Productive Cities: Sorting, Selection, and Agglomeration

Kristian Behrens

Université du Québec à Montréal

Gilles Duranton

University of Pennsylvania

Frédéric Robert-Nicoud

Université de Genève

Large cities produce more output per capita than small cities. This higher productivity 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 stylized facts about sorting, agglomeration, and selection in cities. It also generates Zipf's law for cities under empirically plausible parameter values. Finally, it provides a useful framework within which to reinterpret extant empirical evidence.

I. Introduction

Output per capita is higher in larger cities. For instance, across 276 US metropolitan areas in 2000, the measured elasticity of average city earnings with respect to city population is 8.2 percent. This paper proposes a

We thank Fabien Candau, Arnaud Costinot, Xavier Gabaix, Ed Glaeser, Laurent Gobillon, Delfim Gomes Neto, Tom Holmes, Yannis Ioannides, Sam Kortum, Sanghoon Lee, Diego Puga, Jacques Thisse, Jonathan Vogel, many conference and seminar participants,

[*Journal of Political Economy*, 2014, vol. 122, no. 3]

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.[1]

Integrating these three explanations of the urban premium into a theoretical framework in which cities are determined endogenously is important for three reasons. First, it yields a better understanding of how sorting, selection, and agglomeration interact. Our results emphasize 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 wage premium attracts more individuals and makes these cities larger, thereby strengthening agglomeration economies.

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

---

and especially Vernon Henderson, Esteban Rossi-Hansberg, and two anonymous referees for very helpful discussions and comments. Behrens is holder of the Canada Research Chair in Regional Impacts of Globalization. Financial support from the CRC Program of the Social Sciences and Humanities Research Council of Canada, as well as from the Standard Research Grants Program, is gratefully acknowledged. Behrens further gratefully acknowledges financial support from Fonds de Recherche Société et Culture Québec (grant NP-127178). Robert-Nicoud was a visiting scholar at Princeton University when the first draft of this paper was written; the hospitality of this institution is gratefully acknowledged. Any remaining errors are ours. Data are available as supplementary material online.

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

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

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

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

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

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

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

In more recent work, Eeckhout, Pinheiro, and Schmidheiny (2014; in this issue) show how a mix of skills can occur in cities when skills complement each other. More specifically, they can generate an overrepresentation of high- and low-skill workers in the largest cities when skill complementarities are stronger between more extreme skills. Finally, Davis and Dingel (2012) develop an original model of learning in cities in which more skilled individuals learn more from each other. In equilibrium, the most skilled individuals sort into the largest city, where they devote more time to learning from each other, while less skilled individuals are better off in the small city, where the cost of living is much lower.

We know of only two papers on selection in cities. Behrens and Robert-Nicoud (forthcoming) propose a multiregion 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 urbanization also generates inequalities. Gennaioli et al. (2013) use a simple framework featuring agglomeration, sorting, and selection to assess empirically the effect of human capital on regional development. We return to their findings later in the paper.

A second innovation of our model is to generate Zipf's law within a static setting. Zipf's law and the size distribution of cities have attracted much attention recently. In random growth models, the population of a city reflects its balance of past shocks (see Gabaix and Ioannides [2004] for a review). Our approach is radically different. It builds on a static model of cities. In equilibrium, the population of a city depends on the productivity of its entrepreneurs, magnified by the trade-off between

[4] 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. Baldwin and Okubo (2006) develop a model with immobile workers in which ex ante identical firms can relocate at a cost after receiving their productivity draw. This timing leads to the relocation of the most productive firms from the small market to the large one and incomplete productivity sorting.

agglomeration economies and urban costs. More specifically, city population is a power function of the talent of its residents in which the power is inversely related to the difference between the intensity of agglomeration economies and that of urban costs. When this difference is small, as is the case in the data, small productivity differences caused by sorting lead to large differences in city population sizes, and the resulting size distribution of cities is approximately Zipf.[5]

In what follows, Section II presents the model. Section III solves for its equilibrium taking the distribution of population as given. Section IV solves for location choices, and Section V derives our results about the size distribution of cities. Section VI proposes two extensions of our model, and Section VII discusses its quantitative implications. Finally, Section VIII presents conclusions.

II. The Model

There is a continuum of individuals of mass Λ in the economy, each choosing a location and an occupation. Individuals are identical except for their "talent," t, and their "serendipity," s. There is also a continuum of homogeneous sites that can be used as cities. The number of cities, their population size, and their composition are endogenous.

Each individual initially knows her talent and chooses where to locate. Upon her moving to a city, serendipity occurs. We model it as a draw of luck. The product of an individual's talent and serendipity determines her productivity: $\varphi \equiv t \times s$. "Serendipity" subsumes many local interactions that are uncertain and affect productivity such as being acquainted with the right people at the right time. Knowing her productivity, each individual then selects into an occupation, entrepreneur or worker. An entrepreneur sets up a firm that produces with productivity φ a variety of differentiated intermediate goods using labor. A worker supplies φ^a efficiency units of labor, with $a \geq 0$.[6]

Empirically, there are frictions to mobility. In our static model we formalize these frictions in a parsimonious and tractable way by assuming free mobility before serendipity occurs and prohibitive mobility costs afterward.[7] The knowledge of their talent allows individuals to sort across cities. The full revelation of their productivity after choosing a city leads to their selection into occupations. That is, our two-step revelation process

[5] We know of two other papers that generate Zipf's law from a static model. The argument of Lee and Li (2013) is the static equivalent of random growth models in which population is determined by the multiplicative aggregation of many randomly distributed local characteristics. Hsu (2012) relies on central place theory.

[6] We allow for heterogeneous worker productivity while most models in the traditions of Dixit and Stiglitz (1977) or Lucas (1978), which we relate to, assume homogeneous labor.

[7] In Sec. VI, we develop an extension in which individuals can obtain additional draws of luck at a cost and show that this generalization does not affect the main properties of our model.

enables us to consider both the spatial sorting of individuals and the productivity selection of firms. Selection without sorting would lead all cities to be symmetric in equilibrium. Sorting without selection would imply that all firms in any one city have the same productivity. Both predictions are counterfactual.

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

Individuals consume two goods: a final good and land. For simplicity, they require one unit of land for accommodation and do not increase their utility by consuming more land. They are also risk neutral so that their utility can be taken to be linear in final good consumption.

To produce the final good, competitive final producers in each city use locally produced differentiated intermediate inputs, which enter into their technology with constant elasticity of substitution $1 + 1/\varepsilon$, with $\varepsilon > 0$. Aggregate output in city c is given by

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

where $x_c(i)$ is the amount of variety i used, and Ω_c is the endogenously determined set of varieties of intermediate inputs produced in city c. Without loss of generality and for simplicity, we make the final good Y freely tradable to 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 that employs labor 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 productivity as in Lucas (1978) and Melitz (2003).[8] Output of variety i is

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

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

[8] As in Lucas (1978), firms differ only in the productivity of their entrepreneur. Because differentiated varieties are imperfect substitutes, we do not need to impose limits on entrepreneurial span of control for firms to remain of finite size. We nonetheless consider this extension in Sec. VII below. As in Melitz (2003), we embed heterogeneous firms in a constant elasticity of substitution demand system. In contrast to Melitz's study, there is no sunk cost to create a firm and receive a productivity draw: individuals know their productivity when they decide whether or not to start a firm.

Turning to the urban structure of our model, we assume that each resident of a city of population L pays θL^{γ} as an urban cost to reside in that city. In online Appendix F, we develop microeconomic foundations that justify this functional form. These foundations rely on a standard monocentric urban framework, where an increase in population leads to greater commuting costs. For cities to remain of finite size in equilibrium, we require γ, the elasticity of urban costs, to be larger than ε, which turns out to be the equilibrium elasticity of agglomeration economies.

To use the terminology of Henderson and Becker (2000), cities arise under "self-organization"; that is, they are the outcome of the mutually compatible optimal choices of a continuum of individuals. The talent composition and population size of a city $c \in C \equiv [0, \bar{c}]$ are endogenously determined. The population of each city c, L_c, is given by

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

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

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

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

III. Selection and Agglomeration

In equilibrium, each individual optimally chooses a city on the basis of her talent. After serendipity occurs and she learns her productivity, she optimally chooses an occupation, worker or entrepreneur. Entrepreneurs then maximize their profit with respect to the price of their variety. Markets for intermediate goods, final goods, and labor clear, and the population adding-up constraints are satisfied.

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

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

$$x(i) = \left[\frac{p(i)}{\mathbb{P}}\right]^{-(1+\varepsilon)/\varepsilon} \frac{Y}{\mathbb{P}}, \tag{5}$$

where $\mathbb{P} \equiv [\int_\Omega p(j)^{-(1/\varepsilon)} dj]^{-\varepsilon}$ is the appropriate price index. It is immediate from equation (5) that the own-price elasticity of demand is $-(1+\varepsilon)/\varepsilon$. Hence, the profit-maximizing price for each intermediate displays a constant markup over marginal cost:

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

where w is the wage per efficiency unit of labor. This expression allows us to rewrite the demand (5) as

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

where $\Phi \equiv [\int_\Omega \varphi(j)^{1/\varepsilon} dj]^{\varepsilon}$ is the appropriate 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 market crowding effect. Using expressions (6) and (7), we rewrite the price index $\mathbb{P}$ in (5) as a function of aggregate productivity, Φ, and obtain

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

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

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

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

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

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

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

Labor in a city is supplied by all individuals with productivity below $\underline{\varphi}$. In efficiency units, city labor supply is equal to $L^S \equiv L\int_0^{\underline{\varphi}} \varphi^a dF(\varphi)$. From equation (2), labor demand for an entrepreneur with productivity φ is $l(\varphi) = x(\varphi)/\varphi$. Combining this expression with equations (7) and (8) and aggregating over all entrepreneurs, we obtain city labor demand:

$$L^D = L\int_{\underline{\varphi}}^{+\infty} l(\varphi)\, dF(\varphi) = (1+\varepsilon)^{-1} Y/w.$$

Equating labor supply and demand implies that workers receive a share $1/(1+\varepsilon)$ of city output:

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

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

The marginal cost of final good producers is equal to $\mathbb{P}$. Then, perfect competition among final good producers yields $\mathbb{P} = 1$ by our choice of the final good as the numéraire. Hence, equation (8) implies

$$w = \frac{1}{1+\varepsilon}\Phi, \tag{12}$$

where aggregate productivity, Φ, as defined in equation (7) can be rewritten as

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

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

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

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

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

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

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

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

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

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

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

Aside from existence and uniqueness, proposition 1 shows that the equilibrium selection cutoff does not depend on city population, conditional on the distribution of productivity (ignoring for now any general equilibrium connection between city size and productivity). This result is the outcome of two offsetting forces. Larger cities have both a higher demand (which lowers the selection cutoff) and more entrepreneurs (which raises it). These two effects exactly offset each other in our framework.[9] The reason behind this exact offset can be found in equation (11), which shows that labor market clearing implies a constant aggregate entrepreneurial income as a share of city output. Hence, keeping the distribution of individual productivity constant, a city hosts the same proportion of workers and entrepreneurs regardless of its size. As we show in online Appendix G, equilibrium selection is also optimal as a result of constant markups.

Proposition 1 also demonstrates that scaling up the distribution of productivity by a factor λ scales up the selection cutoff by the same factor. Again, this property occurs because aggregate entrepreneurial income is a constant share of city output. Simply put, a city whose residents are twice as productive has a selection cutoff twice as high. This case is of particular empirical relevance since Combes et al. (2012) find that the distribution of log productivity in larger cities in France is, to a first approximation, a shifted version of its counterpart in smaller cities.[10] In addition, the share of entrepreneurs should also be constant across cities. Empirically, the share of self-employed workers—a proxy for entrepreneurship—is uncorrelated with city population in the United States. Regressing the employment share of self-employed workers on log city population in 276 US metropolitan statistical areas (MSAs) using 2000 census data yields a coefficient of 0.0003 with a standard error of 0.001.

Proposition 1 has two further implications. First, sorting induces selection. If larger cities attract more talented individuals, they will be tougher markets. As emphasized by Frank Sinatra in his 1979 "New York, New York" song, "If I can make it there, I'll make it anywhere." Second, conditional on sorting, there are no differences in selection across cities. This result is also compatible with the findings of Combes et al. (2012) for the productivity of French firms. They find no differences in selection cutoffs across cities of different population sizes after accounting for common productivity differences that affect all firms. Put differently,

[9] There are at least two ways to make the productivity cutoff vary with city population conditional on the distribution of productivity. The first is to impose a different demand structure for varieties. In the spirit of Melitz and Ottaviano (2008), Behrens and Robert-Nicoud (forthcoming) use non–constant elasticity of substitution preferences to generate markups that decrease with the number of local varieties. This feature naturally leads to tougher selection in larger markets. The second possibility is to change the supply side and have the ratio of fixed to variable costs for firms depend on city population. On the one hand, a fixed cost (in addition to the entrepreneur's forgone labor) paid in the numéraire would be relatively less costly in larger cities where productivity is higher and would thus 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 and, in turn, would mean a lower proportion of entrepreneurs in larger cities.

[10] Scaling up the distribution of productivity across cities implies a shift when comparing the distribution of log productivity across cities.

there are no differences in selection between large and small French cities after controlling for sorting and agglomeration.

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

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

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

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

$$Y = \left[\int_{\underline{\varphi}}^{+\infty} \varphi^{1/\varepsilon} dF(\varphi)\right]^{\varepsilon} \left[\int_{0}^{\underline{\varphi}} \varphi^{a} dF(\varphi)\right] L^{1+\varepsilon}, \tag{16}$$

which shows that Y/L is also proportional to L^{ε}.

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

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

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

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

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

Finally, observe that we make the final good tradable to monetize the benefits of agglomeration easily. While this working assumption simplifies the quantitative exercise we conduct below, it is unimportant for our theory. In online Appendix H, we show that our model is isomorphic to one in which individuals consume a continuum of a nontraded local final goods. Hence, our approach also subsumes the "consumer city" model of urban economics in which the benefits from local diversity cannot be traded (as in, e.g., Lee 2010).

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

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

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

Proof. From equations (9) and (10) and the selection cutoff condition $\pi(\underline{\varphi}) = w \times (ts)^a$, the expected indirect utility of an individual with talent t in her city before her productivity is fully revealed is

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

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

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

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

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

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

IV. Sorting and Cities

Until now, we have taken the distribution of talent across cities as given. We now turn to location choices and the sorting of individuals across cities depending on their talent. To this end, we define the assignment function $\mathcal{M} : [\underline{t}, \bar{t}] \rightarrow C$, which maps talents into cities. An equilibrium choice of cities is such that

$$\mathcal{M}(t) = \{c \in C : \mathbb{E}V_c(t) \geq \mathbb{E}V_{c'}(t) \ \ \forall\ c' \in C\} \tag{18}$$

for all $t \in [\underline{t}, \bar{t}]$; that is, each individual is located in the city that maximizes her utility. Since individuals initially differ by their talent only, this location choice is a mapping from $[\underline{t}, \bar{t}]$ to C (a distinctive characteristic of our framework is that the destination set, C, is endogenous). In equilibrium, no individual wants to deviate to another city given the location choices of all other individuals. Once in a city, individuals make their occupational choice as described in Section III. Entrepreneurs choose employment in their firm to maximize profit, and all markets clear. Formally, an equilibrium satisfies the population adding-up constraint (4), selection and agglomeration as described by equations (10)–(13), and optimal location choice as given by equation (18).

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

Symmetric equilibria are both empirically counterfactual and theoretically not very illuminating. From now on, we thus focus on equilibria with sorting. Specifically, we construct an equilibrium with a single type of talent in each city. Because talent is sufficient to characterize a city in that case, we now drop the subscript c and denote cities by their level of talent instead.[11] We refer to cities in this equilibrium as talent-homogeneous cities and note that despite a single type of talent per city, there can still be enormous productivity heterogeneity within cities due to serendipity.

Let S denote the common serendipity threshold to become an entrepreneur and σ the share of efficiency units of labor used in production. By proposition 1, these two quantities are constant across talent-homogeneous cities:

[11] We are implicitly assuming that all cities of talent t are identical, i.e., have the same population size. This is without further loss of generality because it is true in equilibrium.

$$
\begin{aligned}
\underline{\varphi} &= St, \\
\sigma &= \int_0^S (s/S)^a \, dG_s(s) \\
&= \frac{1}{\varepsilon}\int_S^\infty (s/S)^{1/\varepsilon} \, dG_s(s),
\end{aligned} \tag{19}
$$

where the last equality follows from equation (14). When choosing a city of talent t, an individual with talent t' maximizes expected utility, $\mathbb{E}V(t', t)$, since serendipity has not occurred yet. Inserting the first equality of (19) into equation (17), we can rewrite the expected utility of a worker of talent t' in a talent-homogeneous city of size L with wage w where individuals are of talent t as

$$
\mathbb{E}V(t',\ t) = wt'^a\left[\int_0^{St/t'} s^a \, dG_s(s) + \left(\frac{t'}{St}\right)^{(1/\varepsilon)-a}\int_{St/t'}^{+\infty} s^{1/\varepsilon} \, dG_s(s)\right] - \theta L^\gamma. \tag{20}
$$

Each agent takes the selection cutoff, equal to St, the wage w, and the size of the city, L, as given when making her choice of location.

Observe that the support of talent is convex by assumption and that $\mathbb{E}V(t', t)$, given by (20), is continuously differentiable in L, t, and t'. An equilibrium with talent-homogeneous cities is characterized by a function $L(t)$ that assigns one city population to each talent such that all individuals of talent t choose to locate in a city populated by $L(t)$ individuals of the same talent. In equilibrium, two cities with the same talent t will be of identical size so that $L(t)$ is uniquely valued.[12] Individuals thus choose their preferred city from a "menu" of possible combinations of talent and population, knowing that the choice of a city talent t implies the choice of a population $L(t)$. We solve for the number of cities of different talents in the next section. Here, we first derive the equilibrium sizes of talent-homogeneous cities.

Proposition 4 (Equilibrium population of talent-homogeneous cities). If γ/ε is close to unity, the talent-homogeneous equilibrium is unique and such that

[12] For a fixed level of talent, wages are proportional to L^ε from (15), whereas urban costs are proportional to L^γ. Since $\gamma - \varepsilon > 0$, $\mathbb{E}V(t, t)$ is bell shaped. For any constant v, there are at most two solutions in $L(t)$ satisfying $\mathbb{E}V(t, t) = v$. In this case, the first solution is below the level of population that maximizes $\mathbb{E}V(t, t)$ while the second is above. The first solution cannot be stable because an arbitrarily small increase in population leads to an increase in expected utility. Only the second solution, which is in the region of decreasing returns, is stable. See Duranton and Puga (2004) for further discussion.

$$L(t) = \left(\frac{1+\gamma}{1+\varepsilon}\xi t^{1+a}\right)^{1/(\gamma-\varepsilon)}, \tag{21}$$

where

$$\xi \equiv \frac{(\varepsilon\sigma)^{1+\varepsilon} S^{1+a}}{\gamma\theta}.$$

Equilibrium city population increases with city talent t, agglomeration economies ε, and worker heterogeneity a and decreases with urban costs θ and γ.

Proof. In equilibrium, each individual solves a constrained optimization problem that consists in picking the city with talent t that maximizes her expected indirect utility from the menu of possible cities. For an individual of talent t', the first-order condition to the city selection problem (18) with talent-homogeneous cities can be written as

$$\left.\frac{\partial \mathbb{E}V(t',t)}{\partial L}\right|_{t'=t} dL + \left.\frac{\partial \mathbb{E}V(t',t)}{\partial t}\right|_{t'=t} dt = 0. \tag{22}$$

From equations (12)–(14) and (20), equation (22) becomes

$$[(\varepsilon\sigma)^{1+\varepsilon}(St)^{1+a}L^{\varepsilon} - \theta\gamma L^{\gamma}]\frac{dL}{L} + \frac{1+a}{1+\varepsilon}(\varepsilon\sigma)^{1+\varepsilon}(St)^{1+a}L^{\varepsilon}\frac{dt}{t} = 0.$$

The previous equation yields a differential equation that determines the menu of talents and populations that supports the talent-homogeneous equilibrium:

$$\gamma\theta L(t)^{\varepsilon}\left[\frac{\xi t^{1+a} - L(t)^{\gamma-\varepsilon}}{L(t)}dL(t) + \frac{1+a}{1+\varepsilon}\xi t^{a}dt\right] = 0, \tag{23}$$

where ξ is defined in equation (21). To solve this differential equation, we can verify that $L(t)$ is of the form $L(t) = (z\xi t^{1+a})^{1/(\gamma-\varepsilon)}$ for some z. Plugging this into (23) yields an equation involving the parameters of the model that is linear in z. Solving for z then gives

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

which satisfies (23) and allows us to obtain (21) after simplification. The comparative static results also follow directly from $\gamma > \varepsilon$.

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

$$a(\gamma + \varepsilon a) + \gamma(1 + a) > \left(\frac{\gamma}{\varepsilon} - 1\right)\left[1 + \frac{Sg(S)}{\sigma}(1 - a\varepsilon)\right], \qquad (24)$$

which always holds if γ/ε is close to one. QED

As made clear by equation (21), four elements determine equilibrium city population. The first is the standard trade-off between agglomeration economies (as given by ε) and urban costs (as given by γ and θ). Unsurprisingly, equilibrium population size increases with agglomeration economies and decreases with urban costs. The second determinant of city population is talent. More talented cities have a larger population because more talented individuals are more productive as entrepreneurs on average and more efficient as workers. Both features increase productivity and are magnified by agglomeration economies. Note that more talented cities are larger even when the distribution of talent has a thin right tail because, as made clear below, the number of cities also adjusts. More talented cities will be larger, but there will also be fewer of them when G_t, the cumulative distribution of talent, is concave. The third determinant of city population is the distribution of serendipity, which affects both the serendipity threshold to become an entrepreneur (S) and the share of efficiency units of labor used in production (σ). Finally, heterogeneity among workers (a) also affects city population size. With a higher a, more talented individuals are relatively more productive as workers, which, again, gets magnified by agglomeration economies.

We may view the equilibrium function $L(t)$ as describing an envelope of indifference curves in (t, L) space. This function is represented by the bold curve in figure 1. It is convex when $\gamma < 1 + \varepsilon$, which is empirically the case as highlighted in Section VII. Consider an individual with talent t' choosing from the menu of equilibrium cities described by $(t, L(t))$. Assume that she picks city c_1, which offers (t_1, L_1). In that case, this individual faces the indifference curve $\mathbb{E}V(t', t_1)$, which describes all the combinations of talent t and population L that offer her the same expected utility as city t_1 conditional on her talent t'. The lower indifference curve $\mathbb{E}V(t', t')$ describes all the combinations of talent t and population L that offer the same expected utility as city t' conditional on a talent t'.[13] Since expected indirect utility is increasing as indifference curves move down and right, $\mathbb{E}V(t', t')$ maximizes the expected utility of an individual with talent t' subject to the equilibrium menu of cities.[14]

[13] Observe that this curve yields higher utility as it has smaller cities (lower urban costs) and more talent (higher productivity). This feature is confirmed (locally at $t' = t$) by computing $\partial\mathbb{E}V(t', t)/\partial L|_{t'=t} > 0$ and $\partial\mathbb{E}V(t', t)/\partial t|_{t'=t} > 0$, which yields the shape of the indifference curves.

[14] Cities are too large in equilibrium so that a marginally smaller city is better all else equal. Expected utility also increases with talent because more talented cities offer higher wages, and this more than offsets the lower probability of becoming an entrepreneur.

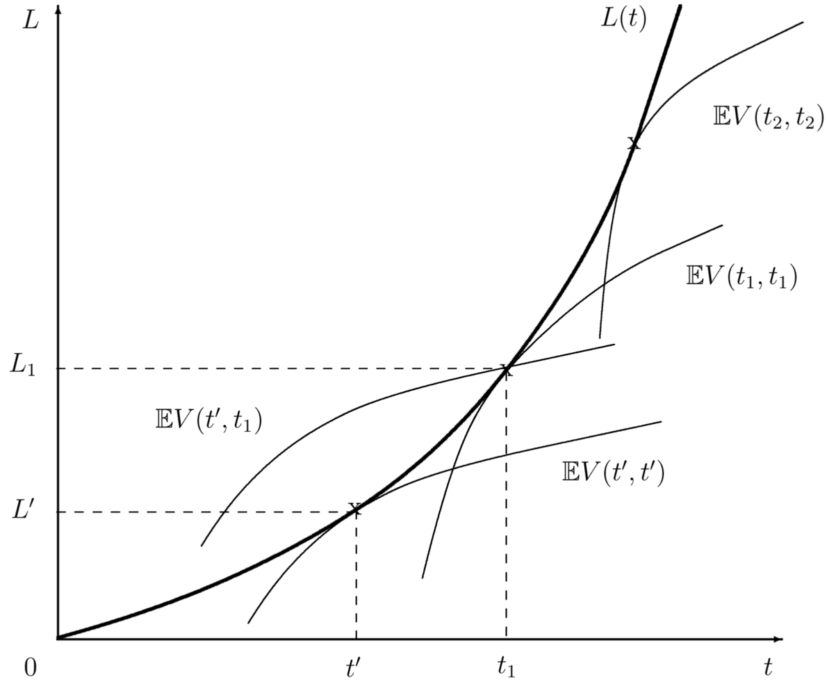

FIG. 1.—Equilibrium with talent-homogeneous cities

Hence, for this individual with talent t', utility is maximized in a city where all individuals have the same talent t' as hers. More generally, the bold curve $L(t)$ is the envelope of indifference curves for all levels of talent. As we move up this curve, we progressively read the optimal choices of individuals with higher talent. These are larger cities. The convexity of the relationship implies that small differences in talent may translate into large differences in city size.

Empirically, our equilibrium matches several of the key features of the data. That larger cities host more talented individuals is documented extensively in the literature (e.g., Berry and Glaeser 2005; Bacolod et al. 2009; Lee 2010; Combes, Duranton, Gobillon, and Roux 2012; Diamond 2013).[15] For 276 US MSAs in 2000, the elasticity of the share of college

[15] Consistent with our model, Diamond (2013) documents that increased sorting across US cities over 1980–2000 also led to higher productivity and higher housing costs in more educated cities. Whereas sorting and higher housing costs are driven by city size in our model, her findings also suggest an important role for amenities that arise endogenously from a greater concentration of skills and that are more highly valued by more educated workers. Although we do not consider them in our model, adding endogenous amenities—fostered by talent and more highly valued by more talented individuals—on top of the talent-size complementarity that we already consider would only reinforce the sorting forces in our model.

graduates with respect to population is 6.8 percent. For more talented individuals to sort into larger cities where urban costs are higher, their rewards must be relatively higher there. This property is exactly what the empirical literature finds (Wheeler 2001; Bacolod et al. 2009; Glaeser and Resseger 2010). It is also the case that more talented individuals migrate to areas that offer them higher rewards (Dahl 2002).

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

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

V. The Size Distribution of Cities

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

To establish this result, we need to impose mild technical restrictions. Namely, we assume that the support of the distribution of talent $g_t(\cdot)$ is compact and includes $\tilde{\xi}^{-1/(1+a)}$, where $\tilde{\xi} \equiv \xi(1+\gamma)/(1+\varepsilon)$. We assume further that the distribution of talent is finite valued and infinitely continuously differentiable around $\tilde{\xi}^{-1/(1+a)}$.

PROPOSITION 5 (Number and size distribution of cities). The equilibrium "number" of cities is proportional to population size Λ. The size distribution of cities converges to Zipf's law as $\eta \equiv (\gamma - \varepsilon)/(1+a)$ goes to zero.

Proof. Let $\mu(t)$ be the measure of cities with talent below t in the talent-homogeneous case that we consider. The assignment of talents to cities is then such that

$$\Lambda G_t(t) = \int_{\underline{t}}^{t} \mu'(\nu) L(\nu) d\nu.$$

Differentiating that expression, we have $\Lambda g_t(t) = \mu'(t)L(t)$; that is, the size of a talent t city times the mass of such cities sums to the density of talent t in the population. Rearranging yields

$$\mu'(t) = \frac{\Lambda g_t(t)}{L(t)}.$$

Solving this differential equation for $\mu(t)$ implies

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

where the second equality holds since the constant of integration κ is equal to $\mu(\underline{t}) = 0$. The measure of the set of cities C is then given by

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

This expression shows that the number of cities increases proportionately with Λ and establishes the first part of the proposition.

From equation (21), equations (25) and (26) may be rewritten as

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

Let $\eta \equiv (\gamma - \varepsilon)/(1 + a)$. A change of variables from talent to equilibrium city size in expression (27) then allows us to derive the probability distribution function for the population size of cities:

$$\begin{aligned} G_L(L;\eta) &\equiv \frac{\mu(L)}{\mu(\bar{L})} = \frac{\int_{\underline{L}}^{L} g_t(\tilde{\xi}^{-1/(1+a)} \ell^{\eta}) \ell^{\eta-2} d\ell}{\int_{\underline{L}}^{\bar{L}} g_t(\tilde{\xi}^{-1/(1+a)} \ell^{\eta}) \ell^{\eta-2} d\ell} \\ \Rightarrow g_L(L;\eta) &= \frac{g_t(\tilde{\xi}^{-1/(1+a)} L^{\eta})}{\int_{\underline{L}}^{\bar{L}} g_t(\tilde{\xi}^{-1/(1+a)} \ell^{\eta}) \ell^{\eta-2} d\ell} L^{\eta-2}, \end{aligned} \tag{28}$$

where $\underline{L} \equiv \underline{t}^{(1+a)/(\gamma-\varepsilon)}\tilde{\xi}^{1/(\gamma-\varepsilon)}$ and $\overline{L} \equiv \overline{t}^{(1+a)/(\gamma-\varepsilon)}\tilde{\xi}^{1/(\gamma-\varepsilon)}$. We note that the size of the largest city, $\overline{L}$, grows arbitrarily large as $\gamma - \varepsilon$ tends to zero but remains finite since $\gamma > \varepsilon$.

Since we assume that the distribution of talent is finite valued and infinitely continuously differentiable around $\tilde{\xi}^{-1/(1+a)}$, the ith derivative of $g_L(\cdot)$ with respect to η, evaluated at $\eta = 0$, satisfies $|g_L^{(i)}| \leq K$ for some $K < \infty$ and for all i. Then, using a Taylor expansion of the second expression of (28) around $\eta = 0$ yields

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

The second-order remainder term above converges to zero as η tends to zero. QED

That the number of cities should be proportional to total population is natural in a context in which cities are endogenous and there are no increasing aggregate returns. The second part of proposition 5 shows that the size distribution of cities converges to Zipf's law for any distribution of talent when η approaches zero. This property is an important result since Zipf's law is a reasonable first-order approximation for observed distributions of city population sizes (Gabaix and Ioannides 2004). As shown in Section VII, the difference $\gamma - \varepsilon$ is empirically small, around 3 percent. We show in Appendix C that for such values of the parameters, the Zipf approximation works extremely well.

To understand the key intuitions behind this result, it is useful to proceed in steps. First, it is easy to see that if talent follows a Pareto distribution, the size distribution of cities is also Pareto because city size in equation (21) is a power function of talent in the city and any power transformation of a Pareto distribution is also a Pareto distribution.

Second, to understand why the Pareto shape parameter of the size distribution of cities is close to one, note that the number of cities of talent t, $g_L(L)$, is given by the number of individuals with this level of talent, $g_t(t)$, divided by the size of those cities, L (where the equilibrium relationship between L and t is given by eq. [21]). This implies $g_L(L)dL = g_t(t)dt/L$. In turn, the size distribution of cities is obtained from a change in variables using the fact that $L^{(\gamma-\varepsilon)/(1+a)}$ is proportional to t by equation (21). Hence, if talent is Pareto distributed with shape parameter m, the size distribution of cities is Pareto with shape parameter $1 + m(\gamma - \varepsilon)/(1 + a)$. Then, when $\gamma - \varepsilon$ is small, the shape parameter of the size distribution of cities is close to one, that is, Zipf's law.

Third, as shown by proposition 5, Zipf's law occurs for any ("regular") distribution of talent provided that $\gamma - \varepsilon$ is small. The reason is that one can write the local Pareto shape of any distribution as $m(t)$, which implies that the "Pareto shape" of the size distribution of cities is $1 + m(t)(\gamma - \varepsilon)/(1 + a)$. Then, provided that $\gamma - \varepsilon$ remains small, any deviation of the distribution of talent from a Pareto distribution can be neglected.[16]

VI. Two Extensions

A. *Discrete Cities with Heterogeneous Talent and Variable Selection*

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

In Appendix D, we investigate an example of equilibrium with discrete cities. In this case, analytical results cannot be obtained in general. Numerical computations are needed. Because cities are in finite number while there is a continuum of talents, cities are no longer talent-homogeneous and selection differs across cities in equilibrium. Despite these important differences, many of the key properties of the equilibrium with talent-homogeneous cities are also properties of this equilibrium with variable selection including the links between city size, productivity, and "city talent." Interestingly, the selection cutoffs across cities differ only marginally. These similarities are a good reason to focus on the simpler and analytically tractable case of talent-homogeneous cities.

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

B. *A Dynamic Version of the Model with Relocations*

So far, our model is static and individuals are stuck in the city they initially chose. Though convenient and useful, this assumption is extreme. As we underscore in Section II, allowing individuals to relocate at no cost

[16] That the distribution of talent should not affect the size distribution of cities is in contrast with existing assignment models. In Gabaix and Landier (2008), for instance, the distribution of earnings of chief executive officers reflects both the size distribution of firms and the distribution of talent. The key difference is the following. In their paper, CEOs with heterogeneous managerial talent are assigned to an exogenous set of firms of heterogeneous size (as measured by their market capitalization). Instead, in our model, entrepreneurs of heterogeneous talent are assigned to an endogenous set of cities, which allows us to obtain Zipf's law under mild conditions on the distribution of talent.

once their productivity φ is fully revealed yields perfect productivity sorting as in Nocke (2006), which is counterfactual. In an extension of the model in which time runs indefinitely, we now allow agents to relocate at a cost.

The setting is as follows. We assume that time T runs discretely. Each individual has a probability of dying $\delta \in (0, 1)$ at each period. A fraction δ of newborns also appear every period so that aggregate population, Λ, is constant. Individuals are endowed with a talent t for life, but serendipity s is modeled as several possible draws of luck. In all other respects the setting is the same as in the model described above.

Consider some arbitrary time T. A newborn observes her talent t, makes a location decision, and receives a first realization of serendipity s for this location. Upon observing her productivity $\varphi = t \times s$, an individual may either "stay" in the chosen city and select an occupation (worker or entrepreneur) or "move" to get another realization of serendipity. Getting another occurrence of serendipity is costly.[17] It involves exiting the current location at time T (hence the term "move"), waiting in the hinterland for one period during which utility is normalized to zero, and picking a new location at time $T + 1$ (possibly different from the one at time T). Likewise, individuals already alive at time $T - 1$ have the choice between staying where they are and sticking with their current productivity, and moving to get a new occurrence of serendipity.

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

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

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

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

Omitting time subscripts to ease notation since we are describing a steady state, we can write the following proposition.

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

[17] If it was not, everybody would keep drawing a new s until getting the upper bound of s.

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

Proof. See Appendix E.

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

VII. Quantitative Implications

We now use our framework to revisit several empirical results. Although our model is highly stylized, it is useful to interpret a variety of empirical findings within a unified framework. Those empirical findings can be seriously misinterpreted using partial equilibrium reasoning.

Equation (16) provides an expression for output in each location. Dividing it by population yields a measure of productivity, namely, output per capita, $y_c \equiv Y_c/L_c$, which depends on local population and a complex function $f(\cdot)$ of the distribution of talent. If we are willing to proxy this complex function of the distribution of talent with average years of education, we obtain the first key estimating equation of Gennaioli et al. (2013), who regress output per capita y_c in 1,499 regions of 105 countries and find

$$\begin{aligned} \ln(y_c) &= \varepsilon \ln L_c + f(G_{t,c}(\cdot), G_s(\cdot)) \\ &\approx 0.068 \log L_c + 0.257\ \mathrm{Educ}_c + \mathrm{controls}_c + v_c, \end{aligned}$$

where $G_{t,c}(\cdot)$ is the distribution of talent in c, $G_s(\cdot)$ is the distribution of serendipity, and $f(\cdot)$ is a function that links them both to productivity. This regression implies a value of 0.068 for ε. It also points at the importance of human capital and education as determinants of output per capita. However, their coefficient of 0.257 on average years of education does not have a structural interpretation in our framework given the complexity of the function $f(\cdot)$ and the unknown mapping between education and talent.

Gennaioli et al. (2013) also work with microdata for 6,314 firms in 76 regions of 20 countries. More specifically, they regress firm revenue Z_i on its employment, the education of its "entrepreneur," the average education of its workers, and, as above, local population and local average education. Our model does not generate this specification. However, we can extend it to allow for limited span of control of entrepreneurs in the wake of Lucas (1978) and obtain this exact specification while leaving its other properties unchanged.[18] We develop this extension in online Appendix I. Gennaioli et al. find

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

where N_i is the employment count of firm i, Educ_i^W the average years of education of its workers, and Educ_i^E the years of education of its entrepreneur. According to our extended model, the coefficient on local population of 0.126 is another estimate of ε, our agglomeration parameter. The coefficient on employment of 0.860 should be equal to $(1-\alpha)/(1+\varepsilon)$, where α is the span of control parameter (corresponding implicitly to the share of the entrepreneur in production). Gennaioli et al. take α to be about 0.1, which implies in our extended model a coefficient on employment very close to the one they measure:

$$(1-\alpha)/(1+\varepsilon) = (1-0.1)/(1+0.126) = 0.800$$

instead of 0.860. Even more interesting, the coefficient on entrepreneurial education is higher than that on worker education. Gennaioli et al. interpret this result as evidence of extremely high returns to education for entrepreneurs.[19] Our model suggests an alternative explanation. Recall that our model indicates that only the most productive individuals become entrepreneurs while the others become workers. Put differently, given talent (or education in this empirical implementation), only individuals with particularly good occurrences of serendipity become entrepreneurs whereas the others (with bad draws) become workers. Put differently, the coefficient on entrepreneurial education is biased upward while that on worker education is biased downward. Whether returns to education are particularly high for entrepreneurs and managers after accounting for positive selection into these occupations is an open question.

[18] By an artifact of the constant elasticity of demand, a linear production function (as we use above) implies that the productivity of entrepreneurs and workers cancels out when computing firm revenue. This knife-edge property is easily avoided by imposing decreasing returns to scale in production using, for instance, a limited span of control argument.

[19] With $\alpha = 0.1$, the returns to education for entrepreneurs in the framework of Gennaioli et al. (2013) are $0.026/0.1 = 26$ percent.

Next, we exploit the restrictions of our original model at the talent-homogeneous equilibrium. Observe that the expected indirect utility (17) can be written as

$$\mathbb{E}V_c(t_c) = \sigma^{1+\varepsilon}(St_c)^{1+a}(\varepsilon L_c)^{\varepsilon} - \theta L_c^{\gamma} = y_c - \theta L_c^{\gamma}.$$

Taking logs, we have

$$\ln y_c = \kappa_1 + (1 + a)\ln t_c + \varepsilon \ln L_c, \tag{29}$$

where κ_1 is a constant term. Expression (29) shows that regressing average earnings on population, while controlling for talent, yields an estimate of agglomeration economies, ε. Now, using the equilibrium relationship linking city population sizes to the distribution of talent (21), $L_c^{\gamma-\varepsilon} = \xi t_c^{1+a}(1 + \gamma)/(1 + \varepsilon)$, to control for the shift, we get

$$\ln y_c = \kappa_2 + \gamma \ln L_c, \tag{30}$$

where κ_2 is another constant term. Hence, regressing average earnings on population without controlling for talent yields an estimate of the urban costs parameter, γ.[20]

We estimate equations (29) and (30) using standard US census data for 276 MSAs 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 labor economics. We obtain[21]

$$\ln y_c = 8.59 + 0.082 \ln L_c, \tag{31}$$

$$\ln y_c = 9.60 + 0.051 \ln L_c + 0.46 \ln t_c. \tag{32}$$

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 0.039 for ε. Using the share of pop-

[20] The result that the elasticity of income with respect to city size equals the elasticity of urban costs seems a priori surprising since utility is not equalized across cities in our framework. Yet, this result holds because in our model at equilibrium

$$\mathbb{E}V_c(t_c) = \sigma^{1+\varepsilon}(St_c)^{1+a}(\varepsilon L_c)^{\varepsilon} - \theta L_c^{\gamma} = \kappa_3 L_c^{\gamma},$$

where κ_3 is a positive bundle of parameters. As can be seen from this expression, the population elasticity of equilibrium utility is γ, which is also the population elasticity of urban costs. We show in online App. J that this result holds beyond our specific model, although it will not be true in general.

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

ulation older than 18 years with a graduate or professional degree to measure t_c in regression (32) yields a coefficient of 0.058 on log population.[22] Note that we refrain from interpreting the coefficient of $\ln t_c$ as providing an estimate of $1 + a$ since we do not know how talent maps into education.

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 et al. (2009) for broader reviews.[23] The sizable drop in the coefficient for log population after adding a measure of city education is also typical (Combes et al. 2008; Glaeser and Resseger 2010). Our favorite 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 one (for a one-dimensional city as we use here). However, recent work on US cities (Albouy 2008; Baum-Snow and Pavan 2012) or French land markets (Combes, Duranton, and Gobillon 2013) reports estimates of γ between 0.04 and 0.12 that are close to ours.[24]

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

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

This coefficient of 0.085 is remarkably close to the coefficient of 0.082 estimated in regression (31). 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

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

[23] We use city aggregated data and few controls. Using microdata and more controls typically results in slightly lower estimates for the coefficient on city size (Combes et al. 2008; Glaeser and Resseger 2010). These small differences are not important for our purpose here.

[24] In a 2013 discussion of our paper, Fabien Candau used Carillo, Early, and Olsen's (2012) data for a panel of 380 US areas and regressed their housing price index on population; he obtains an elasticity of $\hat{\gamma} = 0.077$, which is again very close to our estimate of 0.082.

$$\ln h_c = 8.93 + 0.068 \ln L_c.$$

This estimate of 0.068 for urban costs remains reasonably close to that in (31) despite relying on a different estimating equation.

With equilibrium city size as given by expression (21), the elasticity of talent to city population size should be equal to $(\gamma - \varepsilon)/(1 + a)$. We obtain $\hat{\gamma} - \hat{\varepsilon} = 0.031$ using (31) and (32). Regressing directly the log share of college graduates on log population yields

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

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

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

Our model also predicts that the share of expenditure on housing is independent of city population. To see this, we note that total land rents (TLR) are given by $\theta\gamma L_c^{1+\gamma}$ as shown in online Appendix F whereas aggregate income is equal to Y_c. Taking the ratio of TLR to aggregate income, using equations (16) and (21) for talent-homogeneous cities and the definitions of σ and ξ, we obtain

$$\frac{\mathrm{TLR}_c}{Y_c} = \frac{(1+\gamma)\varepsilon}{1+\varepsilon}$$

after simplification. This quantity does not depend on city population size and is equal to 0.052 for our preferred estimates of ε and γ. This result is important for two reasons. First, it is in line with findings by Davis and Ortalo-Magné (2011). They show that expenditure shares on housing are constant over time and across US MSAs at around 24 percent. If we take a share of land in housing of 25 percent as in Combes et al. (2013), we find an empirical value of TLR/Y equal to $0.24 \times 0.25 = 0.06$, which is close to 0.052. Second, this result of a constant share of land is obtained with an additively separable utility function. Hence, Cobb-Douglas preferences are not required to generate constant expenditure shares on housing across cities.

Finally, we turn to the inefficiency of the talent-homogeneous equilibrium and quantify its economic costs. From equations (12)–(14) and (17), the first-order condition with respect to L only and evaluated at $t = t_c$ implies that the optimal population of talent-homogeneous cities is

$$L^o(t_c) = (\xi t_c^{1+a})^{1/(\gamma-\varepsilon)}. \tag{33}$$

Put differently, optimal city size is the solution to an equation identical to equation (23) for equilibrium city size without the second term in the brackets. A direct comparison between equations (33) and (21) shows that cities are oversized in equilibrium. To understand why, consider the following hypothetical situation of a small isolated city and a mass of individuals outside. Provided that the reservation level of the latter is low enough, they should gradually move to the city. Because $\gamma > \varepsilon$, there exists an optimal city size, and the expected utility of all individuals in the city, as it grows, should first increase before decreasing after the city passes its optimal size. In standard models of urban systems (e.g., Henderson 1974), an equilibrium with cities that are too large is reached when expected utility inside the city is equal to the reservation level.

Because individuals differ in talent, things are more complicated in our case. Heuristically, as shown by proposition 3, more talented workers benefit more from the city being larger. Hence, as the city grows, we reach a situation in which the expected utility of more talented individuals keeps growing while that of less talented individuals declines. This situation leads to further in-migration of more talented individuals and out-migration of less talented individuals. Eventually, the city ends up being too large and populated only by the most talented individuals. Taking these individuals out, we can repeat the same thought experiment for the city with the most talented individuals among those that remain.[25]

[25] Interestingly, city sizes are uniquely determined in equilibrium. The trade-off between agglomeration economies and urban costs leads to net output per resident being a bell-

Using again $\hat{\varepsilon} = 0.051$ and $\hat{\gamma} = 0.082$, we can compute that equilibrium cities are oversized by a factor of

$$\frac{\widehat{L_c}}{\widehat{L_c^o}} = \left(\frac{1+\hat{\gamma}}{1+\hat{\varepsilon}}\right)^{1/(\hat{\gamma}-\hat{\varepsilon})} = 2.55. \tag{34}$$

Figure 2 plots the oversize of cities as computed in equation (34) for varying values of γ and three values of ε. This plot indicates that an oversize of 145–165 percent is to be expected. Consistent with the comparative statics of equation (34), the figure also shows that city oversize decreases in γ and ε ($< \gamma$). Using a first-order linear approximation of equation (34) when $\gamma - \varepsilon$ is small, we obtain $L_c/L_c^o \approx \exp[1/(1+\varepsilon)]$, 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 (21), (33), and (34) to compute an estimate of the indirect utility (consumption) loss:

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

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

The reason why losses from oversized cities are so small is the following. Recall first that cities are oversized by a factor close to $e \approx 2.72$. Imagine next that earnings (Y_c/L_c) are of the same magnitude as urban costs (θL_c^{γ}). Then, the maximum loss from oversize would be $1 - e^{-(\gamma-\varepsilon)}$ or about 2.8 percent for our preferred value of $\hat{\gamma} - \hat{\varepsilon} = 0.031$. However, equilibrium urban costs are only about 5 percent of earnings as shown above. Hence, the actual loss is much smaller than 2.8 percent. These results are consistent with those of Au and Henderson (2006) for Chinese

shaped function of city population. With homogeneous individuals, there is a coordination failure in city formation so that any population size between optimal city size and grossly oversized cities—leaving their residents with zero consumption—can occur in equilibrium (Henderson and Becker 2000). In our model, the sorting of heterogeneous individuals makes this indeterminacy disappear entirely. Formally, this property of our model 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 population sizes are uniquely determined.

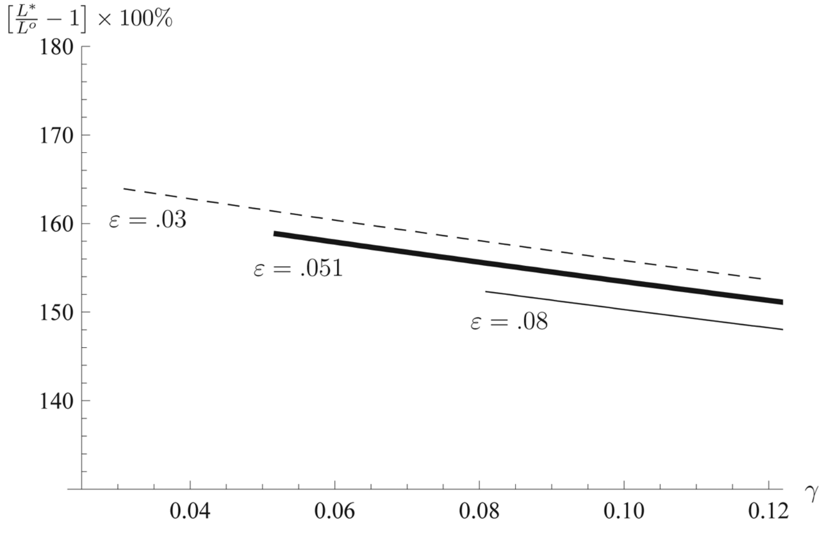

Fig. 2.—Oversize as a function of γ for different values of $\varepsilon < \gamma$

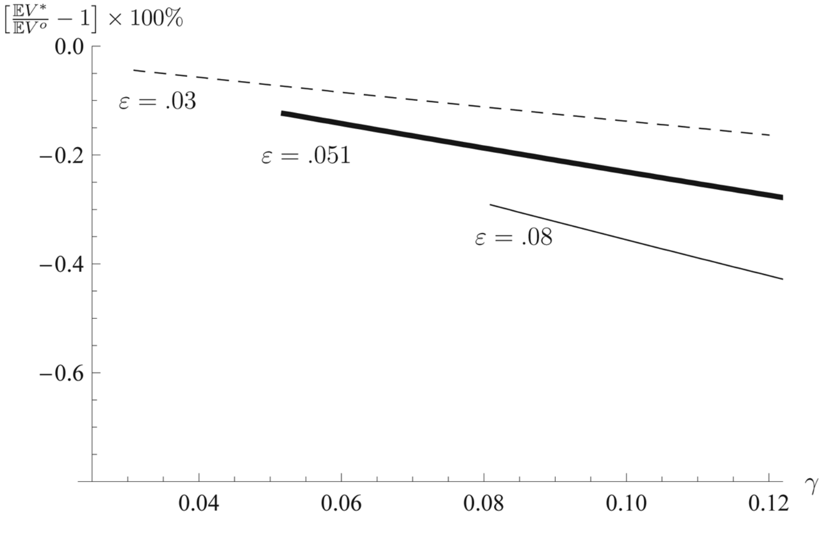

Fig. 3.—Efficiency loss as a function of γ for different values of $\varepsilon < \gamma$

cities. Using the fact that Chinese migration policies have limited the growth of Chinese cities, they estimate the shape of net benefits from cities as a function of their size. Like us, they find a very flat curve past the optimum. Restricting the size and growth of cities is unlikely to deliver substantial welfare improvements.

VIII. Conclusion

Although abundant empirical research in urban economics has substantiated a significant positive correlation between skills and city population, theory has had much less to say about the spatial sorting of heterogeneous individuals across an urban hierarchy until now. Our paper is an attempt to make progress in this direction. We have shown that ex ante sorting along talent and ex post selection along productivity, when coupled with an otherwise standard model of agglomeration economies and monocentric cities, allows us to replicate key stylized facts: larger cities host more talented individuals, have more productive (but not a greater proportion of) entrepreneurs, pay higher wages, and have higher urban costs. Importantly, even though firms in larger cities are, on average, more productive than those in smaller cities, there is considerable heterogeneity in firm productivity within each city. Finally, our sorting mechanism and its interactions with agglomeration economies and urban costs provides a simple static explanation for Zipf's law.

In addition to our theoretical contributions, our model also provides a unified setting within which to interpret quantitative evidence obtained from various standard regressions. It suggests, in particular, how regressions of measures of productivity, skills, and urban costs on log population can be consistently interpreted and how they relate quite naturally to each other. For instance, according to our model, regressing city wages on city size or density alone provides an estimate of the congestion elasticity rather than, as sometimes claimed, an estimate of the agglomeration elasticity. We believe that such an interpretative framework is useful for guiding future empirical analysis.

A number of issues remain open for future work. Cities are essentially passive in our model. In reality, cities, especially the most talented ones, actively limit their population growth, which may foster sorting even further (Gyourko, Mayer, and Sinai 2013). Allowing cities to play a more active role within our framework figures prominently in our research agenda. In addition, we also note that our model departs from the standard modeling assumption whereby utility for a given individual is equalized across all cities. Instead, individuals strictly prefer a specific type of city depending on their talent. We conjecture that this feature has many further implications. For instance, 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 useful for addressing questions related to the local provision of public goods and local taxation. We hope that future work will shed light on these issues.

Appendix A

Symmetric Equilibria of the Model

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

Proposition A1 (Symmetric equilibria). There exists a continuum of stable equilibria such that $F_c(\cdot) = F(\cdot)$ and $L_c = L$ for all cities c only if the variation in talent across the population is small enough.

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

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

$$\begin{aligned}\frac{\varepsilon}{\gamma}\left[\int_{\underline{\varphi}}^{+\infty} \varphi^{1/\varepsilon} dF(\varphi)\right]^{\varepsilon}&\left[\bar{t}^{a} \int_{0}^{\underline{\varphi}/\bar{t}} s^{a} dG_s(s) + \underline{\varphi}^{a}\left(\frac{\bar{t}}{\underline{\varphi}}\right)^{1/\varepsilon} \int_{\underline{\varphi}/\bar{t}}^{+\infty} s^{1/\varepsilon} dG_s(s)\right] \\ &\leq \theta L^{\gamma-\varepsilon}(1+\varepsilon) \\ &\leq \left[\int_{\underline{\varphi}}^{+\infty} \varphi^{1/\varepsilon} dF(\varphi)\right]^{\varepsilon}\left[\underline{t}^{a} \int_{0}^{\underline{\varphi}/\underline{t}} s^{a} dG_s(s)\right. \\ &\quad \left. + \underline{\varphi}^{a}\left(\frac{\underline{t}}{\underline{\varphi}}\right)^{1/\varepsilon} \int_{\underline{\varphi}/\underline{t}}^{+\infty} s^{1/\varepsilon} dG_s(s)\right].\end{aligned} \tag{A1}$$

In addition, it implies that

$$\frac{\varepsilon}{\gamma} < \frac{\underline{t}^a \int_0^{\underline{\varphi}/\underline{t}} s^a dG_s(s) + \underline{\varphi}^a (\underline{t}/\underline{\varphi})^{1/\varepsilon} \int_{\underline{\varphi}/\underline{t}}^{+\infty} s^{1/\varepsilon} dG_s(s)}{\bar{t}^a \int_0^{\underline{\varphi}/\bar{t}} s^a dG_s(s) + \underline{\varphi}^a (\bar{t}/\underline{\varphi})^{1/\varepsilon} \int_{\underline{\varphi}/\bar{t}}^{+\infty} s^{1/\varepsilon} dG_s(s)} \tag{A2}$$

must hold at any stable symmetric equilibrium. Since $\gamma > \varepsilon$, the left-hand side of this expression is smaller than unity. The right-hand side is increasing with $\underline{t}$ and decreasing with $\bar{t}$. Furthermore, it is also smaller than unity, but it tends to one as $\underline{t} \to \bar{t}$. Hence, this condition is fulfilled for a "sufficiently homogeneous" population ($\underline{t} \approx \bar{t}$). Additionally, when $\underline{t} < \bar{t}$, symmetric equilibria are never stable when $\gamma \approx \varepsilon$, that is, when "net urban costs" are small.

Finally, note that condition (A1) bounds the population L of symmetric cities. It is then easy to verify that when (A2) holds, there exists in general a continuum of pairs $(L, \bar{c})$ of city populations, L, and number of cities, $\bar{c}$, such that (A1) and the adding-up constraints $\Lambda g_t(t) = \bar{c}L(t)$, for all $t \in [\underline{t}, \bar{t}]$, hold (the latter implying, of course, that $\Lambda = \bar{c}L$). QED

Proposition A1 establishes that there generally exists a continuum of stable symmetric equilibria when the variation of talent across the population is small enough or when agglomeration economies, ε, are small compared to urban costs, γ. Neither case seems empirically relevant. This fact suggests that ability sorting is a natural equilibrium outcome.

Appendix B

Second-Order Conditions for the Equilibrium with Talent-Homogeneous Cities

Rewrite expression (20) as follows:

$$\mathbb{E}V(t', t) = \theta^{-\varepsilon/(\gamma-\varepsilon)} \left[\frac{1+\gamma}{1+\varepsilon} \frac{(\varepsilon\sigma)^{1+\varepsilon}}{\gamma} (St)^{1+a}\right]^{\gamma/(\gamma-\varepsilon)} \left[\frac{\gamma}{\varepsilon} \frac{1+\varepsilon}{1+\gamma} \frac{\sigma(t', t)}{\sigma} - 1\right], \tag{B1}$$

where

$$\sigma(t', t) \equiv \frac{1}{1+\varepsilon}\left[\left(\frac{t'}{t}\right)^a \int_0^{St/t'} \left(\frac{s}{S}\right)^a dG_s(s) + \left(\frac{t'}{t}\right)^{1/\varepsilon} \int_{St/t'}^{\infty} \left(\frac{s}{S}\right)^{1/\varepsilon} dG_s(s)\right].$$

A sufficient condition for the talent-homogeneous case to be an equilibrium is that $\mathbb{E}V(t', t)$ is quasi-concave in t for all $(t', t) \in T \times T$. Imposing quasi concavity on (B1) yields an expression that is so unwieldy that it is impossible to assess how restrictive it is. Imposing concavity on (B1) at the talent-homogeneous equilibrium, a more stringent sufficient condition than quasi concavity, also yields an expression that is still quite unwieldy in general. By contrast, the following necessary second-order condition is not. That is to say, we require the second derivative of (B1) with respect to t to be negative when evaluated at $t' = t$. Straightforward but tedious algebra yields

$$\frac{d^2}{dt^2}\mathbb{E}V(t',t)\big|_{t'=t} \propto -\, t^{-2+[\gamma/(\gamma-\varepsilon)](1+a)}\sigma\left\{-\left(\frac{\gamma}{\varepsilon}-1\right)\left[1+\frac{Sg(S)}{\sigma}(1-a\varepsilon)\right] + a(\gamma+\varepsilon a)+\gamma(1+a)\right\},$$

where $\sigma \equiv \sigma(t,\ t)$. Hence, if a talent-homogeneous equilibrium exists, then equation (24) in the main text holds. Note that condition (24) is not overly restrictive. Indeed, as shown in Section VII, γ/ε is close to unity so that the right-hand side of the foregoing equation is small.

Appendix C

Zipf's Law

As shown in Section V, the size distribution of cities converges to Zipf's law when $(\gamma - \varepsilon)/(1+a)$ goes to zero, irrespective of the underlying distribution of talent. In this appendix, we quantify the quality of this approximation when $\gamma - \varepsilon$ is within the range of empirically plausible estimates.[26]

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

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

with $k \to 1$. To evaluate the quality of the power law approximation of $g_L(\cdot)$ for a given η, a first natural metric is given by the distance between the two distributions:

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

where $\underline{L} \equiv \underline{t}^{(1+a)/(\gamma-\varepsilon)}\tilde{\xi}^{1/(\gamma-\varepsilon)}$ and $\bar{L} \equiv \bar{t}^{(1+a)/(\gamma-\varepsilon)}\tilde{\xi}^{1/(\gamma-\varepsilon)}$. A second way to judge the quality of the approximation involves generating random samples from the

[26] We assume $a = 0$. Since we cannot reliably estimate this quantity, taking its lower bound allows us to remain conservative when assessing the quality of the Zipf approximation.

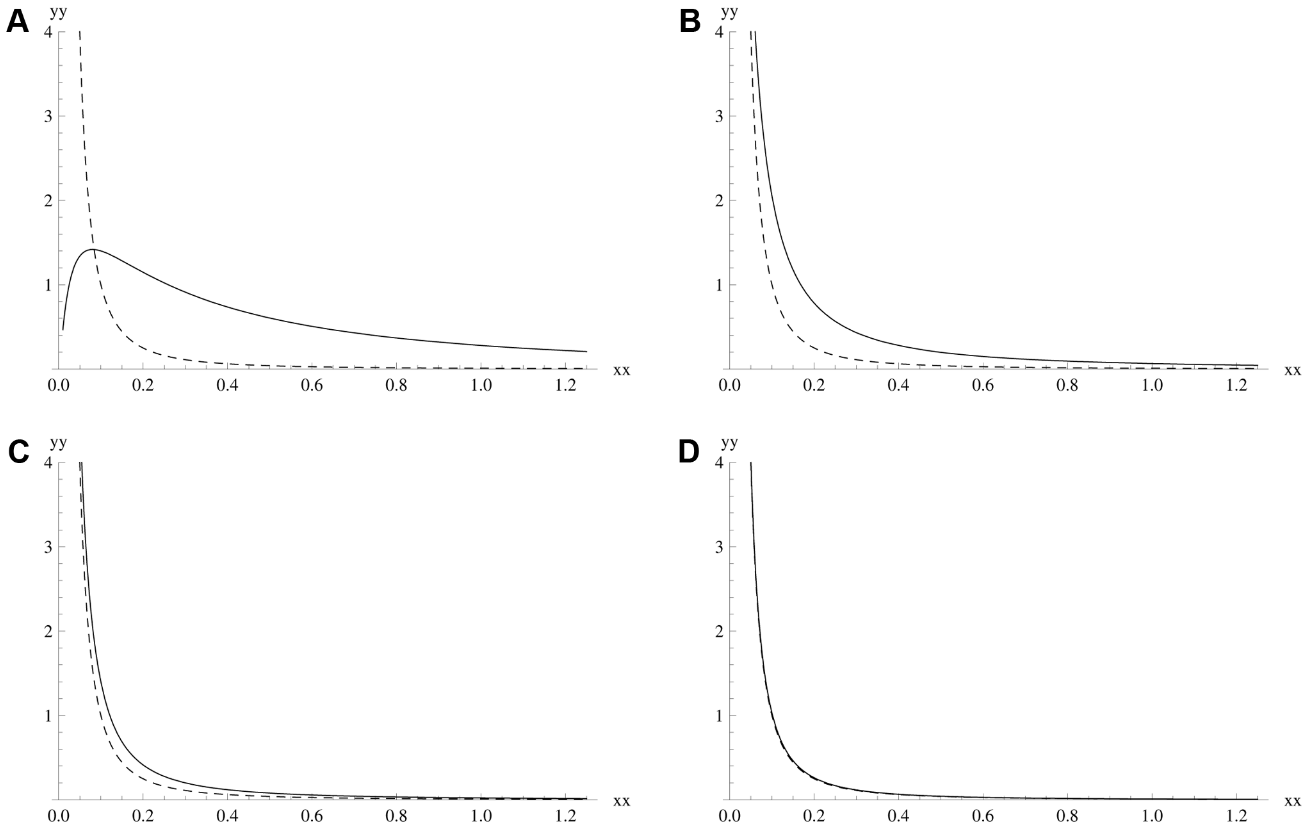

Fig. C1.—Quality of the numerical approximation of the Pareto distribution. *A*, $\eta = 1$; *B*, $\eta = 0.5$; *C*, $\eta = 0.25$; *D*, $\eta = 0.03$. The dashed line is the theoretical Pareto distribution, and the solid line is the approximation.

TABLE C1
QUALITY OF THE NUMERICAL APPROXIMATION OF THE PARETO DISTRIBUTIONS

η	Distance $d(g_L;\eta)$: Truncated Pareto	Distance $d(g_L;\eta)$: Pareto	ML Estimate of $\hat{k}$: Truncated Pareto*	OLS Estimate of $\hat{k}$: Pareto†
1	62.9127	62.9815	−.2521	.7175
1/2	21.0455	21.0858	.3690	.6840
1/4	4.4038	4.4222	.7371	.8362
1/8	.9314	.9399	.8688	.9701
1/16	.2104	.2144	.9514	1.0362
1/32	.0498	.0518	.9843	1.0144
1/64	.0121	.0131	.9403	.9901
1/128	.0030	.0035	.9764	1.0802
1/256	.0007	.0010	1.0216	1.0392
1/512	.0002	.0003	.9926	1.0328

* ML estimates for the truncated Pareto distribution follow Aban et al. (2006) and are computed from random samples of 1,000 observations that are generated using inverse transform sampling.

† OLS estimates of k for the Pareto distribution are computed using the same random samples as for the ML etimates. Following Gabaix and Ibragimov (2011), we use log(rank − 1/2) as the dependent variable in that case.

approximation and to estimate the parameters of the truncated Pareto distribution from those samples. The better the approximation, the closer $\hat{k}$ must be to one.

We illustrate how the approximation converges to both a truncated and a complete Pareto distribution by using for $g_t(\cdot)$ a lognormal distribution with parameters $(\mu_t, \sigma_t) = (1.4, 1.4)$ and with support $[\underline{t}, \bar{t}] = [0.01, 5]$.[27]

We evaluate the approximation using values of $\eta = 1/2^n$, for $n = 1, 2, \ldots, 10$. Figure C1 shows that the approximation rapidly converges to a (truncated) Pareto distribution with shape parameter k close to one. This result can be seen more formally from table C1, which reports the maximum likelihood (ML) and the ordinary least squares (OLS) estimates of the shape parameter k of the truncated and the complete Pareto distributions. To estimate those parameters, we sample 1,000 points from the approximation using inverse transform sampling. The ML estimate of k is then obtained from that sample using the estimator developed by Aban, Meerschaert, and Panorska (2006). As can be seen from table C1, the distance between the approximation and the truncated Pareto distribution vanishes quickly. The same occurs, to almost the same degree, for a complete Pareto distribution. Furthermore, as can be seen from the last column of table C1, the estimates of the shape parameter converge to one relatively quickly. In particular, for a value of $\eta = 1/32 \approx 0.031$, which corresponds closely to our empirical value $\hat{\gamma} - \hat{\varepsilon}$, the approximation is already fairly close to one. Hence, even if the underlying distribution of talent is lognormal, as in our

[27] Our results are robust to the underlying distribution and parameterization. As expected, the approximation is better and convergence is faster when the underlying distribution of talent is more skewed to the right.

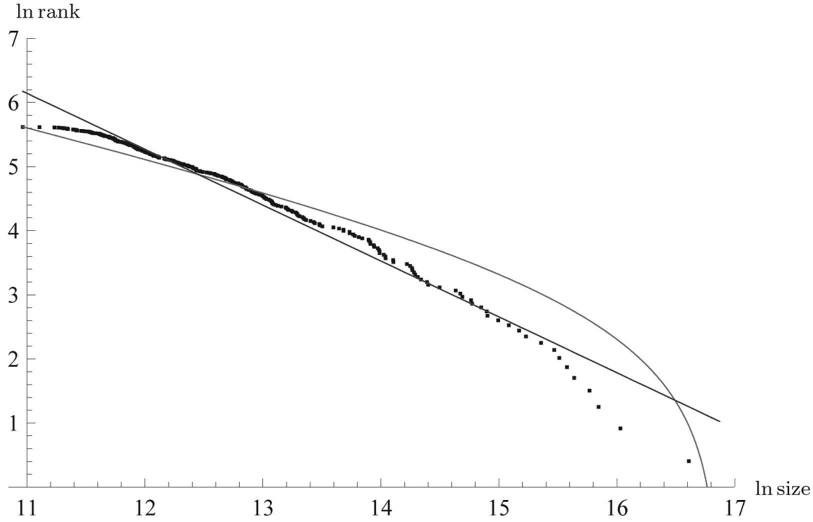

FIG. C2.—Pareto fit (straight line) and truncated Pareto fit (concave curve) for 276 MSAs in 2000.

example, the size distribution of cities will be approximately Zipf with a shape parameter close to one.

Estimating the Pareto parameter from the data on our 276 MSAs yields a coefficient of $\hat{k} = 0.8716$, whereas the parameter for a truncated distribution is $\hat{k}_t = 0.4484$. Figure C2 depicts the observed rank size distribution for 276 MSAs in 2000 (black dots), as well as the fits for the truncated (concave curve) and complete (straight line) Pareto.

Appendix D

Example of Equilibrium with Discrete Cities, Heterogeneous Talent, and Varying Selection

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

To keep things simple, we assume that $a = 0$ so that all workers have the same productivity $\varphi^a = 1$. We also consider only three types of cities, type 1, type 2, and type 3 cities, and subscript variables accordingly. We also specify the distributions of talent and serendipity to be uniform over $T = [\underline{t}, \bar{t}]$ and $\Sigma = [\underline{s}, \bar{s}]$, 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 $\varphi \equiv t \times s$. When we use theorem 1 of Glen, Leemis, and Drew (2004) and assume, without loss of generality for our purpose, that $\underline{t}\overline{s} < \overline{t}\underline{s}$, the product of talent and serendipity is distributed as follows:

$$f(\varphi) = \begin{cases} \dfrac{1}{(\overline{s}-\underline{s})(\overline{t}-\underline{t})}\ln\left(\dfrac{\varphi}{\underline{s}\underline{t}}\right) & \text{if } \underline{t}\underline{s} \le \varphi \le \underline{t}\overline{s} \\ \dfrac{1}{(\overline{s}-\underline{s})(\overline{t}-\underline{t})}\ln\left(\dfrac{\overline{s}}{\underline{s}}\right) & \text{if } \underline{t}\overline{s} \le \varphi \le \overline{t}\underline{s} \\ \dfrac{1}{(\overline{s}-\underline{s})(\overline{t}-\underline{t})}\ln\left(\dfrac{\overline{s}\overline{t}}{\varphi}\right) & \text{if } \overline{t}\underline{s} \le \varphi \le \overline{t}\overline{s}. \end{cases} \tag{D1}$$

Using equation (D1), 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 populations L_i, and their selection cutoffs $\underline{\varphi}_i$ for $i = 1, 2, 3$ are all endogenously determined. Let

$$\Delta\mathbb{E}V_i(t) = \mathbb{E}V_i(t) - \max_{j \neq i}\mathbb{E}V_j(t).$$

A spatial equilibrium is such that every individual with talent t picks the city that maximizes her expected indirect utility. Formally, $\Delta\mathbb{E}V_1(t) \ge 0$ for all $t \in [\underline{t}, t_1)$ (and negative otherwise), $\Delta\mathbb{E}V_2(t) \ge 0$ for all $t \in [t_1, t_2)$ (and negative otherwise), and $\Delta\mathbb{E}V_3(t) \ge 0$ for all $t \in [t_2, \overline{t}]$ (and negative otherwise).

Figure D1 depicts the expected indirect utility differentials for the three types of cities, as well as the two talent cutoffs for $\varepsilon = 0.47$, $\gamma = 0.5$, $\theta = 0.5$, $\Lambda = 5{,}000$, $\underline{t} = \underline{s} = 1$, $\overline{t} = 5$, and $\overline{s} = 2$.[28] In figure D1 we also set the numbers of cities to $n_1 = 20$, $n_2 = 6$, and $n_3 = 2$. We note that the mass of cities of each type is not uniquely determined in equilibrium, as was the case with talent-homogeneous cities. There exists instead a continuum of n_i, $i = 1, 2, 3$, that can be sustained as an equilibrium.

The allocation we have chosen is in equilibrium for t_1 and t_2 as determined in figure D1 since all individuals located in type 1 cities (i.e., to the left of t_1) get an expected utility no smaller than in type 2 or type 3 cities, all individuals in type 2 cities (i.e., between t_1 and t_2) get an expected utility no smaller than in type 1 or type 3 cities, and all individuals in type 3 cities (i.e., to the right of t_2) get an expected utility no smaller than in type 1 or type 2 cities.

In line with the results derived in the case of talent-homogeneous cities, more talented cities are larger and more productive and pay higher wages. We have $L_3 = 1{,}660.42$, $\underline{\varphi}_3 = 6.58$, and $w_3 = 89.46$, whereas the corresponding figures for type 2 cities are $L_2 = 240.38$, $\underline{\varphi}_2 = 3.13$, and $w_2 = 17.14$ and for type 1 cities are L_1

[28] Our values for agglomeration economies, ε, and urban costs, γ, are much larger than empirically reasonable to accentuate differences across cities in the figure.

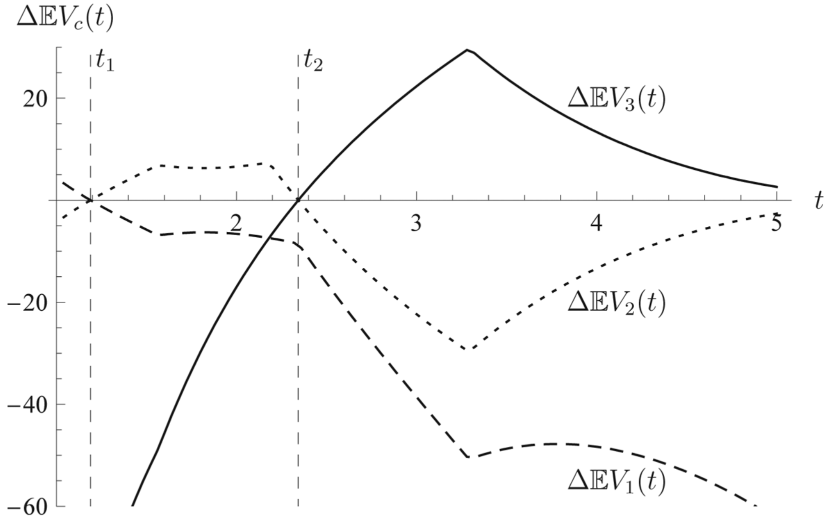

FIG. D1.—Example of spatial equilibrium with imperfect ability sorting

$= 11.84$, $\underline{\varphi}_1 = 1.88$, and $w_1 = 2.45$. In words, type 3 cities are about seven times larger than type 2 cities, which are themselves about 20 times larger than type 1 cities. Furthermore, type 3 wages exceed type 2 wages by a factor of about 5, and type 2 wages exceed type 1 wages by a factor of 7. The selection cutoffs reflect a similar ranking. Importantly, the strong right skew in the size distribution of cities does not stem from the right skew in the distribution of talent. The latter is uniform. Instead, sorting, agglomeration economies, and the population-talent complementarity generate these strong asymmetries.

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

Figure D2 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. In a comparison of 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, which is due to both agglomeration and sorting. Second, there is substantial dilation of profits in large cities relative to small cit-

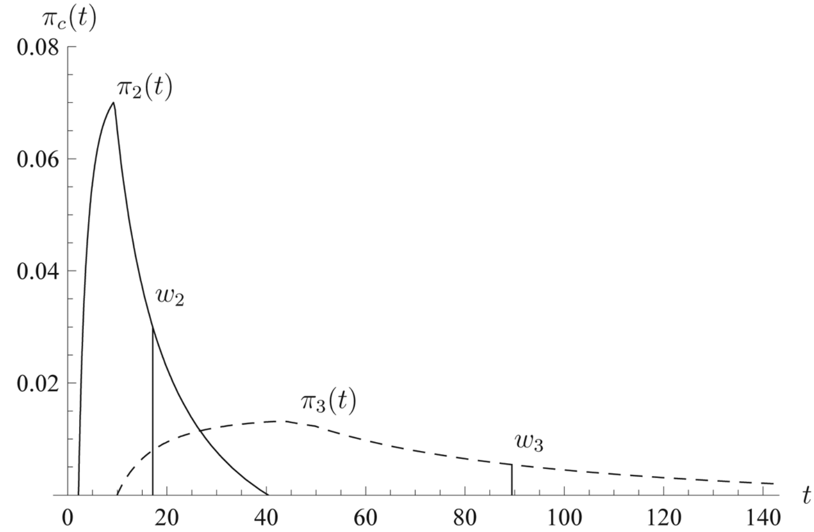

FIG. D2.—Distribution of profits in type 2 and type 3 cities

ies. Large cities host, on average, more productive individuals, but the most productive of them benefit disproportionately from being there. Large cities are thus more unequal than small cities by most conventional measures of inequality. This property is consistent with the findings of the literature on inequalities in cities (Glaeser, Resseger, and Tobia 2009; Baum-Snow and Pavan 2013; Behrens and Robert-Nicoud, forthcoming). 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). These features are clearly reminiscent of the findings of Combes et al. (2012) regarding the distributions of firms' productivities in small and large French cities.

Appendix E

Proof of Proposition 6

Claim 1 holds by construction. We thus prove claims 2–4 imposing claim 1. Let

$$\mathbb{E}_{\hat{s}} V(t', t) \equiv \frac{1}{1 - G_s(\hat{s})} \int_{\hat{s}}^{\infty} V(st', t)\, dG_s(s) \tag{E1}$$

denote the expected utility conditional on drawing $s \geq \hat{s}$, and let $M(s, t', t)$ denote the value of moving from talent-homogeneous city t for an individual with talent t' and current serendipity s, with $M(s, t) \equiv M(s, t, t)$. Then the value of moving is equal to

$$M(s,t) = 0 + (1-\delta)\left\{[1 - G_s(\hat{s})]\frac{\mathbb{E}_{\hat{s}}V(t',\ t)}{\delta} + G_s(\hat{s})\max_{t} M(s,t',t)\right\}$$
$$= (1-\delta)\left\{[1 - G_s(\hat{s})]\frac{\mathbb{E}_{\hat{s}}V(t',t)}{\delta} + G_s(\hat{s})M(s,t)\right\},$$

where the second line follows from the optimal location choice (in steady state, choosing a city with talent $t = t'$ remains optimal at the current period if it was the optimal city to choose at the previous period). Rearranging and using (E1) yields the following expression for the value of moving:

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

that is, $M(s, t)$ does not depend on s. Using this expression and (E1), we may rewrite the condition requiring that the net current value of staying is (weakly) positive for all s greater than $\hat{s}$ as

$$\begin{aligned} 0 &\le V(st,t) - \delta M(t) \\ &= V(st,t) - \frac{1-\delta}{1-(1-\delta)G_s(\hat{s})}\int_{\hat{s}}^{\infty} V(\tilde{s}t,t)\,dG_s(\tilde{s}) \quad \forall\ s \ge \hat{s}. \end{aligned} \tag{E2}$$

By inspection, this value is increasing in s and the cutoff for serendipity $\hat{s}$ is implicitly defined as

$$V(\hat{s}t,t) = \frac{1-\delta}{1-(1-\delta)G_s(\hat{s})}\int_{\hat{s}}^{\infty} V(\tilde{s}t,t)\,dG_s(\tilde{s}); \tag{E3}$$

that is, the opportunity cost of moving is equal to the expected value of moving for some $\hat{s}$. The left-hand side (LHS) of equation (E3) is strictly increasing in $\hat{s}$ from 0 to $+\infty$ over $(0, \infty)$ by inspection. The right-hand side (RHS) of equation (E3) evaluated at $\hat{s} = 0$ is equal to $(1-\delta)\mathbb{E}_0 V(t) = (1-\delta)\mathbb{E}V(t) > 0$ and, at the limit $\hat{s} \to +\infty$, it is equal to zero. Thus, by continuity and the intermediate value theorem, there exists $\hat{s} \in (0, \infty)$ that satisfies equation (E3). To prove the uniqueness of $\hat{s}$, differentiate the RHS of (E3) to obtain

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

where the last inequality follows from (E2). This result establishes the uniqueness of $\hat{s}$ and thus claim 2.

From (E3), the LHS does not depend on δ, whereas the RHS is decreasing in δ by inspection. Hence, $\hat{s}$ must be a decreasing function of δ, which proves the first part of claim 3. The second part of this claim immediately follows from the fact that $\lim_{\delta\to 1} M(s,t) = 0$ for all values of $s > 0$. When death is certain, waiting for better times is worthless.

To obtain claim 4, we use the definition of $M(t)$ to rewrite (E3) as $0 = V(\hat{s}t,t) - \delta M(t)$. We then use the definitions of $V(\cdot)$ and $M(\cdot)$ in order to get

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

Recall that $w(St)^a$ is proportional to $L^\varepsilon t^{1+a}$, where t^{1+a} is itself proportional to $L^{\gamma-\varepsilon}$ at the equilibrium with talent-homogeneous cities (see eq. [33] in the text). Hence, the final condition at the equilibrium with talent-homogeneous cities is given by

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

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

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

References

Aban, Inmaculada B., Mark M. Meerschaert, and Anna K. Panorska. 2006. "Parameter Estimation for the Truncated Pareto Distribution." *J. American Statis. Assoc.* 101 (473): 270–77.

Abdel-Rahman, Hesham M., and Ping Wang. 1997. "Social Welfare and Income Inequality in a System of Cities." *J. Urban Econ.* 41 (3): 462–83.

Albouy, David. 2008. "Are Big Cities Really Bad Places to Live? Estimating Quality of Life across Metropolitan Areas." Working Paper no. 14472, NBER, Cambridge, MA.

Au, Chun-Chung, and J. Vernon Henderson. 2006. "Are Chinese Cities Too Small?" *Rev. Econ. Studies* 73 (3): 549–76.

Bacolod, Marigee, Bernardo S. Blum, and William C. Strange. 2009. "Skills in the City." *J. Urban Econ.* 65 (2): 136–53.

Baldwin, Richard E., and Toshihiro Okubo. 2006. "Heterogeneous Firms, Agglomeration and Economic Geography: Spatial Selection and Sorting." *J. Econ. Geography* 6 (3): 323–46.
Baum-Snow, Nathaniel, and Ronni Pavan. 2012. "Understanding the City Size Wage Gap." *Rev. Econ. Studies* 79 (1): 88–127.
———. 2013. "Inequality and City Size." *Rev. Econ. and Statis.* 95 (5): 1535–48.
Behrens, Kristian, and Frédéric Robert-Nicoud. Forthcoming. "Survival of the Fittest in Cities: Urbanisation and Inequality." *Econ. J.*
Berry, Christopher R., and Edward L. Glaeser. 2005. "The Divergence of Human Capital Levels across Cities." *Papers Regional Sci.* 84 (3): 407–44.
Carillo, Paul E., Dirk W. Early, and Edgar O. Olsen. 2012. "A Panel of Price Indices for Housing Services, Other Goods, and All Goods for All Areas in the United States, 1982–2010." Manuscript, Econ. Dept., Univ. Virginia.
Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. "Spatial Wage Disparities: Sorting Matters!" *J. Urban Econ.* 63 (2): 723–42.
———. 2011. "The Identification of Agglomeration Economies." *J. Econ. Geography* 11 (2): 253–66.
———. 2013. "The Costs of Agglomeration: Land Prices in French Cities." Manuscript, Wharton School, Univ. Pennsylvania.
Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012. "The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection." *Econometrica* 80 (6): 2543–94.
Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2012. "Sorting and Local Wage and Skill Distributions in France." *Regional Sci. and Urban Econ.* 42 (6): 913–30.
Dahl, Gordon B. 2002. "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets." *Econometrica* 70 (6): 2367–2420.
Davis, Donald R., and Jonathan I. Dingel. 2012. "A Spatial Knowledge Economy." Manuscript, Econ. Dept., Columbia Univ.
Davis, Morris A., and François Ortalo-Magné. 2011. "Household Expenditures, Wages, Rents." *Rev. Econ. Dynamics* 14 (2): 248–61.
Diamond, Rebecca. 2013. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980–2000." Manuscript, Econ. Dept., Stanford Univ.
Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *A.E.R.* 67 (3): 297–308.
Duranton, Gilles, and Diego Puga. 2004. "Micro-foundations of Urban Agglomeration Economies." In *Handbook of Regional and Urban Economics,* vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2063–2117. Amsterdam: North-Holland.
Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. "Spatial Sorting." *J.P.E.* 122 (June): 554–620.
Ellison, Glenn D., and Edward L. Glaeser. 1999. "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *A.E.R. Papers and Proc.* 89 (2): 311–16.
Ethier, Wilfred J. 1982. "National and International Returns to Scale in the Modern Theory of International Trade." *A.E.R.* 72 (3): 389–405.
Epple, Dennis, and Thomas Nechyba. 2004. "Fiscal Decentralization." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2423–80. Amsterdam: North-Holland.
Gabaix, Xavier, and Rustam Ibragimov. 2011. "Rank−1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents." *J. Bus. and Econ. Statis.* 29 (1): 24–39.

Gabaix, Xavier, and Yannis M. Ioannides. 2004. "The Evolution of City Size Distributions." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2341–78. Amsterdam: North-Holland.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Q.J.E.* 123 (1): 49–100.

Gennaioli, Nicola, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. "Human Capital and Regional Development." *Q.J.E.* 128 (1): 105–64.

Glaeser, Edward L., and Matthew G. Resseger. 2010. "The Complementarity between Cities and Skills." *J. Regional Sci.* 50 (1): 221–44.

Glaeser, Edward L., Matthew G. Resseger, and Kristina Tobia. 2009. "Inequality in Cities." *J. Regional Sci.* 49 (4): 617–46.

Glen, Andrew G., Lawrence M. Leemis, and John H. Drew. 2004. "Computing the Distribution of the Product of Two Continuous Random Variables." *Computational Statis. and Data Analysis* 44 (3): 451–64.

Gyourko, Joseph, Christopher Mayer, and Todd Sinai. 2013. "Superstar Cities." *American Econ. J.: Econ. Policy* 5 (4): 167–99.

Henderson, J. Vernon. 1974. "The Sizes and Types of Cities." *A.E.R.* 64 (4): 640–56.

Henderson, J. Vernon, and Randy Becker. 2000. "Political Economy of City Sizes and Formation." *J. Urban Econ.* 48 (3): 453–84.

Hendricks, Lutz. 2011. "The Skill Composition of US Cities." *Internat. Econ. Rev.* 52 (1): 1–32.

Hsu, Wen-Tai. 2012. "Central Place Theory and City Size Distribution." *Econ. J.* 122 (563): 903–22.

Lee, Sanghoon. 2010. "Ability Sorting and Consumer City." *J. Urban Econ.* 68 (1): 20–33.

Lee, Sanghoon, and Qiang Li. 2013. "Uneven Landscapes and the City Size Distribution." *J. Urban Econ.* 78 (November): 19–29.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell J. Econ.* 9 (2): 508–23.

Melitz, Marc J. 2003. "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity." *Econometrica* 71 (6): 1695–1725.

Melitz, Marc J., and Gianmarco I. P. Ottaviano. 2008. "Market Size, Trade and Productivity." *Rev. Econ. Studies* 75 (1): 295–316.

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland. 2009. "A Meta-Analysis of Estimates of Urban Agglomeration Economies." *Regional Sci. and Urban Econ.* 39 (3): 332–42.

Monte, Ferdinando. 2011. "Skill Bias, Trade, and Wage Dispersion." *J. Internat. Econ.* 83 (2): 202–18.

Moretti, Enrico. 2004. "Human Capital Externalities in Cities." In *Handbook of Regional and Urban Economics*, vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2243–91. Amsterdam: North-Holland.

Mori, Tomoya, and Alessandro Turrini. 2005. "Skills, Agglomeration and Segmentation." *European Econ. Rev.* 49 (1): 201–25.

Nocke, Volker. 2006. "A Gap for Me: Entrepreneurs and Entry." *J. European Econ. Assoc.* 4 (5): 929–56.

Puga, Diego. 2010. "The Magnitude and Causes of Agglomeration Economies." *J. Regional Sci.* 50 (1): 203–19.

Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies." In *Handbook of Regional and Urban Eco-*

nomics, vol. 4, edited by J. Vernon Henderson and Jacques-François Thisse, 2119–71. Amsterdam: North-Holland.

Sattinger, Michael. 1993. "Assignments Models of the Distribution of Earnings." *J. Econ. Literature* 31 (2): 831–80.

Wheeler, Christopher H. 2001. "Search, Sorting, and Urban Agglomeration." *J. Labor Econ.* 19 (4): 879–99.