shows that city oversize decreases in γ and ε ($< \gamma$). Using a first-order linear approximation of equation (38) when $\gamma - \varepsilon$ is small, we obtain $L^*_c/L^o_c \approx \exp(\frac{1}{1+\varepsilon})$ which tends to Euler's number when ε and γ go to zero. Given that ε and γ 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 (25), (26), and (38) to compute an estimate of

[21] In equation (24), δ the equilibrium size distortion is an alternative measure of oversize. Subtracting the constant of equation (34) from that of (33) yields $\ln(\delta) + \ln(1+\gamma)$. Using $\widehat{\gamma} = 0.082$ and the constants estimated in equations (35) and (36), we obtain $\widehat{\delta} = 2.54$.

the indirect utility (consumption) loss:

$$\widehat{\Delta \mathbb{E}V} \equiv \frac{\mathbb{E}V(\hat{L}^*) - \mathbb{E}V(\hat{L}^o)}{\mathbb{E}V(\hat{L}^o)} = -1 + \frac{\widehat{\gamma}}{\widehat{\gamma} - \widehat{\varepsilon}} \left(\frac{\widehat{L_c^*}}{L_c^o} \right)^{\widehat{\varepsilon}} - \frac{\widehat{\varepsilon}}{\widehat{\gamma} - \widehat{\varepsilon}} \left(\frac{\widehat{L_c^*}}{L_c^o} \right)^{\widehat{\gamma}} = -0.2\%. \tag{39}$$

This loss in consumption is economically small, about one-fifth of a percentage point. To confirm the robustness of this magnitude, figure 4 plots the economic loss associated with this oversize for the same parameter values as figure 3. 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 of at most $e = 2.72$. Imagine next that earnings are of the same magnitude as urban costs. Then, the maximum loss from oversize would be $1 - e^{-(\gamma-\varepsilon)}$ or about 2.8% for our preferred value of $\gamma - \varepsilon = 0.031$. However, equilibrium urban costs are much smaller than earnings so that the actual loss is much smaller than that. These results are consistent with those of Au and Henderson (2006) for Chinese city. 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.[22]

6 Equilibria with varying selection: an example

Equilibria with constant selection across cities seem empirically relevant. They are however special cases. While a general analysis of all equilibria is beyond the scope of this paper, we turn now to a situation 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 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 $T = [\underline{t}, \overline{t}]$ and $\Sigma = [\underline{s}, \overline{s}]$, respectively. Total population is fixed to Λ, 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 $\varphi \equiv ts$. Using theorem 1 of Glen, Leemis, and Drew (2004) and assuming without loss of generality for our purpose that $\underline{t}\overline{s} < \overline{t}\underline{s}$, the product of talent and luck is distributed as follows:

$$f(\varphi) = \begin{cases} \frac{1}{(\overline{s}-\underline{s})(\overline{t}-\underline{t})} \ln\left(\frac{\varphi}{\underline{s}\underline{t}}\right) & \text{if} \quad \underline{t}\underline{s} \leq \varphi \leq \underline{t}\overline{s} \\ \frac{1}{(\overline{s}-\underline{s})(\overline{t}-\underline{t})} \ln\left(\frac{\overline{s}}{\underline{s}}\right) & \text{if} \quad \underline{t}\overline{s} \leq \varphi \leq \overline{t}\underline{s} \\ \frac{1}{(\overline{s}-\underline{s})(\overline{t}-\underline{t})} \ln\left(\frac{\overline{s}\overline{t}}{\varphi}\right) & \text{if} \quad \overline{t}\underline{s} \leq \varphi \leq \overline{t}\overline{s} \end{cases} \tag{40}$$

[22] Figure 4 also shows that the loss from oversize increases with agglomeration economies, ε. This is because a higher ε (given γ) implies larger cities. In turn this magnifies the inefficiency from oversize (that the oversize becomes *relatively* smaller with ε only partially offsets this).

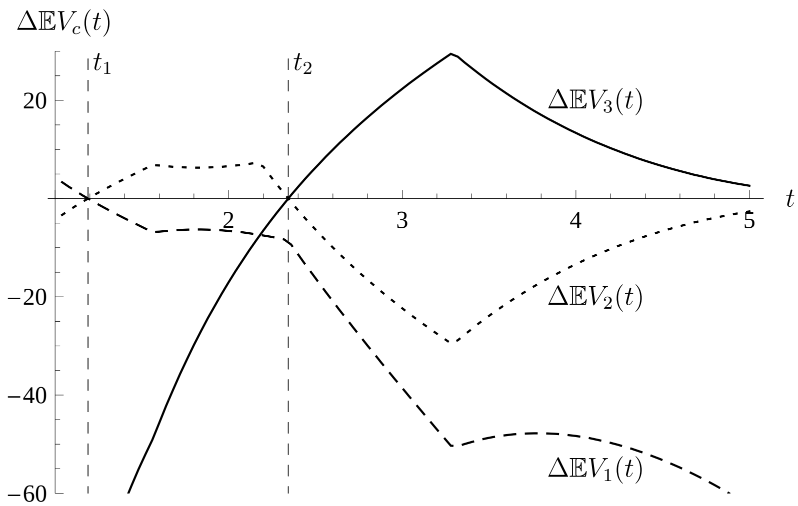

Figure 5. Example of spatial equilibrium with imperfect ability sorting

Using (40), 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 $t \in [t_2, \overline{t}]$ choose to locate in type-3 cities. The thresholds t_1 and t_2, the number of type-i cities, their sizes L_i and their productivity 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 $t \in [t_2, \overline{t}]$ (and negative otherwise).

Figure 5 depicts the expected indirect utility differential for the three types of cities, as well as the two talent cutoffs for: $\varepsilon = 0.47$, $\gamma = 0.5$, $\theta = 0.5$, $\Lambda = 5000$, $\underline{t} = \underline{s} = 1$, $\overline{t} = 5$ and $\overline{s} = 2$. In figure 5 we also set the numbers of cities to $n_1 = 20$, $n_2 = 6$ and $n_3 = 2$. This choice of parameter values calls for two remarks. First, our values for agglomeration economies, ε, and urban costs, γ are much larger than empirically reasonable for robustness purposes. Second, 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, $i = 1, 2, 3$ which can be in equilibrium.

The allocation we have chosen is in equilibrium for t_1 and t_2 as determined on figure 5 since all individuals located in type-1 cities (i.e., 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., 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 $w_3 = 89.46$,

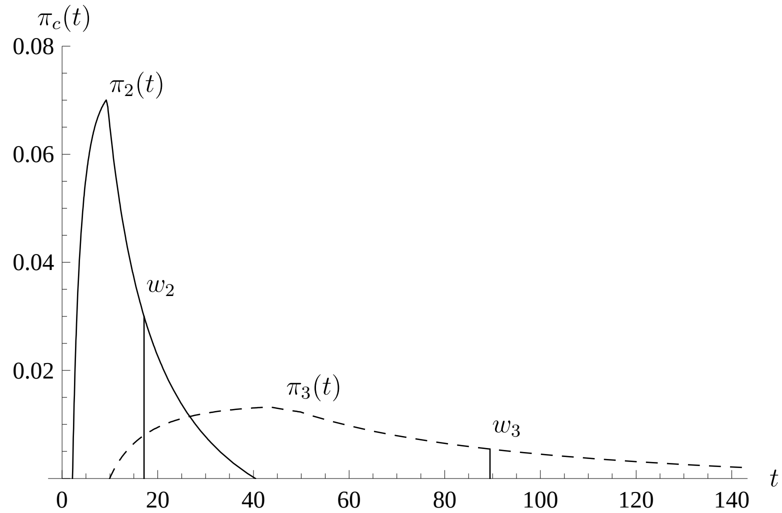

Figure 6. Distribution of profits in type-2 and type-3 cities

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 productivity 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 size-talent complementarity generate these asymmetries.

With talent-homogeneous cities, the degree of selection $F_c(\underline{\varphi}_c)$ is the same for all cities, a knife-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 productivity 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. The intuition is that larger cities provide entrepreneurs with access to more and richer consumers. This almost fully offsets the tougher environment.

Figure 6 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, 2009b; 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 (2009) regarding the distributions of firms' productivities in small and large French cities.

7 Conclusion

Although abundant empirical research in urban economics has substantiated a significant positive correlation between skills and city size, theory had much less to say about the spatial sorting of heterogeneous individuals across an urban hierarchy until now. This 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 more productive on average than in smaller cities, this is not true for all firms. Finally, the distribution of talent maps into the distribution of city sizes and this provides a simple static explanation for why cities' size distribution is approximately a Pareto with a shape parameter close to minus one.

In addition to our theoretical contribution, we believe that our model may be useful to make progress in various directions. First, our framework goes beyond the analysis of the interactions between sorting, selection and agglomeration. It also provides a setting within which to interpret quantitative evidence obtained from standard regressions. It suggests, in particular, how various regressions of measures of productivity, of skills and of 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 may be useful for guiding future empirical analysis.

Second, our concept of spatial equilibrium may be useful to shed new light on spatial arbitrage when static utility levels are not equalised across locations at equilibrium. Most of the literature in urban economics focuses on situations where all individuals are indifferent across all cities (e.g., Glaeser and Gottlieb, 2009). 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 may have 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, Tom Holmes, Yannis Ioannides, Sanghoon Lee, Diego Puga, Jacques Thisse, Jonathan Vogel, and especially Vernon Henderson and Esteban Rossi-Hansberg for very helpful discussions and comments. Conference and seminar participants at Berkeley, Chicago Fed, Columbia, Louvain-la-Neuve, Milan, MIT, Philadelphia Fed, Princeton, Rennes, Rimini, San Francisco, and University of Virginia provided valuable feedback. 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 is a visiting scholar at Princeton University; the hospitality of this institution is gratefully acknowledged. Any remaining errors are ours.

References

[1] 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.

[2] Albouy, David (2009) Are big cities really bad places to live? Improving quality-of-life estimates across cities. *NBER Working Paper* #14472.

[3] Amiti, Mary and Lisa Cameron (2007) Economic geography and wages, *Review of Economics and Statistics* 89(1), 15–29.

[4] Antràs, Pol, Luis Garicano, and Esteban Rossi-Hansberg (2006) Offshoring in a knowledge economy, *Quarterly Journal of Economics* 121(1), 31–77.

[5] Arnott, Richard J. (2004) Does the Henry George Theorem provide a practical guide to optimal city size?, *American Journal of Economics and Sociology* 63(5), 1057-1090.

[6] Au, Chun-Chung and J. Vernon Henderson (2006) Are Chinese cities too small? *Review of Economic Studies* 73(3), 549–576.

[7] Bacolod, Marigee, Bernardo S. Blum, and William C. Strange (2009) Skills in the city, *Journal of Urban Economics* 65(2), 136–153.

[8] 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

[9] Bartelsman, Eric J. and Mark Doms (2000) Understanding productivity: Lessons from longitudinal microdata, *Journal of Economic Literature* 38(3), 569–594.

[10] Baum-Snow, Nathaniel and Ronni Pavan (2009a) Understanding the city size wage gap. Processed, Brown University.

[11] Baum-Snow, Nathaniel and Ronni Pavan (2009b) Inequality and city size. Processed, Brown University.

[12] Behrens, Kristian and Frédéric Robert-Nicoud (2009) Survival of the fittest in cities: Agglomeration, polarization, and income inequality. CIRPÉE Discussion paper 09-19.

[13] Berry, Christopher R. and Edward L. Glaeser (2005) The divergence of human capital levels across cities, *Papers in Regional Science* 84(3), 407–444.

[14] Ciccone, Antonio and Robert E. Hall Hall (1996) Productivity and the density of economic activity, *American Economic Review* 86(1), 54-70.

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

[16] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux (2009) The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *CEPR Discussion Paper* #7191.

[17] Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2010) Estimating agglomeration economies with history, geology, and worker effects. In: Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge, MA: National Bureau of Economic Research, 15–65.

[18] Dahl, Gordon B. (2002) Mobility and the return to education: Testing a Roy model with multiple markets, *Econometrica* 70(6):2367-2420.

[19] Davis, Donald R. (2009) A spatial knowledge economy. Processed, Columbia University.

[20] Dixit, Avinash K. and Joseph E. Stiglitz (1977), Monopolistic competition and optimum product diversity *American Economic Review* 67(3) 297–308

[21] Doms, Mark, Ethan Lewis, and Alicia Robb (2010) Local labor force education, new business characteristics, and firm performance, *Journal of Urban Economics* 67(1), 61–77.

[22] Duranton, Gilles (2007) Urban evolutions: The fast, the slow, and the still, *American Economic Review* 97(1), 197–221.

[23] 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.

[24] Eeckhout, Jan (2004) Gibrat's Law for (all) cities, *American Economic Review* 94(5), 1429–1451.

[25] 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.

[26] Ellison, Glenn D., Edward L. Glaeser, and William R. Kerr (2010) What causes industry agglomeration? Evidence from coagglomeration patterns, *American Economic Review* 100, forthcoming.

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

[28] 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.

[29] Foster, Lucia, John Haltiwanger, and Chad Syverson (2008) Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?, *American Economic Review* 98(1), 394–425.

[30] Gabaix, Xavier (1999) Zipf's law for cities: An explanation, *Quarterly Journal of Economics* 114(3), 739–767.

[31] 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.

[32] Glaeser, Edward L., and Joshua D. Gottlieb (2009) The wealth of cities: Agglomeration economies and spatial equilibrium in the United States, *Journal of Economic Literature* 47(4), 983–1028.

[33] Glaeser, Edward L. and Matthew G. Resseger (2010) The complementarity between cities and skills, *Journal of Regional Science* 50(1), 221–244.

[34] Glaeser, Edward L., Matt Resseger, and Kristina Tobio (2009) Inequality in cities, *Journal of Regional Science* 49(4), 617–646.

[35] 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.

[36] Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010) Identifying agglomeration spillovers, *Journal of Political Economy* 118(3), 536–598.

[37] Gyourko, Joseph, Christopher Mayer, and Todd Sinai (2006) Superstar cities. Processed, The Wharton School, University of Pennsylvania.

[38] Henderson, J. Vernon (1974) The sizes and types of cities, *American Economic Review* 64(4), 640–656.

[39] Henderson, J. Vernon and Randy Becker (2000) Political economy of city sizes and formation, *Journal of Urban Economics* 48(3), 453–484.

[40] Henderson, J. Vernon and Anthony J. Venables (2009) The dynamics of city formation, *Review of Economic Dynamics* 12(2), 233–254.

[41] Holmes, Thomas J. (1999) Localization of industry and vertical disintegration, *Review of Economics and Statistics* 81(2), 314–325.

[42] Hsu, Wen-Tai (2008) Central place theory and city size distribution. Processed, Chinese University of Hong Kong.

[43] Lee, Sanghoon (2010) Ability sorting and consumer city, *Journal of Urban Economics*, forthcoming.

[44] Lee, Sanghoon and Qiang Li (2009) Uneven landscape and city size distribution. Processed, University of British Columbia.

[45] Melitz, Marc J. (2003) The impact of trade on intra-industry reallocations and aggregate industry productivity, *Econometrica* 71(6), 1695–1725.

[46] Melitz, Marc J. and Gianmarco I.P. Ottaviano (2008) Market size, trade and productivity, *Review of Economic Studies* 75(1), 295–316.

[47] 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.

[48] Nocke, Volker (2006) A gap for me: Entrepreneurs and entry, *Journal of the European Economic Association* 4(5), 929–956.

[49] Henry G. Overman and Diego Puga (2009) Labour pooling as a source of agglomeration: An empirical investigation. In: Edward L. Glaeser (ed.) *The Economics of Agglomeration*. Cambridge, MA: National Bureau of Economic Research.

[50] Puga, Diego (2010) The magnitude and causes of agglomeration economies, *Journal of Regional Science* 50(1), 203–219.

[51] 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.

[52] Rossi-Hansberg, Esteban and Mark L. J. Wright (2007) Urban structure and growth, *Review of Economic Studies* 74(2), 597–624.

[53] Sattinger, Michael (1993) Assignments models of the distribution of earnings, *Journal of Economic Literature* 31(2), 831–880.

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

[55] Soo, Kwok Tong (2005) Zipf's Law for cities: a cross-country investigation, *Regional Science and Urban Economics* 35(3), 239–263.

[56] Syverson, Chad (2004) Market structure and productivity: A concrete example, *Journal of Political Economy* 112(6), 1181–1222.

[57] Wheeler, Christopher H. (2001) Search, sorting, and urban agglomeration, *Journal of Labor Economics* 19(4), 879–899.

Appendix A. Second-order conditions for the equilibrium with talent-homogeneous cities

Let $\mathbb{E}V(t_c, t) \equiv \mathbb{E}V_c(t)$ and have $\tilde{\xi}$ such that $\tilde{\xi} = \xi$ at the optimum and $\tilde{\xi} = \frac{1+\gamma}{1+\varepsilon}\xi$ at the equilibrium. Using $L = (\tilde{\xi}t_c)^{\frac{1}{\gamma-\varepsilon}}$, which holds both at the equilibrium and at the optimum by definition of $\tilde{\xi}$, the consumption of the numéraire with talent-homogeneous cities is given by:

$$\mathbb{E}V_c(t_c) \equiv \theta\left(\frac{\gamma\xi}{\varepsilon\tilde{\xi}} - 1\right)(\tilde{\xi}t_c)^{\frac{\gamma}{\gamma-\varepsilon}}.$$

This expression is positive if and only if $\varepsilon < \gamma$, which we assume to hold true for cities to be of finite size. Furthermore, since the elasticities γ and ε are empirically much smaller than unity (see section 5), we impose the additional restriction $\gamma < 1$.

Letting $s' \equiv \min\{\overline{s}, \max\{St_c/t, \underline{s}\}\}$ and using (20), a sufficient condition for the talent-homogeneous case to be an equilibrium is that

$$\begin{aligned}
\mathbb{E}V_c(t) &= \frac{1}{1+\varepsilon}\left[\varepsilon\phi\left(\tilde{\xi}t_c\right)^{\frac{1}{\gamma-\varepsilon}}\right]^{\varepsilon} St_c\left[G_s(s') + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s)\right] - \theta(\tilde{\xi}t_c)^{\frac{\gamma}{\gamma-\varepsilon}} \\
&= \frac{\theta\gamma\left(\tilde{\xi}\right)^{\frac{\gamma}{\gamma-\varepsilon}}\xi}{\phi\varepsilon(1+\varepsilon)\tilde{\xi}}\left\{1 - \frac{\phi\varepsilon(1+\varepsilon)\tilde{\xi}}{\gamma\xi} + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) - \int_{s'}^{\overline{s}}\mathrm{d}G_s(s)\right\} t_c^{\frac{\gamma}{\gamma-\varepsilon}} \qquad \text{(A.1)}
\end{aligned}$$

be quasi-concave in t_c for all $(t, t_c) \in T \times T$. Imposing quasi-concavity on (A.1) yields, however, an expression that is so unwieldy that it is impossible to see how restrictive this condition would be. We follow an alternative route by imposing a more stringent sufficient condition than quasi-concavity. As will be shown, the resulting set of parameter values that satisfy this overly strong condition is non-empty, thereby implying that the set of parameters that satisfy quasi-concavity is non-empty too.

We first derive conditions under which $\mathbb{E}V_c(t)$ is concave in t_c for all $(t, t_c) \in T \times T$. Taking the first derivative of (A.1) with respect to t_c (and disregarding the multiplicative constant in front of the curly bracket) yields

$$\begin{aligned}
\frac{\partial \mathbb{E}V_c}{\partial t_c} &= \frac{\gamma}{\gamma-\varepsilon}t_c^{\frac{\gamma}{\gamma-\varepsilon}-1}\left\{1 - \frac{\phi\varepsilon(1+\varepsilon)\tilde{\xi}}{\gamma\xi} + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) - \int_{s'}^{\overline{s}}\mathrm{d}G_s(s)\right\} \\
&\quad + t_c^{\frac{\gamma}{\gamma-\varepsilon}}\left\{-\frac{1}{\varepsilon t_c}\left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) - \left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial s'}{\partial t_c} + g_s(s')\frac{\partial s'}{\partial t_c}\right\}.
\end{aligned}$$

Tedious computations show that the second derivative is given by

$$\begin{aligned}
\frac{\partial^2 \mathbb{E}V_c}{\partial t_c^2} &= \frac{\gamma\varepsilon}{(\gamma-\varepsilon)^2}t_c^{\frac{\gamma}{\gamma-\varepsilon}-2}\left\{1 - \frac{\phi\varepsilon(1+\varepsilon)\tilde{\xi}}{\gamma\xi} + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) - \int_{s'}^{\overline{s}}\mathrm{d}G_s(s)\right\} \qquad \text{(A.2)} \\
&\quad + \frac{2\gamma}{\gamma-\varepsilon}t_c^{\frac{\gamma}{\gamma-\varepsilon}-1}\left\{-\frac{1}{\varepsilon t_c}\left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) - \left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial s'}{\partial t_c} + g_s(s')\frac{\partial s'}{\partial t_c}\right\} \qquad \text{(A.3)} \\
&\quad + t_c^{\frac{\gamma}{\gamma-\varepsilon}}\left\{\frac{1+\varepsilon}{\varepsilon^2 t_c^2}\left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s) + \frac{1}{\varepsilon t_c}\left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial s'}{\partial t_c}\right. \\
&\quad + \frac{1}{\varepsilon t_c}\left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial s'}{\partial t_c} - \frac{1}{\varepsilon s'}\left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}}\left(\frac{\partial s'}{\partial t_c}\right)^2 g_s(s') - \left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}} g_s'(s')\left(\frac{\partial s'}{\partial t_c}\right)^2 \qquad \text{(A.4)} \\
&\quad \left. - \left(\frac{ts'}{t_cS}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial^2 s'}{\partial t_c^2} + g_s'(s')\left(\frac{\partial s'}{\partial t_c}\right)^2 + \frac{\partial^2 s'}{\partial t_c^2} g_s(s')\right\},
\end{aligned}$$

where (A.2), (A.3) and (A.4) denote the first, the second, and the third term of this second derivative, respectively. To simplify this derivative, note first that (A.2) can be rewritten as:

$$\frac{\gamma\varepsilon}{(\gamma-\varepsilon)^2}t_c^{\frac{\gamma}{\gamma-\varepsilon}-2}\left\{G_s(s') - \frac{\phi\varepsilon(1+\varepsilon)\tilde{\xi}}{\gamma\xi} + \left(\frac{t}{t_c}\right)^{\frac{1}{\varepsilon}}\int_{s'}^{\overline{s}}\left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}}\mathrm{d}G_s(s)\right\}. \qquad \text{(A.5)}$$

Observe next that in (A.3), the second and third term inside the curly bracket are such that:

$$-\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial s'}{\partial t_c} + g_s(s')\frac{\partial s'}{\partial t_c} = -g_s(s')\frac{\partial s'}{\partial t_c}\left[1-\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\right] = 0,$$

where the last equality follows from the definition of s'. Hence, (A.3) boils down to:

$$-\frac{2\gamma}{\varepsilon(\gamma-\varepsilon)} t_c^{\frac{\gamma}{\gamma-\varepsilon}-2} \int_{s'}^{\overline{s}} \left(\frac{ts}{t_c S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s). \tag{A.6}$$

Last, some straightforward rearrangements allow us to make the following terms appear in (A.4):

$$g_s'(s')\left(\frac{\partial s'}{\partial t_c}\right)^2\left[1-\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\right] = 0 \quad \text{and} \quad g_s(s')\frac{\partial^2 s'}{\partial t_c^2}\left[1-\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\right] = 0$$

where the equalities again follow from the definition of s'. Hence, the last four terms in (A.4) cancel out. Regrouping, we then can re-express (A.4) as follows:

$$t_c^{\frac{\gamma}{\gamma-\varepsilon}-2}\left\{\frac{1+\varepsilon}{\varepsilon^2}\int_{s'}^{\overline{s}}\left(\frac{ts}{t_c S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) + \frac{t_c}{\varepsilon}\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}} g_s(s')\frac{\partial s'}{\partial t_c}\left[2-\frac{t_c}{s'}\frac{\partial s'}{\partial t_c}\right]\right\}. \tag{A.7}$$

Adding expressions (A.6), (A.5) and (A.7), and rearranging, we then finally obtain:

$$\begin{aligned}
\frac{\partial^2 \mathbb{E}V_c}{\partial t_c^2} = & -\left(\frac{\varepsilon}{\gamma-\varepsilon}\right)^2 t_c^{\frac{\gamma}{\gamma-\varepsilon}-2}\left[(1+\varepsilon)\phi\frac{\tilde{\xi}}{\xi} - \frac{\gamma}{\varepsilon}G_s(s')\right] \\
& - \frac{[\varepsilon-\gamma(1-\varepsilon)][\gamma-\varepsilon(1+\varepsilon)]}{[\varepsilon(\gamma-\varepsilon)]^2} t_c^{\frac{\gamma}{\gamma-\varepsilon}-2}\int_{s'}^{\overline{s}}\left(\frac{ts}{t_c S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \\
& + \frac{1}{\varepsilon} t_c^{\frac{\gamma}{\gamma-\varepsilon}-2} g_s(s')s'\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\frac{t_c}{s'}\frac{\partial s'}{\partial t_c}.
\end{aligned} \tag{A.8}$$

The *necessary second-order condition* (NSOC) requires this second derivative to be negative at $t = t_c$. Evaluating the NSOC at equilibrium, using $\phi = G_s(S)$, and simplifying, we obtain:

$$\left.\frac{\partial^2 \mathbb{E}V_c}{\partial t_c^2}\right|_{t=t_c} = -\frac{t_c^{\frac{\gamma}{\gamma-\varepsilon}-2}}{\varepsilon}\left[\phi\frac{\varepsilon-\gamma(1-\varepsilon)}{\gamma-\varepsilon} - g_s(s)S\right].$$

Thus, the NSOC may be expressed as

$$\frac{g_s(s)S}{G_s(S)} < \frac{\varepsilon-\gamma(1-\varepsilon)}{\gamma-\varepsilon}, \tag{A.9}$$

This implies that $\varepsilon < \gamma < \frac{\varepsilon}{1-\varepsilon}$. Note next that ϕ and S are two endogenous variables linked by (19). Thus, it is easy to find distributions of luck s that satisfy (A.9) in general. As most distributions are characterised by at least two parameters, the choice of the support gives us an additional degree of freedom.

Since general sufficient second-order conditions for arbitrary distributions of luck and talent are too unwieldy to be revealing, we establish the generic existence of a non-zero measure set of parameter values supporting talent-homogeneous cities as an equilibrium for the special case of Pareto distributions.

Proposition 10 (Equilibrium existence with Pareto distributions) *Assume that s follows a Pareto distribution with shape parameter $k > 1/\varepsilon$ on the support $\Sigma \equiv [\underline{s}, \infty)$, with $\underline{s} > 0$. Then there generically exists a non-empty set of parameter values such that the talent-homogeneous case of Proposition 7 is an equilibrium.*

Proof. If $G_s(s) = 1 - (\underline{s}/s)^k$ for $s \in \Sigma \equiv [\underline{s}, \infty)$ with $\underline{s} > 0$ and $k > 1/\varepsilon$, then $\frac{sg_s(s)}{1-G_s} = k$ for all $s \in \Sigma$. Then, the solution for ϕ and S yields

$$\phi \equiv G_s(S) = \frac{1}{\varepsilon}\int_S^\infty \left(\frac{s}{S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) = \frac{k}{k\varepsilon - 1}\left(\frac{\underline{s}}{S}\right)^k \quad \text{and} \quad \left(\frac{\underline{s}}{S}\right)^k = 1 - \phi. \tag{A.10}$$

Together, these two expressions imply $\frac{\phi}{1-\phi} = \frac{k}{k\varepsilon - 1} > \frac{1}{\varepsilon}$. Plugging this result into (A.9) implies that $\partial^2 \mathbb{E}V_c/\partial t_c^2|_{t=t_c} < 0$ if and only if

$$k < \frac{\gamma}{\gamma - \varepsilon}. \tag{A.11}$$

This simple condition is the NSOC when s is Pareto distributed. Turning now to the *sufficient second-order condition* (SSOC), let us rewrite (A.8) as

$$\frac{\partial^2 \mathbb{E}V_c}{\partial t_c^2} = -t_c^{\frac{\gamma}{\gamma-\varepsilon}-2} Z(t, t_c).$$

Then a SSOC is that $Z(t, t_c) > 0$ for all $(t, t_c) \in T \times T$. With s being Pareto distributed, $Z(t, t_c)$ is equal to

$$\begin{aligned} Z(t, t_c) &= \frac{1}{(\gamma - \varepsilon)^2}\left\{\phi(1+\gamma)\varepsilon^2 - \gamma\left[1 - \left(\frac{\underline{s}}{s'}\right)^k\right]\varepsilon \right. \\ &\left. + \frac{[\varepsilon - \gamma(1-\varepsilon)][\gamma - \varepsilon(1+\varepsilon)]}{\varepsilon^2}\frac{k}{k - \frac{1}{\varepsilon}}\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\left(\frac{\underline{s}}{s'}\right)^k\right\} - \frac{1-\phi}{\varepsilon}k, \end{aligned}$$

where the first term makes use of the definition of G_s and the second term follows from

$$\int_{s'}^\infty \left(\frac{ts}{t_c S}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) = \frac{k}{k - \frac{1}{\varepsilon}}\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\left(\frac{\underline{s}}{s'}\right)^k.$$

Using (A.10), Z can then be rewritten as

$$\begin{aligned} Z(t, t_c) &= \frac{\varepsilon}{(\gamma - \varepsilon)^2}\left\{\phi\varepsilon(1+\gamma) - \gamma\left[1 - \left(\frac{\underline{s}}{s'}\right)^k\right]\right\} \\ &+ \phi\frac{[\varepsilon - \gamma(1-\varepsilon)][\gamma - \varepsilon(1+\varepsilon)]}{\varepsilon(\gamma - \varepsilon)^2}\left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\left(\frac{S}{s'}\right)^k - \frac{1}{\varepsilon}(1-\phi)k, \end{aligned} \tag{A.12}$$

which is clearly decreasing in s' by $k > 1/\varepsilon$. Hence, the term in curly brackets is positive for all $(t, t_c) \in T \times T$ if it is positive at $(t, t_c) = (\underline{t}, \bar{t})$. Let $\kappa \equiv (\bar{t}/\underline{t})^k > 1$. Then for the case $S\bar{t}/\underline{t} > \underline{s}$, which always holds by $\kappa > 1$ and $\phi \in (0, 1)$, we have

$$\left(\frac{\underline{s}}{s'}\right)^k = \frac{1}{\kappa}\left(\frac{\underline{s}}{S}\right)^k = \frac{1-\phi}{\kappa} \quad \text{and} \quad \left(\frac{ts'}{t_c S}\right)^{\frac{1}{\varepsilon}}\left(\frac{S}{s'}\right)^k = \frac{1}{\kappa}.$$

Plugging these expressions into (A.12) provides us with the following SSOC:

$$\begin{aligned} Z(\kappa) &\equiv \frac{\varepsilon}{(\gamma-\varepsilon)^2}\left\{\phi\varepsilon(1+\gamma) - \gamma\left[1 - \frac{1-\phi}{\kappa}\right]\right\} + \phi\frac{[\varepsilon-\gamma(1-\varepsilon)]\,[\gamma-\varepsilon(1+\varepsilon)]}{\varepsilon(\gamma-\varepsilon)^2}\frac{1}{\kappa} - \frac{1-\phi}{\varepsilon}k \\ &= \frac{1}{(\gamma-\varepsilon)^2\,[k(\varepsilon+1)-1]}\left[\gamma - (\gamma-\varepsilon)k - (\gamma-\varepsilon)^2(k\varepsilon-1)\frac{k}{\varepsilon}\right] \\ &\quad + \frac{1}{\kappa}\frac{1}{(\gamma-\varepsilon)^2\,[k(\varepsilon+1)-1]}\left\{\gamma\varepsilon(k\varepsilon-1) + [\varepsilon-\gamma(1-\varepsilon)]\,[\gamma-\varepsilon(1+\varepsilon)]\frac{k}{\varepsilon}\right\} \\ &\equiv \frac{1}{(\gamma-\varepsilon)^2\,[k(\varepsilon+1)-1]}\left(A + \frac{1}{\kappa}B\right) > 0, \end{aligned}$$

Three properties of $Z(\kappa)$ are noteworthy. First,

$$Z(1) = \frac{k}{k(\varepsilon+1)-1}\left(\frac{\gamma}{\gamma-\varepsilon} - k\right) > 0.$$

This inequality holds by (A.11). It is nothing but the NSOC and so was to be expected. Second, $B > 0$ by $\varepsilon < \gamma < \varepsilon/(1-\varepsilon)$ and thus $Z(\kappa)$ is continuously decreasing in κ. Finally, $Z(\kappa)$ is continuous in ε, γ and k. Together, these properties imply that the SSOC holds around $\kappa = 1$ for all triples (ε, γ, k) that belong to some set $\psi_{\text{NSOC}} \subset \mathbb{R}_+^* \times \mathbb{R}_+^* \times (1, \infty)$, where

$$\psi_{\text{NSOC}} \equiv \left\{(\varepsilon, \gamma, k) : \quad \frac{1}{\varepsilon} < k < \frac{\gamma}{\gamma-\varepsilon} \quad \text{and} \quad \varepsilon < \gamma < \frac{\varepsilon}{1-\varepsilon}\right\}$$

is non-empty. By continuity, for any κ arbitrarily close to unity we can find a non-zero measure set of parameters (ε, γ, k) such that $Z(\kappa)$ is positive. Specifically, either of the following two cases applies: (i) if

$$A \equiv \frac{\varepsilon\,[\gamma - (\gamma-\varepsilon)k] - (\gamma-\varepsilon)^2(k\varepsilon-1)k}{\varepsilon} > 0,$$

then $Z(\kappa) > 0$ for all $\kappa \geq 1$; otherwise, (ii) if $A < 0$ there exists a $\bar{\kappa} > 1$ such that $Z(\kappa) > 0$ for all $\kappa \in [1, \bar{\kappa})$ and $\lim_{\kappa \to \bar{\kappa}} Z(\kappa) = 0$. Hence, there generically exists a non-empty set of parameter values

$$\psi_{\text{SSOC}} \equiv \{(\varepsilon, \gamma, k, \underline{s}, \kappa) : (\varepsilon, \gamma, k) \in \psi_{\text{NSOC}}, \quad \underline{s} > 1 \quad \text{and} \quad 1 < \kappa < \bar{\kappa}\}$$

such that all combinations of parameters in ψ_{SSOC} ensure that the equilibrium characterised by Proposition 6 is a global maximum for all $t \in T$ when luck is distributed as a Pareto with shape k over the set $\Sigma \equiv [\underline{s}, \infty)$. ■

Appendix B. Partition equilibria with constant selection

In this appendix, we consider equilibria with constant selection ($F_c(\underline{\varphi}_c) = \phi$ for all cities c) and imperfect talent sorting. Observe that Proposition 3 is a true single-crossing condition when selection is constant across cities, thereby ensuring the existence of a separating equilibrium. In what follows, we focus on partition equilibria only, i.e., equilibria where talent is allocated such that $[\underline{t}, t_1] \cup [t_1, t_2] \cup \ldots \cup [t_{c-1}, t_c] \cup [t_c, t_{c+1}] \cup \ldots \cup [t_{|C|-1}, \overline{t}] = T$ and $t_{c+1} > t_c$ for all $c = 1, ..., |C| - 1$. t_c is the talent indifferent across cities c and $c+1$. Because individuals optimally choose their city, t_c is such that $\mathbb{E}V_c(t_c) = \mathbb{E}V_{c+1}(t_c) > \mathbb{E}V_j(t_c)$, for all $j \neq c, c+1$. We now establish the basic properties of partition equilibria with constant selection when productivity distributions are homogeneous of degree zero (which for instance occurs when G_s and G_t are both Pareto).

Proposition 11 (Partition equilibria with constant selection) *Consider a partition of talent and assume that $F_c(\underline{\varphi}_c) \equiv F(\underline{\varphi}_c, t_{c-1}, t_c)$ is homogeneous of degree zero in all its arguments. A partition of talent such that*

$$\forall c \in C : \text{(i) } \phi_c = \phi, \quad \text{(ii) } \frac{t_{c+1}}{t_c} = \kappa, \quad \text{(iii) } \underline{\varphi}_c = S t_c \quad \text{and, (iv) } \left(\frac{L_{c+1}}{L_c}\right)^{\gamma - \varepsilon} = \kappa,$$

with $\phi \in (0, 1)$, $S \in (\underline{s}, \overline{s})$ and $\kappa > 1$, is an equilibrium candidate. In such an equilibrium, aggregate productivities are linked by $\Phi_{c+1} = \kappa^{\frac{\gamma}{\gamma - \varepsilon}} \Phi_c$.

Proof. Consider cities c and $c+1$. The equilibrium productivity cutoffs are determined by (13). In city c, with talent support $[t_{c-1}, t_c]$, it must hence be that

$$F(\underline{\varphi}_c, t_{c-1}, t_c) = \frac{1}{\varepsilon} \int_{\underline{\varphi}_c}^{t_c \overline{s}} \left(\frac{\varphi}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi, t_{c-1}, t_c).$$

Let $t_c = \kappa t_{c-1}$ and $t_{c+1} = \kappa t_c$. Then $\underline{\varphi}_{c+1} = \kappa \underline{\varphi}_c$ is the equilibrium productivity cutoff in city $c+1$. To see this note that

$$F(\underline{\varphi}_{c+1}, t_c, t_{c+1}) = F(\kappa \underline{\varphi}_c, \kappa t_{c-1}, \kappa t_c) = F(\underline{\varphi}_c, t_{c-1}, t_c) = \frac{1}{\varepsilon \kappa} \int_{\kappa \underline{\varphi}_c}^{\kappa t_c \overline{s}} \left(\frac{\varphi/\kappa}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \mathrm{d}F(\varphi/\kappa, t_{c-1}, t_c),$$

where the second equality is due to F being homogeneous of degree zero, and the last equality because $\mathrm{d}F$ is homogeneous of degree -1. It is then readily apparent that, using the change in variables $\rho = \varphi/\kappa$, that

$$F(\underline{\varphi}_{c+1}, t_c, t_{c+1}) = F(\underline{\varphi}_c, t_{c-1}, t_c) = \frac{1}{\varepsilon} \int_{\underline{\varphi}_c}^{t_c \overline{s}} \left(\frac{\rho}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \mathrm{d}F(\rho, t_{c-1}, t_c),$$

thus establishing that $\underline{\varphi}_{c+1} = \kappa \underline{\varphi}_c$. It then follows immediately that selection is constant across cities, i.e., $F(\underline{\varphi}_c, t_{c-1}, t_c) = \phi$ for all $c \in C$. This establishes (i)–(iii). To prove (iv), note first that by definition of t_c, we have $\mathbb{E}V(t_c, \underline{\varphi}_c) = \mathbb{E}V(t_c, \underline{\varphi}_{c+1})$. A substitution using equation (16) and $F_c(\underline{\varphi}_c) = \phi$ implies

$$\theta L_c^{\gamma} \left\{ \frac{1}{\theta(1+\varepsilon)} (\varepsilon\phi)^{\varepsilon} L_c^{\varepsilon-\gamma} \underline{\varphi}_c \left[G_s\left(\frac{\underline{\varphi}_c}{t_c}\right) + \int_{\underline{\varphi}_c/t_c}^{\overline{s}} \left(\frac{s t_c}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - 1 \right\}$$

$$= \theta L_{c+1}^{\gamma} \left\{ \frac{1}{\theta(1+\varepsilon)} (\varepsilon\phi)^{\varepsilon} L_{c+1}^{\varepsilon-\gamma} \underline{\varphi}_{c+1} \left[G_s\left(\frac{\underline{\varphi}_{c+1}}{t_c}\right) + \int_{\underline{\varphi}_{c+1}/t_c}^{\overline{s}} \left(\frac{s t_c}{\underline{\varphi}_{c+1}}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - 1 \right\}.$$

Using (ii) and (iii), it must also be true that $\mathbb{E}V(t_c, \underline{\varphi}_c) = \mathbb{E}V(\kappa t_{c-1}, \kappa\underline{\varphi}_{c-1})$ and $\mathbb{E}V(t_c, \underline{\varphi}_{c+1}) = \mathbb{E}V(\kappa t_{c-1}, \kappa\underline{\varphi}_c)$, which implies that

$$\theta L_c^{\gamma} \left\{ \frac{1}{\theta(1+\varepsilon)} (\varepsilon\phi)^{\varepsilon} L_c^{\varepsilon-\gamma} \kappa\underline{\varphi}_{c-1} \left[G_s\left(\frac{\underline{\varphi}_{c-1}}{t_{c-1}}\right) + \int_{\underline{\varphi}_{c-1}/t_{c-1}}^{\overline{s}} \left(\frac{s t_{c-1}}{\underline{\varphi}_{c-1}}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - 1 \right\}$$

$$= \theta L_{c+1}^{\gamma} \left\{ \frac{1}{\theta(1+\varepsilon)} (\varepsilon\phi)^{\varepsilon} L_{c+1}^{\varepsilon-\gamma} \kappa\underline{\varphi}_c \left[G_s\left(\frac{\underline{\varphi}_c}{t_{c-1}}\right) + \int_{\underline{\varphi}_c/t_{c-1}}^{\overline{s}} \left(\frac{s t_{c-1}}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - 1 \right\}.$$

By definition, t_{c-1} must also obey $\mathbb{E}V(t_{c-1}, \underline{\varphi}_{c-1}) = \mathbb{E}V(t_{c-1}, \underline{\varphi}_c)$, thus implying that

$$\theta L_{c-1}^{\gamma} \left\{ \frac{1}{\theta(1+\varepsilon)} (\varepsilon\phi)^{\varepsilon} L_{c-1}^{\varepsilon-\gamma} \underline{\varphi}_{c-1} \left[G_s\left(\frac{\underline{\varphi}_{c-1}}{t_{c-1}}\right) + \int_{\underline{\varphi}_{c-1}/t_{c-1}}^{\overline{s}} \left(\frac{s t_{c-1}}{\underline{\varphi}_{c-1}}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - 1 \right\}$$

$$= \theta L_c^{\gamma} \left\{ \frac{1}{\theta(1+\varepsilon)} (\varepsilon\phi)^{\varepsilon} L_c^{\varepsilon-\gamma} \underline{\varphi}_c \left[G_s\left(\frac{\underline{\varphi}_c}{t_{c-1}}\right) + \int_{\underline{\varphi}_c/t_{c-1}}^{\overline{s}} \left(\frac{s t_{c-1}}{\underline{\varphi}_c}\right)^{\frac{1}{\varepsilon}} \mathrm{d}G_s(s) \right] - 1 \right\}.$$

The two expressions are compatible if $L_{c+1}/L_c = L_c/L_{c-1}$ and $(L_c/L_{c-1})^{\gamma-\varepsilon} = \kappa$, that is, if condition (iv) holds. ■

Several comments are in order. First, observe that in a partition equilibrium, we have $\underline{\varphi}_{c+1} > \underline{\varphi}_c$ for all $c \in C$ by Proposition 11. Furthermore, using (16) and because selection is constant, more talented individuals prefer city $c+1$ to city c if and only if:

$$\left(\frac{L_{c+1}}{L_c}\right)^{\varepsilon} \left(\frac{\underline{\varphi}_{c+1}}{\underline{\varphi}_c}\right)^{1-\frac{1}{\varepsilon}} \frac{\int_{\underline{\varphi}_{c+1}/t_c}^{\infty} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s}{\int_{\underline{\varphi}_c/t_c}^{\infty} s^{\frac{1}{\varepsilon}} \mathrm{d}G_s} > 1. \tag{B.1}$$

Since $(L_{c+1}/L_c)^{\varepsilon} = \kappa^{\varepsilon/(\gamma-\varepsilon)}$ and $\underline{\varphi}_{c+1}/\underline{\varphi}_c = \kappa$ for some $\kappa > 1$ by Proposition 11, (B.1) reduces to $1 < \kappa^{\frac{\varepsilon(\gamma-1)-\gamma}{\varepsilon(\gamma-\varepsilon)}}$, thus requiring that $\gamma < \frac{\varepsilon}{1-\varepsilon}$ as in the necessary second-order condition derived in Appendix 7. When this condition is violated, partition equilibria with constant selection cannot occur. This leaves us either with: (i) symmetry; or (ii) a partition with different selection rates (i.e., a different ϕ_c for each c). We discuss the former in section 4.1 and the latter case in section 6.

Second, what is the distribution of population across cities and does the rank-size rule hold? Starting with the full population condition, let $|C|$ denote the equilibrium total number of city types (the cardinal of C), and let n_c denote the equilibrium number of cities of type c. By definition, we have $\frac{n_c L_c}{\Lambda} = G_t(t_c) - G_t(t_{c-1})$, with Λ the total population. Assume that both talent and luck follow truncated power laws with shape parameters m and k and with supports $T = [\underline{t}, \overline{t}]$ and $S = [\underline{s}, \overline{s}]$, respectively. In that case,

$$\frac{n_{c+1} L_{c+1}}{n_c L_c} = \frac{t_c^{-m} - t_{c+1}^{-m}}{t_{c-1}^{-m} - t_c^{-m}}.$$

Using (ii) and (iv) and rearranging yields $n_{c+1}/n_c = \kappa^{-\left(\frac{1}{\gamma-\varepsilon}+m\right)}$ or $n_c = n_{|C|}\kappa^{\left(\frac{1}{\gamma-\varepsilon}+m\right)(|C|-c)}$, where $n_{|C|}$ is the number of the most talented cities which we normalise to unity without loss of generality. Observe that the number of cities naturally decreases as we move up in the distribution of talent. We may express the size of type-c cities as

$$L_c = L_{|C|}\kappa^{-\frac{1}{\gamma-\varepsilon}(|C|-c)}. \tag{B.2}$$

The rank r_c of a type-c city is the number of cities larger than cities of type c plus one, i.e.,

$$r_c \equiv \begin{cases} 1 + \sum_{i=c+1}^{|C|} n_i & \text{if} \quad c = 1, 2, \ldots, |C|-1 \\ 1 & \text{if} \quad c = |C|. \end{cases}$$

For $c \neq |C|$, and since $\kappa^{-\left(\frac{1}{\gamma-\epsilon}+m\right)} < 1$, this may be rewritten as

$$r_c \;=\; 1 + \kappa^{\left(\frac{1}{\gamma-\varepsilon}+m\right)|C|} \sum_{i=c+1}^{|C|} \kappa^{-\left(\frac{1}{\gamma-\varepsilon}+m\right)i} = 1 + \frac{1}{\kappa^{\frac{1}{\gamma-\varepsilon}+m} - 1}\left[\left(\frac{L_c}{L_{|C|}}\right)^{-[1+m(\gamma-\varepsilon)]} - 1\right],$$

where the last equality follows from (B.2). Rearranging and taking logarithms yields $\ln(r_c - 1) \approx \alpha_0 + \beta_0 \ln L_c$, where $\alpha_0 \equiv \ln(L_{|C|}^{1+m(\gamma-\varepsilon)}) - \ln(\kappa^{\frac{1}{\gamma-\varepsilon}+m} - 1)$ and $\beta_0 \equiv -1 - m(\gamma-\varepsilon)$. The coefficient β_0 is again approximatively Zipf's law, as in the case with perfect sorting developed in Section 4.

Appendix C. Optimal talent allocation without selection

In this appendix, we solve for the optimal allocation of talent across cities. Specifically, there is no random and heterogeneous luck, that is, normalising $s = 1$, talent and productivity as entrepreneur are the same thing: $\varphi = t$. Despite this simplification, note that the selection problem is not trivial: the planner still need to allocate individuals across occupations (worker or entrepreneur). The only difference with the framework in the text is that selection and sorting (the allocation of individuals across space) occur simultaneously.

Another difference (and this one not substantial) is that we work here with a discrete space and a discrete set of talents. To simplify the expressions below, we introduce a bit of extra notation.

Let L_t denote the stock of people with talent t in the economy; h_{tc} the mass of individuals with talent t allocated to city c as *entrepreneurs*; l_c the mass of *workers* in city c; and $L_c \equiv l_c + \Sigma_t h_{tc}$, the number of people in city c. Also, $\phi_c \equiv l_c/L_c$ and $\alpha_{tc} \equiv h_{tc}/L_c$ denote the fractions of workers and entrepreneurs with talent t in city c, respectively. We use $\varphi_c \equiv \left[(1-\phi_c)\sum_t \alpha_{tc} t^{1/\varepsilon}\right]^\varepsilon$ for the *average* productivity in city c and n_c for the number of cities of type c.

Planner's problem. The planner seeks to maximise output net of urban cost, in that aim creating cities of different types and allocating individuals among those cities ('sorting') as well as across occupations ('selection'). Obviously, output is maximised if the most talented individuals become entrepreneurs. Thus, the planner chooses a τ such that individuals with talent $t < \tau$ are workers and individuals with $t \geq \tau$ are entrepreneurs.

We may thus write the planer's program as

$$\max_{\tau,\{l_c\}_c,\{h_{t,c}\}_{tc}} \Omega \equiv \sum_c \left(Y_c - \theta L_c^{\gamma+1}\right) n_c \tag{C.1}$$

$$\begin{aligned} \text{subject to:} \qquad Y_c &= l_c \left(\sum_t h_{tc} t^{1/\varepsilon}\right)^\varepsilon = \phi_c(1-\phi_c)^\varepsilon L_c \varphi_c && \text{(C.2)} \\ L_c &= l_c + \sum_t h_{t,c} && \text{(C.3)} \\ h_t &= \sum_c h_{tc} n_c && \text{(C.4)} \\ \Lambda G_t(\tau) &= \sum_c l_c n_c. && \text{(C.5)} \end{aligned}$$

The constraints are as follow. Equation (C.2) is the value of city output once the local markets clear (the second equality follows from the definitions above). Equation (C.3) is the composition of city population. Equation (C.4) is the full-employment condition for entrepreneurs. Finally, equation (C.5) is the full-employment condition for workers (and the full population is redundant).

Planner's solution. We may substitute (C.2) and (C.3) directly into the objective function; let η_t denote the Lagrange multiplier associated with (C.4) and let ω denote the Lagrange multiplier associated with (C.5). The first order conditions (FOCs) of this program are, for all cities of type c such that $n_c, l_c > 0$ and all $t \in T$ (the SOCs are satisfied under positive assortative matching; more on this below):

$$\begin{aligned} \frac{\partial \Omega}{\partial l_c} &= \left[-\omega + \frac{Y_c}{l_c} - (1+\gamma)\theta L_c^\gamma\right] n_c = 0 \\ \frac{\partial \Omega}{\partial h_{tc}} &= \left[-\eta_t + \varepsilon l_c \left(\sum_{\tilde{t}} h_{\tilde{t}c} \tilde{t}^{1/\varepsilon}\right)^{\varepsilon-1} t^{1/\varepsilon} - (1+\gamma)\theta L_c^\gamma\right] n_c \leq 0, \quad h_{tc} \geq 0, \quad \frac{\partial \Omega}{\partial h_{tc}} h_{tc} = 0 \\ \frac{\partial \Omega}{\partial n_c} &= \left(Y_c - \theta L_c^{\gamma+1}\right) - \left(\omega l_c + \sum_t h_{tc}\eta_t\right) = 0. && \text{(C.6)} \end{aligned}$$

Multiplying the first FOC by l_c, the second by h_{tc}, summing and using the third implies $\varepsilon Y_c/L_c = \gamma\theta L_c^{\gamma}$ so that net city output is $Y_c - \theta L_c^{\gamma+1} = \frac{\gamma-\varepsilon}{\varepsilon}\theta L_c^{\gamma+1} = \frac{\gamma-\varepsilon}{\gamma}Y_c$ (the first right-hand side gives net income as a function of city size; the second one, as a function of gross income). Using the expressions for city income (C.2) and the composition of city population (C.3) yields the optimal relationship between city composition and city size: $\varepsilon\phi_c(1-\phi_c)^{\varepsilon}\varphi_c = \gamma\theta L_c^{\gamma-\varepsilon}$. In words, larger cities have more productive entrepreneurs on average.

Now, as may be clear from the first two FOCs, ω and η_t are the shadow 'real prices' of labour and entrepreneurial talent (in the sense that they are net of urban costs). Starting with the second FOC in order to derive the real rewards of talent t in any city where such talent settle, total entrepreneurial profit net of urban costs are equal to $\sum_t h_t\eta_t = \theta L_c^{\gamma+1}\left[\phi_c(1+\gamma)-1\right]$. Labour is homogenous; using the expression above to solve for ω in the first FOC yields $\omega = \frac{1}{\phi_c}\theta L_c^{\gamma}\left[\frac{\gamma}{\varepsilon} - \phi_c(1+\gamma)\right]$. A natural solution to the expression for ω is $\omega = 0$. As is standard in assignment problems, individuals at the bottom of the (talent) distribution get their outside option, which is zero consumption. Using the expression for ω above, this pins down ϕ_c and τ. For all cities we have then $\phi_c = \phi$, where

$$\phi \equiv \frac{1}{1+\gamma}\frac{\gamma}{\varepsilon} \equiv G_t(\tau). \tag{C.7}$$

That is, the share of workers is the same across cities (and thus the same as in the economy as a whole).[23] Loosely speaking, 'selection' is constant at the optimum. Being a share, ϕ must belong to the unit interval. This is the case iff

$$\gamma < \frac{\varepsilon}{1-\varepsilon}. \tag{C.8}$$

We characterise below the optimal allocation when this condition fails. Substituting the common ϕ into the productivity-size relationship, we find

$$\varphi_c = (1+\gamma)^{1+\varepsilon}\left[\frac{\varepsilon}{(1+\gamma)\varepsilon-\gamma}\right]^{\varepsilon}\theta L_c^{\gamma-\varepsilon}. \tag{C.9}$$

This implies that the elasticity of city size with respect to productivity is $\frac{1}{\gamma-\varepsilon}$.

Substituting the common ϕ into the expression for total net profit, we get $\sum_t h_t\eta_t = \frac{\gamma-\varepsilon}{\varepsilon}\theta L^{\gamma+1} = Y_c - \theta L_c^{\gamma+1}$. This confirms that workers get zero consumption. Thus, the nominal wage is a simple function of city size: $w_c = \theta L_c^{\gamma}$. We may also write each talent's net reward as $\eta_t = (1+\gamma)\theta L_c^{\gamma}\left[\frac{\varepsilon\gamma}{(1+\gamma)\varepsilon-\gamma}\left(\frac{t}{\varphi_c}\right)^{1/\varepsilon} - 1\right]$ so that only the people with a talent

$$t \geq \varphi_c\left(1 - \frac{\gamma-\varepsilon}{\varepsilon\gamma}\right)^{\varepsilon} \tag{C.10}$$

make non-negative profit in city c. This establishes that the qualitative relationship between average productivity and city size also holds for individual productivity/talent. In entrepreneurial-talent-homogenous cities (so that $\varphi_c = t$), this is equal to $\eta_t = \left[1 - \frac{1}{1+\gamma}\frac{\gamma}{\varepsilon}\right]^{\frac{\varepsilon}{\gamma-\varepsilon}}\left[(1+\gamma)\theta\right]^{-\frac{1}{\gamma-\varepsilon}+1}\frac{\gamma-\varepsilon}{(1+\gamma)\varepsilon-\gamma}t^{\frac{1}{\gamma-\varepsilon}}$.

[23] The factor real rewards $\sum_t h_t\eta_t$ and ω are both non-negative in this case, as they should be.

Positive assortative matching (SOC). There is positive assortative matching between talent and city productivity (i.e., $\eta(t,\varphi)$ is log supermodular in t and φ) iff

$$\frac{\partial^2 \eta(t,\varphi)}{\partial t \partial \varphi} > 0. \tag{C.11}$$

This single-crossing condition corresponds to the condition in Proposition 3 in the text. Here, it is *unconditional*. Thus, if satisfied, this is a sufficient condition for the solution to (C.6) to be a maximum to the problem in (C.1). Using the expression for η_t above, we get $\frac{\partial \ln \eta}{\partial t} = t^{-1+1/\varepsilon} \varphi^{\frac{\gamma}{\gamma-\varepsilon}-\frac{1}{\varepsilon}}$ so that the single-crossing condition above holds iff the exponent of φ is positive, that is, iff (C.8) holds, which is also the condition that ensures that ϕ belongs to the unit interval.

Symmetric allocation. To complete the analysis, we now show that the optimal allocation when (C.8) fails involves identical cities, all endowed with the same composition of individuals and all of the same size. To see this, note that the non-negative profit condition (C.10) is violated iff (C.8) fails. Could negative assortative matching be a solution? In that case, only the individuals with a talent less than the the city average should join that city. But not everybody can have a productivity below the average. Thus, negative assortative matching would yield a contradiction. To understand why, note that in our case, unlike in traditional assignment problems, φ_c is an endogenous variable. So, when the inequality above holds, the symmetric configuration is optimal. Summarising, we have shown:

Proposition 12 (Optimal allocation in the absence of random and heterogeneous luck) *If* (C.8) *holds, then at the optimal allocation (i) all cities have the same fraction of workers ϕ given by* (C.7)*; (ii) larger cities have more talented entrepreneurs by* (C.9) *and* (C.11)*; (iii) the elasticity of city productivity with respect to city size is $\gamma - \varepsilon$ by* (C.9). *If* (C.8) *fails, then the optimal allocation is a system of symmetric cities.*

By implication of (ii), at the limit of a continuum of cities, there is only one type of entrepreneurial talent per city and this is strictly increasing in city size; thus the relationship illustrated by figure 2 also holds at the optimal allocation. More generally, the qualitative properties of the optimal allocation of this simplified model (as summarised in the proposition above) and those of the talent-homogeneous equilibrium described in the text are *identical*. As to the central quantitative implications, the elasticity of L_c with respect to t_c is the same: it is equal to $\frac{1}{\gamma-\varepsilon}$ in both cases. The same key condition (C.8) also needs to hold in both cases.