On the Distribution of City Sizes

Juan-Carlos Córdoba*
Rice University

July 7, 2004

Abstract

The city size distribution of many countries is remarkably well approximated by a Pareto distribution. We study what constraints this regularity imposes on standard urban models. We find that under general conditions urban models must have (i) a balanced growth path and (ii) a Pareto distribution for the underlying source of randomness. In particular, one of the following combinations can induce a Pareto distribution of city sizes: (i) preferences for different goods follow reflected random walks, and the elasticity of substitution between goods is 1; or (ii) total factor productivities in the production of different goods follow reflected random walks, and increasing returns are equal across goods.

JEL classifications: R11, R12, O41, J10

Keywords: City Size Distribution, Zipf's Law, Rank-Size Rule, Pareto Distribution, Urban Growth, Multisectorial Models, Balanced Growth, Cities

*This paper is based on the first chapter of my Ph.D dissertation at the University of Rochester (July 2001). I thank Per Krusell for his insightful advice, Peter Hartley for the detailed comments, and Alan Stockman, Mark Bils, Michael Cranston, Tony Smith, Esteban Rossi-Hansberg, two anonymous referees, and participants in seminars at the universities of Rochester, Pensilvania, Carnegie Mellon, Pittsburgh, Western Ontario, Cornell, Northwestern, Rice, the London School of Economics, and Tulane. I acknowledge financial support provided by the Central Bank of Colombia. All remaining errors are my own. Email: jcordoba@rice.edu

1 Introduction

A remarkable empirical regularity is that the city size distribution in many countries is well approximated by a Pareto distribution. This claim is so widely accepted among social scientists that it has gained the status of a law, Zipf's Law, or a rule, the Rank-Size Rule[1,2]. It has also inspired extensive research mainly in the fields of urban economics and regional science.

Until recently, there was no satisfactory economic explanation for the regularity. This is at least the conclusion obtained by leading researchers on the topic: "*at this point we have no resolution to the explanation of the striking regularity in city size distribution. We must acknowledge that it poses a real intellectual challenge to our understanding of cities...*" (Fujita, Krugman and Venables [1999, p. 225]); and "*It is therefore no surprise that we still lack such a model... Yet this turns out to be a real embarrassment, because the rank-size rule is one of the most robust statistical relationships known so far in economics*" (Fujita and Thisse, [2000, p. 9]).

This paper studies necessary conditions to produce a Pareto distribution of city sizes within a class of standard urban models. For this purpose, we build on the statistical results of Gabaix [1999] and Córdoba [2004] who find that, under some mild restrictions, a Pareto distribution of city sizes, and Zipf's distribution in particular, can only result from Gibrat's law. Our goal is to translate these statistical results into meaningful economic restrictions about preferences, technologies, and stochastic components in standard economic models of cities.

We focus on a set of urban models in which cities arise due to economies of localization, as in Black and Henderson [1999]. In this formulation, cities emerge due to the presence of scale economies, external to firms but internal to industries. Moreover, the size of cities is determined by one of two forces. One is the presence of negative externalities, such as congestion costs, that limit the gains from agglomeration. This is the standard constraint employed in the urban literature to determine city size (e.g., Black and Henderson 1999, Rossi-Hansberg and Wright 2003). A second force that can also limit city size, when congestion costs are not binding, is the extent of the market. In this case, the economy-wide demand for the city's main tradable

[1] The observation is usually associated to Zipf [1949], but Auerbach [1913] seems to have been the first to uncover it. The large literature on the topic includes, among others, Rosen and Resnik [1980], Carroll [1982], Eaton and Eckstein [1997], Brakman *et al.* [1999], Roehner [1995], Gabaix [1999], Ioannides and Overman [2003], and Soo [2002]. See Ioannides and Gabaix [2003] for a recent survey.

[2] Pareto distributions are defined as

$$\Pr(X \leq x) = 1 - (x/a)^{-\delta}, \ x \geq a.$$

Zipf's law states that the city size distribution satisfies $\delta = 1$. The Rank-Size rule uses the previous formula to provide a deterministic description of the data rather than a probabilistic description.

output determines city size. We study the resulting city size distribution under these two alternative explanations of why city size is limited. The two explanations differ in their implications for the evolution of the urban system as population grows.

Under the first approach, the urban system evolves largely via the extensive margin, that is, by increasing the number of cities. Under the second approach, the urban system grows only via the intensive margin, that is, by increasing the size of the existing cities. Although the first approach is more standard in the literature, we think the evidence better supports the second formulation as new cities rarely arise in a mature urban system, an observation documented by Eaton and Eckstein [1997].

Our formulation allows for a fairly general functional forms for preferences, technologies, and stochastic components. The set of feasible functional forms is then restricted so that cities evolve according to Gibrat's law. In particular, Gibrat's law for cities requires that all cities in the urban system share the same expected growth rate. This single restriction imposes substantial discipline on the analysis. For cities to exhibit identical expected growth rates, fundamental parameters controlling city sizes must be identical across cities. We find that, as a general rule and independently of which force limit city growth, the degree of positive and negative externalities must be identical across cities. Such a restriction is typically not required in urban models. This in part explains their difficulties in accounting for the distribution of city sizes.

For the case in which city size is limited by the extent of the market, we find that city growth is independent of city size only under one of the following three conditions : (i) the elasticity of substitution between goods is equal to one; (ii) positive externalities are equal across goods; or (iii) a knife-edge condition on preferences and technologies is satisfied. Furthermore, under general conditions, the steady state distribution of the fundamental stochastic variables, controlling preferences or technologies, also must be Pareto. Finally, a well-defined relationship links the Pareto exponent characterizing the city size distribution to the degree of increasing returns in the economy and the elasticity of substitution between goods.

For the case in which city size is limited by the negative externalities, we show that in addition to a unitary elasticity of substitution between goods, positive and negative externalities must also be equal across industries. We then show that the distribution of city sizes is Pareto if and only if the underlying distribution of productivity parameters is Pareto too.

The characterization obtained above can then be easily employed to provide economic explanations for the observed distribution of city sizes. We offer at least two explanations. A standard urban model with localization economies can generate a Pareto distribution of city sizes if (i) preference parameters for different goods follow reflected random walks and the elasticity of substitution between goods is 1; or (ii) total factor productivities of different goods follow reflected random walks and increasing returns are equal across industries. These explanations are the first in the literature to be fully consistent with increasing returns to scale, a central component in urban models[3].

In summary, this paper provides general conditions under which standard urban models produce a Pareto distribution of city sizes. It also offers specific examples of fully stochastic urban models based on increasing returns that can support a Pareto distribution of city sizes. Finally, it provides economic interpretations for the Pareto exponent of the distribution of city sizes.

The paper is divided into 5 sections. Section 2 reviews the evidence and the related literature on the topic. Section 3 studies models in which city size is determined by the extent of the market, and Section 4 studies models in which city size is determined by binding negative externalities. Finally, Section 5 concludes.

2 Evidence and Related Literature

There is abundant empirical literature that supports the claim that the city size distribution is well approximated by a Pareto distribution. The original evidence is presented by Auerbach [1913], and Zipf [1949]. A classical empirical paper is Rosen and Resnik [1980] who studied a cross-section of countries. They find that the Pareto coefficients differ across countries, ranging from 0.80 to 1.96. Soo [2002] updates Rosen and Resnik using recent data and confirms their claims. Eaton and Eckstein [1997] analyzes the cases of France and Japan, Brakman *et al.* [1999] the Netherlands, Roehner [1995] several countries, and Ioannides and Overman [2003] study in detail the case of the United States[4]. These exercises usually find the Pareto exponent for the U.S. close to 1, but different from 1 for most other countries. Most of the theoretical literature focuses on the U.S. case assuming an exponent equal to 1. We focus on the more general case of an arbitrary exponent.

Several probabilistic and a few economic models have been proposed to account for this evidence. Among the most prominent probabilistic models are the ones by Champernowne [1953], Simon [1955], Steindl [1965], and more recently, Gabaix [1999], and Córdoba [2001, 2004]. The fundamental insight obtained by these authors is that Gibrat's law, or proportional growth, can lead to Pareto distributions. More precisely, if a stochastic variable follows a growth process that is independent of the position of the variable, then its limit distribution can be Pareto, a result first established by Chapernowne. Simon generalizes the result showing that proportional growth can explain many different skew distributions, such as log-normal, Pareto and Yule. He also derives a very simple formula linking the Pareto exponent with the underlying growth process. It is equal to $\frac{1}{1-\pi}$, where π is, in our case, the probability that

[3] To the extent of my knowledge, these were the only explanations available that were consistent with increasing returns at the time of my dissertation, in July of 2001. Recent papers have tried to provide additional explanations. I review them below.

[4] See also Ionnides and Gabaix [2003] for a recent survey.

new cities emerge. Gabaix [1999] establishes that Gibrat's law can lead to Zipf's distributions if the number of cities is constant, but if new cities emerge then only the upper tail of the distribution is Zipf. Córdoba [2001, 2004] finds that a generalized Gibrat's law process, one that allows the variance, but not the mean of the city growth process to depend on city size, can account for Pareto exponents different from 1 even if the number of cities is constant. He also finds that, under mild conditions, this generalized Gibrat's process is required to produce a Pareto distribution of city sizes.

In contrast to the success of the probabilistic approach, economic models have failed to match the evidence. Krugman [1996] and Fujita *et al.* [1999] conclude that none of the existing economic models can properly explain the data. Most urban models are deterministic which cannot explain the observed movement of cities across the size distribution. Moreover, due to parameter heterogeneity, cities usually do not grow at the same rate. In this respect, Black and Henderson [1999] is an important exception. They offer a deterministic urban model that can display a steady state in which all cities grow at the same rate. The conditions for this result, however, are unappealing. The result requires unusual functional forms for preferences and technologies. Moreover, the growth rate of their economy either increases or decrease through time, a counterfactual.

This paper belongs to a second generation of urban models[5] that explicitly incorporate stochastic elements and seek to explain the distribution of city sizes. Two early works in this genre are Eaton and Eckstein [1997] and Gabaix [1999]. Eaton and Eckstein construct a model in which city size depends on the amount of human capital accumulated in cities. Cities of different sizes coexist because they differ in their productivity as places to acquire capital. There are spillovers across cities in the accumulation of human capital. Eaton and Eckstein generate proportional growth only under the condition of zero discounting. Gabaix [1999] provides a model that produces a Zipf's distribution based on the random supply of amenities in different locations. This approach is unorthodox, as it requires constant returns to scale technologies, and leaves unexplained the nature of the amenities and the underlying source of agglomeration. In contrast with these papers, our paper provides clear, simple, and more general analytical conditions within a class of standard urban models.

Two recent related works are Rossi-Hansberg and Wright [2003] and Duranton [2002.] Rossi-Hansberg and Wright construct a stochastic urban model along the lines of the deterministic model of Black and Henderson [1999.] Like Black and Henderson, they are able to produce proportional growth, and Zipf's distributions, only under very particular and unappealing conditions, as stated in their Proposition 4. Either physical capital cannot enter the production function, or capital can enter, but the capital share must be equal to 1 minus two times the externality of labor. Numerical simulations confirm that large cities in their model are too small compared

[5] I thank an anonymous referee for this expression.

with the predictions of a Zipf distribution. They claim some quantitative success but at the cost of denying that Zipf's law really holds in practice. Moreover, the results are sensitive to the exact calibration. In contrast, we offer models that can produce a Pareto distribution of city sizes under more appealing and general conditions. This is because we use a different formulation of the externalities, and allow more capital mobility across cities than they do.

Finally, Duranton [2002] offers an unorthodox alternative to the models presented in this paper. He introduces the notion of agglomeration into a standard quality-ladder model of growth. In his model, a city is a collection of industries, and industries move randomly to other locations where research has been more successful. His baseline model is unorthodox, as there are no gains from agglomeration, and nothing limits the city size[6]. Duranton determines the number of cities by assuming that each city has one immobile industry. This is also an unorthodox approach because exogenous location heterogeneity is not the standard explanation for the existence of an urban system. His model can only produce an approximate Pareto distribution of city sizes. Moreover, these results are sensitive to the exact calibration and little is known about crucial parameters such as the externalities associated with research. We offer instead analytical results in a rather general framework that can accommodate standard and non-standard models. Finally, we believe that models based on a central role for research and development are of limited use for explaining the cross-country evidence on city size distributions as most countries adopt their technologies, rather than invent them.

3 Cities limited by the extent of the market

The models we study in this paper are based on economies of localization. Cities in these economies emerge due to the presence of scale economies, external to firms but internal to industries, as in Henderson [1988]. In this section we also formulate the existence of negative externalities, such as congestion costs, that induce cities to specialize in production. Contrary to the existing literature, however, negative externalities play no additional role in limiting city size in this section. The size of a specialized city is just limited by extent of the market.

We allow for stochastic technologies and preferences, and look for conditions such that the equilibrium distribution of city sizes is Pareto. There are two main results in this section. First, the same conditions that guarantee the existence of a balanced growth path in multisectorial endogenous growth models are also needed to generate Pareto distributions. The reason is simple. If cities specialize in production, at least to

[6] Duranton considers some extension of his model to include more traditional positive and negative externalities, captured by the functions $A(\cdot)$ and $C(\cdot)$. However, those externalities are made irrelevant as he immediately assumes $A(\cdot)=C(\cdot)$.

some extent, then city growth just mirrors sectoral growth. Second, the distribution of fundamentals, preferences and technologies, must be Pareto under general conditions.

3.1 Baseline economy

Consider an economy inhabited by a large number of workers, firms, and a benevolent government. The number of workers, N_t, grows continuously over time at the exogenous compound rate $\gamma \geq 0$, i.e., $N_t = e^{\gamma t}$. There are S locations and I industries such that $S > I$, and I is large. At the beginning of every period the government announces linear taxes on income for each location and industry[7], τ_{is}, $1 \leq i \leq I$ and $1 \leq s \leq S$. Firms and workers then choose locations and industries to produce and work during the period. A location with a positive mass of workers is called a 'city'. To simplify notation, time subscripts are dropped.

Labor mobility guarantees that the after-tax wage rate, denoted w, is equal across locations and industries with a positive mass of workers. The pre-tax wage rate, w_{is}, must then satisfy $w = (1 - \tau_{is})w_{is}$. Goods can be costlessly transported across locations. Arbitrage then guarantees that the price of good i, q_i, is equal across locations.

3.1.1 Production

Firms are competitive. A firm in location s and industry i chooses the amount of labor, l_{is}, that maximizes profits given by

$$\max_{l_{is}} q_i A_i \varphi^i(L_{is}, L_s) l_{is} - w_{is} l_{is}, \tag{1}$$

where L_s is the size of city s, L_{is} is the size of industry i in city s, and $A_i \varphi^i(L_{is}, L_s)$ is the productivity of labor, exogenous to the firm. Productivity has two components. The first is an idiosyncratic technological shock, A_i, known at the beginning of the period before time t decisions are made. A_i follows a Markov process with transition probability $\phi^A(x_0, t_0; x, t)$, common across industries. Denote Φ_t^A the time t distribution of A_i and assume that ϕ^A has a unique invariant distribution, Φ^A.

The second component is a differentiable function $\varphi^i : \Re_+^2 \rightarrow \Re$ that describes the benefits and costs of agglomeration. φ^i satisfies $\varphi_1^i(L, L_s) > 0$, $\varphi_2^i(L, L_s) < 0$ and $\varphi_1^i(L, L_s) + \varphi_2^i(L, L_s) > 0$. The first condition states that there are economies of scale at the industry (local) level, holding the size of the city constant. The source of these increasing returns may be informational spillovers, search and matching in local labor markets, and/or pecuniary externalities. The second condition introduces negative

[7] Taxes are introduced to induce efficient production. There are other alternatives to induce efficient production. See footnote (10) below.

externalities such as congestion costs. These externalities reduce the productivity of industries in a particular city as city size increases. The third condition states that, from the point of view of an individual industry, the economies of scale are always positive even after accounting for the congestion costs. A particular function that satisfies these conditions is

$$\varphi^i(L_{is}, L_s) = L_{is}^{\tau_i} L_s^{-\beta_i},\ \tau_i > \beta_i > 0. \tag{2}$$

Finally, free entry guarantees zero profits in all locations and industries, and determines the level of employment in every industry, given the size of the city

$$q_i A_i \varphi^i(L_{is}, L_s) = w_{is}. \tag{3}$$

3.1.2 Preferences

Workers seek to maximize their lifetime utility described by

$$U = E_0 \int e^{(\gamma-\rho)t} u\left(c_t; \theta_t\right) dt,$$

where $c_t = [c_{1t}, c_{2t}, ., c_{It}]$ is a vector of consumption goods, $\theta_t = [\theta_{1t}, \theta_{2t}, ..., \theta_{It}]$ is a random vector of tastes for goods, ρ is the rate of discount, and u is a homothetic instantaneous utility function. θ_t is known at the beginning of the period before time t decisions are made. θ_{it} follows a Markov process with transition probability $\phi^\theta(x_0, t_0; x, t)$, common across industries. Denote Φ_t^θ the time t distribution of θ_i, and assume that ϕ^θ has a unique invariant distribution, Φ^θ.

Let W be the total after-tax labor income of the economy (at time t), $q = [q_1, q_2, ..., q_I]$ be vector of prices, and $C_i(q, \theta, W)$ the aggregate demand of good i. Since u is homothetic, the aggregate demand functions have the form[8]

$$C_i(q, \theta, W) = C_i(q, \theta)W$$

Denote ϵ_{ij} the price elasticity of demand of good i with respect to price j, i.e. $\epsilon_{ij} \equiv \frac{\partial \ln C_i(q_t,\theta)}{\partial \ln q_j}$, and denote ϵ the $I \times I$ price elasticity matrix, a matrix with ϵ_{ij} in the i row and j column.

3.1.3 Government

There is a benevolent government that seeks to maximize a utilitarian social welfare function by using lump sum taxes and transfers and location and industry specific

[8] Demand for good i only depends on time t variables q, W and θ since aggregate borrowing and lending is zero.

proportional income taxes. In the absence of externalities in production, it would be optimal to use only lump sum taxes and transfers. In the presence of externalities, however, it is optimal to use income taxes to induce efficient production.

Production is efficient if each city hosts a single industry, and each industry is located in a single city [9]. This type of city specialization avoids unnecessary congestion costs that arise when more than one industry locates in the same city (formally, $\varphi^i(L_{is}, L_{is}) > \varphi^i(L_{is}, L_{is} + L)$ for any $L > 0$). We assume that the government uses the following tax scheme that induces efficient production: zero income taxes on activity i at location i, and confiscatory income taxes on the same activity in any other location. Under this tax scheme, cities specialize in production but no income taxes are paid in equilibrium[10]. The total production of industry i is thus determined by $\varphi^i(L_{is}) \equiv \varphi^i(L_{is}, L_{is})$. The degree of net increasing returns in activity i is measured by the elasticity of average productivity with respect to the agglomeration,

$$a_{is}(L_{is}) := \frac{\varphi^{i\prime}(L_{is})}{\varphi^i(L_{is})} L_{is}. \tag{4}$$

For the particular functional form (2), $\varphi^i(L_{is}) = L_{is}^{a_i}$, where $a_i(L_{is}) = a_i := \tau_i - \beta_i$.

3.1.4 Deterministic balanced growth paths

The model just described is fairly general. It allows the degree of externalities to differ across industries and/or sizes. In addition, it imposes no major restrictions on the instantaneous utility function, like symmetry. This section studies what restrictions on preferences and technologies are required in order for the equilibrium of the deterministic model to exhibit a balanced growth path or parallel growth. According to Gabaix [1999], and Córdoba [2004, Theorem 3] this is a requirement that a model of cities must satisfy in order to produce a Pareto distribution of city sizes.

Consider a deterministic version of the model.

Definition: A *balanced growth competitive equilibrium with optimal taxes,* or *balanced growth path,* are trajectories for prices, q_{it}, wages, w_t, quantities of goods, C_{it}, and labor allocations, L_{it}, such that (i) $w_t = q_{it} A_i \varphi_i(L_{it})$ (profit maximization); (ii) $C_{it} = C_i(q_t, \theta) w_t N_t$ (Utility maximization); (iii) $C_{it} = A_i \varphi^i(L_{it}) L_{it}$ (Goods market

[9] We introduce non-tradable goods below so that efficient cities do not fully specialize. Moreover, diversification in tradable goods can be obtained by allowing complementarities accross goods that outweight the costs. Our results are robust to those changes.

[10] Thus, our solution concept will be a competitive equilibrium with optimal taxation (Ramsey Equilibrium). Alternative, the following solution concept with a large number of competitive 'developers' (as in Henderson 1974) produces the same allocations. Each developer operates one location, and chooses an activity i, and labor input, L_{is}, to maximize profits $q_i A_i \varphi^i(L_{is}, L_s) L_{is} - w L_{is}$ given q_i and w. The developer acts competitively because she takes prices as given. Due to free entry and the large number of locations, profits are zero in equilibrium.

clearing); (iv) $N_t = \sum L_{it}$ (Labor market clearing); (v) $\frac{L_{it}}{N_t} = constant$ (Balanced industry growth); (vi) $q_{1t} = 1$ (the numeraire).

Let $\widehat{q}_j = \frac{dq_j/dt}{q_j}$ be the instantaneous growth rate of the relative price q_j. According to conditions (i), (v) and (vi) of the previous definition, relative prices obey

$$\widehat{q}_{it} = [a_{1t} - a_{it}]\,\gamma \text{ for all } i \text{ and all } t. \tag{5}$$

This equation states that along balanced growth paths changes in relative prices are completely determined by technological factors. Preferences play no role. Define

$$\epsilon_t \equiv \begin{bmatrix} \epsilon_{11t} & \epsilon_{12t} & .. & \epsilon_{1It} \\ \epsilon_{21t} & \epsilon_{22t} & .. & \epsilon_{2It} \\ .. & .. & .. & .. \\ \epsilon_{I1t} & \epsilon_{I2t} & .. & \epsilon_{IIt} \end{bmatrix}, \triangle a_t \equiv \begin{bmatrix} a_{1t} - a_{it} \\ a_{1t} - a_{2t} \\ \cdots \\ a_{1t} - a_{it} \end{bmatrix}.$$

Note that ϵ_t is purely determined by preferences while $\triangle a_t$ is purely determined by technologies. The following is the first main result of the paper:

Proposition 1 *A balanced growth path exists if only and only if*

$$\left(\overline{1} + \varepsilon_t\right) \triangle a_t = 0 \textit{ for all } t. \tag{6}$$

where $\overline{1}$ is the identity matrix.

Proof. Denote $s_{it} \equiv \frac{q_{it}C_{it}}{w_t N_t}$ the share of total expenditure on good i. According to the definition of a balanced growth path, s_{it} must be constant and given by $s_{it} = \frac{L_{it}}{N_t} = s_i$ for all i. Using condition (ii) of the definition above, s_{it} also satisfies $s_{it} = q_{it}C_i(q_t, \theta)$ for all i. Differentiating the two previous conditions and using (5) one obtains the result. ■

According to this Proposition, we can typify three type of sufficient conditions to obtain a balanced growth path.

1. $\triangle a_t = 0$ for all t or $a_i = a$ for all i and t. This is a *technological balanced growth path* because it restricts only technologies but not preferences. For example, preferences can be asymmetric, and/or the elasticities of substitution can vary. A technological balanced growth path exists if increasing returns are identical across all industries.

2. $\overline{1} + \varepsilon_t = 0$, or equivalently, $\epsilon_{ii} = -1$ for all i and $\epsilon_{ij} = 0$ for all $j \neq i$. This is a *demand-driven balanced growth path* because it only restricts preferences, but no technologies. For example, increasing returns may be explosive for some industries while they may die out for other industries. This restriction implies that u is of the Cobb-Douglas type so that the elasticity of substitution between goods is equal to 1.

3. Finally, if $\triangle a_t \neq 0$ and $\overline{1} + \varepsilon_t \neq 0$ then the condition (6) implies that 1 is an eigenvalue of ε_t, and $\triangle \alpha_t$ is an associated eigenvector. This is a knife-edge *balanced growth path* because such path only exists under a delicate combination of technologies and preferences.

The previous results are summarized in the following corollary.

Corollary 2 *A balanced growth path exists if and only if (i) $\epsilon_{ii} = -1$ and $\epsilon_{ij} = 0$ for all i, j and $i \neq j$ (preferences are of the Cobb-Douglas type); or (ii) $a_i = a$ for all i (increasing returns are identical for all industries); or (iii) 1 is an eigenvector of ε_t and $\triangle a_t$ is one of its associated eigenvectors..*

3.1.5 The city size distribution

We now study what constraints on the steady state distribution of the fundamentals, Φ^A and Φ^θ, is implied by the fact that the steady state distribution of city sizes, P, is Pareto with exponent δ. We find that Φ^A and Φ^θ must also be Pareto with exponent $\delta\lambda$, where λ is function of the parameters of the model.

To further simplify the analysis, suppose that $\varphi^i(L_{is}) = L_{is}^{a_i}$ and $u(c, \theta) = \left(\sum_i (\theta_i^\eta c_i)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}$, where η is the elasticity of substitution. Following Corollary 2, we assume that either $a_i = a$ for all i or $\eta = 1$ in order to assure the existence of a steady state distribution of city sizes.

The assumed preferences induce the following demand function for good i,

$$C_i^d = v^{\eta-1} \left(\frac{\theta_i}{q_i}\right)^\eta W, \tag{7}$$

where v is the price index of the consumption good defined as $v := \left(\sum_i \theta_i^\eta q_i^{1-\eta}\right)^{\frac{1}{1-\eta}}$. The aggregate supply of good i is equal to

$$C_i^s = A_i L_i^{1+a_i}$$

under optimal taxation. In addition, the zero profits condition implies that

$$w = q_i A_i L_i^a.$$

Using the last three equations to solve for L_i, it follows that

$$L_i = B_i \left[A_i^{\eta-1} \theta_i^\eta\right]^{\frac{1}{1+a_i(1-\eta)}},$$

where $B_i = \left[v^{\eta-1} w^{-\eta} W\right]^{\frac{1}{1+\alpha_i(1-\eta)}}$. This equation together with the equilibrium condition $N = \sum L_i$, implies that

$$\frac{L_i}{N} = \frac{B_i \left[A_i^{\eta-1}\theta_i^{\eta}\right]^{\frac{1}{1+a_i(1-\eta)}}}{\sum_i B_i \left[A_i^{\eta-1}\theta_i^{\eta}\right]^{\frac{1}{1+a_i(1-\eta)}}}.$$

Finally, under the assumption that either $a_i = a$ for all i or $\eta = 1$, the previous expression could be simplified to, adding time subscripts,

$$\frac{L_{it}}{N_t} = \frac{z_{it}/I}{\sum_i z_{it}/I}. \tag{8}$$

where $z_{it} := \left[A_{it}^{\eta-1}\theta_{it}^{\eta}\right]^{\frac{1}{1+a(1-\eta)}}$. This is the crucial expression for our purposes. The variable on the left hand side is the relative size of city i. According to the evidence, the steady state distribution of this variable is Pareto with exponent δ. It follows that the variable on the right hand side must be distributed Pareto too. Furthermore, since $\sum_i z_{it}/I$ approaches a constant for large I, then z_{it} must be distributed Pareto with exponent δ.

The following is the main result of this section.

Proposition 3 *Suppose I is sufficiently large. Then, the distribution of $\frac{L_{it}}{N_t}$ (city sizes) is Pareto if and only if the distribution of z_{it} is Pareto. Moreover, let δ be the Pareto exponent of the city size distribution. Then, (i) if only preferences are stochastic (so that $A_{it} = A$ for i and t) then Φ_t^{θ} must be Pareto with exponent $\frac{\delta\eta}{1+a(1-\eta)}$ provided that $1 + a(1-\eta) > 0$; (ii) if only technologies are stochastic (so that $\theta_{it} = \theta$ for all i and t) then Φ_t^{A} must be Pareto with exponent $\frac{\delta(\eta-1)}{1+a(1-\eta)}$, provided that $\frac{\eta-1}{1+a(1-\eta)} > 0$; (iii) If $\eta = 1$ then preferences must be stochastic with exponent δ.*

Proof. *(i) In this case $z_{it} = \theta_{it}^{\frac{\eta}{1+a(1-\eta)}}$. The pdf of z is a Pareto density, $\delta z_l^{\delta} z^{-(1+\delta)}$. Since θ is a monotonic transformation of z, then the pdf of θ is given by $\delta|\lambda|\theta_l^{\delta\lambda}\theta^{-(1+\delta\lambda)}$, where $\lambda = \frac{\eta}{1+a(1-\eta)}$ (Hogg and Craig (1995), page 169). This is the pdf of a Pareto distribution with exponent $\delta\lambda$ if $\lambda > 0$. Part (ii) of the Proposition can be proven in the same way. In that case z_{it} becomes $z_{it} = A_{it}^{\frac{\eta-1}{1+a(1-\eta)}}$. Part (iii) is immediate.* ■

This Proposition states a simple but powerful result. Under some general restrictions the fundamentals of the model must be distributed Pareto if the city size distribution is Pareto. This result is very general since only a minimum set of restrictions have been imposed. An additional constraint provided by the Proposition is that η must be bounded above by the degree of increasing returns ($\frac{1+a}{a} > \eta$.) This result indicates that if goods are easily substituted (η is too large), then the economy may end up producing only a single good in order to fully exploit the increasing returns to scale.

The final part of the Proposition states that when preferences are Cobb-Douglas, technological shocks play no role in determining the city size distribution. Instead, the distribution is completely determined by demand side shocks, such as shocks to preferences. In this case, larger cities are cities that produce the more preferred goods in the economy. In contrast, cities that produce goods with better technological levels are neither smaller nor larger. This is the result of two forces that exactly offset each other in the case of Cobb-Douglas preferences. On one hand, cities with better technologies tend to be smaller because less workers are needed to satisfy the demand for their goods. On the other hand, better technologies reduce prices and increase demand.

But why would θ or A be distributed Pareto? Proposition 2 and Lemma 5 in Cordoba [2001] characterizes a class of processes that produce this distribution. A particular case results when the exogenous variables follow proportional growth processes. If the stochastic process determining θ or A satisfies Gibrat's law, then their steady state distribution may be Pareto. More precisely, if θ (or A) follows a 'reflected geometric Brownian motion' process, then the distribution of θ (or A) converges to a Zipf's distribution (Gabaix [1999], Proposition 1). The following two results combine Proposition 1 in Gabaix with the previous Proposition. Denote G the distribution of city sizes.

Proposition 4 *(Random tastes) Suppose $A_{it} = A$ and θ follows a reflected geometric Brownian motion. Then G is Pareto with exponent $\delta = \frac{1+a(1-\eta)}{\eta}$. G is Zipf only if $\eta = 1$.*

This Proposition provides an economic interpretation for a Zipf distribution of city sizes based on stochastic tastes. It arises when $\eta = 1$ and tastes follow reflected random walks.

Proposition 5 *(Random technologies) Suppose $\theta_{it} = \theta$ and A follows a reflected geometric Brownian motion. Then G is Pareto with exponent $\delta = \frac{1}{\eta-1} - a$. G is Zipf only if $\eta - 1 = \frac{1}{1+a}$.*

This Proposition provides a technological interpretation for a Pareto distribution of cities. It arises when technological levels follow reflected random walks, and increasing returns are equal across industries.

3.2 Diversified Cities and Non-tradable goods

Cities are usually regarded as very diversified production entities but the previous model portraits cities as highly specialized. The following is an extension of the model where cities are highly diversified in the production of *nontradables*, although they still specialize in the production of *tradables*. All results from the previous section hold. Relative city sizes are still completely determined by the relative size of their tradable sectors.

Denote the goods in the previous section *tradable* goods, T. They are produced under scale economies and bear no transportation costs. In addition to tradables, there are other types of goods in the economy, called *nontradables*, which are costly to transport and can be produced under the following constant returns to scale technology.

$$y_i = l_i \text{ for } i \in NT,$$

Preferences are similar to the previous section but now they include nontradable goods,

$$u(c) = \left(\sum_{i \in T \cup NT} \left(\theta_i^{\eta} c_i \right)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}, \ \eta > 0.$$

Demand functions are still given by (7) for $i \in T \cup NT$.

We call tradables the goods that are produced under scale economies. Each one of them is produced in a single location, but consumed everywhere. Goods produced under constant returns to scale are non-tradable. No location has a particular advantage producing them and they bear prohibitive transportation costs if traded. To save in transportation costs, these goods are produced at the same place where they are demanded. As a result, cities in this model specialize in producing one tradable, but diversify in producing all nontradables.

It is now established that the relative population of any two cities is completely determined by the extent of their tradable sectors. Let $L^{NT} = \sum_i L_i^{NT}$ be the amount of labor employed in the production of nontradables, and similarly for L^T. The total population in a particular city includes workers in both activities. Let L_{is}^{NT} be the size of workers producing good $i \in NT$ at city s, and let $L_s^{NT} := \sum_i L_{is}^{NT}$ be the total size of worker producing nontradables at city s. Total population at city s is thus given as $X_s := L_s^{NT} + L_s^T$.

Proposition 6 $\frac{X_i}{X_j} = \frac{L_i^T}{L_j^T}$ *along a balanced growth path.*

Lemma 7 ***Proof.*** *Since preferences are homothetic, all demands are linear in income. In addition, relative income between any two cities is just their relative population since wages are equal across cities in a balanced growth path. Thus, relative consumption of good h between cities i and j is*

$$\frac{c_{hi}}{c_{hj}} = \frac{X_i}{X_j} \text{ for all } h.$$

From the supply side, we have $c_{hi} = L_{hi}$ for $h \in A$. Therefore, $\frac{L_{hi}}{L_{hj}} = \frac{X_i}{X_j}$ for all h, which implies,

$$\frac{L_i^A}{L_j^A} = \frac{\sum_{h \in A} L_{hi}}{\sum_{h \in A} L_{hj}} = \frac{X_i}{X_j}.$$

Now, since $\frac{X_i}{X_j} = \frac{L_i^B + L_i^A}{L_j^B + L_j^A}$, *it follows that* $\frac{L_i^A}{L_j^A} = \frac{L_i^B}{L_j^B}$. *Thus,* $\frac{L_i^B}{L_j^B} = \frac{X_i}{X_j}$. ■

According to this Proposition one can safely ignore nontradables when determining relative city sizes, but still can interpret cities as diversified production places.

3.3 A Model with Capital

The previous models have abstracted from capital, either physical or human. Externalities, however, are usually associated with the amount of human capital in the city. There is a simple way to introduce capital in the model that leaves the previous results intact. Suppose the production function for tradable goods is given by

$$y_{is} = \varphi_i(K_{is}, L_{is}) l_{is}^{1-\alpha_i} k_{is}^{\alpha_i} \text{ for } i \in B, \tag{9}$$

where K_{is} is aggregate capital employed in the production of good i at location s, and k_i is individual capital. Suppose there is a rental market for capital and capital can be moved between locations without cost. Let r be the rental rate and w the wage rate. Profit maximization requires the relative prices of capital and labor to be equal to the relative productivities, i.e., $\frac{r}{w} = \frac{\alpha_i}{1-\alpha_i}\frac{l_{is}}{k_{is}}$ or

$$k_{is} = \frac{\alpha_i}{1-\alpha_i}\frac{w}{r} l_{is}$$

Moreover, city specialization implies $K_{is} = \frac{\alpha_i}{1-\alpha_i}\frac{w}{r} L_{is}$. Replacing these two expression into the production function (9), one obtains:

$$y_{is} = A_i \varphi_i\left(\frac{\alpha_i}{1-\alpha_i}\frac{w}{r} L_{is}, L_{is}\right)\left(\frac{\alpha_i}{1-\alpha_i}\frac{w}{r}\right)^{\alpha_i} l_{is} \text{ for } i \in B,$$

or

$$y_{is} = \tilde{\varphi}_i(L_{is}, w/r) l_{is} \text{ for } i \in B$$

which has the same functional form as the one in Section 4.1. The inclusion of w/r into $\tilde{\varphi}_i$ does not affect the previous results because it is identical across cities.

4 Cities limited by negative externalities

In this Section we alter the model so that the limitation on city growth arises from negative externalities rather than the extent of the market. Thus, city size in this section is determined by the optimal trade-off between positive and negative externalities. As in the previous section, we seek conditions so that all efficient city sizes grow at the same rate. However, a feature of these type of models is that not all cities can attain an optimal size, and therefore not all cities can grow at the same rate.

This is because nothing in the model guarantees enough population so that all cities attain their optimal size. Cities that do not attain their optimal size are expected to grow faster until they reach their optimal size. As a result, Gibrat's law cannot hold in these models and, in general, a Pareto distribution for city sizes cannot be obtained. These models, however, can produce an approximate Pareto distribution. To investigate this, we follow the strategy proposed by Rossi-Hansberg and Wright [2003]. They allow for a non-integer number of cities so that all cities can attain an optimal size. Formally, this is only an approximation because if the number of cities is not an integer, then it means that one city cannot attain its optimal size[11]. However, this assumption provides a good approximation for the actual city size distribution if the number of cities specializing in each industry is large.

We introduce two main modifications to the model of Section 3. First, we explicitly introduce capital in the production function. Firm j in industry i at location s now produces output according to the production function

$$y_{jis} = \varphi^i(X_{is}, X_s)x_{jis}$$

where $\varphi^i(X_{is}, X_s)$ is the total factor productivity (TFP) of industry i at location s, $x_{jis} = k_{jis}^{\alpha_i} l_{jis}^{1-\alpha_i}$ for factor inputs k_{jis} and l_{jis} represent the output net of TFP (unaugmented output), $X_{is} = \sum_j x_{jis}$ is the total unaugmented output of industry i in location s, and $X_s = \sum_i X_{is}$ is the unaugmented output in location s. Notice that the production function in equation (1) is a special case where $\alpha_i = 0$. For $\alpha_i > 0$, this formulation generalizes the formulation in Section 3.1, and allows to handle capital or other factors easily. As in Section (3.1), φ^i depends positively on X_{is}, which describes the positive local externalities from industry agglomeration, and negatively on X_s, which describes the negative urban externalities. We further assume the following particular functional form for total factor productivity:

$$\varphi^i(X_{is}, X_s) = A_i X_{is}^{a_i} - \lambda_i X_s^{b_i},\ b_i > a_i, \tag{10}$$

We assume that λ_i is a positive constant, while $A_i > 0$ is the only random variable in the model. As we will see below, in order to allow for economic growth into this model, we assume that $A_i = \widetilde{A}_i e^{gt}$ where g is an instantaneous growth rate and $\widetilde{A}_i$ is a stationary variable[12]. In particular, $\widetilde{A}_i$ follows a Markov process with transition probability $\phi^A(x_0, t_0; x, t)$, common across industries. Denote Φ_t^A the time t distribution of $\widetilde{A}_i$ and assume that that ϕ^A has a unique invariant distribution, Φ^A.

A second change is that we now think of industries as producing intermediate inputs rather than final outputs. Inputs are then transformed into final output using the following constant returns to scale technology

[11] See footnote 13 for an example.

[12] In the previous section, percapita output grew as a result of the positive externalities associated to population growth. In this section, positive and negative externalities offsets each other so that an additional force to produce economic growth is required.

$$Y \equiv \prod_{i=1}^{I} Y_i^{1/I}, \tag{11}$$

where $Y_i = \sum_{js} \varphi^i(X_{is}, X_s)x_{jis}$ is the aggregate supply of input from industry i.

4.1 The Planner's problem

We only consider the efficient allocation of resources in this economy. This allocation can be decentralized via Ramsey taxation, as in the previous section, or via competitive developers, as in Black and Henderson (1999). Efficient cities specialize in the production of a single input to avoid unnecessary negative externalities without any benefit. As a result, the total factor productivity in a location s that specializes in good i is given by $\varphi^i(X_{is}, X_{is})$. Let M_i denote the number of cities producing input i.

The size of an industry in different locations could differ. However, if there is an efficient industry size in terms of X_i, denoted by X_i^*, the planner would like to choose $X_{is}^* = X_i^*$ in all locations producing input i. In that case, the aggregate supply of input i would be $M_i^* \varphi_i(X_i^*)X_i^*$. Following Rossi-Hansberg and Wright [2003], we assume that M_i^* can take a non-integer value. This simplifies the problem as all cities in industry i can attain their efficient size and supply and demand can be equated by choosing the right M_i^*. This assumption is formally inconsistent with the problem because a non-integer number of cities means that there is a city that cannot attain its optimal size[13]. However, this assumption provides a good approximation if the number of cities in each industry is large, since the city of suboptimal size becomes a trivial fraction of all cities.

An efficient allocation is the solution to the following planner's problem.

$$\max_{Y_t, Y_{it}, M_{it}, X_{it}, K_{it}, L_{it}, C_t, K_t} E_0 \int e^{(\gamma-\rho)t} u\left(C_t/N_t\right) dt$$

subject to

$$\prod_{i=1}^{I} Y_{it}^{1/I} \geq \delta K_t + \dot{K}_t + C_t \tag{12}$$

$$Y_{it} = M_{it}\varphi_i(X_{it})X_{it} \tag{13}$$

$$X_{it} = K_{it}^{\alpha_i} L_{it}^{1-\alpha_i} \tag{14}$$

[13] For example, suppose $\varphi(X^*)X^* = 1$ so that each city in a particular industry produces 1 unit of good (a normalization that is possible by picking the right constants in φ). Moreover, suppose that $M^* = 10.5$. (any non-integer will work.) In this case, the total industry supply, $M^*\varphi(X^*)X^*$, is 10.5 units. Thus, the industry has 10 cities of optimal size producing 10 units and one city producing 0.5 units. The size of the last city must satisfy $\varphi(X)X = 0.5$. Note that $X = X^*$ does not satisfy this equation, which implies that the last city cannot have an optimal size.

$$K_t \geq \sum M_{it} K_{it} \tag{15}$$

$$N_t \geq \sum M_{it} L_{it} \tag{16}$$

The first restriction of the problem is the resource constraint. It states that aggregate output can be used for gross investment or consumption. The other restrictions have been already discussed.

Denote p_t, q_{it}, s_{it}, r_t and w_t the Lagrange multipliers associated to the restrictions (12) to (16). The first order necessary conditions with respect to Y_{it}, X_{it}, K_{it}, L_{it} and M_{it} are given by:

$$\frac{p_t Y_t}{Y_{it} I} = q_{it} \tag{17}$$

$$q_{it} M_{it} \varphi_i(X_{it}) + q_{it} M_{it} \varphi_i'(X_{it}) X_{it} = s_{it} \tag{18}$$

$$s_{it} \alpha_i X_{it} = r_t K_{it} M_{it} \tag{19}$$

$$s_{it} \left(1 - \alpha_i\right) X_{it} = w_t L_{it} M_{it} \tag{20}$$

$$q_{it} \varphi_i(X_{it}) X_{it} = r_t K_{it} + w_t L_{it}. \tag{21}$$

4.2 Efficient allocations

The following Proposition describes efficient city sizes, number of cities, and final production.

Proposition 8 *The efficient amount of aggregate production of final goods, industry size, and number of locations in each industry are given by*

$$Y_t^* = K_t^{\alpha} \left(B_t N_t\right)^{1-\alpha} \tag{22}$$

$$L_{it}^* = \left(\frac{N_t}{K_t}\right)^{\alpha} \left[\frac{a_i A_{it}}{b_i \lambda}\right]^{\frac{1}{b_i - a_i}} \tag{23}$$

and

$$M_{it}^* = \frac{N_t / I}{L_{it}^*}. \tag{24}$$

where $B_t = \left[\prod_{i=1}^{I} \left(\frac{b_i - a_i}{b_i I} \left(\frac{a_i A_{it}}{b_i \lambda}\right)^{\frac{a_i}{b_i - a_i}}\right)\right]^{\frac{1}{I(1-\alpha)}}$.

Proof. *See Appendix.* ■

Equation (23) states that the efficient city size, L_{it}^{*}, is a positive function of the technological level of the industry, A_{it}, and the positive externalities, a_i, and a negative function of the aggregate capital-labor ratio, and the negative externalities, b_i.

Now consider the conditions required for the existence of a balanced growth path along which all cities grow at the same rate. For this purpose, suppose momentarily that there is no uncertainty. As argued in Section (3.1.4), the existence of a balanced growth path is a precondition to generate a Pareto distribution of city sizes. It guarantees that Gibrat's law is satisfied in the deterministic version of the model. Along a balanced growth path (K/N) and A grow at constant rates. Thus, according to equation (23), all cities can grow at a common rate only if $a_i = a$ and $b_i = b$, that is, if the strength of the positive and negative externalities are equal across industries[14]. In contrast with the results of the previous section, this parameter restriction is required despite the fact that the elasticity of substitution between goods is 1. This important result is summarized in the following corollary of Proposition 8.

Corollary 9 *A balanced growth path of city sizes exists only if $a_i = a$ and $b_i = b$.*

This corollary complements Corollary 2. Together they provide a sharp restriction on urban models. They state that independently of whether the size is limited by the extent of the market or by congestion costs, the extent of the externalities must be identical across industries for a balanced growth path to exist. Most current urban models do not impose this restriction, which partly explains their difficulties in accounting for the evidence of city size distribution.

The first part of Proposition 8 states that the aggregate production function is constant returns to scale, a result also found by Rossi-Hansberg and Wright [2003] in a related model. The form of the production function implies that economic growth results only from technological progress, but not from population growth. It is a standard exercise to show that the steady state growth rate of per-capita output and per-capita capital is equal to the growth rate of B, the labor augmenting technological level. The growth rate of B in turn is a weighted average of industry specific technological growth. However, under the restriction provided by Proposition 9, one can write $B_t = e^{g\frac{1}{(1-\alpha)}\frac{a}{b-a}t}\widetilde{B}_t$ where $\widetilde{B}_t = \prod\left(\left(\frac{b-a}{bI}\right)\left(\frac{a\widetilde{A}_{it}}{b\lambda}\right)^{\frac{a}{b-a}}\right)^{\frac{1}{(1-\alpha)I}}$. For a sufficiently large number of industries, $\widetilde{B}_t$ is approximately constant so that the instantaneous growth rate of B, and the growth rate of the economy, is $\frac{g}{(1-\alpha)}\frac{a}{b-a}$.

[14]A related expression (23) holds if the parameters $\alpha_i^{\prime}s$ are different across industries. One can then infer from that expression that $\alpha_i = \alpha$ is required for the existence of a balanced growth path. Otherwise, changes in the aggregate capital-labor ratio will affect cities differently.

There is another restriction of parameters that is required in order for cities to grow over time, as is the case in data. From (23), it follows that

$$\begin{aligned}\frac{dL_{it}^{*}/dt}{L_{it}^{*}} &= -\alpha\frac{d\left(K/L\right)/dt}{K/L}+\frac{1}{b_i-a_i}\frac{dA_{it}/dt}{A_{it}}\\ &= -\alpha\frac{g}{(1-\alpha)}\frac{a}{b-a}+\frac{g}{b_i-a_i}\text{ for large } I.\end{aligned}$$

Thus, L_{it}^{*} grows over time only if $\frac{1-\alpha}{\alpha}>a$. This limits the degree of increasing returns. Moreover, this condition shows that existing cities will not grow unless A increases, even if population is increasing.

4.3 The city size distribution

Suppose that $a_i=a$ and $b_i=b$, as required by Proposition 9. We can now state the main result of this section. The following Proposition states that a Pareto distribution of city sizes is equivalent to a Pareto distribution of the underlying productivities. As before, denote P_t the cumulative distribution of city sizes at time t.

Proposition 10 *Suppose I is sufficiently large. Then, P_t is Pareto with Pareto exponent δ if and only if Φ_t^A is Pareto with Pareto exponent $\frac{1+\delta}{2(b-a)}$.*

Proof. *See Appendix.* ■

This Proposition is analogous to Proposition 3. It provides a strong condition on the fundamentals of the model needed to produce a Pareto distribution of city sizes. In particular, Φ^A must be Pareto. One can then proceed, as in the previous section, to provide examples of Markov processes with invariant Pareto distributions.

4.4 Generalizations

At least two generalizations to the previous models are possible. One possibility is to allow $\varphi(X)$ not to depend on the level of production directly but on K and N. This will, for example, allow the externalities to be associated with N more than K. The following functional form provides an additional role for labor externalities and preserves the results of the previous section:

$$\varphi(X,N)=AX^{a}N^{(\alpha+a\sigma)(1-1/\sigma)}-\lambda X^{b}N^{(\alpha+b\sigma)(1-1/\sigma)}$$

where $1\leq\sigma\leq\infty$. Notice that $\sigma=1$ is the case analyzed in the previous section. Moreover, as σ increases, the positive and negative externalities associated with labor become more important.

A second alternative is to allow allocations to be inefficient. For example, we found that if the planner only takes into account the labor externality but not the capital externality, then a more general functional form for φ is possible. In that case, φ can depend on (K,N) in a rather general form.

5 Conclusions

A robust empirical regularity is that the city size distribution in many countries is well approximated by a Pareto distribution. This paper provides a set of restrictions that urban models must satisfy in order to be consistent with this regularity. More precisely, it derives restrictions on preferences, technologies, and on the stochastic properties of the exogenous driving forces of standard urban models.

We find that under general conditions the steady state distribution of the exogenous driving force must be Pareto. Moreover, we find that if city size is determined by the extent of the market, one of two additional conditions is required. Either, the elasticity of substitution between goods must be equal to one and preferences must be stochastic, or externalities must be equal across goods and technologies must be stochastic. Finally, in a model in which city size is determined by binding negative externalities, we find that both a unit elasticity of substitution between goods and equal degrees of positive and negative externalities across industries are required.

We have examined specific urban models with economies of localization. We conjecture that our main results are far more general than this. The reason is that we are looking for balanced growth conditions using a particular multisectorial model. A similar exercise conducted by King, Plosser, and Rebelo (1988) using a single sector representative agent model has proven to be very general and applicable to most economic growth models, even if growth is endogenous and the competitive equilibrium is inefficient.

6 References

Auerbach, F. (1913), "Das Gesezt der Bevölkerungskoncentration," Petermanns Geograpische Mitteilungen, LIX: 73-76.

Black, D. and Henderson, J.V. (1999), "A Theory of Urban Growth," *Journal of Political Economy 107: 252-84.*

Brakman, S., H. Garretsen, Ch. van Marrewijk and M. van den Berg (1999), "The Return of Zipf: Towards a further understanding of the Rank-Size Rule," Journal of Regional Science 39: 183-213.

Carroll, Glenn R. (1982), "National city size distributions: what do we know afeter 67 years of research?," *Progress in Human Geography.* 6(1): March, 1-43.

Champernowne, D. (1953), "A Model of Income Distribution," *Economic Journal*, LXIII: 318-351.

Córdoba, Juan-Carlos (2001), "Cities and Growth." Ph.D Dissertation, University of Rochester, July.

Córdoba, Juan-Carlos (2004), "A Generalized Gibrat's Law for Cities." Mimeo, Rice University.

Dobkins, L. H. and Y. M. Ioannides (2000), "Evolution of Size Distribution: U.S. Cities," in *Economics of Cities*, J. M. Huriot and J-F Thisse, University of Cambridge.

Duranton, G. (2002), "City size distribution as a consequence of the growth process", Mimeo, London School of Economics.

Eaton, J. and Zvi Eckstein (1997), "Cities and growth: Theory and evidence from France and Japan", *Regional Science and Urban Economics 27*: 443-474.

Fujita, Masahisa, Paul Krugman and Anthony J. Venables (1999), *The Spatial Economics: Cities, Regions, and International Trade*, (Cambridge, MA: The MIT Press).

Fujita, Masahisa, J.M. Thisse (2000), "The Formation of Economic Agglomerations:Old Problems and New Perspectives," in *Economics of Cities: Theoretical Perspectives,* Huriot, Jean-Marie and Jacques-François Thisse (Editors), (Cambridge, UK: Cambridge University Press).

Fujita, Masahisa, Mori Tomoya (1997), "Structural Stability and Evolution of Urban Systems," *Regional Science and Urban Economics* 27: 399-442.

Gabaix, Xavier (1999), "Zipf's Law for Cities: An Explanation," *Quarterly Journal of Economics 114(3):* 739-766.

Glen-Mann, M. (1994), *The Quark and the Jaguar* (New York, NY: Freeman)

Gibrat, R.(1931), *Le inégalités économiques* (Paris, France: Librairie du Recueil Sirey).

Henderson, J.V. (1974), "The sizes and types of cities," *American Economic Review 64: 640-56.*

Henderson, J.V. (1988), *Urban Development. Theory, Facts and Illusion,* (Oxford University Press).

Huriot, Jean-Marie and Jacques-François Thisse (Editors) (2000), *Economics of Cities: Theoretical Perspectives* (Cambridge, UK: Cambridge University Press).

Ioannides, Y. M. and X. Gabaix (2003), "The Evoluation of City Size Distributions," in Handbook of Urban and Regional Economics, J.V. Henderson and J-F Thisse (eds).

Ioannides, Y. M. and H. G. Overman (2003), "Zipf's Law for Cities: An Empirical Examination," *Regional Science and Urban Economics* 33: 127-137.

Kalecki, M. (1945), "On the Gilbrat Distribution," *Econometrica* 13: 161-170.

King, R.G, C.I. Plosser and S. Rebelo (1988), "Production, Growth and Business Cycles: I. Basic Neoclassical Model," *Journal of Monetary Economics* 21: 195-232.

Krugman, Paul (1996), *The self-organizing economy,* Blackwell Publisher.

Lucas, Robert (1988), "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22: 3-42.

Roehner, M. B. (1995), "Evolution of Urban Systems in the Pareto Plane," *Journal of Regional Science* 35(2): 277-300.

Rosen, K. T., and M. Resnik (1980), "The Size Distribution of Cities: An Examination of Pareto Law and Primacy," *Journal of Urban Economics* 8: 165-186.

Rossi-Hansberg, E. and M. L. Wright (2003), "Urban structure and growth", Institute for Empirical Economics, Discussion Paper 141.

Segal, T. (1976), "Are There Returns to Scale in City Size?," *The Review of Economics and Statistics 58*: 339-350.

Simon, H. (1955), "On a Class of Skew Distribution Functions," *Biometrica*, XXLII: 425-440.

Simon, H. and C. P. Bonini (1958), "The Size Distribution of Business Firms," *The American Economic Review 48:* 607-617.

Soo, Kwok Tong (2002), "Zipf's Law for cities: A cross country investigation," manuscript, London School of Economics.

Zipf, G.(1949), *Human Behavior and the Principle of Least Effort*, (Cambridge, MA: Addision-Wesley)

Appendix

Proof of Proposition 8. We first prove the second and third result. Adding (19) and (20), and using (21), it follows that $s_{it} = q_{it}M_{it}\varphi_i(X_{it})$. Substituting this result into (18), one obtains the intuitive result $\varphi_i'(X_{it}^*) = 0$ (the efficient city size, in terms of X, is the one that maximizes total factor productivity.) Using (10), and this condition, the efficient size is given by

$$X_{it}^* = \left[\frac{a_i A_{it}}{b_i \lambda}\right]^{\frac{1}{b_i - a_i}}. \tag{25}$$

This solves for the city size in terms of X. To solve for city size in terms of labor, notice that (19) and (20) imply

$$K_{it}^* = \frac{\alpha}{1-\alpha}\frac{w_t}{r_t}L_{it}^*, \tag{26}$$

which together with (14) and (25) produces

$$L_{it}^* = \left(\frac{1-\alpha}{\alpha}\frac{r_t}{w_t}\right)^{\alpha} X_{it}^*.$$

Multiplying (26) by M_{it}^*, adding up across industries, and using the aggregate resource constraints (15) and (16), one can solve for prices

$$\frac{w_t}{r_t} = \frac{1-\alpha}{\alpha}\frac{K_t}{N_t} \tag{27}$$

Finally, substituting this price ratio into the previous equation one obtains the efficient city size in industry i. As for the number of cities, note from (17) and (13) that one obtains

$$M_{it}^* = \frac{p_t Y_t / I}{q_{it}\varphi_i(X_{it}^*)X_{it}^*}.$$

Now, according to (21) and (26), $q_{it}\varphi_i(X_{it})X_{it} = \frac{1}{1-\alpha}w_t L_{it}^*$. Substituting this result into the previous expression gives $M_{it}^* L_{it}^* = (1-\alpha)\frac{p_t Y_t/I}{w_t} = M_{jt}^* L_{jt}^*$. Adding up across industries, it follows that $M_t = \sum_i M_{it}L_{it} = I \cdot M_{jt}^* L_{jt}^*$. Solving for M_{jt}^* gives the result. Finally, to obtain the amount of aggregate production, note that

$$\begin{aligned} Y_{it} &= M_{it}\varphi_i(X_{it}^*)X_{it}^* \text{ (Eq. 13)} \\ &= M_{it}\left(A_iX_i^{*a} - \lambda X_i^{*b}\right)X_{it}^* \text{ (Eq. 10)} \\ &= M_{it}\left(1-\frac{a_i}{b_i}\right)X_{it}^{*1+a_i} \text{ (Eq. 25)} \\ &= \frac{N_t/I}{L_{it}^*}\left(1-\frac{a_i}{b_i}\right)X_{it}^{*1+a} \text{ (Eq.24)} \\ &= \frac{N_t/I}{\left(\frac{N_t}{K_t}\right)^{\alpha}}\left(1-\frac{a_i}{b_i}\right)X_{it}^{*a_i} \text{ (Eq.23)} \\ &= K_t^{\alpha}N_t^{1-\alpha}\left(1-\frac{a_i}{b_i}\right)X_{it}^{*a_i}/I \end{aligned}$$

Aggregate production is thus given, according to (11), by

$$= K_t^{\alpha}N_t^{1-\alpha}\prod\left(\left(\frac{b_i-a_i}{b_iI}\right)\left(\frac{a_iA_{it}}{b_i\lambda}\right)^{\frac{a_i}{b_i-a_i}}\right)^{1/I}.$$

■

Proof of Proposition 10. Suppose P_t is Pareto with exponent δ. We need to prove that Φ_t^A is Pareto. Denote $F_t(\cdot)$ the distribution of L_i^* at time t, and $f_t(\cdot)$ its density. Thus, $F_t(L) \equiv \Pr(L_{it}^* \leq L)$. In addition, $M_{it}^* = \frac{\zeta_t}{L_{it}^*}$ (from 24) where $\zeta = L_t/I$. Denote $H_t(M) \equiv \Pr(M_{it}^* \leq M)$ and h_t its density. Note that

$$\begin{aligned} H_t(M) &= \Pr(\frac{\zeta_t}{L_{it}^*} \leq M) \\ &= \Pr(\frac{\zeta_t}{M} \leq L_i^*) \\ &= 1 - F_t(\frac{\zeta_t}{M}) \end{aligned}$$

and that $h_t(M) = f_t(\frac{\zeta_t}{M})\frac{\zeta_t}{M^2}$. Denote S_{it} the random size of a city and note that $P_t(S) \equiv \Pr(S_{it} \leq S)$. P_t in the model satisfies

$$P_t(S) = \frac{\int^S f_t(L)h_t(\frac{\zeta_t}{L})dL}{\int^{\infty} f_t(L)h_t(\frac{\zeta_t}{L})dL}.$$

Note that

$$h(\frac{\zeta}{L}) = f(\frac{\zeta}{\frac{\zeta}{L}})\frac{\zeta}{\left(\frac{\zeta}{L}\right)^2} = f(L)\frac{x^2}{\zeta}$$

so that $P_t(S) = \int^S \frac{f_t(L)^2 L^2}{\Theta_t \zeta_t} dL$ where $\Theta_t = \int^\infty f_t(L) h_t(\frac{\zeta_t}{L}) dL$. Finally, we also know that P_t is a Pareto distribution, so that $p_t(S) = \delta S_{lt}^{\delta} S^{-(1+\delta)}$ where S_{lt} is the minimum city size at time t. From the above expression, it follows that $\frac{f_t(L)^2 L^2}{\Theta_t \zeta_t} = \delta S_{lt}^{\delta} L^{-(1+\delta)}$ so that solving for $f_t(L)$ one obtains $f_t(L) = (\Theta_t \zeta_t \delta)^{1/2} S_{lt}^{\delta/2} L^{-(1+\frac{1+\delta}{2})}$. This is just the density of a Pareto distribution with exponent $\gamma = \frac{1+\delta}{2}$. Picking Θ properly, we find that

$$f_t(L) = \gamma L_{lt}^{\gamma} L^{-(1+\gamma)}$$

where L_{lt} is the minimum optimal size. Given that $f_t(L)$ is Pareto, we know show that Φ_t^A is Pareto. We can write (23) as $L_{it}^* = D_t A_{it}^{\frac{1}{b-a}}$ where D_t is a deterministic component (given that I is large). Then,

$$\begin{aligned}
\Phi_t^A(A) &= \Pr\left(A_{it} \leq A\right) = \Pr\left(\left(\frac{L_{it}}{D_t}\right)^{b-a} \leq A\right) \\
&= \Pr\left(L_{it} \leq D_t A^{\frac{1}{b-a}}\right) = F_t\left(D_t A^{\frac{1}{b-a}}\right) \\
&= 1 - \left(\frac{D_t A^{\frac{1}{b-a}}}{L_{lt}}\right)^{-\gamma},
\end{aligned}$$

which is a Pareto distribution with exponent $\frac{\gamma}{b-a} = \frac{1+\delta}{2(b-a)}$. This proves the first part of the Proposition. In order to prove that P is Pareto given that Φ^A is Pareto, one just just needs to follow the same steps in reverse order. ■