ZIPF'S LAW FOR CITIES: AN EXPLANATION*

XAVIER GABAIX

Zipf's law is a very tight constraint on the class of admissible models of local growth. It says that for most countries the size distribution of cities strikingly fits a power law: the number of cities with populations greater than S is proportional to $1/S$. Suppose that, at least in the upper tail, all cities follow some proportional growth process (this appears to be verified empirically). This automatically leads their distribution to converge to Zipf's law.

I. INTRODUCTION

Zipf's law for cities is one of the most conspicuous empirical facts in economics, or in the social sciences generally. The importance of this law is that, given very strong empirical support, it constitutes a minimum criterion of admissibility for any model of local growth, or any model of cities. Since George Zipf's[1] original explanation [1949], many explanations have been proposed, but all pose considerable difficulties. The present paper proposes a simple and robust account for the regularity.

To visualize Zipf's law, we take a country (for instance, the United States), and order the cities[2] by population: No. 1 is New York, No. 2 is Los Angeles, etc. We then draw a graph; on the y-axis we place the log of the rank (N.Y. has log rank ln 1, L.A. log rank ln 2), and on the x-axis the log of the population of the corresponding city (which will be called the "size" of the city). We take, like Krugman [1996a, p. 40], the 135 American metropolitan areas listed in the Statistical Abstract of the United States for 1991.

We see a straight line, which is rather surprising (there is no tautology causing the data to generate automatically a straight line). Furthermore, we find its slope is -1. We can run the regression,

* I thank Edward Glaeser and Paul Krugman for their guidance, and Richard Arnott, Robert Barro, Olivier Blanchard, John Campbell, David Cutler, Donald Davis, Aaron Edlin, Murray Gell-Mann, Elhanan Helpman, Yannis Ioannides, Richard Johnson, Jed Kolko, David Laibson, Eric Maskin, Daniele Paserman, José Scheinkman, Andrei Shleifer, Didier Sornette, Olivier Vigneron, Romain Wacziarg, William Wheaton, two anonymous referees, and seminar participants at several universities for their helpful comments. I also thank Jonathan Eaton and Zvi Eckstein for their data.

1. The original discoverer of Zipf's law, however, seems to be Auerbach [1913].

2. The term "city" is, strictly speaking, a misnomer; "agglomeration" would be a better term. So, the "city" of Boston should include Cambridge. Indeed Rosen and Resnick [1980] show that Zipf's law holds better the more carefully agglomerations are constructed.

The Quarterly Journal of Economics, August 1999

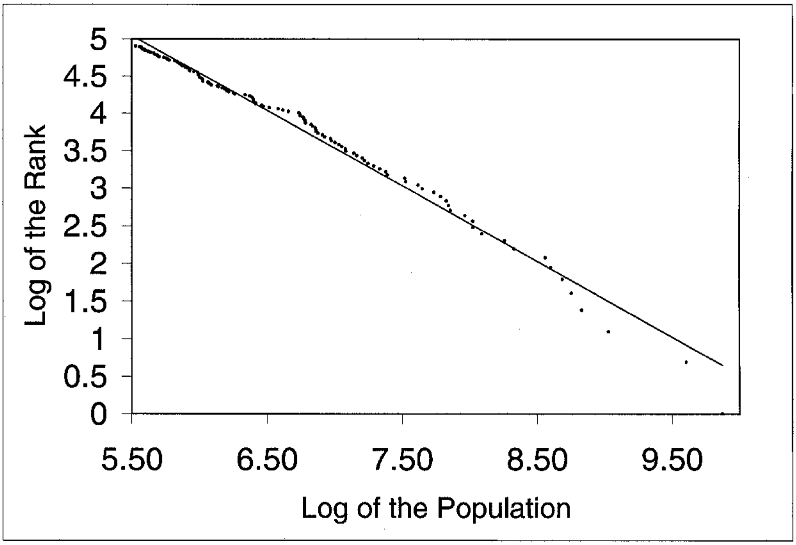

FIGURE I
Log Size versus Log Rank of the 135 largest U. S. Metropolitan Areas in 1991
Source: Statistical Abstract of the United States [1993].

$$\ln \text{Rank} = 10.53 - \underset{(.010)}{1.005} \ln \text{Size}, \tag{1}$$

where the standard deviation is in parentheses, and the R^2 is .986. The slope of the curve is very close to -1. This is an expression of Zipf's law: when we draw log-rank against log-size, we get a straight line, with a slope, which we shall call ζ, that is very close[3] to 1. In terms of the distribution, this means that the probability that the size of a city is greater than some S is proportional to $1/S$: $P(\text{Size} > S) = a/S^{\zeta}$, with $\zeta \simeq 1$. This is the statement of Zipf's law.[4]

3. In fact, the regression above is not quite appropriate. Indeed, Monte-Carlo simulations show that it understates the true ζ by .05 on average, and understates the standard deviation on the estimate, which is around .1. But even given those minor corrections, the estimates of ζ all remain around 1. See Dokkins and Ioannides [1998a] for state-of-the-art measurement of ζ.

4. There are slight variations on the expression of Zipf's law. The most common one is the "rank-size rule," which subsection III.4 discusses. Its expression is less convenient than the above probabilistic representation. Also, Gell-Mann [1994, p. 95] proposes the modification $P(\text{Size} > S) = a/(S + c)^{\zeta}$, where c is some constant. This paper sticks to the traditional representation (with $c = 0$) of Zipf's law, for two reasons. First, there is an immense empirical literature that studies this representation. Second, theory turns out to say that the representation with the constant $c = 0$ is the one we should expect to hold.

We can repeat the exercise, with similar, though less clean results, for other periods in U. S. history [Dobkins and Ioannides 1998a; Krugman 1996a, p. 41; 1996b; Zipf 1949, p. 420], for most countries in the modern period [Rosen and Resnick 1980]—and for even India in 1911 [Zipf 1949, p. 432], and China in the mid-nineteenth century [Rozman 1990, p. 68].

The striking fit of Zipf's law has generated many attempts at an explanation. Section IV will review the various theories that have been proposed, showing that each of them presents considerable difficulties. This paper will provide a very simple reason for the emergence of Zipf's law. It draws on the insights introduced in economics by Champernowne [1953]. Indeed, it has long been noted, in economics since at least Champernowne [1953], in physics before that (see Section VI), and then repeatedly in the literature on Zipf's law (e.g., Richardson's survey [1973]) that random growth processes could generate power laws.[5] But these studies stopped short of explaining why the Zipf exponent should be 1. This paper shows that the most natural conditions on the Markov chain (identical growth process across sizes) necessarily lead to this exponent of 1.

Consider a situation where there are a fixed number of cities (we will see that nothing changes with varying number), and that, over time, their sizes grow (and possibly shrink) stochastically. Assume only that, at least for a certain range of (normalized) sizes, the cities follow similar processes; i.e., their growth processes[6] have a common mean (equal to the mean city growth rate) and a common variance. This homogeneity of growth processes is often referred to as Gibrat's law,[7] after Gibrat[8] [1931]. Then, automatically, in the steady state, the distribution of cities in that range will follow Zipf's law with a power exponent of 1.

This necessary emergence of Zipf's law may sound surprising. An analogy for it would be the central limit theorem:[9] if we take a variable of arbitrary distribution (of finite variance) and calculate

5. These random growth processes have been recently rediscovered by the physicists Levy and Solomon [1996].

6. In percentage, not absolute, terms, of course.

7. Hence, the formal definition of Gibrat's law is that the probability distribution of the growth process $\gamma_{t+1} = S_{t+1}/S_t$ does not depend on the initial size S_t. Note that it can depend on things other than the size S_t, e.g., on the main industrial activity of the city. The requirement is that conditioning *only* on the size of the city does not bring information about the growth process.

8. For a review of the recent literature that started with Gibrat's work, see Sutton [1997].

9. I borrow this analogy from Casti [1995].

the mean of its successive realizations, normalized appropriately, this mean will always have (asymptotically) a normal distribution, independently of the characteristic of the initial process. Likewise, whatever the particulars driving the growth of cities, their economic role etc., as soon as they satisfy (at least over a certain range) Gibrat's law, their distribution will converge to Zipf.

More work is needed to establish this entirely, but it appears that empirical analyses seem to support Gibrat's law. The equality of the average growth rate across sizes has already been studied in the literature: see, Glaeser, Scheinkman, and Shleifer [1995] for the United States in the postwar period, and Eaton and Eckstein [1997] for France and Japan in the twentieth century. For the variance of the growth rate, using Eaton and Eckstein's data, it does not seem different across sizes.[10] This is reassuring, because there is also a sense, analyzed in Section V, in which Zipf exponents allow us to get information about the archeology of the growth processes, and Zipf exponents of 1 suggest that the process followed was close to Gibrat's law.

The proposed interpretation transforms a quite puzzling regularity—Zipf's law—into a pattern much easier to explain, Gibrat's law. On the other hand, the strength of these laws gives guidance to the theorist of city growth: models of city growth should deliver Gibrat's law in the upper tail.[11] In this light, more empirical work lies ahead to explain why Gibrat's law works so well, i.e., why traditional economic forces (e.g., as described in Henderson [1974]) seem to have so little power to statistically

10. Specifically, for each of the countries in their studies, we eliminate the capital to avoid problems that result from its specificity, and divide their sample into two, the upper half of the distribution in the initial period and its lower half, and calculate the growth rates of each half over the sample. To avoid Galton's fallacy problems, the starting date is the year at which the size criterion for the selection of cities has been chosen. An F-test evaluates the equality of the variances (see Hoel [1974, p. 140]). For both countries, actually, the variances of the log-growth rates are slightly *higher* for large cities than for small cities, though this difference is not statistically significant. For Japan 1965–1985, the variances are, respectively, 2.91 and 2.34 percent, and the F-statistic is $F = 1.03$, much below the critical value at the 5 percent level, $F^{.05}(19, 18) = 2.20$. For France 1911–1990, the variances are, respectively, 9.53 and 9.19 percent, $F = 1.04$, when the critical value is $F^{.05}(118, 18) = 2.17$.

11. Hence, economic models that explain city size distribution by relying on characteristics of hierarchies between cities, demand, supply curves, technological considerations, and the like (see, for instance, Henderson [1974, 1988]) are at best incomplete if they fail to satisfy Gibrat's law in the end, at least approximately. Equation (13) below explains quantitatively the sense in which they have to satisfy some approximation of Gibrat's law. Its simplest interpretation is that the averages and variances of growth rates have to be roughly independent of the size of the city. Subsection V.2 will allow us to make this more quantitative.

shape the city size distributions, as compared with the mechanical inertial forces emphasized in subsection II.2.

Section II shows the basic idea in a simple way. Section III states some propositions, and shows how the assumptions can be somewhat relaxed. Section IV contrasts this paper with the related literature. Section V discusses empirical variations around the base line Zipf exponent of 1, and shows how simple deviations from Gibrat's law (the small to medium cities have larger variance) allow us to explain them.

II. The Basic Insight: Zipf's Law as the Steady State Distribution Arising from Gibrat's Law

II.1. The Basic Idea

First, let us establish some notation. If $\tilde{S}$ is the size of a city, Zipf's law can be expressed as $P(\tilde{S} > S) = aS^{-\zeta}$ for some a (and over a large range of sizes S), where ζ denotes the exponent in Zipf's law. Zipf's law corresponds to the assertion that $\zeta = 1$. The corresponding density is $p(S) = bS^{-\zeta-1}$, for $b = \zeta a$. It will be useful to use the local Zipf exponent: $\zeta(S) := -Sp'(S)/p(S)$, where $p(S)$ is the probability distribution[12] of S.

In the basic model, there will be a fixed number of cities (the next section will show that new cities do not change anything), say N. Consider the following baseline situation: start with an initial, arbitrary distribution of cities. Let each city grow at an arbitrary mean rate, say 2 percent (it does not matter if this mean rate is time varying[13]), but around this mean growth cities have year-to-year (decade-to-decade) shocks in their growth: so, their growth is 2 percent, plus or minus .2 percent, say, each year (the standard deviation can also vary with time). Let us allow the cities to evolve freely, and study their limit distribution. To get some normalization, let us note their *normalized* size, $S^i (i = 1, \ldots, N)$. That is, S^i is the population of city i divided by the total urban population.[14] So $\Sigma_{i=1}^{N} S_t^i = 1$ at each date t.

12. An exponent of ζ for the tail distribution is equivalent to an exponent of $\zeta + 1$ for the density.

13. This is a conjecture that we firmly believe to be true. The reason is that the Zipf distribution, with an exponent of 1, is still a solution of the steady state equation (4), even when $f(\gamma)$ is time-varying, i.e., $f(\gamma,t)$: it still satisfies $\int_0^\infty \gamma f(\gamma,t)\, d\gamma = 1$. However, we could not find any argument in the mathematical literature—here we deal with Markov chains with time-varying transition matrices—to help us establish this rigorously.

14. More rigorously, S_t^t should be defined as $S_t^i = P_{it}/$(Total expected urban population), i.e., $S_t^i := P_{it}/(P_0 e^{gt})$, if g is the expected growth rate of the population (g should be understood as continuously compounded). This way we get rid of

Consider that, at least in the upper tail, the process is of the form, $S_{t+1}^i = \gamma_{t+1}^i S_t^i$, where the γ_{t+1}^i's are independent and identically distributed random variables with a distribution $f(\gamma)$. $\gamma_{t+1}^i - 1$ is the growth rate of city i. The average normalized size must stay constant ($\Sigma_{i=1}^N S_t^i = 1$), which requires that $E[\gamma] = 1$ (the mean normalized growth rate is 0), or

$$\int_0^\infty \gamma f(\gamma)\, d\gamma = 1. \tag{2}$$

Let us call $G_t(S) := P(S_t > S)$ the tail distribution of city sizes as time t. The equation of motion for G_t is[15]

$$\begin{aligned} G_{t+1}(S) &= P(S_{t+1} > S) = P(\gamma_{t+1} S_t > S) = E[1_{S_t > S/\gamma_{t+1}}] \\ &= E[E[1_{S_t > S/\gamma_{t+1}} | \gamma_{t+1}]] = E[G_t(S/\gamma_{t+1})] \\ &= \int_0^\infty G_t\left(\frac{S}{\gamma}\right) f(\gamma)\, d\gamma. \end{aligned} \tag{3}$$

Suppose (Section III will give conditions for this to hold) that there is a steady state process $G_t = G$. It verifies that

$$G(S) = \int_0^\infty G\left(\frac{S}{\gamma}\right) f(\gamma)\, d\gamma. \tag{4}$$

Compare this with (2). A distribution of the type $G(S) = a/S$ satisfies the steady state equation (4). So Zipf's law is a very good candidate for steady state distribution. In fact, one can prove (in Section III) that this candidate is the only steady state distribution. This is the proposed explanation for Zipf's law. If cities grow randomly, with the same expected growth rate and the same standard deviation, the limit distribution will converge to Zipf's law.

It is possible to make the result more intuitive. There are two parts to it: first, the existence of a power law; then the existence of a power law of -1. The existence of a power law can be thought of as due to a simple physical principle: scale invariance. Because the growth process is the same at all scales, the final distribution process should be scale-invariant. This forces it to follow a power law. To see why the exponent of the power law is 1, a concrete

---

border effects for very large cities (i.e., S close to 1). The next equation would then read $E[\Sigma_{i=1}^N S_t^i] = 1$.

15. Here 1_A is the indicator function for set A. The expectations are over all random variables S_t, S_{t+1}, γ_{t+1}.

situation might help. Suppose that cities are on a discrete grid, and that at each point in time a city might double, or halve in size. Because we must satisfy the constraint that the average size (understood as share of the total population) be constant, the probability of doubling has to be $\frac{1}{3}$, and the probability of halving $\frac{2}{3}$ (the expected growth is $\frac{1}{3} \cdot 2 + \frac{2}{3} \cdot \frac{1}{2} - 1 = 0$). To see how the number of cities of a given size can be constant, take a size S. One can quickly convince oneself that the number of cities of size $2S$ should be half the number of cities of size S, and the number of cities of size $S/2$ should be double. This is precisely an expression of Zipf's law.[16]

For this explanation to be correct, city size processes must have the time to converage to Zipf's law. Indeed, Krugman [1996a, pp. 96–97] shows that an important problem with Simon's [1995] explanation is that his process takes too much time (at the limit, an infinite time) to converage to Zif's law. It is qualitatively clear that there needs to be enough variance in the city growth rates; indeed, if the variance σ^2 is zero, there is no convergence at all. In fact, the proposed approach passes this test, without difficulties. Empirical estimations[17] put the decennial variances on the order of .1. Monte-Carlo simulations[18] show that, starting from a large range of fairly spread distributions, one is quite close to $\zeta = 1$ in less than a century. Indeed, as soon as the total variance since the initial distribution reaches the order of $\sigma^2 T = .7$, the distribution is very close to the power law, with an exponent of $\zeta = 1.05$; hence seven decades are enough to reach this value of ζ. In twice that time, the Zipf exponent reaches $\zeta = 1.001$.[19] This explains why even very young (but dynamic) urban systems satisfy Zipf's law (with reasonable precision) quickly: for instance, the United

16. The reasoning is made in terms of the tail distribution, which should have an exponent of 1.

17. Dobkins and Ioannides [1998a] give $\sigma^2 = 20$ percent/decade for U. S. cities in the twentieth century. For U. K. cities, 1800–1850, one can calculate, from Bairoch, Batou, and Chèvre (1988): $\sigma^2 = 5$ percent/decade. France and Japan in the twentieth century seem to be relatively ossified—the theme of Eaton and Eckstein [1997]: their σ^2 are 1.26 and 1.32 percent/decade, respectively. Taking into account the positive autocorrelation in growth rates shown by Glaeser, Scheinksman, and Shleifer [1995] would increase these estimates.

18. They use the reflected geometric Brownian motion of the next section. The *Mathematica* program is available from the author upon request.

19. This appears to be right for initial distributions that have from very large tails—e.g., power law up to a Zipf exponent of 1, as we want to a finite mean—to quite thin tails, e.g., initial distributions that are power law with a Zipf exponent much larger than 1 (say of 20), and even that are Gaussian. Finally, by Proposition 1 we know that the mechanism is stable; the relatively quick convergence seen here means that the mechanism will have commensurately high stability.

States as early as 1790 [Zipf 1949, p. 420], and Argentina in 1860 [Smith 1990, p. 23].

II.2. An Economic Model that Predicts Gibrat's Law

Let us turn to a possible model in which the process described arises. It is a very simplistic model, and its only function is to make the point that the process that generated Zipf's law could very well be the product of simple neoclassical economic forces. No doubt it could be enriched in many directions.

Our objective is to have a model where cities' growths have the same mean and same variance (Gibrat's law). To have a common variance, we need to have citywide shocks.[20] Here these citywide shocks will be amenity shocks[21] (the end of this section considers other shocks). These shocks increase the utility derived from consumption in a multiplicative way. If city i has amenities of level a_{it}, an agent with consumption c living in it will receive a utility $u(c) = a_{it}c$. One should think of these amenity shocks as policy shocks (shocks to the level of taxes, to pollution, to the quality of the police, schools, or roads[22]) or natural shocks (earthquakes, diseases, or variations in harvests in less developed economies). The a_{it}'s are independent and identically distributed.

Population growth comes about by migration (this is historically the dominant factor). We have overlapping generation agents, or agents with probability δ of death.[23] The timing is the following: once they are born, the agents migrate to the city of their choice. Once they have chosen the city, and thus have paid the big cost of migration, they do not move any more until they die. In equilibrium the benefit of moving to a city with better amenities at time $t + 1$ would be much lower than the moving cost. Let us see what city an agent will choose when she is born, at time t. The wage in city i is w_{it}, and there is no capital and no social

20. Otherwise, if shocks are industry-specific, large, well-diversified cities will have a much lower variance than smaller cities. This is counterfactual (maybe in the same way that Luxembourg, whose size is one hundred times smaller than the United States, does not have a variance in growth rate 100 times that of the United States).

21. City-specific but not industry-specific "productivity shocks" would produce the same result, but would have a correspondence to reality somewhat more difficult to identify.

22. For instance, the city selects a commissioner for the police department. The technology imposes a unique commissioner per city (otherwise coordination problems are too high). The competence of the commissioner is only revealed after his election. This competence determines that of the police, hence, part of the a_{it} of the city.

23. See Blanchard [1985].

insurance policy. So, if the levels of amenities a_{it} are independent and identically distributed, the decision problem is simply one of short horizon: $\max_i a_{it} w_{it}$. Hence, in equilibrium all "utility-adjusted" wages will be the same: for all cities i we have

(5) $$a_{it} w_{it} = u_t,$$

where u_t is the common value of the utility. This is the first condition of the equilibrium.

Let N_i^y be the number of youngsters who decide to migrate to city i, and N_i^o the population of seasoned agents already in city i. We assume the production technology to be constant return to scale: $F(N_i^o, N_i^y) = N_i^o f(N_i^y/N_i^o)$. Hence, the wages of the young will be $w_i = w_i^y = f'(N_i^y/N_i^o)$. Combining this with the equalization of utilities (5), we get $N_i^y = N_i^o f'^{-1}(u_t/a_{it})$. Given that the increase in population of city i is then $\Delta N_{it} = N_{it}^y - \delta N_i^o$, we get our equation for the growth of city i:

(6) $$\gamma_{it} := \Delta N_{it}/N_{it} = f'^{-1}(u_t/a_{it}) - \delta.$$

We have reached our goal: because the a_{it}'s distribution is independent of the initial size N_{it} of city i, the city-growth processes are identical across sizes (Gibrat's law).[24]

Some comments might help align this very stylized model with reality. In the model above, the variance of growth rate does not decrease with the size of a city. As mentioned in the introduction, this appears to be empirically the case in the upper tail. This fact might run counter to the economic intuition that, because large cities contain more industries, this allows them to diversify the shocks they receive, and their variance should be smaller than for small cities. The explanation for the constant variance in the upper tail is most probably the following. Consider that the shocks to the growth rate are the sum of different shocks, as in

(7) $$\gamma^{it} = \overline{\gamma} + \gamma^{it}_{\text{policy}} + \gamma^{it}_{\text{region}} + \gamma^{it}_{\text{industries}},$$

(these quantities will be defined shortly) so that, assuming for simplicity independence, the variance of the growth rate of γ_{it} of a

24. Incidently, if total population growth is constant, u_t will be constant: $u_t = u$. Indeed, if $N_t = \Sigma_i N_{it}$ is the total urban population, the total growth rates is $\Delta N_t = \Sigma_i N_{it}^o \,(f'^{-1}(u_t/a_{it}) - \delta)$, so that in the limit of a very large number of cities, the law of large numbers give that u_t is the solution of $\Delta N_t/N_t = E[f'^{-1}(u_t/a)] - \delta$, which by strict concavity of f admit a unique solution in u_t, which is independent of t if the growth rate of the total urban population $\Delta N_{it}/N_t$ is constant with time.

typical city of size S is then

$$\sigma^2(S) = \sigma^2_{\text{policy}}(S) + \sigma^2_{\text{region}}(S) + \sigma^2_{\text{industries}}(S). \tag{8}$$

The component $\overline{\gamma}$ is just the mean growth rate in the country. $\gamma^{it}_{\text{policy}(i)}$ represents the shocks due to the quality of the public goods offered by the city, to the level of the taxes, etc. It might be reasonable to think that its variance does not depend on the size of the city: $\sigma^2_{\text{policy}}(S) = \sigma^2_{\text{policy}}$. $\gamma^{it}_{\text{region}}$ is the shock due to the "macroeconomic" performance of the region in which city i is located (for instance, California is in recession when the rest of the United States booms, or the mining regions of Europe receive negative shocks that affect all their cities). It is plausible that the regional shocks affect all cities of the region equally, so that it is scale independent: $\sigma^2_{\text{region}}(S) = \sigma^2_{\text{region}}$. Finally, the term $\gamma^{it}_{\text{industries}}$ represents the shocks to city i's population growth due to the shocks experienced by i's industries. It is the one whose variance may very well decrease with the size S of the city. In the extreme case, assume that the variation is of the form given by the central limit theorem, so that we have $\sigma^2_{\text{industries}}(S) = \sigma^2_{\text{industries}}/S$. Then the total variance of a city of size S is

$$\sigma^2(S) = [\sigma^2_{\text{policy}} + \sigma^2_{\text{region}}] + \sigma^2_{\text{industries}}/S. \tag{9}$$

The assumption that the variance is independent of the size of the city would be true in the upper tail, because the effects of industrial diversification have died out ($\sigma^2_{\text{policy}} + \sigma^2_{\text{region}} \gg \sigma^2_{\text{industries}}/S$), and the variance is due only to the size-invariant shocks,[25] σ^2_{policy}, and σ^2_{region}. More empirical work would be welcome to pin down the determinants of the variance of city size. But this analysis suggests that there is nothing extravagant in the empirical finding that variance reaches a positive floor in the upper tail.

III. The Details of the Mechanism

III.1. Models of Random Growth with Gibrat's Law in the Upper Tail

The previous section gave the essence of the mechanism that generated Zipf's law. This section will expose this mechanism in more detail. The substantive point to make here is that the above mechanism needs a small grain of sand to have a steady state.

25. For instance, in the above model, $\sigma^2_{\text{industries}(i)}(S) = \sigma^2_{\text{region}(i)}(S) = 0$, and $\sigma^2_{\text{policy}(i)}(S) = \sigma^2$.

Namely, it needs a mechanism that prevents the small cities from becoming too small. The clearest version of such a mechanism is given by a "random walk with (lower) barrier," which will be introduced now. A more general version of a similar idea can be found in Appendix 1.

For analytical simplicity, the model works in continuous time. (It will also be useful later, in Section V, to study deviations from Gibrat's law.) To express the size processes, start from (6), and take the continuous limit. Along well-known lines, this gives

$$dP_{it}/P_{it} = gdt + \sigma dB_{it},$$

where B_{it} is a Brownian motion, for some γ, σ, which depends on the discrete parameters of the model (u,a_{it},δ). Likewise, considering the normalized[26] sizes $S_{it} := P_{it}/(\text{total urban population expected at time } t)$, we get

$$dS_{it}/S_{it} = \mu dt + \sigma dB_{it}, \tag{10}$$

where the expected growth in normalized size is $\mu = \gamma(S) - \overline{\gamma}$, the difference between the growth rate $\gamma(S)$ of cities of size S and the mean growth rate $\overline{\gamma}$.

It is instructive to see why some mechanism preventing small cities from becoming too small is necessary.[27] If such a mechanism were not present, the city-size distribution would become degenerate. Indeed, in the continuous-time case,[28] we would get $S_t = S_0 \exp(-\sigma^2 t/2 + \sigma B_t)$, which has no steady state distribution. The city-size distribution would just be a log-normal, where most cities would have infinitesimal size.

To represent the idea that we need a force that prevents small cities from becoming too small, introduce a lower bound $S_{\min}$ on the size of cities. When the size reaches $S_{\min}$, it is prevented from going below it: its increment is $dS_t = S_t \max(\mu dt + \sigma dB_t, 0)$. (For more on reflected Brownian motion, see Harrison [1985].) We have the following proposition.

PROPOSITION 1. Suppose that the normalized sizes S follow the "reflected geometric Brownian motion" process $dS_t/S_t = \mu dt + \sigma dB_t$ for $S_t > S_{\min}$, and $dS_t = S_t \max(\mu dt + \sigma dB_t, 0)$, for $S_t \leq$

26. The remark in footnote 14 applies.
27. Sornette and Cont [1997] discuss these issues in a physics context.
28. The discrete time case would give the same result. The equation of motion $S_{t+1} = \gamma_{t+1} S_t$, with $E[\gamma_{t+1}] = 1$ would give by iteration $\ln S_t = \ln S_0 + \Sigma_{\tau=1}^{t} \ln \gamma_\tau$, and the law of large numbers shows that $\log S_t/t \to E[\log \gamma_{t+1}] < 0$ (by Jensen's inequality) which means that $\log S_t$ is very close to $-\infty$; i.e., S_t is very close to 0. Technically, it converges to 0 in probability.

$S_{\min}$, where $\mu < 0$ is a negative drift, and B_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 distribution with exponent $\zeta = 1/(1 - S_{\min}/\overline{S})$, where $\overline{S}$ is the mean city size. Hence, in the limit where the minimal allowable city size $S_{\min}$ tends to 0, the exponent ζ tends to 1.

Proof. Consider the process of log-sizes $s := \ln S$, which follows $ds = (\mu - \sigma^2/2)dt + \sigma dB_t$ for $s > s_{\min} := \ln S_{\min}$, and is reflected at $s_{\min}$. The explicit calculations in Harrison [1985, p. 15] show that the distribution of s converges to an exponential distribution $P(s > s') = e^{-\zeta(s'-s_{\min})}$ for $s' \geq s_{\min}$, for some ζ. In other words, $P(S > S') = (S'/S_{\min})^{-\zeta}$ for $S' \geq S_{\min}$. The normalization condition $E[S] = S$ gives the value of ζ: because $E[S] = \int_{S_{\min}}^{\infty} \zeta S'^{-(\zeta+1)} S_{\min}^{\zeta}\, dS' = S_{\min}\zeta/(\zeta - 1)$, we get $\zeta = 1/(1 - S_{\min}/\overline{S})$. □

Thus, Proposition 1 says that this lower barrier[29] is enough to induce the power law distribution of city sizes, and, more interestingly, that, as the barrier become lower ($S_{\min}$ tends to 0, hence becomes almost invisible), the exponent converges to 1. So, in essence, all we need is an infinitesimally small (low) barrier, to ensure that the steady state distribution will be Zipf with an exponent ζ very close[30] to 1.

The random walk with a barrier is just the starkest idealization of the situation where some force keeps cities from becoming too small. Any Markov chain with a strong enough repelling force would produce the same result, namely Zipf's law from Gibrat's law. Appendix 1 presents an alternative and analytically tractable variant of this idea of "random growth with an infinitesimal barrier," which is given by the "Kesten" processes, after the work of the mathematician Harry Kesten [1973].

The moral of this section is that quasi-Zipf's law distribution (with $\zeta \simeq 1$) can be obtained by adding some small impurity (a

29. An additional comment about the barrier may be useful. Above, the barrier has a fixed relative size, which means that it grows at a rate γ if the urban population grows at this rate. In fact, one can readily establish from equation (4) in Harrison [1985, p. 15], that the condition is slightly less stringent: the barrier has to grow at some rate $\gamma_0 > \gamma - \sigma^2/2$. (The reason is that the diffusion growth rate of the population of city size s minus the barrier s_0 is $\gamma - \sigma^2/2 - \gamma_0$, and has to be negative).

30. The reader should note that the formula $\zeta = 1/(1 - S_{\min}/\overline{S})$, while it helps clarify the analytical issues, is not suitable for empirical purposes. The reason is that, as Section V develops, the exponent ζ is lower than 1 for smaller cities. This strongly influences the impact of smallest city size on the Zipf exponent. In particular, with $\zeta(S)$ increasing in S, the ζ of the upper tail will satisfy $\zeta^{\text{upper tail}} < 1/(1 - S_{\min}/\overline{S})$.

lower bound $S_{\min}$ to the size of cities, a small repelling force of magnitude $\bar{\varepsilon}$ for Kesten processes) to the Gibrat assumption. The pure Zipf's law with the $\zeta = 1$ corresponds to the case where the friction becomes infinitesimal (the lower bound $S_{\min}$ tends to 0, the repelling force of size tends to 0). The rest of this section will assume that these frictions have become infinitestimal, and reason with the limit distribution with $\zeta = 1$.

III.2. Countries Formed of Heterogeneous Regions

One can relax the hypotheses leading to Zipf's law in another relevant direction. Consider the hypothesis of the common growth process. Suppose that, in the country, there are regions that behave quite differently. The outcome will not invalidate Zipf's law.

PROPOSITION 2. Suppose that the country is formed of R regions, and that in each region the hypotheses of Proposition 1 are satisfied, so that Zipf's law with exponent ζ is verified in each region. In particular, the growth processes are (in the upper tail) identical within each region, but not necessarily across regions. Then the asymptotic national city distribution exists and satisfies Zipf's law, with the same exponent ζ.

The proof is very simple. Consider that a region is a country, and apply Proposition 1. In region r we have $P(S > s \mid \text{the city is in region } r) \sim a_r/s^{\zeta}$. Now, note by λ_r the probability that a city is in region r ($\sum_{r=1}^{R}\lambda_r = 1$). Then, at the national level we have $P(S > s) = \sum_{r=1}^{R} P(S > s$ and the city is in region $r) = \sum_{r=1}^{R}\lambda_r P(S > s \mid \text{the city is in region } r) \sim \sum_{r=1}^{R}\lambda_r a_r/s^{\zeta} = a/s^{\zeta}$, with $a := \sum_{r=1}^{R}\lambda_r a_r$. □

Hence, Europe should follow Zipf's law if each European country follows it, and likewise the United States should follow Zipf's law if it can be decomposed into regions in each of which Gibrat's law is satisfied.

III.3. Urban Dynamics with New Cities

We state here that the appearance of new cities does not affect our prediction for Zipf's law, so long as it is not too important. More precisely, Proposition 3 states that Zipf's law still holds when the appearance rate of new cities υ is lower than the growth rate of existing cities γ. It also shows how the Zipf exponent becomes larger than 1 in the (much less relevant) case where $\upsilon > \gamma$.

PROPOSITION 3. Assume that the number of cities increases at a rate v that is not greater than the growth rate γ of existing cities: $v \leq \gamma$. Then, there is still a steady state distribution that satisfies Zipf's law with an exponent of 1 in the upper tail. When $v > \gamma$, in the continuous-time case, the steady-state distribution has an exponent ζ which is the positive root of $\zeta^2 - (1 - 2\gamma/\sigma^2)\zeta - 2v/\sigma^2 = 0$. In particular, it is greater than 1.

Proof. The case $v < \gamma$ of the proposition is proved in Appendix 2. The idea is that new cities are born too far from the upper tail to influence its distribution. To study the case $v \geq \gamma$, go to the continuous-time representation. Call $p(S,t)$ the distribution of S at time t. The evolution of $p(S,t)$ is given by the forward Kolmogorov equation,[31] modified to accommodate the appearance of new cities:

$$\frac{\partial}{\partial t} p(S,t) = -\frac{\partial}{\partial S}(\gamma S p(S,t)) + \frac{1}{2}\frac{\partial^2}{\partial S^2}(\sigma^2 S^2 p(S,t)) - vp(S,t),$$

where the term $-vp(S,t)$ reflects the fact that there are new cities. We get in the steady state $\sigma^2/2(S^2p(S))'' - \gamma(Sp(S))' - vp(S) = 0$, which leads ζ to be the positive solution of $h(\zeta) = \zeta^2 - (1 - 2\gamma/\sigma^2)\zeta - 2v/\sigma^2 = 0$. In particular, it is greater than 1 (observe that $h(0) < 0$, so that there is only one positive root, and $h(1) < 0$ when $v < \gamma$, which implies that it is greater than 1). $\square$

III.4. The Rank-Size Rule

Finally, we can explore the issue of the "rank-size rule," a sister of Zipf's law. Zipf's law, again, states that the probability that a city has a size greater than S decreases as $1/S$. We should expect the size $S_{(i)}$ of a city of rank i to follow a power law: the size of the city of rank i varies as $1/i$, and the ratio of the second largest city to the largest city should be 1/2, the ratio of city 3 and city 2, 2/3, and so on, which is the statement of the rank-size rule. These size ratios are often used to compare actual urban patterns with "ideal" (Zipf) patterns (e.g., for a recent instance see Alperovitch [1984] or Smith [1990]). In fact *even if Zipf's law is verified exactly, the rank-size rule will be verified only approximately,* if our probabilistic interpretation of Zipf's law is correct. We should not

31. See any book on stochastic calculus, e.g., Karatzas and Shreve [1991, p. 282]. For a pedagogical derivation see Dixit and Pindyck [1994, pp. 88–90].

expect the actual values of these ratios to be very close to their "ideal" values. This is formalized in the following proposition (the regularized beta function is defined in equation (19) of Appendix 2).

PROPOSITION 4. Order the cities by size ($S_{(1)} \geq S_{(2)} \geq \ldots$), and suppose that the steady-state distribution satisfies Zipf's law with an exponent of 1. Then, for $i < j$, the mean of $S_{(j)}/S_{(i)}$ is i/j, its standard deviation $\sqrt{(1 - i/j)i/[j(j + 1)]}$, and its median $B-1(\frac{1}{2}, i, j - i)$, the inverse at ½ of the regularized beta function. The mean of $S_{(i)}/S_{(j)}$ is $(j - 1)/(i - 1)$. These results hold even for a finite number of cities (i.e., they are exact, not asymptotic).

This teaches us two things. First, the formulation of the rank-size rule is in a sense correct, but only in expected values and then one considers ratios of the type "size of the smaller city/size of the larger city." The mean of the ratios $S_{(j)}/S_{(i)} < 1$, for a city i larger than city j, is i/j, but the expected ratio of the ratio $S_{(i)}/S_{(j)}$ is larger than j/i (a Jensen's inequality effect). Second, the reader can see in Table I that the standard deviations of these ratios are quite high. For instance, the size ratio between city 2 and city 1, has a mean of ½, but has a quite high standard deviation, namely .2887. Even the ratio between city 100 and city 10, which has an expected value of ⅒, has a standard deviation of .03. So, for instance, the fact that in a given country the second and third cities are quite close in size does not disprove Zipf's law.

IV. PREVIOUS ATTEMPTS AT EXPLAINING ZIPF'S LAW FOR CITIES

Let us contrast the present proposal with the vast range of previous explanations. We will just show the main directions, and

TABLE I
STATISTICS OF THE SIZE RATIO $S_{(j)}/S_{(i)}$ BETWEEN THE jTH LARGEST CITY AND THE iTH LARGEST CITY ($i < j$)

	$S_{(j)}/S_{(i)}$	$S_{(2)}/S_{(1)}$	$S_{(3)}/S_{(1)}$	$S_{(3)}/S_{(2)}$	$S_{(10)}/S_{(5)}$	$S_{(100)}/S_{(10)}$
Mean	i/j	.5	.333	.6667	.5	.1
Standard deviation	$\sqrt{\frac{(1 - i/j)i/j}{j + 1}}$	.2887	.2357	.2357	.1508	.0299
Median[a]	$B^{-1}(1/2, i, j - i)$	.5	.2929	.7071	.5	.0973

a. $B^{-1}(1/2, i, j - i)$ denotes the inverse at 1/2 of the regularized beta function with parameters i and $i - j$. (See Appendix 2 for a definition.)

their limitations here. The reader interested in this massive literature can consult Carroll [1982], a fairly comprehensive review of the literature. Suarez-Villa [1988] contains an update, and Krugman [1996b] describes the principal models. Economic models[32] have been inadequate. Arguably, the two most successful models have been Steindl's and Simon's.

Steindl's growth model. In Steindl's model [1965, 1968] new cities are born at a rate v, and existing cities grow at a rate of γ. The result is that the distribution of new cities will be in the form of a power law, with an exponent $\zeta = v/\gamma$, as a quick derivation shows.[33] However, this is quite problematic. It does not deliver the result we want, namely the exponent of 1. It delivers it only by assuming that historically $v = \gamma$. This is quite implausible empirically, especially for mature urban systems, for which $v < \gamma$.

Stochastic growth models. The most successful model of stochastic growth is Simon's[34] [1955]. But it has quite a few daunting problems. (For an exposition of Simon's model, see Krugman [1996a, 1996b].) In this model new migrants arrive at each period, and with a probability π they will form a new city, while with probability $1 - \pi$ they will go to an existing city. The probability that they choose to locate in a given city is proportional to its population. Then this model generates a power law, with exponent $\zeta = 1/(1 - \pi)$. Thus, the exponent of 1 has a very natural explanation: the probability π of new cities is small. This seems quite successful. However, this approach has at least two major drawbacks. First, as Krugman [1996a, pp. 96–97; 1996b] shows, there is a degeneracy at $\pi = 0$. To get an exponent $\zeta = 1$, one needs π very close to 0, and then the process converges infinitely slowly. A second problem is that it is essential to this model that the rate of growth of the number of cities has to be

32. These economic models [Losch 1954; Hoover 1954; Beckman 1958] rely on the idea of urban hierarchy. Imagine a pyramidal system of the functions of cities, with small cities containing only the most basic services, cities of the next level containing a few more (a doctor for example), and larger ones having yet further services (specialized doctors), etc. Under mild conditions, this will imply that small cities are more numerous than big ones. However, one does not see how this could lead to a power law, much less one with the slope of -1. The same problem arises with Henderson's [1974] model of the optimal size of cities.

33. The cities of size greater than S are the cities of age greater than $a = \ln S/\gamma$. Because of the form of the birth process, the number of these cities is proportional to $e^{-va} = e^{-v \ln S/\gamma} = S^{-v/\gamma}$, which gives the exponent $\zeta = v/\gamma$.

34. An ancestor of Simon's model is Yule [1924]. This type of model has been revived by Hill [1974] and Hill and Woodroofe [1975]. However, these models are tailored to their original object, the distribution of biological genera and species, and do not explain why the exponent of the distribution should be close to 1.

greater than the rate of growth of the population of the existing cities—a historical counterfactual.

The explanation proposed in the present paper resembles Simon's. But it should be emphasized that although the two models are mathematically similar, they are *economically* completely different. Simon's explanation, although stochastic in appearance, is fundamentally deterministic, and boils down to Steindl's model (because of the historical importance of Simon's model, Appendix 3 develops this point). In Simon's model the exponent of Zipf's law is close to 1 because the growth rate of the number of cities is close to the growth rate of existing cities. (Another way to put this is to say that Simon's model is unable to account for Zipf's law when the growth rate of the number of cities is lower than the growth rate of existing cities.) The present paper's model does not suppose (directly or as a consequence) this counterfactual. Here the source of Zipf's law is Gibrat's law, not assumptions on the emergence rate of new cities.[35] Besides, most readings of the Simon models had (mistakenly, as we see from Proposition 3) inferred that, to get the Zipf exponent of 1, the appearance rate of cities had to be close to 0. The present explanation does not have to make this assumption. This appearance rate so long as it is below the growth rate of existing cities, does not affect the Zipf exponent.

V. Deviations from an Exponent of 1

V.1. Empirical Facts on the Deviations from the Exponent of 1

The bulk of the upper tail satisfies Zipf's law with an exponent of 1. Otherwise, there are two deviations from this exponent of 1. The first one is simple: in most countries Zipf plots usually present an outlier, the capital, which has a bigger size

35. Another attempt by Curry [1964], works through entropy maximization. It happens that if one carefully designs a measure of "entropy" of city sizes, the distribution that maximizes it is the Zipf distribution. However, although this emergence of power laws is a convenient result of statistical mechanics, we are given no explanation of why the entropy should be maximized in the first place. More recently, Alperovitch [1982] proposes a more economic explanation, but has to calibrate the production functions of the model to get Zipf's law. Likewise, Marsili and Zhang [1998] propose a model with interactions between cities that they calibrate to get Zipf's law. Finally, Krugman [1996b] proposes a "percolation" model, where city sizes are driven by quality of the environment (e.g., the city is a port). The physical structure of the environment exhibits some power law features (because of the percolation phenomena), which drives the power law for the city distribution. However, as Krugman points out, this type of approach has difficulties explaining why the exponent should be so close to 1.

than Zipf's law would warrant. There is nothing surprising there, because the capital is indeed a peculiar object, driven by unique political forces [Ades and Glaeser 1995].

The second[36] deviation is less particular. When one starts from the upper tail and extrapolates to the middle of the tail according to Zipf's law, one sees that there are too few medium-to-small cities (100,000 inhabitants or less in the United States). Their Zipf exponent ζ is lower than 1 [Dobkins and Ioannides 1998a].[37]

The approach developed above gives an explanation for the lower Zipf exponent for small to medium cities. The reason is that they do not perfectly follow Gibrat's law: smaller cities have a bigger variance, as will be seen shortly. It is therefore natural to extend the theory to the case where Gibrat's law is not satisfied.

V.2. These Deviations Can Be Explained by a Deviation from Gibrat's Law

In order to preserve tractability, it is useful to use the continuous-time representation of the growth processes introduced in Section III. If $\mu(S)$ is the expected growth rate of a city of normalized size S, and $\sigma(S)$ its standard deviation, the normalized city size (dropping the index i of the city for convenience) will follow a process of the form,

$$dS_t/S_t = \mu(S)dt + \sigma(S)dB_t. \tag{11}$$

Call $p(S, t)$ the distribution of s at time t. The forward Kolmogorov equation (introduced in subsection III.1) gives its

36. An additional deviation rests on more uncertain historical data—hence is more elusive. Europe prior to the sixteenth century does not seem to obey Zipf's law. The distribution of cities is too flat [de Vries 1984, p. 94]. It is only subsequently that it has shifted toward a Zipf pattern. One can conjecture many explanations for this. It might be, for instance, that cities did not grow much above a certain level, because diseases spread too quickly in large cities, the transportation of food was imperfect, or rulers did not want to have cities that were too large, because they would be tempting prey. Finally, the link between the Zipf exponent and the level of development appears to be too debated to be a firmly established fact. Wheaton and Shishido [1981] find a cross-sectional, U-shaped relationship; whereas Parr [1985] sees such a pattern only for developed countries; he finds a monotonic, decreasing pattern for developing countries.

37. Dobkins and Ioannides [1998a] find that the Zipf exponent ζ seems to have decreased slightly over a century. The fact that smaller cities have a smaller Zipf exponent suggests a simple explanation for their finding: because they used a fixed cutoff (50,000 inhabitants), later decades have more cities in their sample. Hence, they contain more "small" (in relative size the object that matters here) cities, which have a lower ζ (a $\zeta < 1$) than large ones. Mechanically, this composition effect makes the aggregate Zipf exponent decrease over time. Of course, more work is needed to assess the truth of this explanation.

equation of motion:

$$(12)\quad \frac{\partial}{\partial t}p(S,t) = -\frac{\partial}{\partial s}(\mu(S)Sp(S,t)) + \frac{1}{2}\frac{\partial^2}{\partial s^2}(\sigma^2(S)S^2p(S,t)).$$

Consider the baseline case where μ and σ are independent of S. Because $E[S]$ must stay constant at $E[S] = 1/N$, this implies $\mu = 0$. (Otherwise, we have $E_0[S_t] = E_0|S_0|\ e^{\mu t}$, which either explodes or shrinks to 0.) Consider the steady state $(\partial/\partial t)p(S,t) = 0$, so that we can write $p(S,t) = p(S)$ (in the neutral case, where $\mu = 0$):

$$\frac{1}{2}\frac{\partial^2}{\partial s^2}(\sigma^2 S^2 p(S,t)) = 0.$$

This can be integrated $p(S) = aS^{-2}$, for some a, which is Zipf's law with exponent 1.

Consider now the steady state distribution $p(S,t) = p(S)$ in the more general case where μ and σ can depend on the size S. The Kolmogorov equation (12) then can be integrated into $-\mu(S)Sp(S,t) + \partial/\partial S(\sigma^2(S)S^2\ p(S,t))/2 = 0$. The Zipf exponent $\zeta := -Sp'(S)/p(S)$ is then

$$(13)\qquad \zeta(S) = 1 - 2\frac{\zeta(S) - \bar{\gamma}}{\sigma^2(S)} + \frac{S}{\sigma^2(S)}\frac{\partial\sigma^2(S)}{\partial S},$$

where μ was replaced with its expression $\gamma(S) - \bar{\gamma}$.

Hence, deviations from a Zipf exponent of 1 can be due to two causes: the means and the standard deviations. The directions are those expected a priori. If a range of city sizes has a high growth rate $(\gamma(s) - \bar{\gamma})$, its distribution will decay less quickly than in the pure Zipf case (ζ will be less than 1), because small cities constantly feed the stock of bigger cities. If it has a high variance, its distribution will likewise be flatter because of the higher mixing of small and big cities.

As mentioned in Section I, studies of modern data[38] support Gibrat's law for means: both Glaeser, Scheinkman, and Shleifer [1995] for the United States between 1950 and 1990 and Eaton and Eckstein [1997] for France and Japan in the twentieth century show that there is no difference in the mean growth rate

38. Robson [1973, pp. 80–81] for England in the nineteenth century, and de Vries [1984, p. 106] for Europe since 1500 find that small cities tend to grow more slowly. However, note that their data are by necessity of lesser quality than the data available for the modern period.

of large and smaller cities. Hence, 0 is a fair baseline estimate for $\gamma(s) - \overline{\gamma}$.

From economic considerations, we would expect $(\partial/\partial s)\sigma(S)^2 < 0$ for small to medium cities (e.g., for cities of rank > 100). If the growth of a city is driven by the performance of the industries it hosts, smaller cities, which contain fewer industries, will have a larger variance than bigger cities. Hence, we see why the higher variance of small cities explains their lower Zipf exponent. A quantitative investigation of this, i.e., measures of $(\partial/\partial S)\sigma(S)^2$, should be high on the empirical agenda.

Finally, formula (13) helps assess how far economic models can deviate from Gibrat's law in their predictions. Of course, it gives only constraints on the quantities $\sigma^2(S)$ and $\gamma(S)$ together, not on them individually. Let us examine what light it sheds on the issues of convergence (as in Barro and Sala-i-Martin [1995]). Say that empirically $\zeta(S)$ is between .8 and 1.2, and that $|[S/\sigma^2(S)][\partial\sigma^2(S)/\partial S]| < .2$ (a guesstimate on the deviations from Gibrat's law for variances, educated by the calculations mentioned in footnote 8). Then, taking, as in subsection II.1, an estimate of $\sigma^2 = .1$/decade, (13) indicates that for the growth rates $|\gamma(s) - \gamma| < .4 \cdot (\sigma^2/2) = 2$ percent per decade. This is a small convergence. Hence, the deviations from Gibrat's law for the mean growth rates are quite small. Another way to put this would be to say that the ($\beta-$) convergence process, if any, is very slow. But this is not surprising, because the good fit of Zipf's law over centuries is the indication that no σ-convergence is taking place for cities, as opposed to what happens for countries and regions.

VI. Other Power Laws in Economics

One finds in economics[39] few other relationships as accurate as Zipf's law for cities. The closest ones are the distribution of income and the distribution of firms, explored, respectively, by Pareto [1896] and Gibrat [1931]. Pareto showed that distribution of income in the upper tail has a power law distribution. Gibrat

39. In the physical or biological sciences, however, power laws are rather common. See Gell-Mann [1994] for an exposition. After considerable effort has been devoted to explaining them, there is no consensus. Among the leading candidates is the recent theory of self-organized criticality (Bak, Tang, and Wiesenfeld [1987]; see Bak [1996] for a popular exposition; Raup [1997] for a more skeptical assessment). Some advocate chaotic processes (see, e.g., Schuster [1995, pp. 92–97]). Processes of random multiplicative growth have also enjoyed renewed popularity (see Sornette [1998] for a survey).

was interested in firms, and offered for the first time—in economics—a model of random growth. He observed that the distribution of the size (as measured by sales or number of employees) of firms tends to be lognormal.[40] He gave the simple explanation that the growth process of firms could be multiplicative—of the form $S_{t+1} = \gamma_{t+1}S_t$—and independent of firm size, hence, of the form of "Gibrat's law." Indeed, as we saw in subsection III.1, taking logs and using the central limit theorem delivers asymptotically a log-normal distribution.

This explanation has the drawback that the resulting process does not converge to a steady state distribution: as seen in subsection III.1, the resulting asymptotic distribution is degenerate with an infinite variance for the log-size. There are two remedies for this. One is to use some force that prevents small sizes from becoming too small. This is the assumption used by Champernowne [1953], who shows that this leads the distribution to converge to a power law distribution. The other one, explored by Rutherford [1955], is to introduce a birth and death process.[41] The present paper relies on their and Gibrat's insight. Mandelbrot [1960] makes the interesting observation that all these models can be viewed as employing processes used long before in physics—to explain the exponential barometric density of the atmosphere.[42] This paper's approach could help to revisit the issue of income distribution. In particular, the formula in Proposition 3 could be used, using v as the death rate of high-income individual instead of the birthrate of new cities. This would give the predicted exponent of income distribution, and be useful in calibrations.

40. Simon and coauthors [Ijiri and Simon 1977] argue that the upper tail of the size distribution of firms looks more like a power law than a lognormal, and thus decreases less rapidly than a lognormal. They propose Simon's [1955] model as an explanation. However, more recent evidence by Stanley et al. [1995] shows that the distribution of the firm listed in Compustat indeed looks like a lognormal, except that it seems to decrease *more* rapidly in the upper tail than a lognormal, rather than less. So, at this stage, the stylized facts on the empirical distribution of firms are still a debated issue.

41. His process does not lead to a power law distribution in his assumptions, but would lead to one if his birthrate had bounded support.

42. Namely, Laplace's law states that the barometric density distribution in the atmosphere varies in $e^{-\zeta s}$, where s is altitude. It is known from physics that this density is the outcome of two forces. Consider that atmosphere is made of Brownian particles that diffuse because of heat motion (the equivalent of randomness in our models), and the gravity that pulls them to the bottom (the negative drift in the model of Proposition 1). The earth (altitude 0) plays the role of the lower barrier of the process. This leads to the exponential Laplace's law for the atmospheric density. The log-sizes in the random growth models and the altitude in the atmospheric model play analog roles.

Finally, note an important difference between Pareto's law for the income distribution and Zipf's law for cities, which can be readily explained here. For cities the ζ exponent is always very close to 1 in the upper tail, whereas the ζ exponent for the income distribution seems both to vary cross-sectionally and to be quite unstable from year to year. (For instance, Feenberg and Poterba [1993] calculate the exponent ζ for the United States, and find that it oscillates between 1.59 and 2.46 between 1970 and 1990). Proposition 3 can offer an explanation for this. The key is the birthrate of new cities v, which can be reinterpreted as the death rate of individuals in the case of incomes. For cities we have $v < \gamma$, so that the resulting exponent does not depend on the details of the country's situation: it is just 1, or very close to it. For incomes we have $v > \gamma$, in which case the exponent depends finely on the situation's parameters, v, γ, σ, which explains why ζ loses its constancy across economic structures and has cross-sectional and possibly time series variations.

VII. Conclusion

The present approach explains why we have Zipf's law across countries with very different economic structures and histories (China in the mid-nineteenth century, India in the early twentieth century, the early and modern United States, and indeed most countries for which we have data). This phenomenon is simply due to Gibrat's law—the fact that cities in the upper tail follow similar growth processes, although these assumptions can be somewhat relaxed. Thus, the task of economic analysis is reduced from explaining the quite surprising Zipf's law to the investigation of the more mundane Gibrat's law. The simplest reason[43] for the latter, the one expressed in the minimalist model of this paper, is that, above a certain size, most shocks stop declining with size, such as regional shocks (shocks to the regional activity, or taste

43. Gabaix [1999] discusses this further and presents a dichotomy: either cities behave like constant-return-to-scale economies in the upper tail (the route pursued in the present paper), or Gibrat's law would be due to endogenously counterbalancing effects of unbounded differences in externalities and productivities (e.g., large cities would have unboundedly larger productivity than small cities, but would suffer from unboundedly larger disamenities). An important question for future empirical research is to determine in which one of those two worlds we are. If we are in the first world, the consequences for local growth modeling would be stark because they would drastically simplify existing models. Seeing how this is compatible with the new economic geography [Fujita, Krugman, and Venables 1999] would also be extremely interesting.

shock to its climate), or municipal policy shocks (more efficient police, or education, or higher taxes). The variance of city growth reaches a positive floor in the upper tail of the size distribution, which makes Gibrat's law hold in this upper tail. Both this phenomenon (Gibrat's law) and its causes (regional versus city-specific shocks, why the diversification effect seems so small) could be fruitfully explored (for recent work see Glaeser, Scheinkman, and Shleifer [1995], Eaton and Eckstein [1997], and Dobkins and Ioannides [1998a, 1998b]). The same approach also proposes a quantitative explanation of why smaller cities have a lower Zipf exponent: these cities have a larger variance than do bigger ones.

APPENDIX 1: ALTERNATIVE DERIVATION OF THE MECHANISM WITH "KESTEN" PROCESSES

Consider indeed that the evolution of the normalized size S_t of a city follows

$$S_{t+1} = \gamma_{t+1} S_t + \varepsilon_{t+1}, \tag{14}$$

where ε_{t+1} is a small, positive increment, with mean $E[\varepsilon_{t+1}] =: \bar{\varepsilon} \geq 0$. In the limit $\bar{\varepsilon} = 0$, we get the "pure" Gibrat process where city growth is independent of city size. Consider the case $\bar{\varepsilon} > 0$. When the city size is big, the ε_{t+1} term is negligible, and we are very close to the pure "Gibrat" case: $S_{t+1} \simeq \gamma_{t+1} S_t$. The term in ε in (14) matters when the size of the city is small. It represents the force that prevents cities from becoming infinitesimally small, analogous to the lower barrier in the random walk above. From Kesten [1973], we can derive easily the following proposition.

PROPOSITION 5. Suppose that each normalized city size follows the growth process $S_{t+1} = \gamma_{t+1} S_t + \varepsilon_{t+1}$, with $E[\varepsilon_{t+1}] = \bar{\varepsilon} > 0$. Then, whatever the initial distribution, the city-size distribution converges to a power law distribution with positive exponent ζ such that $E[\gamma^{\zeta}] = 1$. For small $\bar{\varepsilon}$, $\zeta = 1 + O(\bar{\varepsilon})$, so that when $\bar{\varepsilon} \to 0$, the city-size distribution converges to a Zipf distribution with exponent 1. When γ_{t+1} is lognormally distributed with $\sigma^2 = \text{var} \ln \gamma_{t+1}$, we have the exact relation $\zeta = 1 - 2 \ln (1 - \bar{\varepsilon}/\bar{S})/\sigma^2$ for all $\bar{\varepsilon}$, where $\bar{S}$ is the mean size of a city.

Proof. Because we work with normalized sizes, the normalization $E[S_t] = \bar{S}$ implies $E[\gamma_{t+1}] = 1 - \varepsilon$, where $\varepsilon := \bar{\varepsilon}/\bar{S}$. By Jensen's inequality, this implies that $E[\ln \gamma] \leq \ln (1 - \varepsilon) < 0$, and we can apply Kesten [1973], which shows that the distribution converges to a Zipf distribution with an exponent ζ that is the positive root of $E[\gamma_{t+1}^{\zeta}] = 1$.

A heuristic rederivation of the value of ζ might be of interest; it mimics the one given in Section II. Suppose that the steady state distribution $G(S) := P(S_t > S)$ is of the form, $G(S) \sim aS^{-\zeta}$. Because $S_{t+1} \sim \gamma_{t+1}S_t$ in the upper tail, the same reasoning as equation (3) again gives equation (4): $G(S) \sim E[G(S/\gamma)]$, which leads to $E[\gamma^{\zeta}] = 1$ by using this form of $G(S)$.

In the case, where γ_{t+1} is lognormal with $\sigma^2 = \text{var} \ln \gamma_{t+1}$, we have $\gamma_{t+1} = \exp(\ln(1-\varepsilon) - \sigma^2/2 + \sigma u)$, where u is a standard normal of mean 0 and variance 1. So $E[\gamma_{t+1}] = \exp((\ln(1-\varepsilon) - \sigma^2/2)\,\zeta + \zeta^2\,\sigma^2/2)$, so the positive root ζ of $E[\gamma_{t+1}^{\zeta}] = 1$ is $\zeta = 1 - 2\ln(1-\varepsilon)/\sigma^2$.

In the general case, where γ_{t+1} is not necessarily lognormal, but with ε in a neighborhood of 0, define $\gamma_0 := \gamma/(1-\varepsilon)$. ζ is the positive root of $f(\zeta,\varepsilon) = 1$ for $f(\zeta,\varepsilon) := E[\gamma_0^{\zeta}\,(1-\varepsilon)^{\zeta}] = 1$. Let us use the implicit function theorem around $(\zeta, \varepsilon) = (1, 0)$. We have $f(0,1) = 1$, $f_{\varepsilon}(0,1) = -1$, and $f_{\zeta}(0,1) = E[\gamma_0 \ln \gamma_0]$, which is positive by Jensen's inequality ($x \ln x$ is convex, and $E[\gamma_0] = 0$). The implicit function theorem gives $\zeta(\varepsilon) = 1 + \varepsilon E[\gamma_0 \ln \gamma_0] + o(\varepsilon)$. Because $E[\gamma_0 \ln \gamma_0] = E[\gamma \ln \gamma] + O(\varepsilon)$, we get a statement in fact slightly stronger than the proposition: $\zeta(\varepsilon) = 1 + a\varepsilon + o(\varepsilon)$ with $a = 1/E[\gamma \ln \gamma]$. □

Appendix 2: Proofs

Proof of Proposition 3

Consider the limit where the time at which we count the cities, t, tends to $+\infty$. Fix a large time T, such that $T \to +\infty$ (for instance, $T = t/2$). We will differentiate between old cities, founded before T and new cities, founded after T. Consider a large city size $S = e^s$ in the upper tail. The probability of finding a city of size greater than S is $P(\tilde{S} > S) = P(\tilde{S} > S, b < T) + P(\tilde{S} > S, b \geq T)$, where b represents the random variable noting the birth date, and, as usual, the comma in the probabilities signifies "and." We will show that the first term dominates the second one, so that new cities do not matter for Zipf's law in the upper tail.

First, note that for cities born at a date τ we have

$$P(\tilde{S} > S \mid b = \tau) \sim e^{-(s - \gamma(t-\tau))} \tag{15}$$

because those cities have an age $t - \tau > t - T$, which tends to $+\infty$, so that Proposition 1 on Zipf's law without new cities applies. The term $e^{-\gamma(t-\tau)}$ is the appropriate normalization factor for cities founded at date $b = \tau$, for they have a mean size of $e^{\gamma(t-\tau)}$ at date t—remember that $E[dS_t/S_s] = \gamma$). Given the birth process of new cities, at time t there are $N_t = \int_0^t ve^{v\tau}\,d\tau = e^{vt} - 1$ cities, and the density at time t of cities born at date τ is $ve^{v\tau}/N_t = ve^{v\tau - n_t}$ where $n_t := \ln N_t$. By integration (15) gives

$$P(\tilde{S} > S, b < T) \sim \int_0^T e^{-s+\gamma(t-\tau)} ve^{v\tau - n_t}\,d\tau = e^{-s+\gamma t - n_t} v \int_0^T e^{-(\gamma - v)\tau}\,d\tau,$$

so that

$$P(\tilde{S} > S, b < T) \sim e^{-s+\gamma t - n_t} v/(\gamma - v) \tag{16}$$

as $T \to +\infty$, because $\gamma - v > 0$. The cities born at a date $\tau > T$ did not necessarily have the time to converge to Zipf's law, but it still will be the case that

$$P(\tilde{S} > S \mid b = \tau) \leq E[\tilde{S}/S \mid b = \tau] = e^{\gamma(t-\tau)-s},$$

so that

$$P(\tilde{S} > S, b \geq T) \leq \int_T^t e^{\gamma(t-\tau)-s} ve^{v\tau - n_t}\,d\tau \sim \frac{e^{-s+\gamma t-(\gamma-v)T-n_t}v}{\gamma - v}. \tag{17}$$

So, combining (16) and (17), $P(\tilde{S} > S, b \geq T) = O(P(S_t > S, b < T)e^{-(\gamma - v)T}) = o(P(S_t > S, b < T))$, for T large, and $P(S_t > S) = P(S_t > S, b < T) + (S_t > S, b \geq T) \sim P(S_t > S, b < T) \sim e^{-s+\gamma t - n_t} v/(\gamma - v)$, which is the form $P(S_t > S) \sim a_t e^{-s} = a_t/S$. This is the expression of Zipf's law.

Proof of Proposition 4

Order the cities by size ($S_{(1)} \geq S_{(2)} \geq \ldots$), call the corresponding log-sizes $s_{(1)} \geq s_{(2)} \geq \ldots$ $(s_{(i)} := \ln S_{(i)})$. By normalization, under Zipf's law the log sizes follow an exponential distribution: $P(s > t) = e^{-t}$ for $t \geq 0$. Then, the Rényi representation theorem on ordered statistics (see, e.g., Reiss [1989, pp. 36–37]) gives that, for $i < j$, the difference $s_{(i)} - s_{(j)}$ can be written as

$$s_{(i)} - s_{(j)} = \sum_{k=1}^{j-1} \frac{x_k}{k}, \tag{18}$$

where the x_k are independent draws of an exponential distribution $P(x_k > x) = e^{-x}$ for $x \geq 0$. This allows us to calculate conveniently the statistics we need:

$$E\left[\frac{S_{(j)}}{S(i)}\right] = E[\exp(s_{(j)} - s(i)] = E\left[\exp\left(-\sum_{k=i}^{j-1}\frac{x_k}{k}\right)\right]$$

$$= \prod_{k=1}^{j-1} E\left[\exp\left(-\frac{x_k}{k}\right)\right] = \prod_{k=1}^{j-1}\frac{k}{k+1} = \frac{i}{j}.$$

Note that the same procedure would give $E[S_{(i)}/S_{(j)}] = (j-1)/(i-1)$. The same type of calculation gives very simply

$$E\left[\left(\frac{S_{(j)}}{S(i)}\right)^2\right] = \prod_{k=i}^{j-1} E\left[\exp\left(-\frac{2x_k}{k}\right)\right] = \prod_{k=i}^{j-1}\frac{k}{k+2} = \frac{i(i+1)}{j(j+1)}.$$

Finally, somewhat longer calculations give

$$\text{(19)} \qquad P\left(\frac{S_{(j)}}{S_{(i)}} \leq \mu\right) = P(\exp(s_{(i)} - s_{(j)}) \leq \mu)$$

$$= \frac{\int_0^{\mu} u^{i-1}(1-\mu)^{j-i}\,du}{\int_0^1 u^{i-1}(1-\mu)^{j-i}\,du} := B(\mu,i,j-i),$$

which is sometimes called the regularized beta function with parameters $i, j-i$. The median μ is then given by its inverse at the point ½, $B^{-1}(½,i,j-i)$.

Appendix 3: Simon's Model Boils Down to Steindl's Model

Given its importance in the literature on Zipf's law, it might be of interest to shed some new light on Simon's model. This appendix shows how Simon's explanation of Zipf is in fact a particular case of Steindl's, and makes the point that its economic nature is essentially different from the present paper's. Call τ the time in the model,[44] so that, given a new inhabitant per period, the urban population is $U_\tau = \tau$ (we can start at $\tau > 0$). We seek a correspondence between the "time of the model" τ and the "real time," noted t. If the real time is such that the population at real time t is $U_t = e^{\gamma_0 t}$, the correspondence between real time t and model time τ is

$$\text{(20)} \qquad \tau = e^{\gamma_0 t}.$$

44. For a clear exposition of Simon's model, see Krugman [1996a, 1996b].

In model time, the number of new cities follows: $E[N_{\tau+1} - N_\tau] = \pi$, or by the law of large numbers: $N_\tau \sim \pi\tau = \pi e^{\gamma_0 t}$. So, in real time, the rate of growth of the number of cities is

$$v = \gamma_0. \tag{21}$$

We now look for the growth rate of an individual city. Call its size S_t. Because of the way the process is set up, we have $E[dS_\tau/d\tau] = E[S_{\tau+1} - S_\tau] = (1 - \pi)S_\tau/U_\tau$, or because $d\tau = \gamma_0 e^{\gamma_0 t} dt = \gamma_0 U_t dt$,

$$E[dS_t] = \frac{(1-\pi)S_\tau}{U_\tau} d\tau = \frac{(1-\pi)S_\tau}{U_\tau} \gamma_0 U_\tau dt = (1-\pi)\gamma_0 S_t dt,$$

which gives the rate of growth of an individual city: $\gamma := E[dS_t/S_t]/dt$:

$$\gamma = (1 - \pi)\gamma_0. \tag{22}$$

Hence, as in the Steindl model, we get the steady state Zipf exponent:

$$\zeta = v/\gamma = 1/(1 - \pi). \tag{23}$$

We see how stringent (indirect) assumptions on the growth rate of the number of cities govern Simon's explanation. They are almost certainly counterfactual, at least in the long run, for many urban systems, e.g., in Europe.

DEPARTMENT OF ECONOMICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

REFERENCES

Ades, A., and E. Glaeser, "Trade and Circuses: Explaining Urban Giants," *Quarterly Journal of Economics,* CX (1995), 195–228.

Alperovitch, G., "Scale Economies and Diseconomies in the Determination of City Size Distribution," *Journal of Urban Economics,* XII (1982), 202–213.

——, "The Size Distribution of Cities: On the Empirical Validity of the Rank-Size Rule," *Journal of Urban Economics,* XVI (1984), 232–239.

Auerbach, F., "Das Gesetz der Bevölkerungskonzentration," *Petermanns Geographische Mitteilungen,* LIX (1913), 73–76.

Bairoch, P., J. Batou, and P. Chèvre, *The Population of European Cities: 800–1850* (Geneva, Switzerland: Droz, 1988).

Bak, P., *How Nature Works: The Science of Self-Organized Criticality* (New York: Copernicus Springer Verlag, 1996).

Bak, P., C. Tang, and K. Wiesenfeld, "Self-Organized Criticality. An Explanation of 1/*f* Noise," *Physical Review Letters,* LIX (1987), 381–385.

Barro, R., and X. Sala-i-Martin, *Economic Growth* (New York, NY: McGraw-Hill, 1995).

Beckman, M., "City Hierarchies and the Distribution of City Size," *Economic Development and Cultural Change,* VI (1958), 243–248.

Blanchard, O., "Debt, Deficits, and Finite Horizons," *Journal of Political Economy,* XCIII (1985), 223–247.

Carroll, G., "National City-Size Distribution: What Do We Know After 67 Years of Research?" *Progress in Human Geography,* VI (1982), 1–43.
Casti, J., "Bell Curves and Monkey Languages," *Complexity,* I (1995), 12–15.
Champernowne, D., "A Model of Income Distribution," *Economic Journal,* LXIII (1953), 318–351.
Curry, L., "The Random Spatial Economy: An Exploration in Settlement Theory," *Annals of the Association of American Geographers,* LIV (1964), 138–146.
De Vries, J., *European Urbanization, 1500–1800* (Cambridge, MA: Harvard University Press, 1984).
Dixit, A., and R. Pindyck, *Investment under Uncertainty* (Princeton, NJ: Princeton University Press, 1994).
Dobkins, L., and Y. Ioannides, "Dynamic Evolution of the U. S. City Size Distribution," in J.-M. Huriot and J.-F. Thisse eds., *The Economics of Cities* (New York, NY: Cambridge University Press, 1998a).
Dobkins, L., and Y. Ioannides, "Spatial Interactions among U. S. Cities," mimeo, Tufts University, 1998b.
Eaton, J., and Z. Eckstein, "City and Growth: Theory and Evidence from France and Japan," *Regional Science and Urban Economics,* XXVII (1997), 443–474.
Feenberg, D., and J. Poterba, "Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns," in *Tax Policy and the Economy,* vol. 7, J. Poterba ed. (Cambridge, MA: MIT Press, 1993), pp. 145–177.
Fujita, M., P. Krugman, and A. Venables, *The Spatial Economy* (Cambridge, MA: MIT Press, 1999).
Gabaix, X., "Zipf's Law and the Growth of Cities," *American Economic Review Papers and Proceedings,* LXXXIX (1999), 129–132.
Gell-Mann, M., *The Quark and the Jaguar* (New York, NY: Freeman, 1994).
Gibrat, R., *Les inégalités économiques* (Paris, France: Librairie du Recueil Sirey, 1931).
Glaeser, E., J. Scheinkman, and A. Shleifer, "Economic Growth in a Cross-Section of Cities," *Journal of Monetary Economics,* XXXVI (1995), 117–143.
Harrison, J., *Brownian Motion and Stochastic Flow Systems* (Malabar, FL: Krieger, 1985).
Henderson, J., "The Sizes and Types of Cities," *American Economic Review,* LXIV (1974), 640–656.
——, *Urban Development: Theory, Fact, and Illusion* (Oxford University Press, 1988).
Hill, B., "The Rank-Frequency Form of Zipf's Law," *Journal of the American Statistical Association,* LXIX (1974), 1017–1026.
Hill, B., and M. Woodroofe, "Stronger Forms of Zipf's Law," *Journal of the American Statistical Association,* LXX (1975), 212–219.
Hoel, P., *Introduction to Mathematical Statistics* (New York, NY: John Wiley & Sons, 1974).
Hoover, E. M., "The Concept of a System of Cities: A Comment on Rutledge Vining's Paper," *Economic and Cultural Change,* III (1954), 196–198.
Ijiri, Y., and H. Simon, *Skew Distributions and the Sizes of Business Firms* (Amsterdam, Netherlands, North-Holland, 1977).
Karatzas, I., and S. Shreve, *Brownian Motion and Stochastic Calculus,* 2nd edition (Berlin, Germany: Springer-Verlag, 1991).
Kesten, H., "Random Difference Equations and Renewal Theory for Products of Random Matrices," *Acta Mathematica,* CXXXI (1973), 207–248.
Krugman, P., *The Self-Organizing Economy* (Cambridge, MA: Blackwell, 1996a).
——, "Confronting the Urban Mystery," *Journal of the Japanese and International Economies,* X (1996b), 399–418.
Levy, M., and S. Solomon, "Dynamical Explanation for the Emergence of Power Law in a Stock Market Model," *International Journal of Modern Physics* C, VII, No. 1 (1996), 65–72.
Losch, A., *The Economics of Location* (translation) (New Haven, CT: Yale University Press, 1954).
Mandelbrot, B., "The Pareto-Levy Law and the Distribution of Income," *International Economic Review,* I (1960), 79–106.
Marsili, M., and Y.-C. Zhang, "Interacting Individuals Leading to Zipf's Law," *Physical Review Letters,* LXXX (1998), 2741–2744.

Pareto, V., *Cours d'Economie Politique* (Geneva, Switzerland: Droz, 1896).
Parr, J., "A Note on the Size Distribution of Cities over Time," *Journal of Urban Economics,* XVIII (1985), 199–212.
Raup, E., "Review of: *How Nature Works,*" *Complexity,* III (August 1997), 30–33.
Reiss, R., *Approximate Distributions of Order Statistics* (Berlin, Germany: Springer-Verlag, 1989).
Richardson, H., "Theory of the Distribution of City Sizes: Review and Prospects," *Regional Studies,* VII (1973), 239–251.
Robson, B., *Urban Growth: An Approach* (London, UK: Methuen, 1973).
Rosen, K., and M. Resnick, "The Size Distribution of Cities: An Examination of the Pareto Law and Primacy," *Journal of Urban Economics,* VIII (1980), 165–186.
Rozman, G., "East Asian Urbanization in the Nineteenth Century," in Van der Woude et al. (1990), pp. 61–73.
Rutherford, R., "Income Distributions: A New Model," *Econometrica,* XXIII (1955), 277–294.
Schuster, H., *Deterministic Chaos* (Weinheim, Germany: VCH, 1995).
Simon, H., "On a Class of Skew Distribution Functions," *Biometrika,* XLII (1955), 425–440.
Smith, C., "Types of City-Size Distributions: A Comparative Analysis," in Van der Woude et al. (1990), pp. 20–42.
Sornette, D., "Linear Stochastic Dynamics with Nonlinear Fractal Properties," *Physica A,* CCL (1998), 295–314.
Sornette, D., and R. Cont, "Convergent Multiplicative Processes Repelled from Zero: Power Laws and Truncated Power Laws," *Journal de Physique I France,* VII (1997), 431–444.
Stanley, M., et al., "Zipf Plots and the Size Distribution of Firms," *Economics Letters,* XLIX (1995), 453–457.
Statistical Abstract of the United States 1993, Bureau of the Census (Washington, DC: Government Printing Office, 1993).
Steindl, J., *Random Processes and the growth of Firms* (New York, NY: Hafner, 1965).
Steindl, J., "Size Distributions in Economics," in *International Encyclopedia of the Social Sciences,* Vol. 14, Silks, ed. (New York, NY: Macmillan and the Free Press, 1968).
Suarez-Villa, L., "Metropolitan Evolution, Sectoral Economic Change, and the City Size Distribution," *Urban Studies,* XXV (1988), 1–20.
Sutton, J., "Gibrat's Legacy," *Journal of Economic Literature,* XXXV (1997), 40–59.
Van der Woude, A., A. Hayami, and J. de Vries, editors, *Urbanization in History* (New York, NY: Oxford University Press, 1990).
Wheaton, W., and H. Shishido, "Urban Concentration, Agglomeration Economies and the Level of Economic Development," *Economic Development and Cultural Change,* XXX (1981), 17–30.
Yule, G., "A Mathematical Theory of Evolution Based on the Conclusions of Dr. J. C. Willis, F.R.S.," *Philosophical Transactions B,* CCXIII (1924), 21–87.
Zipf, G., *Human Behavior and the Principle of Last Effort* (Cambridge, MA: Addison-Wesley, 1949).