

Zipf's Law and the Growth of Cities

By XAVIER GABAIX *

Economic growth has been studied at many levels, including at the national (Robert Barro, 1991), the regional (Barro and Xavier Sala-i-Martin, 1992), and the local (Edward Glaeser et al., 1992)(i.e., at the level of the agglomeration). A striking pattern of agglomerations is Zipf's law for cities, which may well be the most accurate regularity in economics. It appears to hold in virtually all countries and dates for which there are data, even the United States in 1790 and India in 1911 (George Zipf, 1949; Kenneth Rosen and Mitchell Resnick, 1980). This paper will argue that satisfying Zipf's law should be a prerequisite for taking a model of local growth seriously, and it will provide some guidance for this.

To visualize Zipf's law, one may take a country, for instance, the United States in 1991, and order the largest cities by population: number 1 is New York, number 2 is Los Angeles, and so on. Then draw a graph: along the y -axis, plot the log of the rank (New York has log rank $\ln 1$, Los Angeles has log rank $\ln 2$); and along the x -axis plot the log of the population (i.e., "size") of the corresponding city. The resulting graph, shown in Figure 1, shows a straight line, which is rather surprising. Furthermore, its slope is -1 , based on the regression

$$\ln(\text{Rank}) = 10.53 - 1.005 \ln(\text{Size})$$

$$[0.010]$$

[†] *Discussants:* José Scheinkman, University of Chicago; Susanto Basu, University of Michigan; Gary S. Becker, University of Chicago.

* Harvard University and Department of Economics, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: xgabaix@kuznets.fas.harvard.edu). I thank Richard Baldwin, Robert Barro, Olivier Blanchard, Edward Glaeser, José Scheinkman (my discussant), and Jaume Ventura for helpful discussions and comments.

where the standard deviation appears in brackets and the R^2 is 0.986. The slope of the curve is very close to -1 . This is an expression of Zipf's law; when one plots log-rank against log-size, the result is a straight line, with a slope, which I shall call $-\zeta$, that is very close 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 $1/S$: $\Pr(\text{Size} > S) = a/S^\zeta$, with $\zeta \simeq 1$.

Many explanations of Zipf's law have been proposed, all of which present considerable difficulties. They can be divided into two strands: one economic and one that relies on random processes. The economic explanations rely on a very delicate balancing of transport costs, positive and negative externalities, and productivity differences. The basic problem with this strand is that it is hard to see how radically different economic structures (say, the United States in 1991 and India in 1911) would produce the same delicate balance of forces. Looking at the economy this way does not give any reason why Zipf's law should hold. The second line of research relies on random processes borrowed from the physical sciences, probably best exemplified by Herbert Simon's (1955) model. The basic problem with Simon's model is that, in it, the exponent in Zipf's law equates the ratio of the growth rate of the number of cities to the growth rate of the population of existing cities. Though Simon presents theoretical reasons why this ratio might be 1 (thus explaining Zipf's law), this ratio is in reality significantly less than 1 (Duncan Black and Vernon Henderson, 1999a), and probably tends to 0 in the long run (say, in Europe). This means that Simon's mechanism does not explain Zipf's law.

I. Explaining Zipf's Law by Gibrat's Law

In Gabaix (1999), I propose the following explanation: Zipf's law derives from Gibrat's

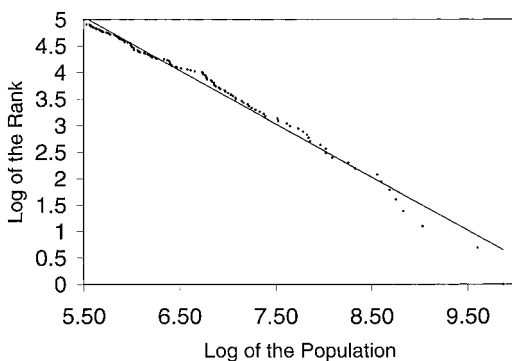


FIGURE 1. A PLOT OF LOG RANK (BY POPULATION) AGAINST LOG POPULATION FOR U.S. CITIES, 1991

law, where Gibrat's law means that the growth process is independent of size. In the basic model, start with an initial, arbitrary distribution of N cities. Let each city grow at some common mean rate, but with some idiosyncratic shocks to the growth rate. Allow the cities to evolve freely, and study the limit distribution of their normalized size, S^i ($i = 1 \dots N$). That is, S^i is the population of city i divided by the total urban population. Thus, $\sum_{i=1}^N S^i = 1$ at each date t .

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 independently and identically distributed city growth factors. Since $\sum_{i=1}^N S_t^i = 1$, the average normalized size must stay constant, so $E[\gamma] = 1$. Call $G_t(S) \equiv \Pr(S_t > S)$ the tail distribution of city sizes at time t . The equation of motion for G_t is $G_{t+1}(S) = E[G_t(S/\gamma)]$. Suppose (and one can give simple conditions for this to hold) that there is a steady-state process $G_t = G$. It verifies $G(S) = E[G(S/\gamma)]$. But because $E[\gamma] = 1$, a distribution like Zipf's law of the type $G(S) = a/S$, satisfies this steady-state equation. In fact, one can prove that this Zipf's law is the necessary 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 must be scale-invariant. This forces it to follow a power law. To see why the exponent of the power law is 1, visualizing a concrete 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 of the constraint that the average size (understood as share of the total population) must be constant, the probability of doubling has to be $1/3$, and the probability of halving $2/3$ (the expected growth is $1/3 \times 2 + 2/3 \times 1/2 - 1 = 0$). To see how the number of cities of a given size can be constant, take S -sized cities. Given the doubling and halving probabilities, the number of $2S$ -sized cities should be half the number of S -sized cities, and the number of $S/2$ -sized cities should be double. This is Zipf's law.

Two more remarks on this may be useful. First, while more work is needed, preliminary evidence presented in Gabaix (1999) supports Gibrat's law. Second, empirically, there are "too few" small cities (say, fewer than 100,000 inhabitants) compared to Zipf's law (i.e., the slope of the graph in Figure 1 becomes shallower for small cities). It turns out that the same theory provides a simple explanation of this: this phenomenon is what should be expected if smaller cities have a larger variance than big cities, as is plausible if one thinks that their industrial base is less diversified than that of big cities.

II. Which Economic Models Are Compatible with Zipf's Law?

Can one reconcile constant-returns-to-scale (CRS)-like behavior of cities with another bit of intuition about cities by saying that they are all about externalities and increasing returns to scale? We will observe here that there are two possibilities: (i) the externalities have bounded effects, and cities are asymptotically CRS, with small and large cities having bounded productivity differences; or (ii) they are unbounded (say, big cities have very large negative externalities), and one must assume that technological differences are very big

with big cities having unbounded technological superiority over small cities. The two effects endogenously compensate each other to give rise to Zipf's law.

I consider an extension of the model of Gabaix (1999) which allows for externalities. This model differs from those of Vernon Henderson (1974), Jonathan Eaton and Zvi Eckstein (1997), and Black and Henderson (1999b) in several respects, notably for the accent on randomness. In the present model, there are overlapping-generation agents, with probability δ of death. Once they are born, agents migrate to the city of their choice. Once they have chosen the city, and thus paid the high cost of migration, they do not move 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. I now ask what city an agent will choose when she is born at time t . The wage in city i is w_{it} . There are N_{it}^o old agents and N_{it}^y new agents. Because the time intervals are brief, $N_{it}^y/N_{it}^o \ll 1$. Given that the increase in population of city i is then $\Delta N_{it} = N_{it}^y - \delta N_{it}^o$, the growth rate for the city is $\gamma_{it} \equiv N_{it}^y/N_{it}^o - \delta$. The production function is $A_i \pi(N_i^o) N_i^o f(N_i^y/N_i^o)$, where A_i is the total productivity in the city, and f is an increasing, concave function. Here $A_i N_i^o f(N_i^y/N_i^o)$ represents the CRS production function of the city, and $\pi(N_i^o)$ represents a technological spillover. Because economically new and old agents are rather good substitutes, I assume that both f and f' have values bounded and away from 0. The nominal wage of new agents is competitive, at $w_{it}^y = A_i \pi(N_i^o) f'(N_i^y/N_i^o)$.

Finally, the utility derived from a wage w_{it} is $a_{it} \lambda(N_i^o) w_{it}$. Here $a_{it} \lambda(N_i^o)$ is the total level of amenities; $\lambda(N_i^o)$ represents the benefits and costs of having lots of people in one's city (e.g., the possibility of varied social contact and sharing of public goods with fixed costs, as well as traffic jams and high rent prices); and a_{it} is the level of the amenities (quality of schooling, policing, taxation). Assuming that the a_{it} are bounded away from 0 and independently and identically distributed, the decision problem of choosing one's city is simply $\text{Max}_i a_{it} \lambda(N_i^o) w_{it}$. Because young agents arbitrage across cities, in equilibrium, all "utility-adjusted" wages have

the same value u_t . Thus $u_t = a_{it} \lambda(N_i^o) w_{it}^y$, or $u_t = a_{it} A_{it} \lambda(N_{it}^o) \pi(N_{it}^o) f'(N_{it}^y/N_{it}^o)$, so

$$\frac{u_t}{a_{it} f'(N_{it}^y/N_{it}^o)} = A_{it} \lambda(N_{it}^o) \pi(N_{it}^o).$$

This is the key equation. It says that, as new agents arbitrage across cities, lower productivity A_{it} in a city must be compensated by higher amenities, or favorable externalities $\lambda(N_{it}^o) \pi(N_{it}^o)$. Given that the left-hand side has bounded variations across cities (i.e., the ratio of its values for two different cities is bounded), the right-hand side must similarly have bounded variations. This shows two basic routes to Zipf's law. The first route is Gibrat's law growth and CRS, and the second is Gibrat's law growth with offsetting unbounded externalities and productivity differences.

Following the first route, variation of $\lambda(N_{it}^o) \pi(N_{it}^o)$ is bounded. This means that externalities do matter, but (plausibly assuming monotonicity for large N) they die off for big cities. Asymptotically, cities are CRS. The necessary correlate of this (because $A_{it} \lambda(N_{it}^o) \pi(N_{it}^o)$ has bounded variation) is that the productivities A_{it} have bounded variation across sizes, implying that productivities do not vary wildly across cities. Thus, as in Gabaix (1999), the process gives rise to Gibrat's law: the randomness of the amenity and productivity shocks a_{it} and A_{it} create shocks to N_{it}^y/N_{it}^o (hence to the growth rate of the population), which do not depend on the size of the city for large N_{it}^o because the externalities $\lambda \pi$ have reached their asymptotic values.

With the second route, one allows $\lambda(N_{it}^o) \pi(N_{it}^o)$ to have unbounded variations. For instance, say that it becomes very close to 0 for large N_{it}^o , since big cities have extremely (at the limit, infinitely) bad amenities or social externalities. In this case A_{it} must have unbounded variations too: to compensate for their infinitely bad amenities, very big cities must have extremely high productivities. For instance, if A_{it} evolves according to some Brownian multiplicative process, and $\lambda(N_{it}^o) \pi(N_{it}^o)$ behaves asymptotically like a power law of N_{it}^o , then N_{it}^o will also follow a Brownian

multiplicative process. Hence Gibrat's and Zipf's law will be verified.

III. Conclusion

Using a fairly general model of city growth (introducing specialization in goods and capital does not change the conclusions of the analysis), I have shown that essentially two very different routes to Zipf's law are possible. In the first route, cities are approximately CRS in the upper tail. They can behave differently from small cities (e.g., have traffic jams), but the externalities do not affect medium-size and big cities differently (the problems associated with traffic jams reach an upper bound). Their productivity is not unboundedly different than that of smaller cities. Big cities are big because they embody crystallized randomness. Across sizes, cities do have differences in amenities, desirability due to externalities, but these differences are not enormous; their impact is bounded. The reason why cities became large is essentially because of inertia in the creation of jobs: the number of new jobs is roughly proportional to the number of existing jobs. With the effect of externalities being bounded, wages adjust to compensate moderate differences in amenities.

In the second route, cities experience externalities that do not die off in the upper tail: they become infinitely important as cities grow. Furthermore, cities have very different levels of productivity. Indeed, those differences are unboundedly large. Big cities are big because they experienced unboundedly large positive productivity increases, and their size creates very negative social externalities and spillover (so as to be equally attractive as small cities for new workers) or very negative productivity shocks with unboundedly large and positive spillovers or externalities. The persistence of these high productivity differences explains why cities have very different sizes.

An important question for future empirical research is to determine which one of those two worlds we are in. If we are in the first world, the consequences for city growth modeling would be stark; they would suggest that, asymptotically, cities behave like CRS economies. Seeing how this is compatible with the

new economic geography (Masahisa Fujita et al., 1999) would also be highly interesting. In any case, satisfying Zipf's law is a minimal criterion of admissibility for any model of city growth. This law is likely to guide empirical and theoretical work in the future.

REFERENCES

- Barro, Robert.** "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics*, February 1991, 106(2), pp. 407–43.
- Barro, Robert and Sala-i-Martin, Xavier.** "Convergence." *Journal of Political Economy*, April 1992, 100(2), pp. 223–51.
- Black, Duncan and Henderson, Vernon.** "Urban Evolution in the USA." Mimeo, Brown University, 1999a.
- . "A Model of Urban Growth." *Journal of Political Economy*, April 1999b, 107(2), pp. 252–84.
- Eaton, Jonathan and Eckstein, Zvi.** "City and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics*, August 1997, 27(4–5), pp. 443–74.
- Fujita, Masahisa; Krugman, Paul and Venables, Anthony.** *The spatial economy*. Cambridge, MA: MIT Press, 1999.
- Gabaix, Xavier.** "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 1999 (forthcoming).
- Glaeser, Edward; Kallal, Hedi; Scheinkman, José and Shleifer, Andrei.** "Growth in Cities." *Journal of Political Economy*, December 1992, 100(6), pp. 1126–52.
- Henderson, Vernon.** "The Sizes and Types of Cities." *American Economic Review*, September 1974, 64(4), pp. 640–56.
- Rosen, Kenneth and Resnick, Mitchell.** "The Size Distribution of Cities: An examination of the Pareto Law and Primacy." *Journal of Urban Economics*, September 1980, 8(2), pp. 165–86.
- Simon, Herbert.** "On a Class of Skew Distribution Functions." *Biometrika*, June 1955, 42(3), pp. 425–40.
- Zipf, George.** *Human behavior and the principle of last effort*. Cambridge, MA: Addison-Wesley, 1949.