Balanced City Growth and Zipf's Law

Juan Carlos Córdoba M.*
University of Rochester

July 3, 2001

Abstract

The evidence on city size distribution is remarkable. In many countries it obeys a Pareto distribution. Although several explanations have been proposed for the regularity, none of them is completely convincing, and the exact economic mechanism at work remains largely a puzzle. Current explanations seem to be inconsistent with scale economies, a central element in leading theories of cities.

This essay studies what conditions models of cities must satisfy in order to be match the evidence. According to the results, cities must exhibit (i) an expected growth rate that is independent of their size; and (ii) a growth variance that is proportional to $size^{\delta-1}$, where δ is the Pareto exponent found in the data. This characterization has important implications. First, it means that the scale of a city can play a role for the evolution of cities but only if it affects the variance of city growth, but not its mean. Second, it shows that *Zipf's law* – the case in which of $\delta = 1$ – must result from *Gibrat's law*. Third, this characterization rationalizes the diversity of Pareto exponents in the data. Conditional on size, city growth must be more stable in countries with lower δ. Finally, the essay provides the first economic model with increasing returns to scale that is able to generate Zipf's distributions.

**Preliminary. Comments very welcome. First version, november 2000. I am indebted to my advisor Per Krusell. I thank Alan Stockman, Mark Bils, Werner Ploberger, Shakeeb Khan, Michael Cranston, Tony Smith and participants in the Applied Workshop at the University of Rochester for helpful comments. I specially thank my wife for her enourmous support. I acknowledge financial support provided by el Banco de la Republica de Colombia. All remaining errors are my own. Email: jcco@troi.cc.rochester.edu.

1 Introduction

The evidence on city size distribution is remarkable. In many countries it obeys a Pareto distribution[1]. For some countries, the regularity is even more remarkable. The empirical Pareto distribution has an exponent close to one, a finding referred to as Zipf's law, after G. Zipf (1949). This subject has a long tradition in the fields of urban economics and regional science, but it is little known in mainstream economics.

Although several explanations have been proposed for the regularity, none of them is completely convincing. The exact economic mechanism at work remains largely a puzzle. Two recent books on the topic conclude that "*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, page 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* [2] *is one of the most robust statistical relationships known so far in economics*" (Fujita and Thisse, 2000, page 9).

This chapter provides a compelling solution to the puzzle. Our first result is to show that under general conditions Zipf's law *must* result from Gibrat's law - the condition that cities follow the same growth process regardless of their size. This result relates to a similar but weaker finding in the literature that claims that Zipf's law *could* result from Gibrat's law. The advantage of our formulation is that it provides a necessary rather than a sufficient condition for city dynamics to be consistent with the evidence of Zipf's distributions.

Our more general result is a sharp characterization of city dynamics consistent not only with Zipf's distributions but, more importantly, with Pareto distributions. In particular, to match the evidence, cities must exhibit (i) an expected growth rate that is independent of their size; and (ii) a growth variance that is proportional to $size^{\delta-1}$, where δ is the Pareto exponent in the data. Two additional pieces of evidence allow us to pin down such a parsimonious process. The fact that cities are growing over time and that the number of cities remains nearly constant.

This characterization has important implications. First, it means that the scale of a city can play a role for the evolution of cities and yet not contradict the fact that the city distribution is Pareto. However, it requires scale to only matter in a limited way: city size may affect the variance of city growth, but not its mean. This result also rationalizes the diversity of Pareto exponents in the data. According to our findings, conditional on size, city growth must be more stable in some countries

[1] The claim usually excludes small cities.

[2] Rank-size rule is another name for Zipf's law.

than in others. Finally, it shows the strong relationship between Zipf's and Gibrat's laws already stated.

This rather statistical characterization of city dynamics must be regarded as a constraint on the reduced form of economic models of cities, in particular models with a Markov representation of the equilibrium path. This representation typically arises along balanced growth paths of economies with idiosyncratic risk at the city level and no aggregate uncertainty.

We use the characterization to analyze a particular but popular model of city formation due to economies of localization. The need to choose a particular model renders the results in this section less general than the previous ones, but still they are interesting. We first ask under what conditions a deterministic version of the model can display *parallel city growth* – the requirement that all cities in the system grow at the same rate. We find a simple and clear characterization. Negative externalities should not limit the size of cities, and the model must possess a balanced growth path. In particular, our model requires either Cobb-Douglas preferences or identical strength of the externalities in all industries and cities. We then introduce randomness into this economy and provide the first example of an economic model with increasing returns to scale able to generate Zipf's distributions.

Finally, we relax some assumptions made along the way. Our previous analysis on city dynamics assumes that the relevant state is the *absolute* size of a city as measured by its population, an assumption consistent with most models of cities. However, one can also hypothesize that the relevant state is the *relative* size of a city with respect to the total population. This alternative formulation provides a whole new class of growth processes consistent with the evidence. One important feature of these processes is that cities with more stable growth must also be cities with lower expected growth. The possible economics underlying this kind of process is left for future research.

The chapter is divided into 6 sections. Section 2 presents relevant evidence for the United States and other countries. Section 3 reviews related literature on the topic. Section 4 is the main section of the paper. It sets up the statistical model and obtains the sharpest results. Section 5 uses the statistical results to evaluate the ability of economic models to fit the evidence. It finds that economic models with scale economies require very special assumptions to fit the evidence. Section 6 develops an alternative (statistical) model of scale economies capable of fitting the evidence, and Section 7 presents some final conclusions.

2 Evidence

We consider the following to be the main features of the distribution of population across cities:

1. It is well described by a Pareto Distribution.
2. It *moves* toward the right over time (due to population growth).
3. The Pareto exponent remains stable over time.
4. Individual cities move across the distribution over time.

To illustrate these features, consider the following exercise. Select a country with a large number of metropolitan areas – the equivalent of cities in our study –, rank them according to their population starting with the largest one, and then plot population against rank in log scale. Figure 1, 2 and 3 shows the results for different periods for three dissimilar countries: United States, Colombia and India. A remarkable pattern emerges. In all cases, the graph exhibits a linear shape, it moves towards the right over time, and the slope remains quite stable as the graph moves. There is no tautology. Although the relationship between size and rank must be negative by construction, nothing in the exercise imposes linearity or a stable slope.

These are not isolated cases. Rather, extensive evidence support an inverse log-linear relationship between rank and size in many countries. The classical study on this topic, by Rosen and Resnik (1980), provides supportive evidence using data for 44 countries. More recent studies include Eaton and Eckstein (1997) for France and Japan, Brakman *et al.* (1999) for Netherlands, and Roehner (1995) for several countries.

The evidence in the graphs is well described by the following *rank-size rule*: $rank = \lambda \cdot size^{-\delta}$. Thus, if $\delta = 1$, for example, this rule states that the largest city in a country is twice the size of the second largest, three times the third largest, etc.

An alternative probabilistic interpretation of the evidence regards it just as a sample distribution of the random variable 'city population'. In that case, the inverse log-linear relationship between size and rank suggests that the random variable is distributed Pareto. To see why, consider a vector of M random draws, $(x_1, ..., x_M)$, from an arbitrary distribution, $F(x)$, ranked from the largest to the smallest so that the index i is the rank of x_i. In this way, i/M is the sample countercummulative distribution of x, i.e. $i/M = 1 - \widehat{F}(x_i)$. If $F(x)$ is Pareto, i.e. $F(x) = 1 - A \cdot x^{-\delta}$, and M is large, then $i/M \approx A \cdot x_i^{-\delta}$, or equivalently, rank and size display a log-linear relationship with slope $-\delta$. Using this interpretation, the slopes in Figures 1, 2, and 3 have a very simple statistical interpretation: they estimate the exponent of the underlying Pareto distribution.

The case of Zipf's distributions has received considerable attention in the literature. The evidence, however, offers no particular support to the view that the underlying distribution is Zipf for most countries. For example, the slope of the graph is significantly below 1 for India and significantly above 1 for Colombia. The results of Rosen and Resnick also help to illustrate this point. They estimate Pareto exponents statistically different from 1, ranging from 0.81 (Morocco) to 1.97 (Australia), with an average of 1.14 (See Figure 4).

A final significant feature of the data is the mobility of cities across the distribution. In the United States, for example, there are well known cases of rising cities, particularly in the south, like Atlanta, Dallas, Houston or Miami, accompanied by the decay of historically important cities like Cleveland, St. Louis, Pittsburgh, or Buffalo. Similar evidence can be found for other countries. In India for example, Bombay overtook Calcutta as the largest city in the country during the eighties.

3 Related Literature

Several probabilistic and few economic models have been proposed to account for the evidence. Among the most prominent probabilistic models are the ones by Champernowne (1953), Simons (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. Simons 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. An evaluation of leading theories of cities carried out by Krugman (1996) and Fujita *et. al.* (1999) concludes 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 (See Eaton and Eckstein, 1997).

The fact that urban models as a rule do not display proportional growth is usually not regarded as a shortcoming. This is because alternative growth processes can also lead to Pareto distributions. We show, however, that models that cannot reproduce Gibrat's law for cities are, as a rule, flawed.

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. Instead, we provide clear and simple 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 in Figure 4, and the fact that new cities hardly arise in many of these countries. We also offer an explanation for these cases.

4 A Reduced Form Model

We interpret the evidence of city-size distribution as describing an urban system that is evolving along a balanced growth path characterized by Pareto distributions. It is important to point out that balanced growth in this context refers to population growth rather than income growth, although they may be associated. Our goal in this section is to characterize the dynamic systems, or more precisely, the Markov processes, that can preserve such equilibrium path.

Consider an economy composed by a large but given number of cities at time t, M_t, and a total urban population, N_t, that grows continuously over time at the exogenous compound rate γ, i.e., $N_t = e^{\gamma t}$. We distinguish between the absolute and the relative size of a city. Let $X_t \in \Re_+^M$ be a vector of population sizes at time t, and define $Y_t := X_t e^{-\gamma t}$ to be the vector of relative sizes.

Assumption 1: $M_t = M$ *and* $\gamma > 0$.

The assumption of a fixed number of cities is not essential; but it is consistent with the evidence, particularly from France and Japan during the last century

(Eaton and Eckstein, 1997). The results below still hold as long as M_t grows slowly[3]. In contrast, the assumption of a growing urban population is crucial. Together with the other assumptions of the model, this forces the scale of all cities to grow (a.s.) over time, allowing us to test for 'scale effects'.

The next is our fundamental assumption. It states that X_t follows a particular Markov process.

Assumption 2: X_{it} follows an i.i.d. diffusion process with stationary transition. In particular, the expected growth rate of X_{it}, $\mu(x)$, and the variance of the growth rate, $\sigma^2(x)$, depend only on the size of X_{it} but not on its identity, i.

We need to carefully motivate the different components of this assumption. As was already mentioned, we do not spell out any particular economic model in this section, but rather assume a reduced form for the law of motion of the population in a city. Different economic models may give rise to this particular reduced form, and the next section analyzes one of them. The advantage of studying reduced forms is that they can provide more general conclusions than any particular economic model.

The first statement in Assumption 2 is that X_{it} follows a Markov process. This is not a strong assumption. Individual state variables usually follow Markov processes along balanced growth paths of general equilibrium models. To see this, consider an economy where the state variable is the distribution of population across cities. Let $F(x,t)$ be such distribution at time t, i.e., $F(x,t)$ is the fraction of cities of size less or equal to x. Suppose in addition that cities face idiosyncratic i.i.d. shocks, $\varepsilon_t \in R^M$. In this context, the law of motion of X_{it} is a function of the individual and aggregate states of the economy,

$$x_{it+1} = H(x_{it}, \varepsilon_{it}; \varepsilon_t, F(X,t)).$$

If the number of cities is sufficiently large, all idiosyncratic noise cancels out so that there would be no aggregate uncertainty in this economy. The motion of x_{it} would not depend on the whole vector ε_t but only on its individual shock, ε_{it}. In addition, along a balanced growth path $F(X,t)$ evolves deterministically depending only on γ and t. Thus, along that path x_{it+1} follows a Markov process,

$$x_{it+1} = H(x_{it}, \varepsilon_{it}; t)).$$

This exposition makes clear that city dynamics may depend not only on the individual population size, but also on the whole distribution of the population.

[3]The results of Gabaix (1990, section III.3) about new cities apply for our model.

However, along a balanced growth path the aggregate distribution becomes irrelevant, and the evolution of an individual state variable depends only on its own current position.

The assumption also states that the process is a diffusion. This process requires, among other things, continuity of the sample paths, a property that significantly reduces the dimensionality of our problem, as we see below. Although this assumption is made mainly for convenience, we do not think it is crucial. In addition, we provide some indirect empirical evidence supporting it.

Assumption 2 also asserts that $\mu(x)$ and $\sigma^2(x)$ depend on X rather than Y. This is a crucial assumption because it determines what is the relevant state variable and how scale effects enter in our model. The dynamics of a city is determined by its actual size. Alternatively, one may hypothesize that changes in size due to aggregate population growth should not affect the relevant scale of a city so that μ and σ^2 must depend on Y. We think, however, that current models of cities strongly support the choice of X as the relevant state variable. In these models, both positive and negative externalities of agglomeration depend on cities actual size. Section 6.2 below studies the case where Y is the relevant state.

Finally, Assumption 2 states that the identity of a city plays no role on its growth process. This is clearly required if one hope 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.

Our problem at this point is to characterize the transition stationary Markov processes consistent with the non-stationary or *moving* Pareto distributions that we observe in the data. It is precisely this interplay between the non-stationarity of the distribution with the stationarity of the transition what allows to sharply characterize the Markov process.

4.1 The Role of Continuity

To better understand the solution procedure and the assumptions behind it, suppose momentarily that there is no population growth and that X_{it} follows a Markov process with finite state space as 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. All we know about the Markov chain is its associated invariant cumulative distribution, $F(x)$, a Pareto distribution in our case. By invariant distribution we mean that the following relationship holds: $f = f\Pi$ holds, where f is the density of F. The goal is to characterize Π given f, and some plausible assumptions about Π. Clearly, we cannot identify the chain only with information about f : 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 follows a diffusion process, such as Brownian motions or Ito processes[4]. Diffusions are Markov processes that change continuously over time[5]. Intuitively, the random variable can only move to states adjacent to the current position or remain in the same state. In terms of the Markov chain, the assumption requires only strictly positive probabilities around the main diagonal, and zeros elsewhere, as illustrated in the following matrix:

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

It is clear now that assuming continuity of the sample paths dramatically reduce 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, we have $D-1$ equations (obtained from the relation $f = f\Pi$), so that at most we can hope to solve $\theta(\cdot)$ 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 the process, $\mu(\cdot)$ and $\sigma^2(\cdot)$, and solve the problem in terms of $\mu(\cdot)$ and $\sigma^2(\cdot)$. All higher conditional moments are completely determined by the first two moments.

Is city-size dynamics reasonably well described by a diffusion process? On purely theoretical grounds, the assumption of continuity is problematic because important economic models predict discontinuities in the size of cities, particularly as new cities arise in the urban system. For example, in Henderson's models (Henderson 1974, 88) new cities require a positive mass of workers to move from old

[4] Dixit and Pindick (1994) provide a variety of applications of those processes to economics.

[5] The discussion here is heuristic. More precise arguments are given in Cox and Miller, page 213, about the representation of a diffusion process as the limit of a random walk with variable transition probabilities. Exact mathematical arguments require to take suitable limits in time and space.

cities. This creates discontinuities in the size of existent and new cities. Similarly, in Krugman-type of models new cities can emerge as result of discontinuous –catastrophic– bifurcations (Fujita and Mori, 1997). These considerations suggest that discontinuities are unimportant in mature urban systems, like the ones we dealing with, where new cities play only a marginal role.

The assumption of continuity can be better justified on empirical grounds. One can compute a transition matrix from the data and check if it looks diagonal like the theoretical one. Figure 5 shows a transition matrix between 1980 and 1990 for U.S. cities computed by Ioannides and Dobkins (2000). The matrix certainly has the required diagonal form which supports the idea that size changes slowly over time, and major jumps are infrequent. Matrices from France and Japan also exhibit similar shape (See Eaton and Eckstein (1997)).

4.2 Balanced Growth

The main analytical instrument that characterizes the probability distribution of a diffusion process is the Forward Kolmogorov Equation – *FKE*– or Fokker-Planck diffusion equation. Let $p(x_0, x; t)$ be the probability density function of x_t, given that at an earlier time t_0 we have $x = x_0$. The *FKE* is law of motion of $p(x_0, x; t)$[6]:

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

The *FKE* is usually employed to solve for the probability density given a particular diffusion process with drift $\mu(x)$ and diffusion $\sigma(x)$. For our problem we flip the procedure around. Following our interpretation of the evidence, we impose the density of X to be a Pareto density and ask what the mean and variance of the diffusion process must look like. This procedure allows us to characterize all transition *stationary* diffusion processes consistent with *non-stationary* Pareto distributions.

Let $P(y) = 1 - y_l^{\delta} y^{\delta}$ and $p(y) = \delta y_l^{\delta} y^{-\delta-1}$ be the cumulative distribution and density function of a Pareto distribution ($\delta > 0$, $y_l > 0$), and let $P(x, t)$ and $p(x, t)$ be the corresponding distribution and density functions of X_t. Using the definition

[6] 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.

of y, one can easily show that $P(x,t)$ is also a Pareto distribution with a growing minimum size, $P(x,t) = 1 - x_{lt}^{\delta} x^{\delta}$ where $x_{lt} := y_l e^{\gamma t}$. The corresponding density is

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

Thus, a stationary Pareto distribution for y, as suggested by the evidence, implies a *non-stationary* or *moving* Pareto distribution for x. Along a balanced growth path, equation (1) becomes[7]:

$$\frac{\partial}{\partial t} p(x,t) = \frac{1}{2} \frac{\partial}{\partial x^2} \left[x^2 \sigma^2(x) p(x,t) \right] - \frac{\partial}{\partial x} \left[x \mu(x) p(x,t) \right], \tag{3}$$

Substituting (2) into this equation we obtain the first restriction on $\mu(x)$ and $\sigma(x)$.

4.2.1 Aggregate Equilibrium

The Markov process governing x_t is really a closed-form solution of an underlying economic model. An equilibrium condition in such model must be that the total population across cities must be equal to the total urban population available. In a discrete time version of our model, this condition would imply that

$$(1+g) \sum_i^M x_{it} = \sum_i^M x_{it+1} \text{ for all } t$$

where g is the growth rate of population per period ($g := e^{\gamma \triangle t} - 1$). The left hand side can be considered as the total population (or labor) supply, and the right hand side is the total demand of population $t+1$. One can rewrite this condition as $g \sum_i^M x_{it} = \sum_i^M x_{it} g(x_{it})$ where $g(x_{it})$ is the expected growth rate of a city of size x_{it}. Alternatively, it can be stated as:

$$\sum_i^M x_{it} \left(g(x_{it}) - g \right) = 0 \text{ for all } t.$$

[7] One can alternatively formulate the problem in terms of y as one with a non-stationary transition but a stationary density, $p(y)$. In that case, the FKE along a balanced growth path would read:

$$0 = \frac{1}{2} \frac{\partial}{\partial y^2} \left[y^2 \sigma^2(y e^{\gamma t}) p(y) \right] - \frac{\partial}{\partial y} \left[x \mu(e^{\gamma t}) p(y) \right].$$

All the results are identical under this formulation.

For M sufficiently large and by the law of large numbers, the previous condition becomes

$$E_t \left[x \left(g \left(x \right) - g \right) \right] = 0 \text{ for all } t,$$

or, returning to our continuous time model, it translates into the key equation

$$E_t[x\left(\mu(x) - \gamma\right)] = 0 \text{ for all } t \tag{4}$$

where $E_t[\cdot]$ is the expected value associated to $P(x,t)$, the Pareto measure. This condition is intuitively clear. 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 taken with respect to the probability density of x, $p(x,t)$, a density that moves to the right over time; second, the previous condition must hold for all t.

The importance of population growth can be better understand if we momentarily assume $\gamma = 0$. In that case, $E_t[x\left(\mu(x) - \gamma\right)] = E[x\left(\mu(x) - \gamma\right)]$ so that the condition (4) only provides a single constraint on $\mu(x)$. But in order to completely identify $\mu(x)$ and $\sigma^2(x)$, we need a continuum of constraints. This is precisely what equation (4) provides, if $\gamma > 0$.

Finally, a subtle point about constraint (4) is that we require it to hold even in the case of $E_t[x] = \infty$, a case that arises when $\delta \leq 1$. The following theorem is the main result of the paper.

Theorem 1 *Let x follows a diffusion process satisfying equations (3), (2) and (4). 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 the expected growth must be equal for all possible sizes, consider the stylized case where $\lim_{x\to\infty} \mu(x) = \mu$, i.e., all very large cities 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 the 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 (4), 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 (4) to hold when medium and large cities co-exist. By backward induction, it also follows that $\mu(S) = \gamma$ since (4) must also hold when the three types of cities co-exist.

Theorem 1 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 opposite extreme, where everyone lives only in one city, occurs when $\delta = 0$. One should expect that a more unequal distribution of population would arise from a more unequal growth process. For example, a process where small cities face higher risks than large cities, but the same expected growth. This is exactly the result in Theorem 1. 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 $\delta > 1$.

4.3 IMPLICATIONS

Theorem (1) 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:. *growth cannot depend on 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 certainly an important generalization with respect to previous results. In particular, our 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.* page 224). According to Theorem 1, such belief can be true; but only if the exponent in the Pareto distribution is below 1. The evidence in Table 3 suggests that this is in fact the case for many countries.

For countries where 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. This result provides a strong case against the importance of scale effects for the U.S., where the evidence supports Zipf's law particularly well.

Theorem 1 also leaves a puzzle. If $\delta > 1$, as the evidence in Figure 3 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, our results only apply once 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.

Alternatively, one of our assumptions may not be granted. It seems that our strongest assumption is to require stationarity of the Markov process. We relax this requirement in Section 6 of the paper. As a result, what we obtain is a much richer variety of growth processes consistent with the evidence where city size plays a larger role. However, the basic conclusion still holds: if $\delta > 1$ then variance must eventually increase with city size! (see Lemma 6). Section 6 also explores the case of a decreasing rate of growth of urban population. Our results are robust to that modification.

Thus, $\delta > 1$ turns out to be a puzzle for our model because it implies that larger cities exhibit larger growth volatility. Unexplored alternatives include relaxing the reasonable assumptions of cities following a diffusion process or the Markov property.

5 Economic Models

In view of the strong results of the previous section, the natural next step is to determine conditions under which agglomeration models exhibit proportional growth. So far as we know, Gabaix (1999) has provided the only *stochastic* economic model able to display proportional city growth. Gabaix, however, follows the unorthodox approach of explaining agglomeration by preferences rather than technologies. The advantage of this approach is that constant returns technologies can be employed, and proportional growth follows naturally from that technology.

Some *deterministic* models of cities with scale economies have also been proposed to explain some features of the evidence about city-size distribution (Eaton and Eckstein (1997), and Black and Henderson (1999)). Those models are able to display *parallel city growth*, or cities growing at the same rate, which is just the deterministic counterpart of proportional growth. To obtain the result, however, they require unappealing assumptions about the parameters of the model. This leaves the impression that proportional growth is generally incompatible with scale economies. For example, Eaton and Eckstein require a discount factor equal to zero, and Black and Henderson need unusual functional forms for preferences and technologies.

In this section we study general conditions required to produce parallel city growth in deterministic models and proportional growth in stochastic models. For that purpose, we study a standard agglomeration model which can sustain an urban system in equilibrium. Cities emerge in this economy due to the presence of scale economies, external to firms but internal to industries, as in Henderson (1988). It turns out that the same conditions that guarantee the existence of a balanced growth path in multisectorial endogenous growth models can also generate parallel city growth. The reason is simple. If cities specialize in production, at least at some extent, then city growth just mirrors sectorial growth.

In contrast to existent literature, negative externalities play no role in limiting city size in our model. As noted by Eaton and Eckstein (1997), any upper bound to city size is inconsistent with parallel growth. Once a city reaches that bound, or an 'optimal size', 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 city 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.

Once we derive conditions for parallel growth using a deterministic framework, we proceed to make the model stochastic and provide the first example of a model with scale economies able to account for Zipf's law. Finally, we elaborate some extensions of the model.

5.1 Basic Deterministic Model

Consider an economy where production and consumption must take place in locations defined by the set $S := \{1, ..., \overline{S}\}$. We call 'city' a location with a positive mass of population. Consumers in this economy have preferences over varieties in the set $I \equiv \{1, ..., \overline{I}\}$, where $\overline{S} > \overline{I}$. The state of the economy at time t is completely described by the distribution of population across locations and activities, $L_{si}(t)$, $s \in \overline{S}, i \in \overline{I}$. Goods are transported without cost and firms are competitive. As before, the total population, N_t, grows exogenously at the continuous compound rate γ. To simplify notation, we drop time subscripts but all variables must be considered as time dependent.

5.1.1 Production of Final Goods

Firms choose labor, l_{is}, to maximize profits:

$$\max_{l_{is}} q_i \varphi_i(L_{is}) l_{is} - w_{is} l_{is}, \;\; s \in \overline{S}, i \in \overline{I}$$

where q_i is the price of good i (equal across locations due to arbitrage), w_{si} is the wage rate at location s and activity i, L_{is} is the mass of workers at location s producing good i, and $\varphi_i : \Re_+ \rightarrow \Re$ is a differentiable function describing the gains of agglomerating. We assume $\varphi_i(L) > \varphi_i(0)$ for $x > 0$ and $\varphi_i'(L) > 0$. Agglomeration economies are those of location. Per-worker productivity at activity i and location s increases with the quantity of workers at the same activity and location.

The degree of increasing returns in activity i is measured by the elasticity of the average productivity with respect to the agglomeration.

$$\alpha_i(L_i) := \frac{\varphi_i'(L_i)}{\varphi_i(L_i)} L_i. \tag{5}$$

The source of this effect can be either informational spillovers, search and matching in local labor markets, or pecuniary externalities. The typical case in the literature entails a constant elasticity, i.e., $\varphi_i(L) = A_i L^{\alpha_i}$ (e.g., Black and Henderson (1999), Henderson (1974, 88), Krugman (1991), Duranton (1998), Lucas (2001)). Notice that we allow the degree of increasing returns to be different across goods. For example, think about the potential externalities associated with the production of entertainment services in Orlando or Hollywood versus the ones associated with the production of cars in Detroit. There are not apparent reasons to expect similar externalities in the production of all goods. Indeed, empirical studies assessing α_i find that it differ significantly across goods (Henderson 1988, chapter 5).

5.1.2 Preferences and demand

Agents in the economy seek to maximize their utility defined as

$$u(c) = \left(\sum_{i \in I} \left(\theta_i^{\eta} c_i \right)^{\frac{\eta - 1}{\eta}} \right)^{\frac{\eta}{\eta - 1}},$$

where c_i is consumption of good $i \in I$, $\eta > 0$ is the elasticity of substitution between goods, and $\theta_i > 0$ is a preference parameter determining the taste for good i. To compute the demand functions, we can use the fiction of a representative agent because preferences are homothetic, and prices are equal across locations in equilibrium, as we see below. Let W be the total income in the economy and q_i be the price of good i. Utility maximization subject to the proper budget constraint gives rise to the following demand functions:

$$c_i = W v^{\eta - 1} \left(\frac{\theta_i}{q_i} \right)^{\eta}, \; i \in I \tag{6}$$

where v is the price index of the consumption good defined as $v := \left(\sum_{i \in I} \theta_i^{\eta} q_i^{1-\eta} \right)^{\frac{1}{1-\eta}}$.

5.1.3 Dynamics

Models with increasing returns like this one usually possess multiple equilibrium in a forward looking economy. To choose a particular equilibrium path we follow the tradition of the field by assuming an ad-hoc but plausible adjustment process: workers move toward activities and locations with currently higher wages[8]. Specifically,

$$\frac{\dot{L}_{is}}{L_{is}} = \gamma + \tau \left(w_{is} - w \right)$$

where w is the average wage in the economy defined as

$$w := \sum_{s,i} w_{is} \left(\frac{L_{is}}{N} \right)$$

[8] This strategy was initially employed by Krugman (1990), and extensively used in Fujita et. al. (1999).

5.1.4 Competitive Equilibrium

A *competitive equilibrium* is a set of trajectories of prices $q_i(t)$, quantities, $x_i(t)$, wages $w_{si}(t)$, and labor allocation, $L_{si}(t)$, such that: (i) $w_{si}(t) = q_i(t)\varphi_i(L_{is})$ (Profit maximization); (ii) $x_i(t) = W(t)v(t)^{\eta-1}\left(\frac{\theta_i}{q_i(t)}\right)^{\eta}$ (Utility maximization); (iii) $x_i(t) = \sum_s \varphi_i(L_{is})L_{is}(t)$ (Goods market clearing); (iv) $N = \sum_{is} L_{is}$ (Labor market clearing); and (iv) $\frac{\dot{L}_{is}}{L_{is}}(t) = \gamma + \tau\left(w_{si}(t) - w(t)\right)$ for $i \in I$ and $s \in S$.

A feature of the equilibrium path is that individual industries tend to concentrate in only one place. This result follows directly from the existence of increasing returns at the industry level. Average productivity and wages are larger in larger industrial concentrations. The location preferred by a particular industry is completely determine by the initial distribution of population. The location with larger initial concentration of workers of one industry eventually attracts the whole industry. More precisely, if $L_{is}(0) > L_{is'}(0)$ for $s, s' \in S$, then the adjustment dynamics imply that $\frac{\dot{L}_{is}}{L_{is}}(t) > \frac{\dot{L}_{is'}}{L_{is'}}(t)$ for all $t > 0$. We cannot, however, exclude the possibility that an industry remains permanently spread out in different locations. This happens when more than one location has the largest initial industry concentration.

As mentioned before, cities are locations with positive population. A particular city may host one or several industries. The total population in a particular location is given by $L_s = \sum_i L_{is}$.

5.1.5 Balanced Growth

Along a balanced growth path wages are equal in all locations and activities. As a result, $\frac{\dot{L}_{is}}{L_{is}} = \gamma$ for all $(i,s) \in I \times S$, and by the definition of L_S, all cities grow at the rate γ. Thus, along balanced growth paths, cities also exhibit parallel growth. We now proceed to find conditions for the existence of a balanced growth path in this economy.

Let n^i be the number of cities producing good i along a balanced growth path and let L^i be the total labor force in industry i. Therefore, L^i/n^i is the size of the industry in each city where the industry locates, and therefore the productivity of industry i is $\varphi_i(L^i/n^i)$.

From (6), $\frac{c_i}{c_j} = \left(\frac{\theta_i}{\theta_j}\frac{q_j}{q_i}\right)^{\eta}$. In addition, total production of good i in the country is $x_i = \varphi_i(L^i/n^i))L^i$. These two results together lead to

$$\frac{\varphi_i(L^i/n^i)L^i}{\varphi_j(L^j/n^j)L^j} = \left(\frac{\theta_i}{\theta_j}\frac{q_j}{q_i}\right)^{\eta} \text{ for } i,j \in I.$$

In addition, along a balanced growth path wages are equal across locations, and relative prices only reflect differences in relative productivities, $\frac{q_j}{q_i} = \frac{\varphi_i(L^i/n^i)}{\varphi_j(L^j/n^j)}$. Plugging this result into the previous equation, we obtain (adding up time subscripts) the key equation

$$\frac{L_t^i}{L_t^j} = \left(\frac{\theta_i}{\theta_j}\right)^{\eta} \left(\frac{\varphi_i(L_t^i/n^i)}{\varphi_j(L_t^j/n^j)}\right)^{\eta-1} \text{ for } i, j \in B. \tag{7}$$

At this point it is convenient to define the function $\Psi^{ij} : \Re_+^2 \rightarrow R_+$ as:

$$\Psi^{ij}(z_i, z_j) := \frac{\varphi_i(z_i)}{\varphi_j(z_j)}$$

We are ready to establish our main result of this section:

Proposition 2 *Balanced growth only exists in the following two cases: (i) $\eta = 1$ (preferences are Cobb-Douglas); (ii) $\alpha_i = \alpha$ for all i (External effects are identical for all goods). In particular, balanced growth exist if $\varphi_i'() = 0$ for all $i \in I$. Balanced growth paths display parallel city growth.*

Proof. *Along a balanced growth path, labor in all industries grow at the same rate. Thus, the left hand side of Equation (7) remains constant as population grows, for all $i, j \in I$. On the right hand side, the expression $(\Psi^{ij}(z_i, z_j))^{\eta-1}$ must be constant as x and y grow at the same rate, a result that transpires if $\eta = 1$ and/or if $\varphi_i(x)$ is independent of x. The result also transpires if $\Psi^{ij}(tz_i, tz_j) = \Psi^{ij}(x, y)$, i.e., Ψ^{ij} is homogenous of degree zero, or, by Euler Theorem, $(\partial\Psi^{ij}/\partial z_i)\, z_i + (\partial\Psi^{ij}/\partial z_j)\, z_j = 0$. Using the definition of Ψ^{ij}, this condition implies*

$$\begin{aligned} 0 &= \frac{\varphi_i'}{\varphi_j} z_i - \frac{\varphi_j' \varphi_i}{(\varphi_j)^2} z_j \\ &= \frac{\varphi_i}{\varphi_j} \left[\frac{\varphi_i'}{\varphi_i} z_i - \frac{\varphi_j'}{\varphi_j} z_j\right]. \end{aligned}$$

Finally, using the definition of $\alpha_i(\cdot)$, this requires $\alpha_i(\cdot) = \alpha_j(\cdot)$ for all $i, j \in I$. ■

Our highly tractable equilibrium show how hard it is to reconcile parallel growth in cities with scale economies. It requires strong constraints on preferences or on the strength of the external effects.

If those constraints are not satisfied, then constant returns to scale become the natural alternative to explain parallel growth. But if technologies are linear, why do cities arise in the first place? A sensible possibility is that scale economies eventually die out. This is the case, for example, if $\alpha_i(L) = \alpha/L^\rho$, for $\rho > 0$ large. Cities arise in this case because larger agglomerations display higher productivity, but scale economies disappear for L sufficiently large. This technology is, in fact, supported by evidence. Henderson (1988, chapter 5) finds that "Localization economies appear to have declining elasticities, or to peter out as scale increases (page 97)." See also Segal (1976).

5.2 A Stochastic Model that Produces Zipf's Law

We now introduce randomness into the model of the previous section so that the relative size of cities changes along a balanced growth path. To simplify the exposition, we choose to introduce stochastic preferences into the model, but similar results can be obtained with stochastic technologies. The rise and fall of cities in this version of the model is driven by the stochastic changes in consumer tastes for goods.

Suppose that there are exactly $\overline{I}$ cities, and each city specializes in the production of exactly one good. In that case, L_t^i is the size of city i. Suppose in addition $\eta = 1$ so that equation (7) now states that $\frac{L_t^i}{L_t^j} = \frac{\theta_i}{\theta_j}$. Using labor market clearing, it follows that

$$\frac{L_t^i}{N_t} = \frac{\theta_{it}}{\left(\sum_j \theta_{jt}\right)} = \frac{\theta_{it}/\overline{I}}{\left(\sum_j \theta_{jt}\right)/\overline{I}}$$

Finally suppose that the preference parameters, θ's, are stochastic. They are drawn from a finite-mean distribution, F. If $\overline{I}$ is sufficiently large, the denominator of the last expression is constant by the law of large numbers. In that case, the distribution of relative city-sizes is just F. Thus, if F is Pareto, so is the distribution of city sizes.

But why would θ be distributed Pareto? As we argued before, Pareto distributions can result from proportional growth processes. Thus, if the growth process of θ satisfies Gibrat's law, then its steady state distribution may be Pareto. The following lemma is an application of Proposition 1 in Gabaix (1999).

Lemma 3 *Suppose θ_{it} follows the "reflected geometric brownian motion" process $d\theta_{it}/\theta_{it} = \sigma dB_{it}$ for $\theta_{it} > \theta_{\min}$ and $d\theta_{it}/\theta_t = \max\{\sigma dB_{it}, 0\}$ for $\theta_{it} < \theta_{\min}$ where B_{it} is a Brownian motion, and $\theta_{\min}$ is the barrier of the process. Then, the distribution of θ_{it} converges to a Zipf's distribution.*

Corollary 4 *The distribution of city-sizes is Zipf.*

5.3 Diversified Cities and Non-tradable goods

Cities are usually regarded as very diversified production entities but our model in the previous section portraits cities as highly specialized. We now develop an extension of the model where cities are highly diversified in the production of *non-tradables*, 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 *non-tradables*, that 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 as before but now they include non-tradables 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 (6) 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 non-tradables.

We now establish that the relative population of any two cities is completely determined by the extent of its tradable sector. We first need to introduce some new notation. Let $L^{NT} = \sum_i L_i^{NT}$ be the size of labor employed in the production of non-tradables, 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 non-tradables at city s. Total population at city s is thus given as $X_s := L_s^{NT} + L_s^T$.

Lemma 5 $\frac{X_i}{X_j} = \frac{L_i^T}{L_j^T}$ *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 non-tradables when determining relative city-sizes, but still can interpret cities as diversified production places.

5.4 A Model with Capital

We 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 our model that leaves our results intact. Suppose the production function for tradable goods is given by

$$y_{is} = \varphi_i(K_{is}, L_{is}) l_{is}^{\alpha_i} k_{is}^{1-\alpha_i} \text{ for } i \in B, \tag{8}$$

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 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 (8), 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 the previous section. The inclusion of w and r into $\tilde{\varphi}_i$ does not affect the previous results because they are equal across cities.

5.4.1 Other Models in the Literature

To the extent of our knowledge, there are currently in the literature only two other models capable of producing parallel growth, in addition to Gabaix's model. The first model is by Black and Henderson (1999). They device an economy where cities arise due to location economies. 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 non-convex technologies into a growth model (See Romer, 1986).

An alternative model was proposed by Eaton and Eckstein (1997). In their model, city size depends on the amount of human capital accumulated cities. Cities of different sizes co-exist 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.

6 Extensions

In this section we study two possible modifications to the basic model of section 4. The first 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 show that only non-stationary diffusion processes can account for such a fact, but under the results of Theorem 1 remain intact. The second extension allows the conditional moments of the Markov process to depend not on the actual size of cities but in their relative size. This extension could be important to reconcile scale economies with the evidence. The underlying economic forces behind such a process is left unexplored.

6.1 Time Dependent Urban Growth

Suppose now that the urban population, $N_t := N(t)$, grows continuously overtime 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 suggest that the diffusion process must be time dependent in order to equilibrate the labor market 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 in 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, we require a condition to assure equilibrium in total population. The analogue to equation (4) is given by

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

The following Proposition summarizes our general result:

Proposition 6 *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. ■

6.2 MODELS WITH RELATIVE SCALE ECONOMIES

In Section 4 we constrained the Markov process to be transition stationary. City dynamics was only allowed to depend on the scale of the city, x, but not on time. Stationarity together with population growth turned out to deliver very sharp predictions about the Markov process.

An alternative way to introduce scale effects into the model would be to allow the Markov process to depend not on the real size of a city, x, but on its relative size, y. This is equivalent to allowing the transition of x to be not just state dependent but also time dependent, i.e., non-stationary. On the other hand, the process for y would be completely stationary, only state dependent. Although it seems a sensible alternative in statistical terms, its economic microfoundations are unclear. Most models of cities postulate positive and negative externalities that derive directly from the actual city size rather than a relative size. For example, in Henderson models, negative external effects of agglomeration such as congestion and commuting costs are attributed to the real size of the population. That said, we now study the implications of this alternative.

The relevant random variable in this section is y rather than x. After imposing stationarity and the Pareto density, the Forward Kolmogorov Equation reads

$$\frac{1}{2}\frac{\partial}{\partial y^2}\left[y^{1-\delta}\sigma^2(y)\right] - \frac{\partial}{\partial y}\left[\mu(y)y^{-\delta}\right] = 0 \tag{11}$$

and after integrating once and solving for $\mu(y)$ [9], we obtain:

$$\mu(y) = \frac{1}{2}\left[y\frac{\partial}{\partial y}\sigma^2(y) + (1-\delta)\sigma^2(y)\right]. \tag{12}$$

As in Section 4, the FKE provides a system of equations that can be solved for $\mu(y)$ in terms of $\sigma^2(y)$. Now consider the mean condition, $E\left[y\mu(y)\right] = 0$. Applying this condition to the previous equation, it becomes:

$$E\left[y^2\frac{\partial}{\partial y}\sigma^2(y)\right] = E\left[(\delta-1)y\sigma^2(y)\right] \tag{13}$$

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

The left hand side of the last expression can be re-expressed as:

$$
\begin{aligned}
E[y^2 \frac{\partial}{\partial y}\sigma^2(y)] &= \delta y_l^\delta \int_{y_l}^{\infty} v^{1-\delta} \frac{\partial}{\partial v}\sigma^2(v)dv \\
&= \delta y_l^\delta \left[v^{1-\delta}\sigma^2(v)\right]_{y_l}^{\infty} + \delta\,(1-\delta)\, y_l^\delta \int_{y_l}^{\infty} v^{-\delta}\sigma^2(v)dv \\
&= \delta y_l^\delta \left[v^{1-\delta}\sigma^2(v)\right]_{y_l}^{\infty} + (1-\delta) \int_{y_l}^{\infty} \delta y_l^\delta v^{-\delta-1} v\sigma^2(v)dv \\
&= \delta y_l^\delta \left[v^{1-\delta}\sigma^2(v)\right]_{y_l}^{\infty} + (1-\delta)\, E\left[y\sigma^2(y)\right].
\end{aligned}
$$

Finally, replacing this result into (13), we obtain $\delta y_l^\delta \left[v^{1-\delta}\sigma^2(v)\right]_{y_l}^{\infty} = 0$ or,

$$
y_l^{1-\delta}\sigma^2(y_l) = \lim_{v\to\infty} v^{1-\delta}\sigma^2(v). \qquad (14)
$$

Equations (12) and (14) are the only constraints we have to identify the underlying process. In contrast to Section 4, these two conditions are not enough to completely pin down the Markov process. However, we can still derive some important qualitative results about the underlying process. The next lemma follows from direct observation of equations (14) and (12):

Lemma 7 *(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.*

These properties are analogous to the ones found in Section 4; but in the present case we can only characterize the limit behavior of the diffusion coefficient, instead of its whole shape. The next Proposition collects our main results of the section.

Proposition 8 *Let $\sigma^2(y)$ be a positive function satisfying (14) and let $\mu(y)$ be a function defined by (12). Then, any diffusion process with drift $\mu(y)$ and diffusion $\sigma^2(y)$ satisfying the Forward Kolmogorov Equation has a stationary Pareto distribution with exponent δ.*

Proof. Replacing (12) into (11), we obtain

$$\frac{\partial}{\partial y}\left[y^2\sigma^2(y)p(y)\right] = \left[y^2\frac{\partial}{\partial y}\sigma^2(y) + (1-\delta)y\sigma^2(y)\right]p(y)$$

or

$$\left[2y\sigma^2(y)g(y) + g(y)y^2\frac{\partial}{\partial y}\sigma^2(y) + y^2\sigma^2(y)g'(y)\right] = \left[y^2\frac{\partial}{\partial y}\sigma^2(y) + (1-\delta)y\sigma^2(y)\right]g(y),$$

and after some simplification, one obtains

$$\frac{g'(y)}{g(y)} = -(1+\delta)\frac{1}{y}.$$

The solution to this differential equation is $g(y) = Ay^{-(1+\delta)}$, i.e., the density of a Pareto distribution with exponent δ. ■

The following lemma characterizes a very general class of diffusion process that possesses a stationary Pareto distribution.

Lemma 9 *Let $\sigma^2(y) = \beta y^{1-\delta} + m(y)$, and define $\mu(y) = \frac{1}{2}\left[ym'(y) + (1-\delta)m(y)\right]$, where $m(y)$ satisfies $m(y_l) = 0$, and $\lim_{y\to\infty} y^{1-\delta}m(y) = 0$. Then, the diffusion with drift $\mu(y)$ and variance $\sigma^2(y)$ has a stationary Pareto distribution.*

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

Note that the process found in the previous section, with $\mu(y) = 0$ and $\sigma^2(y) = \beta y^{1-\delta}$, satisfies the previous lemma. But in contrast with Section 4, there are many other processes that can deliver the required result.

6.2.1 Zipf's Law with Relative Scale Effects

Now consider the particular case of Zipf's law. In that case $\delta = 1$, and (12) and (14) read

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

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

These two equations provide a very parsimonious and sharp characterization of the diffusion processes associated to a stationary Pareto distribution with exponent 1.

Proposition 10 *(Zipf's distribution) Suppose* $\delta = 1$ *in Proposition (8). Then,* $\mu(y) \lesseqgtr 0$ *if and only if* $\sigma'(y) \lesseqgtr 0$. *(Thus, if large cities exhibit more stable growth, then they also must exhibit lower mean growth.)*

Thus, the fact that cities are Zipf's distributed 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.

6.2.2 Dynamics

Up to this point we have not discussed the issue of convergence of the probability measures to its stationary distribution. In particular we would like to know whether $P(y_0; y, t)$ converges to a Pareto distribution regardless of the initial distribution and given that the coefficients of the process satisfy (12) and (14).

Convergence demands more structure within the problem. We now consider a specific but well known diffusion process to study its dynamic properties. Suppose city growth follows an univariate Ito Process. In particular, suppose $y \in \Re^+$ follows the "reflected Ito process":

$$\frac{dy}{y} = \left\{ \begin{array}{c} \max\{\mu(y)dt + \sigma(y)dz, 0\} \text{ if } y \leq y_l \\ \mu(y)dt + \sigma(y)dz \text{ if } y_l < y \end{array} \right\} \tag{15}$$

where dz is the increment of a Wiener process, and $\mu(y)$ and $\sigma(y)$ are the drift and the diffusion coefficients respectively as defined before. One special case of the previous process is the "reflected geometric Brownian motion with drift", in which $\mu(y) = \mu$ and $\sigma(y) = \sigma$.

It can be easily shown that the Forward Kolmogorov Equation corresponding to this particular diffusion process is given by equation (11). Therefore, if $\mu(y)$ and $\sigma(y)$ satisfies the assumptions of Proposition (8), the diffusion process has a stationary Pareto distribution. We assume this is the case.

The process in (15) makes it explicit that the diffusion process considered up to now requires two assumptions in order to possess a stationary Pareto distribution. First, the process must be reflected at a strictly lower positive barrier, y_l. This guarantees that the support of the distribution is a strictly positive interval of the type $[y_l, \infty)$. Second, the barrier must be constant in terms of the normalized size, Y, which means a growing barrier for the actual size of cities, X.

For some results it is convenient to work with $\ln(y)$ rather than with y. Define v as $v := \ln(y)$. It is a standard result in stochastic calculus that:

$$dv = \left\{ \begin{array}{c} \max\{\left(\mu(y) - \frac{1}{2}\sigma^2(y)\right) dt + \sigma(y)dz, 0\} \text{ if } v \leq v_l \\ \left(\mu(y) - \frac{1}{2}\sigma^2(y)\right) dt + \sigma(y)dz \text{ if } v_l < v \end{array} \right\} \tag{16}$$

where $v := \ln(y_l)$.

At this point, we would like to argue that the distribution of y finally decays to this stationary distribution regardless of the initial distribution. However, for this to hold, we need to impose additional structure on the problem. We require the diffusion process to be strongly recurrent so that the probability mass cannot escape to the boundaries. According to Risken (1988, pages 134-37), we need constraints on the coefficients of the Ito process to avoid singularities or y_t escaping to infinity. A singularity arises if the diffusion coefficient tends to zero, or if a coefficient does not approach a finite number (Feller, page 469). At this point, the representation in equation (16) turns out to be more convenient[10]. In terms of that equation, we require $\mu(y) - \frac{1}{2}\sigma^2(y)$ and $\sigma^2(y)$ to approach finite numbers for y large and $\sigma^2(y)$ to be bounded away from zero. These two conditions immediately imply that $\mu(y)$ must also approach a finite number. In economic terms, the convergence of $\mu(y)$ and $\sigma^2(y)$ to finite values implies that Gibrat's law must hold for large cities.

Given than $\mu(y)$ and $\sigma(y)$ are linked by (12), the following lemma establishes a necessary and sufficient condition for $\mu(y)$ and $\sigma(y)$ to converge to finite values, to avoid singularities, and y escaping to infinity:

Lemma 11 *$\mu(y)$ and $\sigma(y)$ satisfying (12) converge to finite values if and only if*

$$\lim_{y\to\infty} |y\frac{\partial}{\partial y}\sigma^2(y)| = \kappa$$

Proof. *First we prove sufficiency. Clearly, $\lim_{y\to\infty} |y\frac{\partial}{\partial y}\sigma^2(y)| = \kappa$ implies*

$$\lim_{y\to\infty} \frac{\partial}{\partial y}\sigma^2(y) = 0$$

so that $\sigma^2(y)$ approaches a constant. These two facts imply than $\mu(y)$ satisfying (12) also approaches a constant. Next we prove necessity. Suppose $\mu(y)$ and $\sigma(y)$ converge to finite values. Then by (12), $y\frac{\partial}{\partial y}\sigma^2(y) = 2\mu(y) + (\delta - 1)\sigma^2(y)$, so that $y\frac{\partial}{\partial y}\sigma^2(y)$ also converge to a finite value. ■

[10] In terms of equation (15), the coefficients of the process are $y\mu(y)$ and $y\sigma(y)$. Thus, for those coefficient to converge to finite numbers we would require $\mu(y)$ and $\sigma(y)$ to approach zero for y large.

Define $\mu := \lim_{x\to\infty} \mu(y)$ and $\sigma := \lim_{x\to\infty} \sigma(y)$. The previous lemma motivates the following assumption on $\sigma^2(y)$.

Assumption 1: $\lim_{y\to\infty} y\frac{\partial}{\partial y}\sigma^2(y) = \kappa < \delta\sigma^2$ *and* $\sigma^2(y)$ *is bounded away from zero for* $y \geq y_l$.

The additional condition that $\kappa < \delta\sigma^2$ is required for the following lemma.

Lemma 12 *Suppose the process described by (15) has coefficients satisfying (12), (14) and assumption 1. Then, y does not escape to infinity (is bounded a.s.).*

Proof. *Assumption 1 implies that* $\sigma^2(y)$ *approaches a constant for large y. Coupled with (12), it also implies that for large y,* $\mu(y)$ *approaches the constant* $\frac{1}{2}\left[\kappa + (1-\delta)\sigma^2\right]$. *Therefore, for large y,* $\frac{dy}{y} \simeq \frac{1}{2}\left[\kappa + (1-\delta)\sigma^2\right]dt + \sigma dz$, *i.e., for large y, y follows a geometric Brownian motion. In terms of y,*

$$\begin{aligned} y_t &\simeq y_0 \exp\left\{\sigma Z_t + \left(\frac{1}{2}\left[\kappa + (1-\delta)\sigma^2\right] - \frac{1}{2}\sigma^2\right)t\right\} \\ &= y_0 \exp\left\{\sigma Z_t - \frac{1}{2}\left(\delta\sigma^2 - \kappa\right)t\right\} \end{aligned}$$

By Assumption 1, this process drifts toward zero. ■

The following is the main result of this section.

Proposition 13 *Suppose y follows a diffusion process that satisfies the conditions of the previous lemma. Then* $P(y_0, y, t)$ *converges to a Pareto Distribution with exponent* δ.

Proof. *See Risken, pages 134-137, and 98-99.* ■

In the standard problem, one has a particular growth process for cities, as the one defined by (15), with some drift and diffusion coefficients, $\mu(y)$ and $\sigma(y)$. The question now is how ca we use the previous results to determine if the process has a limit Pareto distribution, and if so, what the exponent in the distribution will be? The following Proposition establishes the result.

Proposition 14 *Consider an Ito process as defined by (15), and suppose* $\sigma(y)$ *satisfies assumptions 1. Define* $\widetilde{\delta}(y)$ *as*

$$\widetilde{\delta}(y) := 1 + \frac{1}{\sigma^2(y)}\left[y\frac{\partial}{\partial y}\sigma^2(y) - 2\mu(y)\right] \tag{17}$$

Then, y converges (in distribution) to a Pareto distribution with coefficient $\widetilde{\delta}(y)$ *if and only if* $\widetilde{\delta}(y)$ *is constant and larger than zero.*

Proof. *First we prove sufficiency. Suppose* $\widetilde{\delta}(y)$ *is a constant larger than zero. Then, (17) is just a restatement of (12). Therefore, Proposition 4 applies and the required result follows. Now, we prove necessity, i.e., that a Pareto distribution requires* $\widetilde{\delta}(y)$ *constant and larger than zero. It follows from the fact that (12) is a necessary condition for an Ito process to have a Pareto Distribution, and 12 can be only equal to (17) if* $\widetilde{\delta}(y)$ *is constant and larger than zero.* ■

7 A Corrigendum

Finally, we want to point out and correct an important inconsistency in the main Proposition of Gabaix's (1999) paper. Our correction keeps the essence of Proposition 1 in Gabaix's paper intact (Gibrat's law produces Zipf's law) but dispenses with an ungranted assumption made by Gabaix.

According to page 749, city size follows a Geometric Brownian Motion (using our notation):

$$dX_{it}/X_{it} = \gamma dt + \sigma dB_{it}.$$

where B_{it} is a Brownian motion. In addition, according to the definition in footnote 14, $S_{it} := \frac{X_{it}}{X_0 e^{\gamma t}}$ is the normalized city size. One can use Ito's lemma to easily establish that

$$dS_{it}/S_{it} = \sigma dB_{it} \tag{18}$$

i.e., detrended city growth has no drift. This is a natural consequence of assuming that the drift of X being independent of size and equal to γ. However, Gabaix finds instead that $dS_{it}/S_{it} = \mu dt + \sigma dB_{it}$. What is particularly disturbing is that $\mu < 0$ is needed for the proof of Proposition 1. Otherwise, the unconditional expected value of S, $\overline{S}$, would be ∞, but Gabaix's arguments depend on the condition $\overline{S} < \infty$, (in particular he needs $\overline{S} = 1/N$ where N is the number of cities). The fact that Gabaix needs $\mu < 0$ is natural and it is equivalent to the stability condition imposed by Champernowne (1953) (see Simons (1955), page 438 for a derivation).

What if we drop the requirement of $\overline{S} < \infty$? Consider for a moment the process described by equation (18) and suppose, as Gabaix, that the process is reflected at some lower size S_{min}. What is the limit distribution of S? A easy way to proceed is to consider the limit distribution of $s := ln\ S$ instead. Ito's lemma establishes that $ds = -\left(\sigma^2/2\right) dt + \sigma dB_t$ for $s > s_{min} := \ln S_{\min}$. We can use Harrison's computation (Harrison, 1990, page 15), the ones used by Gabaix, to find that the tail distribution of s converges to the exponential function $P(s > s') = e^{-(s'-s_{\min})}$. Harrison's results apply because the drift of ds is negative $(-\sigma^2/2)$ even though dS/S has no drift. Equipped with the tail distribution for s, it is straightforward to show that the tail distribution of S is $P(S > S') = S'/S_{min}$, i.e., Zipf's' law. The result holds true regardless of the specific value of S_{min}. This is in sharp contrast to Proposition 1 in Gabaix paper, where S_{min} needs to converge to zero in order to obtain a Zipf's distribution.

The result seems a little bit magical. Why does the process drift toward the origin if $\mu = 0$? To understand why, consider the simple Brownian motion with lower barrier at zero and no drift (the pure Gibrat's process). This process automatically drifts toward the origin making S log-normally distributed with mean that goes to zero. The lower barrier at $S_{min} > 0$ is what induces some probability mass to be accumulated near S_{min}, originating a Pareto distribution rather than the log-normal distribution.

If S is Zipf's distributed, why do we care about $\overline{S} < \infty$? According to Gabaix, this condition is needed to guarantee aggregate population equilibrium, at least in expected value. But an alternative condition to guarantee equilibrium is to assume $E[\mu S] = 0$ (zero expected growth of detrended cities), which is guaranteed by the fact that $\mu = 0$. Naturally, $E[S] = \infty$ makes no sense for normalized (or unnormalized) city size. That is an undesired consequence of approximating the discrete distribution of S by a continuous distribution with support on the whole real line (or at least above a minimum size).

The following Proposition summarizes the corrected result.

Proposition 15 *Suppose the normalized sizes S follows the "reflected geometric Brownian motion" process $dS_t/S_t = \sigma dB_t$ for $S_t > S_{min}$ and $dS_t = S_t \max(\sigma dB_t, 0)$, for $S_t \leq S_{\min}$ where $_t$ is a Brownian motion, S_{min} is the barrier of the process, i.e., the minimal normalized city size. Then the distribution converges to a Zipf's distribution with exponent 1 regardless of the size of the lower barrier, S_{min}.*

8 Conclusion

This chapter characterizes growth processes consistent with the evidence of Pareto distributions for cities under quite general conditions. They include not only processes that satisfy Gibrat's law, but more importantly, processes that are size dependent. We obtain a sharp characterization. The growth process must be such that expected city growth is independent of its size and the variance of city growth must have the form $A \cdot Size^{\delta-1}$, where δ is the Pareto exponent.

This characterization has powerful implications. First, it means that under general conditions, Zipf's law can only result from Gibrat's law. Growth must be independent of size. Thus, Gibrat's law is not just *an* explanation of Zipf's law, as argued in the literature, but it is *the* explanation. Second, it provides a rationalization of how the scale of a city may matter for its growth. It affects the stability of growth but not its mean. Finally, it also provides a rationalization for the diversity of exponents found in the data. Cities in different countries face different growth stability.

We also offer simple and clear conditions under which a standard model of cities reproduces Gibrat's law. Negative externalities should not limit the size of cities, and the model must possess a balanced growth path. In particular, our model requires either Cobb-Douglas preferences or equal externalities across goods. Along the way, we provide the first economic model with increasing returns able to generate Zipf's law.

9 References

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

Abdel-Rahman, H. M. (2000), "City Systems: General Equilibrium Approaches", in *Economics of Cities*, Cambridge University Press.

Akerlof, George A. (1997), "Social Distance and Social Decisions", *Econometrica*, Vol 65, No. 5, September, 1005-1027.

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

Becker, Gary S. and Kevin M. Murphy (1992), "The Division of Labor, Coordination Cost, and Knowledge", *Quarterly Journal of Economics*, Volume 107, Issue 4, Nov., 1137-1160.

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.

Davis, E. G. and J.A. Swanson (1972), "On the Distribution of City Growth Rates in a Theory of Regional Economic Growth", *Journal of Economic Development and Cultural Change*, Vol. 20, No. 3, April, 495-503.

Dixit, A. and R. S. Pindyck (1994), *Investment Under Uncertainty*,Princeton University Press.

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).

Hall, Robert E. (1997), “Macroeconomic Fluctuations and the Allocation of Time”, *Journal of Labor Economics* v15, n1, Part 2 January, S223-50.

‘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).

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.

Jovanovic, B. (1996), "Learning and Growth", in Advances in Economics, Edited by David Kreps and Kenneth Wallis, CambridgeUniversity Press.

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 E. (1967), "Adjustment Costs and the Theory of Supply", *The Journal of Political Economy 75,* 321-334.

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

____ (2001), "Externalities and Cities", *Forthcoming*, *Review of Economic Dynamics.*

Mills, E.S. (1972), *Studies on the Structure of Urban Economy*, (Baltimore: Johns Hopkins University Press).

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.

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

Simons, 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).

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 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

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

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

integrating once (with respect to x)

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

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

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

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

Now, according to the condition (4), 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

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

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 (4)). 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,

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

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,

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

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 (4) 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

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

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

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

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

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

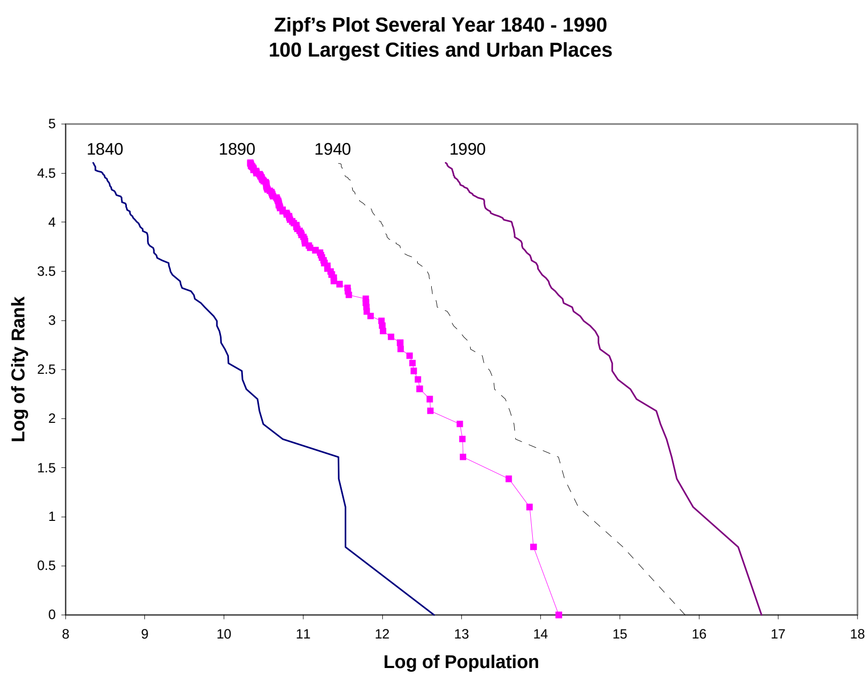
where $\theta := \lim_{x\to\infty} x\left[x^{-\delta}\sigma^2(x) - h\right]$, a finite number according to (A8). Finally, solving for $\sigma^2(x_{lt})$ from the previous equation we obtain

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

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

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

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). ■



Population of 100 Largest Cities and Other Urban Places in the U. S. 1790 TO 1990, U.S. Census Burea.
For 1990 we use Metropolitan Area Population, Census Bureau.

Figure 1:

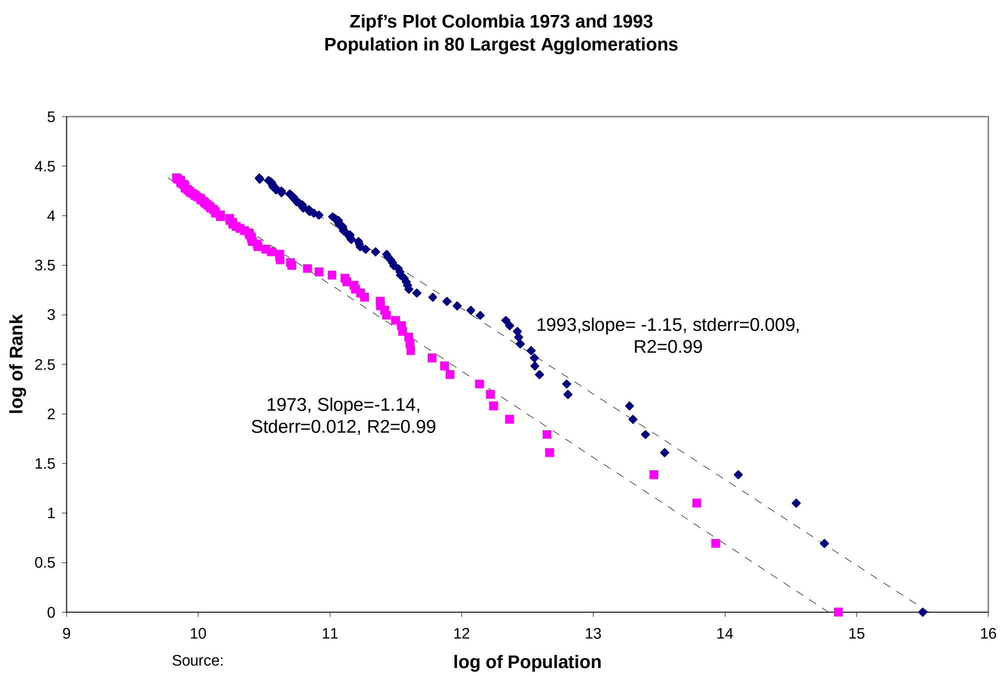

Zipf's Plot Colombia 1973 and 1993
Population in 80 Largest Agglomerations
log of Rank
log of Population
Source:
1993,slope= -1.15, stderr=0.009,
R2=0.99
1973, Slope=-1.14,
Stderr=0.012, R2=0.99
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
9
10
11
12
13
14
15
16

Figure 2:

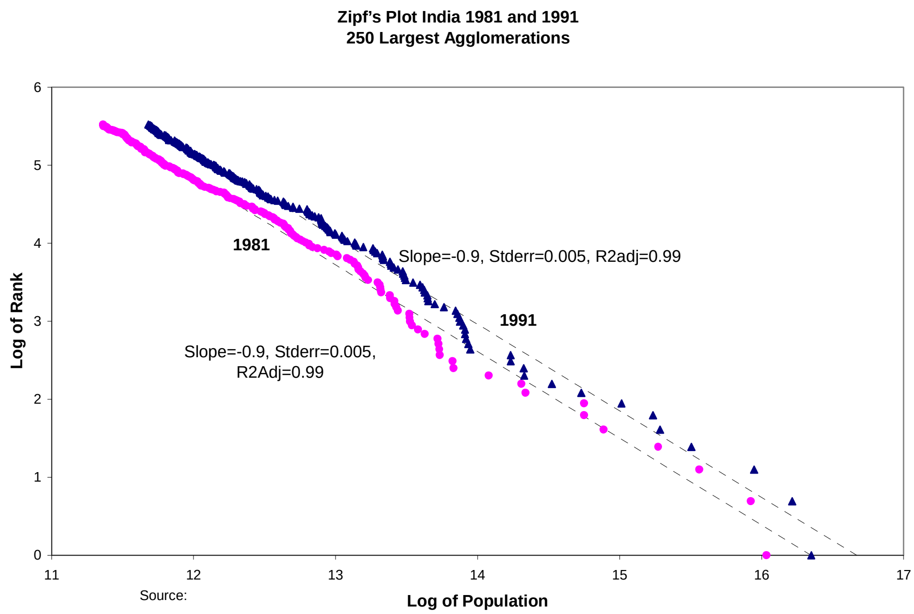

Zipf's Plot India 1981 and 1991
250 Largest Agglomerations
Log of Rank
Log of Population
1981
1991
Slope=-0.9, Stderr=0.005, R2adj=0.99
Slope=-0.9, Stderr=0.005, R2Adj=0.99
Source:
0
1
2
3
4
5
6
11
12
13
14
15
16
17

Figure 3:

PARETO COEFFICIENTS
Several countries

Argentina	0.933	Malaysia	0.968
Australia	1.963	Mexico	1.153
Austria	0.875	Morocco	0.809
Brazil	1.153	Netherlands	1.266
Canada	1.132	Nigeria	1.537
Colombia	0.847	Norway	1.265
Czechoslavakia	1.107	Philippines	1.253
Denmark	1.374	Poland	1.127
Ethiopia	0.97	Romania	1.085
Finland	1.084	S. Africa	0.997
France	1.325	Spain	1.133
E. Germany	1.125	Sri Lanka	1.13
W.German	1.171	Sweden	1.41
Ghana	1.104	Switzerland	1.095
Greece	1.138	Thailand	0.961
Hungary	1.092	Turkey	1.077
India	1.204	U.K.	1.178
Indonesia	0.967	U.S.S.R	1.278
Iran	0.993	U.S.A.	1.184
Israel	0.983	Venezuela	1.106
Italy	1.046	Yugoslavia	1.186
Japan	1.289	Zaire	0.93

Source: Rosen & Resnik 1980

Figure 4:

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**
	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					
1	**0.4**			12.90	38.71	48.39					
9	**0.5**		3.13	9.38	21.88	37.50	25.00	3.13			
8	**0.6**					9.09	60.61	30.30			
0	**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 5: