On the Distribution of City Sizes

Juan-Carlos Córdoba[*]
Rice University

February 14, 2003

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

** First version, November 2000. This paper initially circulated under the title "Zipf's Law: A Case Against Scale Economies?" I thank Per Krusell for his insightful advice, Alan Stockman, Mark Bils, Michael Cranston, Tony Smith, Esteban Rossi-Hansberg, and participants in seminars at the universities of Rochester, Pensilvania, Carnegie Mellon, University of Pittsburgh, Western Ontario, Cornell, Northwestern, and the London School of Economics. I acknowledge financial support provided by el Banco de la Republica de 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.

The literature, however, has been unsuccessful in providing a 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]).

The objective of this paper is to find restrictions that standard urban models must satisfy in order to generate Pareto distributions for city sizes, and then use these restrictions to provide economic explanations for the regularity. The restrictions are obtained in a two-step process. The first step characterizes Markov processes that can preserve Pareto distributions. The second step uses this characterization to find restrictions on preferences, technologies, and on the stochastic properties of the exogenous driving forces in a standard urban model.

As expected, there are many Markov processes that can preserve Pareto distributions. Among them are the well-known cases of processes that satisfy Gibrat's law, or the law of proportionate effects[3] (Champernowne [1953], Gabaix, [1999]). More importantly, they also include non-proportional growth processes where size affects growth.

The important result is that, under plausible assumptions about population growth and the number of cities, only a very specific Markov process can support Pareto

[1] The observation is usually associated to Zipf [1949], but Auerbach [1913] seems to be 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], and Gabaix [1999], Ioannides and Overman [2000], and Soo [2002].

[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.

[3] Gibrat's law describes processes that are size independent in the sense that the expected growth rate and the variance of the growth rate are independent of the position of the random variable.

distributions. Such process is proportional in mean but not in variance. The non-proportionality of the variance is required to preserve the Pareto exponent. In particular, Pareto distributions with larger exponents (more unequal distributions) require more volatile growth processes. This provides a new interpretation for the Pareto exponent, one that complements Simon [1955] who links the Pareto exponent to the fraction of new agents in the system (new cities).

The finding also allows to generalize other important result in the literature. Previous studies (Champernowne [1953], Simon [1955], Gabaix [1999].) have shown that proportional growth can explain Zipf's distributions -Gibrat's law implies Zipf's law-. This paper shows the reverse: under plausible conditions, Zipf's law implies Gibrat's law.

Armed with this statistical characterization, the paper moves to study under what conditions a standard urban model with localization economies can generate Pareto distributions for city sizes. The key statistical result employed is that all cities must have the same expected growth rate. This property is equivalent to require the model to have a balanced growth path. There are three main results in the second part of the paper. First, only under one of the following three conditions city growth is independent of city size: (i) the elasticity of substitution between goods is equal to one; (ii) externalities are equal across goods; or (iii) a knife-edge condition on preferences and technologies is satisfied. Second, under general conditions, the steady state distribution of the fundamentals, preferences and technologies for different goods, must be Pareto too. Third, the Pareto exponent depends on the degree of increasing returns in the economy and the elasticity of substitution between goods.

The statistical and economic characterization obtained in the paper can then be easily used to provide economic explanations for the observed distribution of city sizes. We provide at least two explanations. A standard urban model with localization economies can generate a Pareto distribution of city sizes if (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. These explanations are the first in the literature to be fully consistent with increasing returns to scale, a central component in models of agglomeration.

To summarize, the main contributions of the paper are the following: (i) it provides general conditions under which standard urban models can generate Pareto distributions for city sizes; (ii) it provides the first fully stochastic urban models in the literature based on increasing returns that can replicate Pareto distributions for city sizes; (iii) it characterizes a very parsimonious stochastic process that can preserve Pareto distributions; (iii) it provides statistical and economic interpretations for the Pareto 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 sets up the statistical model and obtains the statistical results of the paper. Section 4 sets up an urban model and uses the statistical results to characterize preferences, technologies and the random properties of the model required to account for the evidence. Section 5 concludes.

2 Evidence and Related Literature

There is an 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 confirm their claims. Eaton and Eckstein [1997] analyses the cases of France and Japan, Brakman *et al.* [1999] the Netherlands, Roehner [1995] several countries, and Ioannides and Overman [2000] study in detail the case of the United States. These exercises usually find the Pareto exponent for the U.S. close to 1, but different from 1 for most other countries.

Several probabilistic and 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]. 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 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.

In contrast to the success of this probabilistic approach, economic models have failed to match the evidence. Krugman [1996] and Fujita *et al.* [1999] conclude that none of them can properly explain the data. Most city models are deterministic which cannot account for the observed mobility of cities. In addition, these models usually predict that cities attain an equilibrium size, as a result of the interplay between positive and negative externalities. The models also predict that urban growth mainly occurs through the increase in the number of cities. This prediction conflicts both with the idea of proportional growth, as older cities must grow at a lower rate, and with the observation that the number of cities stabilizes as the urban system matures (Eaton and Eckstein, [1997]).

Some success in matching the evidence is obtained in two recent works by Eaton and Eckstein [1997] and Black and Henderson [1999]. They offer deterministic urban models that display a steady state in which all cities grow at the same rate. These works, however, require unappealing assumptions on the primitives of their models. Eaton and Eckstein require a discount factor equal to zero, and Black and Henderson need unusual functional forms for preferences and technologies. This paper provides clear, simple and general conditions.

Another drawback in the current literature is that it cannot account for Pareto exponents different from 1 when the number of cities remains constant. These cases seem relevant in light of the evidence presented by Rosen and Resnik [1980], and the fact that new cities hardly arise in many of these countries. This paper offers an explanation for these cases.

3 Reduced Form Models

We interpret the city size distribution evidence as describing an urban system that is evolving along a balanced growth path. This section seeks to characterize Markov processes that can preserve such equilibrium path. Readers can jump to the next section without major disruption.

3.1 Basic Assumptions

Consider an economy composed by a continuum of cities (locations) of size S_t, and a total population, N_t, that grows continuously over time at the exogenous compound rate $\gamma \geq 0$, i.e., $N_t = e^{\gamma t}$. Let $X_t(i)$, $i \in [0, S_t]$ be a collection of city sizes, defined as population size, at time t.

The first assumption states that the urban system evolves via the intensive margin, i.e. larger cities, rather than the extensive margin, i.e. more cities. It is consistent, for example, with evidence that suggests that the number of cities remains roughly constant in mature urban systems like Japan, France, and England (Eaton and Eckstein, 1997).

Assumption 1: $S_t = S$.

This assumption is convenient but it is not essential. The results still hold as long as S_t grows slowly[4].

The second assumption states that $X_t(i)$ follows a particular Markov process. This is in fact a statement about the equilibrium dynamics of an unspecified economic model. The next section studies an economic model with this reduced form, but the results in this section apply to any economic model with this reduced form.

Assumption 2: $X_t(i)$ *is an i.i.d diffusion process with stationary transition. In*

[4] The results of Gabaix [1990, section III.3] about new cities apply for our model. The idea is that urban system still grows mainly in the intensive margin than the extensive margin.

particular, the drift, $\mu(x)$, and the diffusion, $\sigma^2(x)$, coefficients only depend on the size of $X_t(i)$ but not on its identity, i.

This assumption has several components. First, it asserts that the dynamics of $X_t(i)$ depend only on the current position of $X_t(i)$. These dynamics could in principle depend on the whole distribution of city sizes at time t. However, such distribution is constant along a balanced growth path since we assume no aggregate shocks and S large (a continuum). It could also depend on the current realization of the idiosyncratic shock. The assumption of i.i.d shocks is made mainly for tractability, and we conjecture that the results do not depend critically on this assumption.

Assumption 2 also states that the identity of a city plays no role on its dynamics. This seems clearly required if one hopes to find a general theory of cities. The alternative is somewhat arbitrary. One would need to pose a theory for each city and explain why and how cities move across the distribution. Given that identity is irrelevant, denote X_t the size of a representative city.

Finally, assumption 2 asserts that X_t is a diffusion. This assumption dramatically simplifies the problem. Suppose momentarily that population is constant and X_t follows a discrete Markov process described by the following Markov chain

$$\Pi = \begin{bmatrix} \pi_{00} & \pi_{01} & \pi_{02} & . \\ \pi_{10} & \pi_{11} & \pi_{12} & . \\ \pi_{20} & \pi_{21} & \pi_{22} & . \\ . & . & . & . \end{bmatrix},$$

where π_{ij} is the transition probability from state i to state j. The goal is to characterize Π given that $p = p\Pi$ holds, where p is the density of the Pareto distribution. Π cannot be fully identified with only this information since the dimensionality of Π is the square of the dimensionality of p.

A way to reduce the number of unknowns is to assume that X is "continuous" in the sense that it can only move to states adjacent to the current position in one period. In that case, Π would look like

$$\Pi = \begin{bmatrix} 1-\theta_0 & \theta_0 & 0 & 0 & . \\ \phi_1 & 1-\theta_1-\phi_1 & \theta_1 & 0 & . \\ 0 & \phi_2 & 1-\theta_2-\phi_2 & \theta_2 & . \\ 0 & 0 & \phi_3 & 1-\theta_3-\phi_3 & . \\ . & . & . & . & . \end{bmatrix} \quad ((*))$$

Imposing this type of continuity dramatically reduces the dimensionality of the problem. The number of unknowns in Π is now $(2D-1)$, where D^2 is the dimension of Π, rather than $D^2 - D$. On the other hand, there are $D-1$ equations (obtained from the relation $p = p\Pi$), so that at most $\theta(\cdot)$ can be solved as function of $\phi(\cdot)$, or vice versa. Alternatively, one can use the analytical probabilities to compute the conditional mean and variance of the *growth rate* of X, $\mu(\cdot)$ and $\sigma^2(\cdot)$, and solve the problem in terms of $\mu(\cdot)$ and $\sigma^2(\cdot)$. Taking proper limits in time and space, the previous procedure is equivalent to assume that the Markov process is a diffusion (Cox and Miller [1965, p. 213]).

Are city size dynamics reasonably well described by diffusion processes? On purely theoretical grounds, continuity may be problematic because economic models predict discontinuities, particularly when new cities arise in the urban system. For example, in Henderson [1974, p. 88] new cities arise as positive mass of workers that move from old cities. This creates discontinuities in the size of existent and new cities. Similarly, new cities can also emerge from discontinuous –catastrophic– bifurcations (Fujita and Mori [1997]). However, these considerations suggest that discontinuities are unimportant in mature urban systems, as the ones we dealing with, where new cities play only a marginal role.

The assumption of continuity could also be justified on empirical grounds. One can construct a transition matrix using data from a particular country, and compare it with the theoretical one illustrated by (*). Figure 1 shows a transition matrix between 1980 and 1990 for U.S. cities computed by Ioannides and Dobkins [2000]. It has the required diagonal form, which supports the idea that city size changes slowly over time, and that major jumps are infrequent. Matrices from France and Japan also exhibit similar shape (See Eaton and Eckstein [1997]).

3.2 Steady State Distribution

Let $p(x_0, x; t)$ be the probability density function of x_t, given that at an earlier time, t_0, $x = x_0$. The Forward Kolmogorov Equation - *FKE* - describes the motion of $p(x_0, x; t)$[5] as

$$\frac{\partial}{\partial t}p(x_0; x, t) = \frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x_0; x, t)\right] - \frac{\partial}{\partial x}\left[x\mu(x)p(x_0; x, t)\right].$$

Along a steady state this equation becomes

$$\frac{\partial}{\partial t}p(x, t) = \frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x, t)\right] - \frac{\partial}{\partial x}\left[x\mu(x)p(x, t)\right], \tag{1}$$

where $p(x, t)$ is the unconditional probability density.

The next assumption incorporates the evidence on city size distribution into the model. It states that the unconditional probability distribution of city sizes is Pareto

[5] The FKE is almost an exact characterization of the conditional probability for diffusion processes. It is not a complete characterization for cases where a positive probability mass can be accumulated on a boundary, i.e., when boundaries are accessible. In our case, boundaries are not accessible by assumption: we know that the probability distribution has no positive mass at any point. Feller [1952] is the classic on the topic. Bharucha-Reid, [1960, pages 142-47] provides a pedagogical introduction.

at all times. Let $P(x,t)$ denote the probability distribution of X at time t, and $p(x,t)$ its corresponding density.

Assumption 3: $P(x,t) = 1 - x_{lt}^{\delta} x^{-\delta}$ and

$$p(x,t) = \delta x_{lt}^{\delta} x^{-(1+\delta)}, \tag{2}$$

for $t \geq 0$, where $x_{lt} = x_{l0} e^{\gamma t}$, $\delta > 0$, and $x_l > 0$.

Note that the minimum city size, x_{lt}, increases at the rate of population growth, γ. Thus, population growth shifts the distribution toward larger values. This is required in order to preserve a stationary distribution for the relative city sizes, x_t/N_t, as observed in the data. Using Equation (2) into (1), one obtains the first restriction on $\mu(x)$ and $\sigma(x)$.

3.3 Equilibrium

The Markov process governing x_t is really a closed-form solution of an underlying economic model. An equilibrium requirement in such model is that the total population across cities equals to the total urban population available. Alternatively, cities must grow in average at the same rate as the population. Since the number of cities is large (a continuum), this requirement can be stated as follows

$$E_t[x\,(\mu(x) - \gamma)] = 0 \text{ for all } t, \tag{3}$$

where $E_t[\cdot]$ is the expected value with respect to $P(x,t)$. Intuitively, this condition states that on *weighted* average cities must grow at the same rate as the urban population. There are two key considerations regarding this equation. First, the expected value is with respect to $P(x,t)$, a time-dependent distribution when $\gamma > 0$; Second, the previous condition must hold for all t.

Equations (1), (2) and (3) are the only three restrictions available to identify $\mu(x)$ and $\sigma^2(x)$. The following is the equilibrium (solution) concept:

Definition 1 *A diffusion process, described by $\mu(x)$ and $\sigma^2(x)$, supports a Pareto equilibrium if it satisfies equations (1), (2), and (3).*

It turns out that if $\gamma > 0$, $\mu(x)$ and $\sigma^2(x)$ can be sharply identified. In that case, $P(x,t)$ is non-stationary and equation (3) provides constraints for every period t. This is equivalent, as shown below, to have constraints for every x. Thus, (3) provides the continuum of equations needed to fully solve the system in this case. In contrast, if $\gamma = 0$, $P(x,t)$ is stationary and equation (3) becomes $E[x\mu(x)] = 0$. This equation provides only a single constraint on $\mu(x)$.

We consider separately the cases $\gamma = 0$ and $\gamma > 0$. The first case represents an economy without population growth, or alternatively, an economy with positive population growth but in which the relevant city size, the one that determines $\mu(x)$

and $\sigma(x)$, is the size of the city relative to the total population. The second case is more plausible: population grows and the relevant scale of the city is its absolute size.

3.4 The Stationary Case: $\gamma = 0$

Substituting (2) into (1) and dropping time subscripts produces

$$\frac{1}{2}\frac{\partial}{\partial x^2}\left[x^{1-\delta}\sigma^2(x)\right] - \frac{\partial}{\partial x}\left[\mu(x)x^{-\delta}\right] = 0 \tag{4}$$

Integrating this equation once and solving for $\mu(x)$[6], one obtains

$$\mu(x) = \frac{1}{2}\left[x\frac{\partial}{\partial x}\sigma^2(x) + (1-\delta)\sigma^2(x) + Ax^{\delta}\right], \tag{5}$$

where A is a constant of integration. This equation characterizes the drift, $\mu(x)$, as a function of $\sigma^2(x)$. Substituting this result into (3) produces, after some simplifications (see Appendix),

$$\sigma^2(x_l) = Ax_l^{\delta} + x_l^{\delta-1}\lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x) - A\right]. \tag{6}$$

The following is the first main result of the paper.

Proposition 2 *A diffusion process with drift $\mu(x)$ and diffusion $\sigma^2(x)$ supports a Pareto equilibrium if and only if $\sigma^2(x)$ is a positive differentiable function satisfying (6), and $\mu(x)$ satisfies (5).*

Proof. For sufficiency, notice that equations (1), (2), and (3) are satisfied once (5) and (6) are used. Necessity has already been established since (5) and (6) were obtained from (1), (2), and (3). ■

A more parsimonious characterization could be obtained if $A = 0$. This is the case if $\sigma^2(x)$ does not increase too fast with x.

Assumption 4: $\lim_{x\to\infty}\frac{\sigma^2(x)}{x^{\delta}} = 0$.

Lemma 3 *Suppose (5), (6) and Assumption 4 hold. Then $\mu(x)$ and $\sigma^2(x)$ in Proposition 2 satisfy*

$$\mu(x) = \frac{1}{2}\left[x\frac{\partial}{\partial x}\sigma^2(x) + (1-\delta)\sigma^2(x)\right], \tag{7}$$

$$\sigma^2(x_l) = x_l^{\delta-1}\lim_{x\to\infty}\left[x^{1-\delta}\sigma^2(x)\right] \tag{8}$$

Proof. *Since the left-hand side of equation (6) is finite,* $\lim_{x\to\infty}\left[x^{-\delta}\sigma^2(x) - A\right] = 0$, *or* $A = \lim_{x\to\infty} x^{-\delta}\sigma^2(x) = 0$ *(by Assumption 4).* ■

[6] Since our interest is to find alternative forms in which scale economies are consistent with the evidence, we assume $\mu(x) \neq 0$ when we integrate.

Thus, equations (7) and (8) fully characterize the diffusion processes consistent with Pareto distributions under the additional Assumption 4. A particular process that satisfies this Lemma is $\mu(x) = 0$ and $\sigma^2(x) = \beta x^{1-\delta}$. Notice that a process that satisfies Gibrat's law does not satisfies this condition unless $\delta = 1$.

The following lemma characterizes a very general class of diffusion process that can support a Pareto equilibrium.

Lemma 4 *Let $m(x)$ be a function that satisfies $m(x_l) = 0$, and $\lim_{x\to\infty} x^{1-\delta} m(x) = 0$. Then, a diffusion process with drift $\mu(x) = \frac{1}{2}[xm'(x) + (1-\delta)m(x)]$ and variance $\sigma^2(x) = \beta x^{1-\delta} + m(x)$ can support a Pareto equilibrium.*

Proof. This process satisfies the conditions of Proposition 2. ■

The next Lemma follows from equations (7) and (8) for large x:

Lemma 5 *(i) If $\delta = 1$, then very large cities share the same the diffusion coefficient (variance) and their mean growth is zero; (ii) If $\delta < 1$ then variance must eventually decrease with size; (ii) if $\delta > 1$ variance must eventually increase with size.*

3.4.1 Zipf's Law

Consider the particular case of Zipf's law. In that case $\delta = 1$, and (7) and (8) read

$$\mu(x) = x\sigma(x)\sigma'(x)$$

$$\sigma^2(x_l) = \lim_{v\to\infty} \sigma^2(v)$$

These two equations provide a very parsimonious characterization of the diffusion processes associated to an invariant Zipf distribution.

Proposition 6 *(Zipf distribution) Suppose $\delta = 1$ in Proposition (2). Then, $\mu(x) \lesseqqgtr 0$ if and only if $\sigma'(x) \lesseqqgtr 0$. (If large cities exhibit more stable growth, then they also must exhibit lower mean growth.)*

Thus, the fact that cities are distributed Zipf give the following strong predictions about city growth: (i) growing cities must have more unstable growth; (ii) more stable cities must be decaying cities. Hence, Zipf's law translates into a surprising interpretation of city growth. High growth is necessarily risky, and low growth is stable.

3.5 The Non-Stationary Case: $\gamma > 0$

In this case one can identify a unique diffusion process that supports Pareto Equilibria. The following is the second main result of the paper.

Theorem 7 *Let x follows a diffusion process satisfying equations (1), (2) and (3). Then, $\mu(x) = \gamma$ and $\sigma^2(x) = Ax^{\delta-1} + Bx^{\delta}$ for all x, where A and B are positive constants.*

Proof. *(See Appendix)* ■

To gain some understanding about why expected growth must be equal for all possible sizes, consider the stylized case in which $\lim_{x\to\infty} \mu(x) = \mu$, i.e., very large cities all grow at the same rate, μ. Suppose also that there are only three types of cities: small (S), medium (M), and large cities (L). The urban system initially includes all three types of cities, but small and medium cities eventually become large as population grows. Thus, all cities eventually grow at the rate μ, which implies, by equation (3), that $\mu = \gamma$. In words, large cities must grow at the same rate as the urban population. But this implies that $\mu(x) = \gamma$ for all x too. Why? Because from the previous discussion $\mu(L) = \gamma$ so that medium cities must grow also on average at the rate γ in order for (3) to hold when only medium and large cities coexist. By backward induction, it also follows that $\mu(S) = \gamma$ since (3) must also hold when the three types of cities coexist.

Theorem 7 provides also a functional form for the diffusion coefficient. If $B > 0$, this coefficient eventually increases with size, a prediction that conflicts with the economic intuition: one would expect growth in large cities to be more stable as large cities are more diversified. On this basis, one may choose $B = 0$ as the plausible option.

The result about the variance has a straightforward intuition, at least for the case $B = 0$. Notice first that δ measures how spread the Pareto distribution is, or alternatively, the degree of inequality. For example, $\delta = \infty$ represents an extremely even distribution as all cities have equal size. The other extreme, only one city hosting all the population, occurs when $\delta = 0$. It seems natural to expect that a more unequal distribution of population would arise from a more unequal growth process. For example, a process such that small cities face higher risks than large cities, but the same expected growth. This is what Theorem 7 states. It says, among other things, that the growth process associated to a Pareto distribution with $\delta < 1$ requires that smaller cities face more unstable growth than larger cities. The opposite occurs if $\delta > 1$.

3.5.1 IMPLICATIONS

Theorem 7 provides a very parsimonious characterization. One could have expected a richer class of Markov processes, even among the diffusion processes, to be consistent with the evidence. However, the interplay between a growing population and the requirement of a stationary growth process singles out a very parsimonious Markov process.

The result regarding the expected growth rate is strong: *mean city growth cannot depend on city size*. This single finding casts serious doubts on most economic models of cities. In particular, models where cities attain an optimal size as a result of the trade-off between positive and negative spillovers. City growth rate in these models depends on whether the city has reached its optimal size or not. In the extreme case, a city stops growing once it attains that size.

The second component of the Theorem shows that the scale of a city can affect the *stability* rather than the mean of its growth process. The result about the variance, $\sigma^2(x) = Ax^{\delta-1} + Bx^{\delta}$, is an important generalization with respect to previous results. In particular, this new characterization can account for the view held by some authors who argue that larger cities must display more stable growth just as a matter of diversification (Fujita *et al.* [1999, p. 224]). According to Theorem 7, such belief can be true but only if the exponent in the Pareto distribution is below 1. The evidence provided by Rosen and Resnik [1980, Table 3], suggests that this is in fact the case for many countries.

For countries for which Zipf's law holds, our result is quite surprising. Gibrat's law must hold there. Neither the mean nor the variance of growth can depend on size.

Theorem 7 also leaves a question. If $\delta > 1$, as the evidence suggests is the case for some countries, then the variance of growth must eventually increase with size, a counterintuitive result. This suggests a problem with our interpretation of the data, with the data, or with our formulation of the problem. As for the data, it could be that Pareto distributions with $\delta > 1$ are not really stable over time. More careful analysis of the data may indicate that the distribution really converges to a Pareto distribution with $\delta \leq 1$. This is an argument advanced by Brakman *et al.* (1999) for the case of the Netherlands. In that case, the results of the paper only apply after the distribution becomes stable. Another important problem with the data is the definition of cities. According to Rosen and Resnik [1980], when a metropolitan definition of cities is used, rather than a political definition, the estimated exponent of the Pareto distribution decreases substantially.

It seems, however, that the assumption of a strictly constant number of cities may be too strong for certain countries. If the number of cities could increase the Pareto exponent would increase too, a result described by Simon [1953].

3.6 Time Dependent Urban Growth

The following extension allows for a time dependent growth rate of urban population. This extension is important because it allows for a decreasing growth rate of urban population, a more realistic description of the urban process in many developed countries. We now show that non-stationary diffusion processes can account for such a fact, but under the results of Theorem 7 remain intact.

Suppose now that the urban population, $N_t := N(t)$, grows continuously over time and define $\gamma_t := N'(t)/N(t) \geq \underline{\gamma}$ to be the instantaneous growth rate of urban population at time t. Assume $\underline{\gamma} > 0$. The fact that γ_t changes through time suggests that the diffusion process must be time dependent in order for the labor market to clear at every t. It is natural to require the drift of the process to be time dependent, $\mu(x,t)$. On the other hand, changes in the deterministic growth of urban population are unlikely to affect the volatility of growth. Thus, we retain our assumption about the variance being only state dependent, i.e., $\sigma^2(x,t) = \sigma^2(x)$ for all t.

The following is the corresponding *FKE* for this process along a balanced growth path

$$\frac{\partial}{\partial t}p(x,t) = \frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x,t)\right] - \frac{\partial}{\partial x}\left[x\mu(x,t)p(x,t)\right], \tag{9}$$

where $p(x,t) = \delta\left(y_l N_t\right)^{\delta} x^{-\delta-1}$. In addition, a condition to assure equilibrium in total population is required. The analogue to equation (3) is given by

$$E_t[x\left(\mu(x,t) - \gamma_t\right)] = 0 \text{ for all } t. \tag{10}$$

The following Proposition generalizes Theorem 7.

Proposition 8 *Let x follow a diffusion process satisfying equations (9) and (10), and suppose the stationary distribution of $y := x/N_t$ is $P(y)$ (the Pareto distribution). Then, $\mu(x,t) = \gamma_t$ and $\sigma^2(x) = Ax^{\delta-1} + Bx^{\delta}$ for all x, where A and B are positive constants.*

Proof. Let $\gamma = \gamma_t$ in Appendix. All results follow. ■

4 Economic Models

This section studies a standard urban model based on economies of localization. Cities emerge in this economy due to the presence of scale economies, external to firms but internal to industries, as in Henderson [1988]. We allow for stochastic technologies and preferences, and look for necessary conditions such that the model can generate a Pareto distribution of city sizes. This section uses results derived in the previous section. In particular, according to Theorem 7, cities must growth at the same rate in a deterministic version of the model. Thus, cities must exhibit parallel growth.

Negative externalities in the model induce cities to specialize in production. However, in contrast to existent literature, negative externalities play no role in limiting city size in our model. City size is just limited by size of the market in which the city specializes. As noted by Eaton and Eckstein [1997], any upper bound to city size is inconsistent with parallel growth. Once a city reaches that bound, its growth rate slows down or becomes zero. It could also be the case that the factors limiting city growth (like transportation or pollutions technologies) evolve over time allowing cities to grow, as in Black and Henderson [1999]. Parallel growth can arise in this case only if the bounds grow at least as fast as the urban population. In this case the bounds become irrelevant because, as a general rule, they do not bind. In addition, such feature is hard to justify because it requires unusual assumptions about the underlying parameters.

There are two main results in this section. First, that 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 at some extent, then city growth just mirrors sectorial growth. Second, the distribution of fundamentals, preferences and technologies, must be Pareto too under general conditions.

4.1 Basic Model

Consider an economy inhabited by a large number of workers, $N_t = e^{\gamma t}$, firms, and a benevolent government. There are S locations and I industries to produce such that $S > I$. At the beginning of every period the government announces linear taxes on income for each location and industry, 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 wages 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.

4.1.1 Production

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

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

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 average 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. 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 $\varphi^i_1(L, L_s) > 0$, $\varphi^i_2(L, L_s) < 0$ and $\varphi^i_1(L, L_s) + \varphi^i_2(L, L_s) > 0$. 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 congestion costs: the productivity of industries in a particular city decreases with the size of the city. 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

$$\varphi^i(L_{is}, L_s) = L_{is}^{\tau_i} L_s^{-\beta_i},\ \tau_i > \beta_i > 0. \tag{11}$$

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

$$q_i A_i \varphi^i(L_{is}, L_s) = w_{is}. \tag{12}$$

4.1.2 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. 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[7]

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

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

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.

4.1.3 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 proportional income taxes. Without externalities in production, it would be optimal to use only lump sum taxes and transfers. With 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 [8]. This type of city specialization avoids unnecessary congestion costs that arise when industries locate next to each other (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[9]. 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,

$$\alpha_{is}(L_{is}) := \frac{\varphi^{i\prime}(L_{is})}{\varphi^i(L_{is})} L_{is}. \tag{13}$$

For the particular functional form (11), $\varphi^i(L_{is}) = L_{is}^{\alpha_i}$, where $\alpha_i(L_{is}) = \alpha_i := \tau_i - \beta_i$.

4.1.4 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. According to Theorem 7, this is a minimum requirement that a model of cities must satisfy in order to be consistent with the evidence of city size distribution.

[8] We introduce non-tradable goods below so that efficient cities do not fully specialize.

[9] 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.

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 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 $\widehat{q}$. According to conditions (i), (v) and (vi) of the previous definition, relative prices obey

$$\widehat{q}_{it} = [\alpha_{1t} - \alpha_{it}]\,\gamma \text{ for all } i \text{ and all } t. \tag{14}$$

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

Define

$$\epsilon_t := \begin{bmatrix} \epsilon_{11t} & \epsilon_{12t} & .. & \epsilon_{1It} \\ \epsilon_{21t} & \epsilon_{22t} & .. & \epsilon_{2It} \\ .. & .. & .. & .. \\ \epsilon_{I1t} & \epsilon_{I2t} & .. & \epsilon_{IIt} \end{bmatrix}, \triangle\alpha_t := \begin{bmatrix} \alpha_{1t} - \alpha_{it} \\ \alpha_{1t} - \alpha_{2t} \\ \cdots \\ \alpha_{1t} - \alpha_{it} \end{bmatrix}.$$

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

Proposition 9 *A balanced growth path exists if only and only if*

$$\left(\overline{1} + \varepsilon_t\right) \triangle\alpha_t = 0 \textit{ for all } t. \tag{15}$$

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

$$s_{it} = \frac{L_{it}}{N_t} = s_i \text{ for all } i.$$

Using condition (ii) of the definition above, s_{it} also satisfies

$$s_{it} = q_{it} C_i(q_t, \theta) \text{ for all } i.$$

Log-linearizing the two previous conditions and using (14) one obtains the result. ■

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

1. $\triangle\alpha_t = 0$ for all t or $\alpha_i = \alpha$ 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. The condition states that 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 $j \neq i$. We call this case a *demand-driven balanced growth path* because it only restrict 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 $\triangle\alpha_t \neq 0$ and $\overline{1} + \varepsilon_t \neq 0$ then the condition (15) implies that 1 is an eigenvalue of ε_t, and $\triangle\alpha_t$ is an associated eigenvector. We call this case 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 10 *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) $\alpha_i = \alpha$ for all i (increasing returns are identical for all industries); or (iii) 1 is an eigenvector of ε_t and $\triangle\alpha_t$ is one of its associated eigenvectors..*

4.1.5 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, F, 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. This result is powerful: one only needs to choose the stochastic processes governing θ and A so that their steady state distributions are Pareto. Examples of these processes are provided in Section 3 and below.

To further simplify the analysis suppose that $\varphi^i(L_{is}) = L_{is}^{\alpha_i}$ and $u(c,\theta) = \left(\sum_i \left(\theta_i^\eta c_i\right)^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}$, where η is the elasticity of substitution. Following Corollary 10, we assume that either $\alpha_i = \alpha$ 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,

$$C_i^d = v^{\eta-1}\left(\frac{\theta_i}{q_i}\right)^\eta W, \tag{16}$$

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

$$C_i^s = A_i L_i^{1+\alpha_i}$$

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

$$w = q_i A_i L_i^\alpha.$$

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

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

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

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

Finally, under the assumption that either $\alpha_i = \alpha$ for all i or $\eta = 1$, the previous expression could be simplified to

$$\frac{L_{it}}{N_t} = \frac{z_{it}/I}{\sum_i z_{it}/I}. \tag{17}$$

where $z_{it} := \left[A_{it}^{\eta-1}\theta_{it}^\eta\right]^{\frac{1}{1+\alpha(1-\eta)}}$. This is the crucial expression for our purposes. The variable on the left hand side is the relative size of city i. 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 fourth main result of the paper.

Proposition 11 *Suppose the steady state distribution of $\frac{L_{it}}{N_t}$ is Pareto with exponent δ. Then z_{it} is also distributed Pareto with exponent δ for large I. In addition, (i) if only preferences are stochastic (so that $A_{it} = A$ for i and t) then Φ^θ must be Pareto with exponent $\frac{\delta\eta}{1+\alpha(1-\eta)}$ provided that $1 + \alpha(1-\eta) > 0$; (ii) if only technologies are stochastic (so that $\theta_{it} = \theta$ for all i and t) then Φ^A must be Pareto with exponent $\frac{\delta(\eta-1)}{1+\alpha(1-\eta)}$, provided that $\frac{\eta-1}{1+\alpha(1-\eta)} > 0$; (iii) If $\eta = 1$ then preferences must be stochastic with exponent δ.*

Proof. *(i) In this case $z_{it} = \theta_{it}^{\frac{\eta}{1+\alpha(1-\eta)}}$. 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* $\lambda = \frac{\eta}{1+\alpha(1-\eta)}$ *(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 proved in the same way. In that case* z_{it} *becomes* $z_{it} = A_{it}^{\frac{\eta-1}{1+\alpha(1-\eta)}}$*. 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: $\frac{1+\alpha}{\alpha} > \eta$. 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. In one hand, cities with better technologies tend to be smaller because less workers are needed to satisfy the demand for their goods. In the other hand, better technologies reduce prices and increases demand.

But why would θ or A be distributed Pareto? Proposition 2 and Lemma 5 characterizes a class of processes that could induce 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.

Proposition 12 *(Random tastes) Suppose* $A_{it} = A$ *and* θ *follows a reflected geometric Brownian motion. Then* F *is Pareto with exponent* $\delta = \frac{1+\alpha(1-\eta)}{\eta}$*.* F *is Zipf only if* $\eta = 1$*.*

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

Proposition 13 *(Random technologies) Suppose* $\theta_{it} = \theta$ *and* A *follows a reflected geometric Brownian motion. Then* F *is Pareto with exponent* $\delta = \frac{1}{\eta-1} - \alpha$*.* F *is Zipf only if* $\eta - 1 = \frac{1}{1+\alpha}$*.*

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

4.2 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 nontradables goods,

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

Demand functions are still given by (16) 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 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 $X_s := L_s^{NT} + L_s^T$.

Lemma 14 $\frac{X_i}{X_j} = \frac{L_i^T}{L_j^T}$ *along a balanced growth path.*

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, $\frac{L_i^B}{L_j^B} = \frac{X_i}{X_j}$. ■

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

4.3 A Model with Capital

The previous models have abstracted from capital, either physical or human, in the previous model. 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

$$y_{is} = \varphi_i(K_{is}, L_{is}) l_{is}^{\alpha_i} k_{is}^{1-\alpha_i} \text{ for } i \in B, \tag{18}$$

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. Profits 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

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

In addition, $K_{is} = \frac{\alpha_i}{1-\alpha_i} \frac{w}{r} L_{is}$. Replacing these two expression into the production function (18), one obtains:

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

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 and r into $\tilde{\varphi}_i$ does not affect the previous results because they are equal across cities.

4.3.1 Other Models in the Literature

To the extent of our knowledge, there are currently in the literature only three other models capable of producing parallel growth. The first is by Gabaix [1999] who induces city growth and mobility due to the random nature of the amenities provided by cities. This unorthodox approach requires constant returns to scale technologies, which leave unexplained the source of agglomeration and the nature of the amenities.

The second model is by Black and Henderson (1999). Cities arise in their economy due to economies of localization. Cities attain an optimal size due to the existence of commuting costs that limit the gains from the positive externalities. Furthermore, optimal city sizes grow due to human capital accumulation. Cities specialize either in the production of intermediate goods or final goods. Parallel growth occurs because the final goods production function is Cobb Douglas, a property consistent with our results in the previous section. This suggests, however, that their result about parallel growth is not robust to the following natural generalization. Several cities specializing in different intermediate inputs, one city specializing in the production of final goods, and elasticity of substitution between inputs different from 1.

There is, however, a more serious problem with their model. Except under knife-edge parametrization, the growth rate of their economy either increases or decrease through time, a counterfactual. Thus, the scale effect does not show up in the growth rate of cities, but in the growth rate of the economy. This is a natural consequence of introducing two engines of growth into the model: capital accumulation and population growth.

An alternative model is the one by Eaton and Eckstein [1997]. In their model, city size depends on the amount of human capital accumulated 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. They are able to generate proportional growth only under the condition of zero discounting.

5 Conclusions

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

The first part of the paper characterizes Markov processes that preserve Pareto distributions. They include the well-known cases of proportional growth processes, but also processes that are size dependent. The paper obtains a sharp characterization when additional plausible assumptions are made about the number of cities and population growth. Pareto distributions can only arise from processes that are proportional in mean but not in variance. Specifically, the expected growth rate of a city must be independent of its size but the variance of the growth rate must have the form $A \cdot Size^{\delta-1}$, where δ is the Pareto exponent. This characterization has several implications. First, it means that under general conditions, Zipf's law can only result from Gibrat's law: growth must size independent. Thus, Gibrat's law is not just *an* explanation of Zipf's law, as argued in the literature, but it is *the (statistical)* explanation. Second, it provides a rationalization of how the scale of a city matters for its growth. Size affects growth volatility but not mean growth. Finally, it also provides a rationalization for the diversity of exponents found in the data. Cities in different countries have different growth volatilities. In particular, Pareto distributions with larger exponents (more unequal distributions) require more volatile growth processes.

The second part of the paper uses the statistical findings to characterize the set of feasible preferences, technologies, and exogenous shocks in a standard urban model. The paper finds that under general conditions the steady state distribution of the exogenous driven force must be Pareto, and one of two conditions must be satisfied: (i) the elasticity of substitution between goods must be 1 and preferences must be stochastic; or (ii) externalities must be equal across goods and technologies must be stochastic. The paper also provides examples of urban models that generate Pareto distributions and Zipf's laws for city sizes.

6 References

Abdel-Rahman, H. M. , and M. Fujita (1990), "Product variety, Marshallian externalities, and city sizes," *Regional Science and Urban Economics, 18:69-86.*

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

Bharucha-Reid. A. T. (1960), *Elements of the Theory of Markov Process and Their Applications*, McGraw-Hill.

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, vol. 39,pp. 183-213.

Carroll, Glenn R. (1982), "National city size distributions: what do we know afeter 67 years of research?," *Progress in Human Geography*, Vol. 6, No. 1, March, 1-43.

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

Cox D. R. and H. D. Miller (1965), *The Theory of Stochastic Processes,* Methuen & Co. LTD.

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, Gilles (1998), "Labor Specialization, Transport Costs, and City Size", *Journal of Regional Science*, Vol. 38, No. 4, 553-573.

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

Feller, William (1952), "The Parabolic Differential Equations and the Associated Semi-groups of Transformations", *Annals of Mathematics,* Vol. 55, No. 3.

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* v27: 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).

‘Harrison, J. Michael (1990), *Brownian Motion and Stochastic Flow Systems*, (Malabar, FL: Robert E. Krieger Publishing Company).

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).

Hogg, Robert V. and Allen T. Craig (1995), *Introduction to Mathematical Statistics (New Jersey: Prentice-Hall).*

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

Ioannides, Y. M. (1994), “Product differenciation and growth in a system of cities,” *Regional Science and Urban Economics 24:461-84.*

Ioannides, Y. M. and H. G. Overman (2000), “Zipf’s Law for Cities: An Empirical Examination,” mimeo.

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

Karlin, S. and H. M. Taylor (1965), *A Second Course in Stochastic Processes*, Academic Press.

Karatzas, I. and S. E. Shreve (1988), *Brownian Motion and Stochastic Calculus,* Springer-Verlag, Berlin.

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

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

Revuz, D. and M. Yor (1991), *Continous Martingales and Brownian Motion,* Springer-Verlag, Berlin.

Risken, H. (1989), *The Fokker-Planck Equation: Methods of Solution and Applications,* Springer-Verlag, Berlin.

Roehner, M. B. (1995), “Evolution of Urban Systems in the Pareto Plane,” *Journal of Regional Science* 35, No. 2, 277-300.

Romer, Paul M.(1986), “Increasing Returns and Long-Run Growth,” *The Journal of Political Economy* 94, No. 5, 1002-1037.

Scitovsky, T. (1954), "Two Concepts of External Economies," *The Journal of Political Economy 62*:143-51.

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.

Steindl, J., *Random Proceses and the growth of Firms* (New York, NY: Hafner, 1965).

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

Sornette, D., and R. Cont. (1997), "Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws," *Journal of Physique I France,* VII, 431-44.

Stanley, M.H *et al. (1995)*, "Zipf plots and the size distribution of firms," *Economic Letters 49:453-457.*

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

Appendix

Proof Equation (6). Multiplying (5) by x, taking expected value with respect to $P(\cdot)$ and using condition (3) one obtains,

$$E\left[x^2\frac{\partial}{\partial x}\sigma^2(x)\right]+(1-\delta)E\left[x\sigma^2(x)\right]+AE\left[x^{1+\delta}\right]=0 \tag{A0}$$

The first term of the last expression can be re-expressed as

$$\begin{aligned}
E[x^2\frac{\partial}{\partial x}\sigma^2(x)] &= \delta x_l^\delta\int_{x_l}^{\infty}x^{1-\delta}\frac{\partial}{\partial x}\sigma^2(x)dx\\
&= \delta x_l^\delta\left[x^{1-\delta}\sigma^2(x)\right]_{x_l}^{\infty}-\delta\left(1-\delta\right)x_l^\delta\int_{x_l}^{\infty}x^{-\delta}\sigma^2(x)dx\\
&= \delta x_l^\delta\left[x^{1-\delta}\sigma^2(x)\right]_{x_l}^{\infty}-\left(1-\delta\right)\int_{x_l}^{\infty}\delta x_l^\delta x^{-\delta-1}x\sigma^2(x)dx\\
&= \delta x_l^\delta\left[x^{1-\delta}\sigma^2(x)\right]_{x_l}^{\infty}-\left(1-\delta\right)E\left[x\sigma^2(x)\right].
\end{aligned}$$

Plugging this result into (A0), it follows that

$$\delta x_l^\delta\left[x^{1-\delta}\sigma^2(x)\right]_{x_l}^{\infty}+AE\left[x^{1+\delta}\right]=0.$$

Now, $Ex^{1+\delta}=\int_{x_l}^{\infty}x^{1+\delta}\delta x_l^\delta x^{-\delta-1}dx=\delta x_l^\delta\int_{x_l}^{\infty}dx=\delta x_l^\delta\left[x\right]_{x_l}^{\infty}$. Thus, we can write the previous equation as $\delta x_l^\delta\left[x^{1-\delta}\sigma^2(x)+Ax\right]_{x_l}^{\infty}=0$, or

$$x_l^{1-\delta}\sigma^2(x_l)-Ax_l=\lim_{x\to\infty}x\left[x^{-\delta}\sigma^2(x)-A\right].$$

which is the equation in the text. ■

Proof of the Main Theorem. In our case $p(x,t)=\delta\left(x_{lt}\right)^\delta x^{-\delta-1}$. Then $\frac{\partial}{\partial t}p(x,t)=\gamma\delta p(x,t)$. The KFE reads

$$\frac{\partial}{\partial t}p(x,t)=\frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x,t)\right]-\frac{\partial}{\partial x}\left[x\mu(x)p(x,t)\right]$$

$$\gamma\delta^2\left(x_le^{\gamma t}\right)^\delta x^{-\delta-1}=\frac{1}{2}\frac{\partial}{\partial x^2}\left[x^2\sigma^2(x)p(x,t)\right]-\frac{\partial}{\partial x}\left[x\mu(x)p(x,t)\right]$$

integrating once (with respect to x)

$$-\gamma xp(x,t)+A(t)=\frac{1}{2}\frac{\partial}{\partial x}\left[x^2\sigma^2(x)p(x,t)\right]-\delta\mu(x)xp(x,t)$$

$$[\mu(x)-\gamma]xp(x,t)=\frac{1}{2}\frac{\partial}{\partial x}\left[x^2\sigma^2(x)p(x,t)\right]-\frac{1}{2}A(t) \tag{A1}$$

Now, integrating in the interval $[x_{lt}, \infty)$ we have

$$\int_{x_{lt}}^{\infty} [\mu(x) - \gamma] x p(x,t) dx = \frac{1}{2} \left[x^2 \sigma^2(x) p(x,t) - A(t) x \right]_{x_{lt}}^{\infty} \tag{A2}$$

Now, according to the condition (3), the left hand side of the previous equation must be zero for all t. Below we show that there are only two possible cases: either $A(t) = 0$ for all t or $A(t) \neq 0$ for all t.

Consider first the case $A(t) = 0$ for all t. Then the following equality must hold for all t.

$$\frac{1}{2} \left[x^2 \sigma^2(x) p(x,t) \right]_{x_{lt}}^{\infty} = 0 \text{ for all } t$$

or

$$\left[\sigma^2(x) \left(x_{lt} \right)^{\delta} x^{1-\delta} \right]_{x_{lt}}^{\infty} = 0 \text{ for all } t$$

or

$$\sigma^2(x_{lt}) x_{lt} = x_{lt}^{\delta} \lim_{v \to \infty} v^{1-\delta} \sigma^2(v) \text{ for all } t \tag{A3}$$

Define $\beta = \lim_{v \to \infty} v^{1-\delta} \sigma^2(v)$ (we require the limit to exist and be bounded to assure a solution satisfying (3)). Then,

$$\sigma^2(x_{lt}) = \beta x_{lt}^{\delta - 1} \text{ for all } t \geq 0$$

We can replace the previous condition "for all t" by the expression "for all x_{lt}", but then it is the same as "for all x" since x_{lt} grows continuously and unboundedly overtime. Thus, we conclude,

$$\sigma^2(x) = \beta x^{\delta - 1} \text{ for all } x \tag{A4}$$

Now, replacing this expression into (A1) given that $A(t)$ is zero, we get,

$$\begin{aligned} [\mu(x) - \gamma] x p(x,t) &= \frac{1}{2} \frac{\partial}{\partial x} \left[x^2 \sigma^2(x) p(x,t) \right] \\ &= \frac{1}{2} \frac{\partial}{\partial x} \left[x^2 \beta x^{\delta - 1} \delta \left(x_{lt} \right)^{\delta} x^{-\delta - 1} \right] \\ &= \frac{1}{2} \frac{\partial}{\partial x} \left[\beta \delta \left(x_{lt} \right)^{\delta} \right] = 0 \end{aligned}$$

therefore,

$$\mu(x) = \gamma \text{ for all } x. \tag{A5}$$

Thus, (A4) and (A5) describe one possible solution. Now consider the case $A(s) \neq 0$ for some $s \geq 0$. In that case, condition (3) imposes

$$\left[x^{1-\delta} \sigma^2(x) \delta x_{lt}^{\delta} - A(t) x \right]_{x_{lt}}^{\infty} = 0 \ \text{ for all } t$$

or

$$x_{lt}\left[\sigma^2(x_{lt})\delta - A(t)\right] = \lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x)\delta x_{lt}^{\delta} - A(t)\right] \text{ for all } t. \tag{A6}$$

This condition requires, among other things $\lim_{x\to\infty} x^{-\delta}\sigma^2(x)\delta x_{lt}^{\delta} - A(t) = 0$ for all t, or

$$A(t) = \delta h x_{lt}^{\delta} \text{ for all } t, \tag{A7}$$

where $h := \lim_{x\to\infty} x^{-\delta}\sigma^2(x)$, a finite number. Substituting (A7) into (A6), we obtain

$$x_{lt}\left[\sigma^2(x_{lt}) - h x_{lt}^{\delta}\right] = \theta x_{lt}^{\delta} \text{ for all } t$$

where $\theta := \lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x) - h\right]$, a finite number according to (A8). Finally, solving for $\sigma^2(x_{lt})$ from the previous equation we obtain

$$\sigma^2(x) = h x^{\delta} + \theta x^{\delta-1} \tag{A8}$$

Now, substituting (A7) and (A8) into (A1),

$$\begin{aligned}
[\mu(x) - \gamma]\delta x_{lt}^{\delta} x^{-\delta} &= \frac{1}{2}\frac{\partial}{\partial x}\left[\left(h x^{\delta} + \theta x^{\delta-1}\right)\delta x_{lt}^{\delta} x^{1-\delta}\right] - \frac{1}{2}\delta h x_{lt}^{\delta} \\
&= \frac{1}{2}\frac{\partial}{\partial x}\left[(hx + \theta)\,\delta x_{lt}^{\delta}\right] - \frac{1}{2}\delta h x_{lt}^{\delta} \\
&= \frac{1}{2}\delta h x_{lt}^{\delta} - \frac{1}{2}\delta h x_{lt}^{\delta} \\
&= 0
\end{aligned}$$

Thus, (A5) also holds if $A(s) > 0$. In any solution, the drift must be γ (expected mean growth must be independent of size). The diffusion coefficient, in the other hand, can either have the form (A4) or (A8), but (A4) is a particular case of (A8). ■

Transition Matrix U.S. Metropolitan Areas
322 cities, 1980 - 1990

		1990									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1980	0.1	82.35	11.76	5.88							
	0.2	9.38	50.00	31.25	6.25	3.13					
	0.3		34.38	37.50	25.00	3.13					
	0.4			12.90	38.71	48.39					
	0.5		3.13	9.38	21.88	37.50	25.00	3.13			
	0.6					9.09	60.61	30.30			
	0.7					3.13	15.63	62.50	18.75		
	0.8							6.25	75.00	18.75	
	0.9								12.50	81.25	6.25
	1									3.13	96.88

Source. Dobkins and Ioannides (2000) p. 258

Figure 1: