Gibrat's Law for (All) Cities

By Jan Eeckhout*

Two empirical regularities concerning the size distribution of cities have repeatedly been established: Zipf's law holds (the upper tail is Pareto), and city growth is proportionate. Census 2000 data are used covering the entire size distribution, not just the upper tail. The nontruncated distribution is shown to be lognormal, rather than Pareto. This provides a simple justification for the coexistence of proportionate growth and the resulting lognormal distribution. An equilibrium theory of local externalities that can explain the empirical size distribution of cities is proposed. The driving force is a random productivity process of local economies and the perfect mobility of workers. (*JEL* D30, D51, J61, R12)

The law of proportionate effect will therefore imply that the logarithms *of the variable will be distributed following the [normal distribution].*

—Robert Gibrat (1931)

The way the population is distributed across geographic areas, while continuously changing, is not random. In fact, there is a strong tendency toward agglomeration, i.e., the concentration of the population within common restricted areas like cities. And while physical geography—rivers, coasts, and mountains—may have played a crucial role in early settlements, in the current day and age, the evolution of the population across geographic locations is an extremely complex amalgam of incentives and actions taken by millions of individuals, businesses, and organizations. Most people will agree that economic factors are the principal determinant of the dynamics of city populations. In the last decade, Detroit, for example, experienced a decline in population as the manufacturing industry in the area suffered a severe downturn. At the other extreme, when the high-technology industry was booming, villages, towns, and cities in the San Francisco Bay area experienced higher-than-average population growth. Increased productivity due to technological progress in the e-business sector led to the creation of such new companies as Yahoo! and the expansion of such existing companies as HP and Apple. This in turn increased labor demand and wages, which induced many individuals to relocate to the Bay area. No doubt an exodus from the Bay area has been at work since the technology market crashed at the beginning of the current decade. This confirms that agglomeration and residential mobility of the population between different geographic locations are tightly connected to economic activity.

Given this direct connection between economic activity and population mobility, it has long been recognized that fully understanding geographic economic activity involves understanding population mobility *and* economic driving forces.[1] A crucial *first* step is to provide an accurate description of agglomeration and population mobility. This involves accounting for the way the population is distributed over different geographic locations and accounting for the evolution over time. Once population mobility is understood, the *second* step involves analyzing the underlying economic mechanisms. Because economic factors are of paramount importance in providing incentives for individuals and businesses to move to different locations, being able to model the economic

* Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104 (e-mail: eeckhout@ssc.upenn.edu; http://www.ssc.upenn.edu/~eeckhout/). I am grateful for discussion and comments from Pol Antràs, Jesús Fernández-Villaverde, Xavier Gabaix, Gregory Kordas, Robert Lucas, Samuel Pessoa, Esteban Rossi-Hansberg, Stephen Yeaple, the editor, and two referees. Jesús Fernández-Villaverde provided the code for the nonparametric kernel estimator.

[1] See, among others, George K. Zipf (1949), J. Vernon Henderson (1974) and Paul Krugman (1996).

forces is of direct importance, especially since different cities are subject to different types of government policies, both within a city and between cities. The motives for intervention often depend on externalities (see Robert E. Lucas and Esteban Rossi-Hansberg, 2002, for a discussion). Through their interventions, policymakers affect economic factors, in particular equilibrium prices of land and labor and, therefore, decisions by individuals and businesses on where to locate. For example, city-specific income-tax incentives will affect after-tax wages and will make certain locations more attractive than others. This in turn will lead to a change in the number of people deciding to establish residence in certain locations. Other examples include transportation taxes and subsidies within and between cities (for example the subsidization of roads, railways, and airports),[2] regional subsidies, and agriculture subsidies that benefit companies in rural towns. An equilibrium theory of choice of geographic location (city, town, or village) driven by market wages and property prices is necessary for the optimal design and evaluation of such policies.

Unfortunately, the literature has faced substantial difficulties in the description of population mobility. The difficulty derives from a puzzle caused by two robust empirical regularities. The *first* empirical regularity is that the largest cities satisfy Zipf's law. Despite the apparent chaotic evolution of city populations, surprising regularities have been observed in the size distribution of cities. As early as 1682, Alexandre Le Maître observed a systematic pattern of the size distribution of cities in France. He describes how the size of Paris related to two groups of cities, each of them proportionally smaller than Paris. But it was not until 1913 that Félix Auerbach, and then George Kingsley Zipf in 1949, formally established the first empirical regularity. They show that within a country, the size of the largest cities is inversely proportional to their rank. For example, in the United States, New York City is roughly twice the size of Los Angeles, the second largest city, and about three times the size of Chicago, the third largest city. The proportionality of rank and size implies that the upper truncated distribution is the Pareto distribution[3] (or power distribution) with exponent equal to one. Zipf's finding has been shown to be robust, both over time and across countries, though with varying Pareto exponents. The *second* empirical regularity is that the growth rate of city populations does not depend on the size of the city. Even though growth rates between different cities vary substantially, there is no systematic pattern with respect to size, i.e., the underlying stochastic process is the same for all cities. This is labeled the proportionate growth process. Empirical research[4] has repeatedly shown that city growth is proportionate: larger cities on average do not grow faster or slower than smaller cities.

While it is surprising that such regularities emerge from a highly intricate underlying mechanism, there is also a puzzle: the two regularities cannot easily be reconciled.[5] In particular, the proportionate growth process (the second regularity) gives rise to the lognormal distribution, not the Pareto distribution (i.e., Zipf's law, the first regularity). This is a well-known proposition established by Gibrat (1931) and originally formulated by the astronomer Jacobus C. Kapteyn (1903): a stochastic growth process that is proportionate gives rise to an asymptotically lognormal distribution.[6] This is not to say that a proportionate growth process plus "something else" cannot give rise to the Pareto distribution or another distribution. There is a long tradition in the economics of income inequality starting with David G. Champernowne (1953) and industrial organization

[2] Transportation expenditure in 2002 was \$62 billion (3.1 percent of total government outlays; 9 percent of outlays excluding transfers) (www.whitehouse.gov/omb/budget).

[3] Vilfredo Pareto (1896) is credited with the discovery that the distribution of individual income satisfies a power law, the Pareto distribution.

[4] See, among others, Edward Glaeser et al. (1996), Jonathan Eaton and Zvi Eckstein (1997), and Yannis M. Ioannides and Henry G. Overman (2003).

[5] Krugman (1995) writes: "We have to say that the rank-size rule is a major embarrassment for economic theory: one of the strongest statistical relationships we know, lacking any clear basis in theory."

[6] Kapteyn (1903) studies skew distributions, mainly in biology, and establishes that they are driven by a simple Gaussian process. If a variable y is generated by additive random shocks which give rise to an asymptotically normal distribution, then given a transformation $y = f(x)$, the variable x has a skew distribution, derived from the transformed stochastic process. One such transformation is $y = \ln x$.

(see John Sutton, 1997, for an overview and Boyan Jovanovic, 1982) studying the relation between proportionate growth and size distributions different from the lognormal. With respect to the size distribution of cities, Xavier Gabaix (1999) and Aharon Blank and Sorin Solomon (2000) propose a resolution of the puzzle and show that proportionate growth processes can generate Zipf's law at the upper tail.[7]

The purpose of this paper is twofold. First, a new resolution of the puzzle is uncovered regarding the two empirical regularities, thus providing an accurate description of population mobility. While an accurate description of population mobility per se may not be of primary interest, it does have fundamental implications for the underlying economic mechanism, which in turn drives the population mobility. The second purpose is to propose and solve an equilibrium theory of local externalities. The equilibrium theory provides an analysis of the underlying economic mechanisms that is consistent with the empirically observed population mobility. This approach of providing an empirically consistent theory is in line with the central thesis in this paper: population mobility is driven by economic forces. Such an empirically consistent equilibrium theory is novel because heretofore the literature[8] has focused on solving the puzzle concerning population mobility. The main interest of this empirically consistent equilibrium theory is that it facilitates the evaluation of government policies that affect citizens' mobility decisions. Is it efficient to provide federal subsidies to small cities to attract residents? What is the effect of government-financed local transportation in large cities?

The breakthrough in the current resolution of the puzzle (the first purpose of this paper) derives from the availability of Census 2000 data. The new dataset is substantially larger than those of earlier censuses. The current data include observations on the entire size distribution of geographic locations, referred to in the Census as "places." For the year 2000, there are observations on 25,359 places, including cities, towns, and villages,[9] ranging in population from 1 to over 8 million. Previously in the literature, only the truncated distribution, i.e., the upper tail of the distribution of the 135 largest cities, or metropolitan areas (MAs), was considered, i.e., 0.5 percent of the current sample and 30.2 percent of the sample population. Using the new data, it is shown that the size distribution of the entire sample is lognormal and not Pareto. Moreover, for those observations for which 1990 data also exist, the growth rate of cities is calculated, and the second regularity, that growth is independent of city size, is confirmed. As a result, the growth process is shown to be proportionate. The proportionate growth process, together with the lognormality of the size distribution, establishes that when considering *all* cities and not just the upper tail of the distribution, Gibrat's prediction concerning the stochastic process holds.

The second purpose of this paper is to analyze an equilibrium model consistent with the empirically observed population mobility. A theory of local externalities is proposed. Like those in Lucas and Rossi-Hansberg (2002), the cities in this model are characterized by local externalities—both positive production externalities (spillovers from nearby factors of production) and negative consumption externalities (lost leisure time from traffic congestion). Those externalities are local, which means they affect the population within a city only, and typically they depend on the size of the city's population. In large cities, for example, firms and workers benefit more from the availability of deep markets for employees and jobs, and those cities also have larger "knowledge spill-

[7] They consider random growth processes with "something else"—the entry of new cities—and apply a process developed in Champernowne (1953) and Harry Kesten (1973). While these processes do generate Pareto distributions, Blank and Solomon (2000) point out that the details specifying and enforcing the smallest size of the cities are crucial, as are the rules for creating new entering cities. Whether or not the resulting limiting distribution is Pareto with exponent equal to one is very sensitive to this entry process. Moreover, testing whether the entry process satisfies the exact and detailed requirements for the Pareto distribution is a challenging empirical endeavor (for the metropolitan areas in the United States, for example, there has been no entry or exit in the set of MAs between 1990 and 2000; the 276 MAs in 1990 are identical to those in 2000). To date, no such evidence has been provided.

[8] Some notable exceptions are discussed below.

[9] In what follows, the term "city" will be used to indicate a place. When a city (as opposed to a town or a village) is referred to, this will be made explicit.

overs." Information concerning new technologies and products spills over faster in markets with high degrees of local interaction, like those of large cities. Simultaneously, workers in larger cities also impose negative externalities on each other because commuting times are longer. The economy differs from the one in Lucas and Rossi-Hansberg (2002) because of the explicit mobility *between* cities, rather than within cities. The aim is to capture the notion of competition between geographic locations, i.e., perfectly mobile citizens making location decisions between different cities. Local externalities within cities regulate the mobility of citizens between different cities (i.e., there are no externalities between cities). It is shown that the local externality model economy predicts behavior that is consistent with the empirical city growth process.

The only remaining issue to resolve is how it is possible that Zipf's law is repeatedly confirmed in the literature, while the underlying distribution is lognormal. The Pareto distribution is very different from the lognormal, so it is obvious that if the true distribution is lognormal, the entire distribution can never be fit to a Pareto distribution at the same time. Consider Figure 1 with a plot of the density function of the lognormal and that of the Pareto distribution (both on a ln scale); observe that the lognormal on a log scale is the normal density function. The density of the Pareto distribution is downward sloping, whereas the lognormal density is initially increasing and then decreasing (given symmetry, half the observations are in the increasing part). If the underlying distribution is lognormal, then goodness of fit tests will categorically reject the Pareto distribution. Still, when regressing log rank on log size for the entire distribution,[10] the coefficient comes out significant. Estimating a linear coefficient when the underlying empirical distribution is not Pareto (i.e., the relation is nonlinear) can obviously produce a significant estimate. This regression test merely confirms that there is a relation between size and rank, but it does not provide a test for the linearity of this relation. As such, testing the significance of the linear coefficient

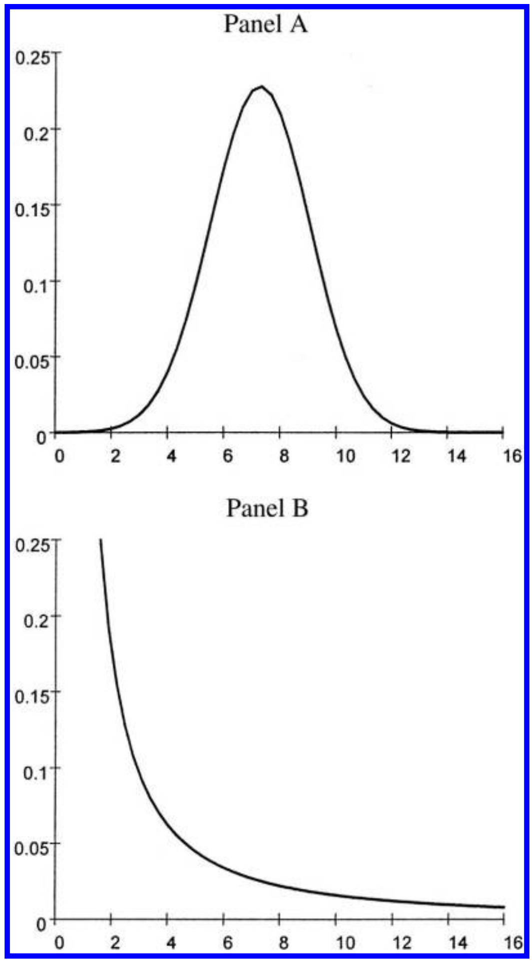

FIGURE 1. DENSITY OF LOGNORMAL (PANEL A) AND PARETO (PANEL B) DISTRIBUTION

is not the equivalent of a goodness-of-fit test for the Pareto distribution.[11]

More important though is that until now the literature considered the truncated distribution (typically, the truncation point is at ln size equal to 12 on the horizontal axis, i.e., for only 135 cities). At the very upper tail of the distribution, there is no dramatic difference between the density function of the lognormal and the Pareto. Now both the truncated lognormal and the Pareto density are downward sloping and similar (the Pareto is slightly more convex). As a result, both the Pareto and the truncated lognormal trace the data relatively closely. The problem is

[10] This is the standard procedure in the literature to verify for Zipf's law.

[11] See also Gabaix and Ioannides (2003) on the shortcomings of OLS.

that the estimated coefficient on the Pareto distribution is extremely sensitive to the choice of the truncation point: as the truncation point increases on the horizontal axis, the estimated Pareto coefficient increases, while the estimated lognormal coefficients remain unchanged. Moreover, for lower truncation points, the Pareto fits the data less and less well. In this paper, we show that these observed empirical changes in the estimated Pareto coefficient are theoretically consistent with the comparative static of a changing truncation point of the lognormal distribution.[12]

Finally, there is a growing literature proposing equilibrium models of economic activity with mobility of citizens that can account for Zipf's law.[13] Rossi-Hansberg and Mark Wright (2004) propose a dynamic general equilibrium theory with population mobility and balanced growth driven by industry-specific shocks. While their theory can explain Zipf's law for the size distribution, the model can also explain deviations of the empirical size distribution from Zipf's law.[14] This attempt to account for empirically observed differences from Zipf's law using growth theory is novel. The results they find concerning the truncated size distribution are consistent with those found in the current paper, confirming the importance of deviations from Zipf's law.

This paper is organized as follows. In Section I, the Census 2000 data are described in detail. The size distribution is shown to be lognormal, and the growth process proportionate. In Section II, the implications, both empirical and theoretical, for estimation of Zipf's law are analyzed when the true underlying distribution is lognormal. In Section III, a theory of local externalities is proposed, consistent with Gibrat's proposition that proportionate growth leads to a lognormal distribution. Finally, some concluding remarks are made in which the parallel is drawn between our results and findings in the exact sciences.

[12] The sensitivity of the Pareto coefficient to the truncation point has been observed in the literature (for an overview, see Gabaix and Ioannides, 2003). Explanations offered for the sensitivity differ, however, from the explanation proposed here, i.e., that the underlying true distribution is lognormal.

[13] See, among others, Rossi-Hansberg and Wright (2004) and Gilles Duranton (2002).

[14] Unfortunately, when predicting a size distribution that is different from the Pareto distribution, their model no longer satisfies proportionate growth.

I. The Empirical Size Distribution of Cities

A. *The Data*

Newly available data from Census 2000 are used.[15] The dataset for deriving the distribution of cities is novel. The units of account are denoted by the Census Bureau as "places." Places are either legally incorporated under the laws of their respective state or are Census Designated Places (CDP). Incorporated places have political/statistical descriptions of city, town (except in New England, New York, and Wisconsin), borough (except in Alaska and New York), or village. People living in locations that are not incorporated are legally resident in the respective counties. Incorporated places can cross county boundaries. Because a considerable fraction of the population lives in places that are not incorporated,[16] the Census Bureau designates such places CDPs. According to the Census Bureau, a CDP is a "statistical entity that serves as a statistical counterpart of an incorporated place for the purpose of presenting census data for a concentration of population, housing, and commercial structures that is identifiable by name, but is not within an incorporated place."[17] In the new census data, the CDPs are included for the first time without any restrictions.[18] The data on places for all U.S. states (including Hawaii and Alaska, and the commonwealth of Puerto Rico) will be used. In what follows, place (whether a city, town, or village) and city will be used interchangeably.

The main advantage of using these census data is that they cover the entire population size distribution. Moreover, with the inclusion of the CDPs in 2000, this new source of data represents the entire geographic concentration of the U.S. population. In the year 2000, 208 million of 281 million individuals (74 percent) were living in a total of 25,359 places. Table 1 reports on the population size in 2000 of the 10 largest "places."

[15] Source: www.census.gov/main/www/cen2000.html.

[16] Very often, whether a place is incorporated depends on state law. For example, under state law in Hawaii, there exist no incorporated places.

[17] Source: www.census.gov.

[18] For Census 2000, CDPs did not have to meet a population threshold to qualify for the tabulation of census data.

TABLE 1—TEN LARGEST CITIES IN THE UNITED STATES

Rank	City	Population S	S_{NY}/S
1	New York, NY	8,008,278	1.000
2	Los Angeles, CA	3,694,820	2.167
3	Chicago, IL	2,896,016	2.753
4	Houston, TX	1,953,631	4.099
5	Philadelphia, PA	1,517,550	5.277
6	Phoenix, AZ	1,321,045	6.062
7	San Diego, CA	1,223,400	6.546
8	Dallas, TX	1,188,580	6.738
9	San Antonio, TX	1,144,646	6.996
10	Detroit, MI	951,270	8.419

Note: S_{NY}/S denotes the ratio of population size relative to New York.
Source: Census Bureau, 2000.

A substantial portion of research into the size distribution of the U.S. population has been done using the MA[19] as the unit of measurement (see, for example, Krugman, 1996; Gabaix, 1999; Ioannides and Henry G. Overman, 2003). An MA typically covers one (or several) large cities. The largest metropolitan area is New York-Northern New Jersey-Long Island, including the cities of New Haven, Connecticut, Newark and Trenton, New Jersey, and several smaller towns in eastern Pennsylvania. The ten largest MAs and their population size are listed in Table 2.

The total number of MAs in the United States is 276, the smallest of which is Enid, Oklahoma, with a population of 57,813. In 2000, 80 percent of the entire U.S. population lived in MAs. At first sight, it may seem surprising that 80 percent lived in the 276 MAs, while only 73 percent lived in 25,359 places. The reason is that MAs cover huge geographic areas. For example, Trenton, New Jersey, is 64 miles from New York City and 144 miles from New Haven, Connecticut. As a result, MAs include a large population living in rural areas which are not counted as places. Consider, for example, Mercer County, New Jersey, in the MA of New York-Northern New Jersey-Long Island, which includes Princeton and Trenton. In 2000, Mercer County had a population of 350,761, of which only about 31 percent lived in incorporated places.

[19] According to the Census Bureau definition, an MA "must include at least one city with 50,000 or more inhabitants, or a Census Bureau–defined urbanized area (of at least 50,000 inhabitants) and a total metropolitan population of at least 100,000 (75,000 in New England)."

B. *The Size Distribution*

Over the entire size distribution, the median city has a population of 1,338. Figure 2 plots the empirical density function on a natural logarithmic (ln) scale, together with the theoretical lognormal density for the empirically observed mean and variance. Figure 3 plots the cumulative density function. The sample mean (in ln, standard error in brackets) is $\hat{\mu} = 7.28$ (0.01) and the standard deviation is $\hat{\sigma} = 1.75$. The theoretical density function of the lognormal size distribution is normal in ln S and given by $\phi(\hat{\mu}, \hat{\sigma})$:

$$\phi(\hat{\mu}, \hat{\sigma}) = \frac{1}{\hat{\sigma}\sqrt{2\pi}} e^{-(\ln S - \hat{\mu})^2/2\hat{\sigma}^2}. \tag{1}$$

A Kolmogorov-Smirnov (KS) test of goodness of fit of the empirical density function against the lognormal with sample mean $\hat{\mu} = 7.28$ and sample standard deviation $\hat{\sigma} = 1.75$ generates the KS test statistic $D = 0.0189$, and the corresponding p-value obtained is 1 percent. This is supporting evidence in favor of lognormality of the size distribution. Though the fit is remarkable, it is not perfect. There seems to be some skewness (third moment is 0.21) and the median value is 7.20 (with mean of 7.28). On the other hand, there is hardly any kurtosis (the fourth moment is 0.03). Possibly there is some censoring (most likely at the bottom of the distribution). The data collected may be contaminated by differences between state legislation with respect to legal incorporation, in particular for small places. In addition, since the data contain CDPs, the decision procedure by the Census Bureau to designate a nonincorporated place may depend on the size of the place and, as a result, it will affect the size distribution of places, in particular at the bottom end. Furthermore, given the extremely large sample size of $n = 25,359$, small deviations from the theoretical distribution are exaggerated in goodness of fit tests. It is surprising that, despite some potential shortcomings of the data, the empirical size distribution fits the lognormal distribution that well.

Before analyzing the properties of the city growth process, a fundamental issue remains: what is the appropriate economic unit that should be studied? As Tables 1 and 2 highlight,

TABLE 2—TEN LARGEST METROPOLITAN AREAS IN THE UNITED STATES

Rank	MA	Population S	S_{NY}/S
1	New York-Northern New Jersey-Long Island, NY-NJ-CT-PA	21,199,865	1.000
2	Los Angeles-Riverside-Orange County, CA	16,373,645	1.295
3	Chicago-Gary-Kenosha, IL-IN-WI	9,157,540	2.315
4	Washington-Baltimore, DC-MD-VA-WV	7,608,070	2.787
5	San Francisco-Oakland-San Jose, CA	7,039,362	3.012
6	Philadelphia-Wilmington-Atlantic City, PA-NJ-DE-MD	6,188,463	3.426
7	Boston-Worcester-Lawrence, MA-NH-ME-CT	5,819,100	3.643
8	Detroit-Ann Arbor-Flint, MI	5,456,428	3.885
9	Dallas-Fort Worth, TX	5,221,801	4.060
10	Houston-Galveston-Brazoria, TX	4,669,571	4.540

Note: S_{NY}/S denotes the ratio of population size relative to New York.
Source: Census Bureau, 2000.

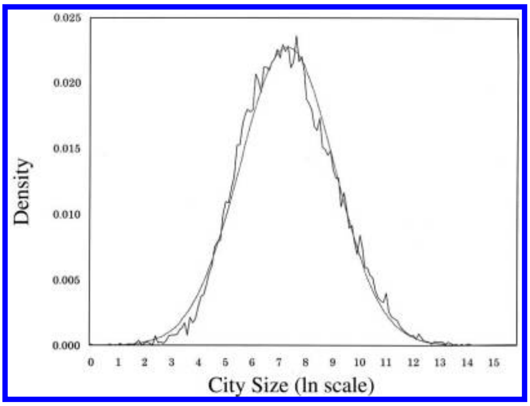

FIGURE 2. EMPIRICAL AND THEORETICAL DENSITY FUNCTIONS

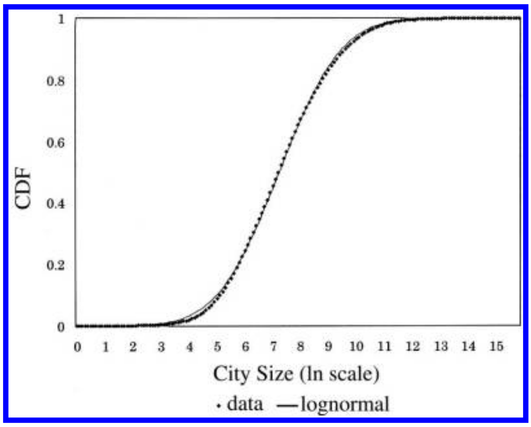

FIGURE 3. EMPIRICAL AND THEORETICAL CUMULATIVE DENSITY FUNCTIONS

cities and MAs represent different notions about the corresponding theory of an economic unit. And depending on the definition, we are studying different objects and therefore different distributions. As is the case with comparisons of countries, we do not have a perfect justification for using a particular unit of account when comparing cities. In our theory below, we consider local externalities that do not affect agents outside the economic unit as the defining characteristic of a city. In reality of course, no externality is purely local. One may therefore want to interpret this assumption as a matter of the extent to which externalities do or do not affect agents outside a given city. The danger is that the partition into economic units is either too fine or, at the other extreme, too coarse. The externalities for some agents in one part of a given economic unit (say those living in New Haven) may not have an impact on those living in different parts of the same unit (say Princeton). Moreover, different research objectives may call for the use of different units of account. For example, if one is interested in analyzing the economic impact of airports, the MA seems a natural unit of account, while cities may be more appropriate when studying schools, public transportation, or waste collection. In past research, both MAs and cities have proven to be useful and relevant economic units, and both have been studied extensively.

In this paper, cities are chosen for several reasons. In addition to the fact that cities are a natural economic unit for studying the local externalities that are modeled in Section III, there is a practical reason: the availability of data. We want to use data that cover the entire range of the populations, in particular the smaller ones. Because MAs are defined by the Census Bureau only for large populations (MAs must include "at least one city with 50,000 or more inhabitants"), the MA dataset does not cover the entire size distribution. And even if the dataset spans the entire domain of the size distribution of all cities, not all inhabitants live in cities, towns, or villages.[20] Unfortunately, these restrictions do not allow for the possibility of augmenting the dataset to include populations that are currently not covered.[21] It should be noted that the current dataset of all cities has already been augmented to form the largest possible dataset that is feasible, with the inclusion of the census-defined CDPs. This increases the number of cities by 31 percent, from 19,361 to 25,359.

The fact that part of the population is not covered is potentially a cause for concern, because rather than capturing deep patterns of populations and population dynamics, we may merely be describing the idiosyncrasy of the jurisdictional formation in the United States. The population that is not covered may be distributed in a completely different way from the lognormal distribution. And since we cannot assign that population to any geographic area comparable to a city, there is no hope of knowing how the remainder is distributed. The lognormality seems to be a strong regularity, however, from whichever perspective population dynamics is considered. First, while we have no way of showing that the distribution of MAs is lognormal given the truncation by definition, we show below that even for MAs, changes in the truncation point produce changes in the estimated Zipf coefficient that are consistent with the fact that the underlying upper tail is derived from the lognormal. Second, the size distribution of CPDs is pretty close to the entire distribution of cities and hence the lognormal. And finally, in the Appendix we show the results of further analysis using additional data that are available from the Census. We plot the size distribution of counties, which covers the entire U.S. population (see Figure A-1 and Table A-1 in Appendix A for the ten largest counties). While it is hardly convincing to make a case for counties as the relevant economic unit, it is surprising that even the size distribution of counties is close to the lognormal. Looking at population dynamics from the perspective of different economic units and including as large a fraction as possible of the U.S. population, there is a strong pattern that is consistent with lognormality.

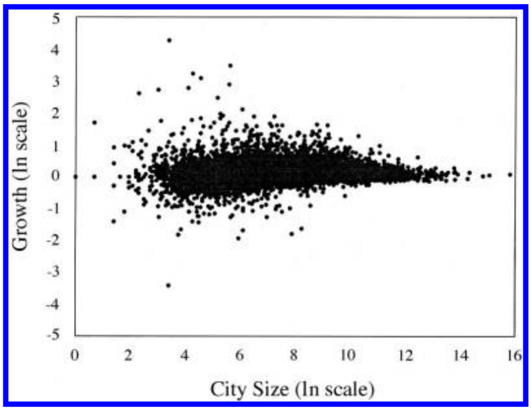

FIGURE 4. SCATTER PLOT OF CITY GROWTH AGAINST CITY SIZE

C. *Proportionate City Growth*

For the cities in the upper tail of the size distribution, population growth has repeatedly been shown to satisfy constant proportionate growth.[22] These findings can be extended beyond those for the upper tail of the distribution. We therefore use the data on population size for places in the United States from both the 1990 and 2000 Censuses. Unfortunately, 1990 Census data do not include the CDPs. As a result, the sample size is significantly smaller (19,361 instead of 25,359). Figure 4 shows the scatter plot of growth against city size (on ln scales). Mere observation of the scatter plot seems to

[20] All citizens belong to a county, which is the primary legal division and the functioning governmental unit.

[21] Those residual populations are included in the counties, and after accounting for the cities, residual populations very often are located in different geographic areas, separated by cities. To make things even worse, many cities extend over different counties, therefore guaranteeing that parts of the residual populations are counted twice.

[22] Glaeser et al. (1996) have shown this to be true for the largest cities in the United States. Eaton and Eckstein (1997) have confirmed this for the largest cities in France and Japan. In a detailed investigation, Ioannides and Overman (2003) nonparametrically estimate the mean and variance of growth rates conditional on size for the largest MAs in the United States. They accept the hypothesis that the city-size growth rate is constant across cities of different sizes, i.e., population growth is proportionate.

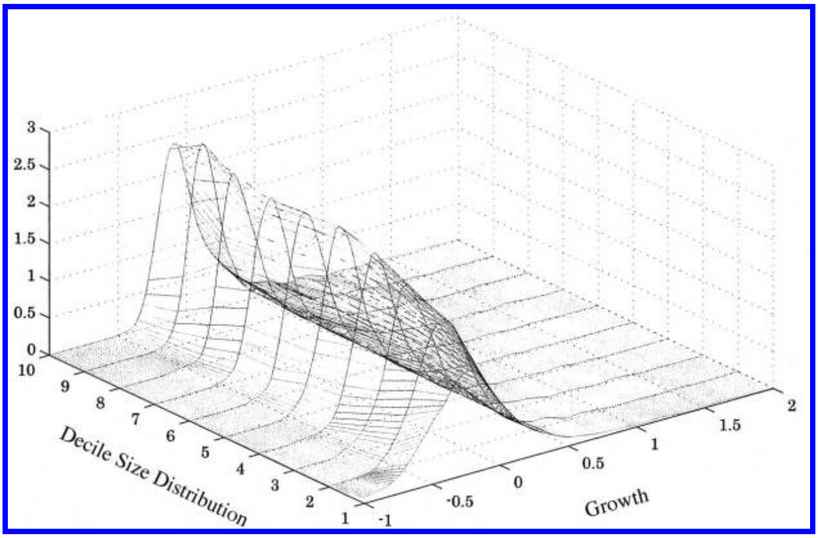

FIGURE 5. SURFACE PLOT: KERNEL DENSITY ESTIMATION OF NORMALIZED GROWTH RATES BY DECILE OF THE SIZE DISTRIBUTION

support that growth is independent of size. In what follows, the dependence relation of growth on size is analyzed in greater detail. We perform both nonparametric and parametric regressions of growth on size.

First, we perform a nonparametric regression of growth on size.[23] The standard parametric regressions as performed below provide us only with an aggregate relationship between growth and size, which is constrained to hold over the entire support of the distribution of city sizes. In contrast, the nonparametric estimate allows growth to vary with size over the distribution. The regression relationship we model is therefore

$$g_i = m(S_i) + \varepsilon_i$$

for all $i = 1, \ldots, 19361$. The objective is to provide an approximation of the unknown relationship between growth and size using smoothing, without making parametric assumptions about the functional form of m. Before estimating m, we report the distribution of growth rates for each decile of the size distribution. Following Ioannides and Overman (2003), we use the normalized growth rate (the difference between the growth rate and the sample mean divided by the standard deviation). In Figure 5, the stochastic kernel density[24] is plotted for each of the 10 deciles. Fixing a particular decile in the distribution, we can observe the distribution of growth rates within that decile. Figure 6 reports the contour plot of the same stochastic kernel, i.e., the vertical projection of the density function. Both figures illustrate that the distribution of growth rates is strikingly stable over different deciles. The best illustration of the size independence is the fact that the contour lines are parallel. The distribution is slightly skewed (the mode is just below zero), and the mode appears fairly constant over different deciles. The same is true for the variance. While the variance of the lowest decile seems to be somewhat higher (the contour lines fan out somewhat), there seems to be little change in the spread of the distribution for higher deciles.

We now proceed to estimate the regression relationship $g_i = m(S_i) + \varepsilon_i$, $i = 1, \ldots, 19361$, where g_i is the normalized growth rate, i.e., the

[23] This section on the nonparametric analysis follows closely the analysis in Ioannides and Overman (2003). We derive a sequence of results for our dataset of all cities similar to theirs, obtained for a time-series dataset on the largest MAs.

[24] Each stochastic kernel is calculated using the bandwidth derived with the automatic method corresponding to the Gaussian distribution (see Bernard W. Silverman, 1986).

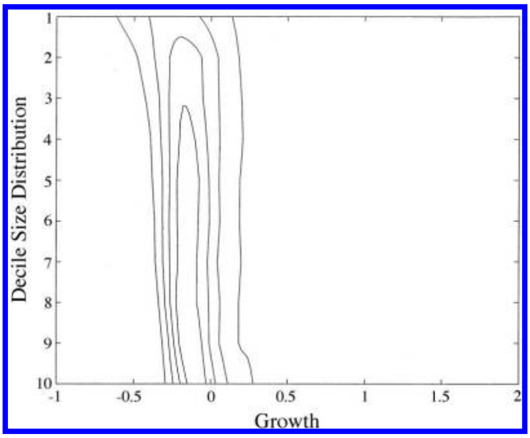

FIGURE 6. CONTOUR PLOT: KERNEL DENSITY ESTIMATION OF NORMALIZED GROWTH RATES BY DECILE OF THE SIZE DISTRIBUTION

difference between growth and the sample mean divided by the sample standard deviation, and S_i is the log of the population size of a city. We will approximate the true relationship by the regression curve $m(s)$ for all s in the support of S_i. The estimate of $m(s)$ will be denoted $\hat{m}(s)$ and is a local average around the point s. This local average smooths the value around s, and the smoothing is done using a kernel, i.e., a continuous weight function symmetric around s. The kernel K used in the remainder of the paper will be an Epanechnikov kernel.[25] The bandwidth h determines the scale of the smoothing, and K_h denotes the dependence of K on the bandwidth h. With the kernel weights, we calculate the estimate of m using the Nadaraya-Watson method,[26] where

$$\hat{m}(s) = \frac{n^{-1} \sum_{i=1}^{n} K_h(s - S_i) g_i}{n^{-1} \sum_{i=1}^{n} K_h(s - S_i)}.$$

In Figure 7 there is a plot of $\hat{m}(s)$ calculated for a bandwidth of $h = 0.5$ (see Silverman, 1986). The Figure also shows the bootstrapped

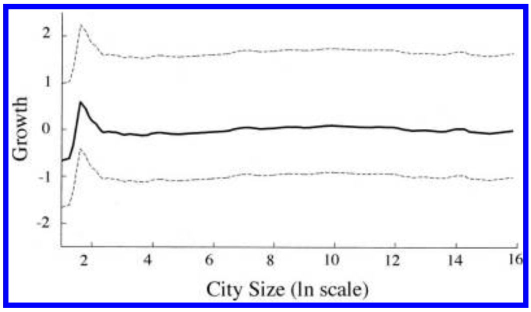

FIGURE 7. KERNEL ESTIMATE OF POPULATION GROWTH (BANDWIDTH 0.5)

95-percent confidence bands (calculated from 500 random samples with replacement). In line with the earlier results, the nonparametric estimate of the conditional mean is stable across different population sizes, except for the very bottom of the distribution.[27] The estimate seems to exhibit some slightly inverted U-shape, with somewhat higher growth rates in the middle range of population sizes and lower growth at the ends. If the underlying relation between growth and size is constant, then the estimate will lie in the 95-percent confidence bands. This seems to suggest that, except for some values near the lower boundary, we cannot reject that growth is independent of size. Observe that because the kernel is a fixed function and boundary observations have support only on one side of the kernel, the kernel estimates near the boundaries must be read with caution.

In Table B-1 in Appendix B, some further descriptive statistics are reported for growth rates over the entire support of the distribution. Consistent with the kernel estimates, average growth rates seem to be constant, except at the very bottom of the distribution. We also calculate the standard deviation and the Interquartile Range (IQR) of the growth rate. The IQR is defined as the difference between the seventy-fifth and twenty-fifth percentiles ($Q3 - Q1$). This provides an indication of the variation in growth rates. For the largest 100 cities, growth rates vary less, whereas the smallest 100 cities exhibit higher variation in growth rates. The

[25] Results below have been replicated using the Gaussian kernel and reveal no differences with those using the Epanechnikov kernel.

[26] See Wolfgang Härdle (1990).

[27] At the bottom of the distribution there is also more variation in growth rates (see IQR calculations below). Because the confidence bands impose a requirement over the entire domain of the size distribution, the width of the bands is likely to be affected by the variation at the bottom.

standard deviation of the growth rate of the largest 100 cities is an order of 4 to 5 times smaller compared to the entire sample (0.158 versus 0.729). Also for the IQR there is a decrease at the top of the distribution (0.154 versus 0.199), but to a lesser extent than in the case of the standard deviation. This seems to indicate that the tails of the distribution of growth rates of the top 100 cities are not as fat. For the smallest 100 cities, the variation in growth rates as measured by the IQR increases[28] 2.5 times relative to the IQR for the entire sample (0.493 versus 0.199). For the remainder of the support of the size distribution, the IQR of growth rates is more or less constant for all sizes, except for the bottom decile of the size distribution. Figure B-1 in Appendix B plots the IQR for each decile. Observe the sharp increase in the IQR at the bottom decile of the distribution (0.297 versus 0.199 for the entire sample).

The proportionate growth process that satisfies Gibrat's law and that gives rise to a lognormal distribution is also characterized by a size-independent variance. The kernel estimate of the variance $\hat{\sigma}^2(s)$ (see Härdle, 1990) is calculated as

$$\hat{\sigma}^2(s) = \frac{n^{-1} \sum_{i=1}^{n} K_h(s - S_i)(g_i - \hat{m}(s))^2}{n^{-1} \sum_{i=1}^{n} K_h(s - S_i)}.$$

As in Ioannides and Overman (2003) for MAs, we find that at the boundaries the variance of growth rates of cities is dependent on size.[29] In particular, for very small cities with population size around 10 inhabitants (with ln size between 2 and 3) and for very large cities, the variation in growth rates is markedly different, as reported in the IQR calculations above. Figure 7 plots the estimated variance[30] (bandwidth 0.5) for 95 percent of the cities in the sample, i.e., excluding the top and bottom 2.5 percent. This corresponds to all cities larger than 65 (ln is 4.1) and smaller than 56,000 (ln is 10.9). We find that some outliers have an enormous impact on the variance. For example, Eagle Mountain, Utah, the fastest growing city in the sample, has grown at a rate of 7,090 percent. These outliers alone cause spikes in the variance, which can be seen from observation of the dotted line, representing the kernel estimate of the variance for all observations (for example, around ln size equal to 7; observe also that given the bandwidth of 0.5, the effect of the outliers is constrained to a distance of 0.5). The solid line represents the kernel estimate of the variance for all observations excluding 9 outliers (observations have been dropped with growth rates above 1,000 percent). Without the outliers, the variance is remarkably stable across different sizes of cities.

Consider now the parametric growth regressions. For the entire size distribution, no significant effect of the size of a city is found on the growth, as confirmed by the following regression:

$$\frac{S_{00}}{S_{90}} = \underset{(0.005)}{1.102} - \underset{(7E(-08))}{3.75E(-08)} \frac{S_{90} + S_{00}}{2}$$

(n = 19361), where S_{00}/S_{90}, the ratio of the population size in 2000 and 1990, is the gross growth rate of the population, and $S_{90} + S_{00}/2$ is the average of the 1990 and 2000 populations. The coefficient on size is clearly insignificant (standard errors in parentheses). Note that the intercept—a net rate of 10.2 percent—is the country-wide growth for the entire sample population between 1990 and 2000 and corresponds to an annual population growth rate of $(1 + g_a)^{10} = 1.103$ or $g_a = 1$ percent. The lack of significance of city growth on size is further confirmed when the dependent variable is the population size in 1990:

$$\frac{S_{00}}{S_{90}} = \underset{(0.005)}{1.103} + \underset{(7.3E(-08))}{2.3E(-09)} S_{90}$$

(n = 19361). Finally, also when using logarithm of gross growth between 2000 and 1990

[28] Though this is not the case for the standard deviation.

[29] See also Härdle (1990) for a discussion on the reliability of the estimates at the tails of the distribution where the density takes on very small values.

[30] Observe in Figure 7 that the variance of the full sample is equal to one. This is due to the fact that we calculate the variance of normalized growth rates.

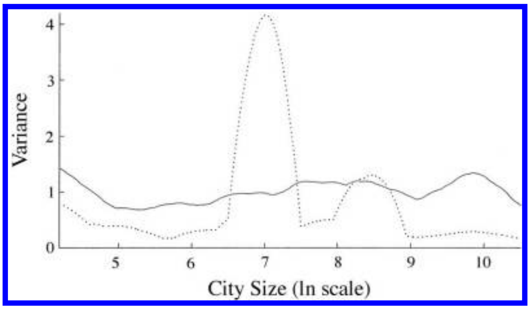

FIGURE 8. KERNEL ESTIMATE OF THE VARIANCE OF POPULATION GROWTH (BANDWIDTH 0.5)

as the dependent variable, the coefficient on size in 1990 remains insignificant. In the latter regression, the *p*-value is 7 percent. When regressing ln of the ratio of population sizes on ln average size, the coefficient comes out significant and positive: 0.0223 (0.001) and with a negative intercept −0.104 (0.007). As can be deduced from Figures 5, 6, and 7, this seems to indicate that the size dependence of growth rates at the very bottom of the distribution affects the nonparametric estimate.

In summary, all these results seem to provide support for the fact that city growth is independent of population size. Some caution is due, however. Growth rates could be calculated only for a sample of 19,361 cities, i.e., those cities for which there is an observation in 1990, and those observations exclude all CDPs. The lognormal distribution in Figure 2 was derived from the size distribution of 25,359 cities in 2000, i.e., a distribution with an additional 31 percent observations. This unfortunate limitation of the data does not permit us to make any definite statement about growth over the entire distribution, most likely until the Census 2010 data become available. If the size distribution of CDPs in the 2000 data can provide any indication (the distribution of CDPs is close to the distribution of all cities), one may expect the CDP distribution of growth rates not to differ too much from the rest of the cities.[31]

II. Zipf's Law

The question remains: what is the relation between Zipf's law for the truncated distribution and the nontruncated lognormal distribution? As argued in the introduction, the entire size distribution cannot possibly fit the Pareto distribution. In what follows, the aim is to establish Zipf's law for the truncated distribution. It will be shown that the estimated coefficient on the Pareto distribution is systematically sensitive to the choice of the truncation point. This will be confirmed to be consistent with the fact that the underlying distribution is lognormal.

In the literature on Zipf's law, the truncation point has repeatedly been chosen around 135 cities,[32] i.e., the 135 largest cities out of the total 25,359 cities are included in the census sample. This implies that 99.5 percent of the sample of cities is dropped, and only the upper 0.5 percentile of the size distribution is considered (this corresponds to 30.2 percent of the population in the sample, and 22.4 percent of the U.S. population).[33] It is well known from the literature that the upper tail of the distribution of cities fits the Pareto distribution extremely well. The objective of this section is to investigate how the estimated coefficient of the Pareto distribution changes as the truncation point changes. Consider therefore the following analysis of Zipf's law.

Zipf's law for cities states that the population size of cities fits a power law with exponent approximately equal to one: the population size of a city is inversely proportional to the rank of the size of the city. The law has been shown to hold for different definitions of cities, including both places and MAs. A city of rank r in the (descending) order of cities has a size S equal to $1/r$ times the size of the largest city in that country. For U.S. cities, the size S of Los Angeles, the second largest, should be ½ the size of New York. The tenth-ranked city, Detroit, should have a size $1/10$ of New York. Above in

[31] It is worth noting that even if growth is shown not to be exactly proportionate, the limit distribution can still be the lognormal. Michael Kalecki (1945) generalizes Gibrat's law for growth processes that are not exactly proportionate.

[32] The 135th largest city in the Census 2000 sample is Chattanooga, Tennessee, with a population of 155,554.

[33] It is of interest to provide some further descriptive statistics of the distribution of cities relative to the sample population of 208 million. Half the sample population lives in the largest 647 cities, three quarters in the largest 2,678 cities, 95 percent in the largest 10,255 cities, and 99 percent in the largest 17,425 cities.

Tables 1 and 2, $S/S_{NY} \approx r$ is reported for the ten largest cities and MAs respectively. The implication of Zipf's law is that when the population size is plotted against their rank on a logarithmic scale, an approximately straight line is obtained.

To see that a distribution that satisfies Zipf's law is the Pareto distribution, consider a variable S, distributed according to the Pareto distribution. Then the density function $p(S)$ and the cumulative density function $P(S)$ satisfy

$$p(S) = \frac{a\underline{S}^a}{S^{a+1}}, \qquad \forall\, S \geq \underline{S}$$

$$P(S) = 1 - \left(\frac{\underline{S}}{S}\right)^a, \qquad \forall\, S \geq \underline{S}$$

where a is a positive coefficient. Strictly speaking, Zipf's law satisfies Pareto with $a = 1$. Note that the rank in the empirically observed distribution is given by

$$r = \underline{N} \cdot (1 - P(S))$$
$$= \underline{N} \cdot \left(\frac{\underline{S}}{S}\right)^a$$

where $\underline{N}$ is the number of cities above the truncation point. Taking natural logs, we get that rank is inversely proportional to size

$$\ln r = K - a \ln S$$

where $K = \ln \underline{N} + a \ln \underline{S}$ is a constant.

Typically, Zipf's law is verified by regressing $\ln r$ on $\ln S$. For the upper truncated city size distribution, the regression gives a highly significant estimate of $\hat{a}$ equal to 1.354:

$$\ln r = \underset{(0.144)}{21.099} - \underset{(0.011)}{1.354} \ln S$$

($\underline{N} = 135$, $\underline{S} = 155,554$, $R^2 = 0.991$). In Figure 9, a scatter plot is presented of $\ln r$ against $\ln S$ and, in addition, the linear regression line estimated above is plotted. This plot can be interpreted as a transformation of the cumulative density function, where on the Y-axis we have the natural logarithm of the survival function $(1 - P(S))$ multiplied by $\underline{N}$.

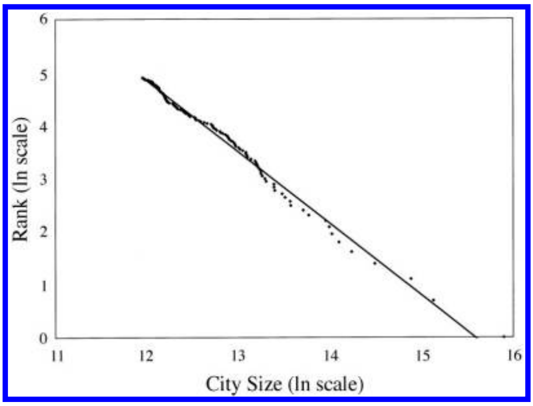

FIGURE 9. CITY SIZE DISTRIBUTION AND LINEAR REGRESSION LINE

Before considering the sensitivity of the estimated Pareto coefficient to the truncation in the size distribution of cities, consider the size distribution of MAs. Performing the same regression on the truncated distribution of MAs, where the MA at the truncation point is Erie, Pennsylvania, with a population of 280,843, we get

$$\ln r = \underset{(0.147)}{17.568} - \underset{(0.011)}{0.999} \ln S$$

($\underline{N} = 135$, $\underline{S} = 280,843$, $R^2 = 0.985$). Observe that for MAs, the estimated coefficient $\hat{a}$ is nearly exactly equal to 1, as originally described by Zipf (1949). Unfortunately, the fact that $\hat{a}$ is equal to 1 is highly sensitive to the choice of the truncation point in either direction: for $\underline{N} = 276$ (the entire MA sample and roughly double the original), $\hat{a} = 0.850$, and for $\underline{N} = 67$ (half the original sample), $\hat{a} = 1.114$). Figure 10 reports a scatter plot of the MA size distribution and the regression lines for the different sample sizes. At the truncation point of $\underline{N} = 135$, the sample ensures a perfect fit with Zipf's original observation.[34]

[34] A referee pointed out that the pervasiveness in the literature of the truncation point at $\underline{N} = 135$ and the resulting estimate of a power coefficient exactly equal to 1, as predicted by Zipf's law, is due to a remarkable historical coincidence. The literature, starting with Krugman (1996), used the *Statistical Abstract of the United States*, which shows only the data for the top 135 cities.

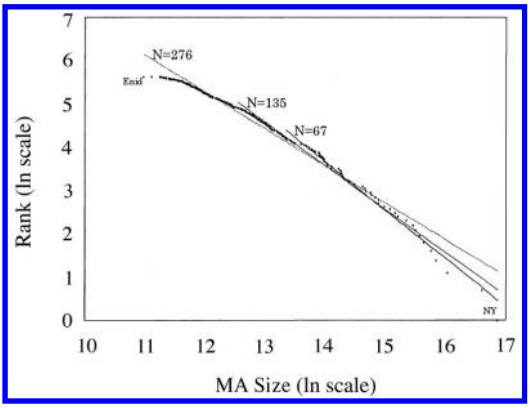

FIGURE 10. MA SIZE DISTRIBUTION AND LINEAR REGRESSION LINES FOR DIFFERENT TRUNCATION POINTS

The estimated coefficient on the Pareto distribution is clearly sensitive to the choice of the truncation point. Moreover, the dependence of the estimate is systematic: the lower the truncation point (i.e., the larger the sample size), the lower the estimated coefficient of the Pareto distribution.[35] The same is expected to be true for the size distribution of cities. In what follows, it is shown that a theoretical justification for the fact that the estimated Pareto coefficient is increasing for an increasing truncation point is given by the fact that the underlying sample is distributed lognormal.

Consider the lognormal density function $\phi(\cdot)$ as given in equation (1). To simplify notation, let $x = \ln S$, and denote the normal cumulative density function by $\Phi(x)$. Now consider the truncated lognormal distribution at truncation point $\underline{x} = \ln \underline{S}$. Then the cdf of the truncated lognormal is

$$\frac{\Phi(x) - \Phi(\underline{x})}{1 - \Phi(\underline{x})}.$$

As before, let $\underline{N}$ be the sample size of the truncated distribution. Then the rank can be written as

$$r = \underline{N} \cdot \left(1 - \frac{\Phi(x) - \Phi(\underline{x})}{1 - \Phi(\underline{x})}\right)$$

$$= \underline{N} \cdot \left(\frac{1 - \Phi(x)}{1 - \Phi(\underline{x})}\right)$$

and taking logs

$$\ln r = \ln\left[\underline{N} \cdot \left(\frac{1 - \Phi(x)}{1 - \Phi(\underline{x})}\right)\right]$$

or

$$\text{(2)} \qquad \ln r = \ln \frac{\underline{N}}{1 - \Phi(\underline{x})} + \ln(1 - \Phi(x)).$$

If the underlying true distribution is the lognormal, then from the last equation, the relation between $\ln r$ and $\ln S$ will not be linear. As a result, the hypothesis that size is everywhere inversely proportional to rank (Zipf's law) is not correct. In particular, $\ln(1 - \Phi(x))$ is not linear in $x = \ln S$. Calculating the derivative of the term that depends on x in equation (2) gives

$$\frac{d}{dx} \ln(1 - \Phi(x)) = -\frac{\phi(x)}{1 - \Phi(x)}$$

$$= -h(x)$$

which is the negative of the hazard rate. It is easily verified that the hazard rate for the corresponding lognormal distribution with sample mean and variance $\hat{\mu} = 7.28$, $\hat{\sigma} = 1.75$ is strictly increasing over the entire domain (and positive by definition). The plot of the hazard function $h(x; \hat{\mu}, \hat{\sigma})$ is given in Figure 11.

A strictly increasing hazard rate implies that the second derivative of the term $\ln(1 - \Phi(x))$ is strictly concave, i.e., $d^2/dx^2 \ln(1 - \Phi(x)) = -h'(x) < 0$. Now, given a decreasing, strictly concave function in x, the linear estimate of this function will systematically depend on the truncation point: the higher the truncation city size, the higher the estimate of the linear regression. Because an increase in the truncation size implies a decrease in the truncated sample population, the estimate will be decreasing as the sample population increases. This establishes the following proposition:

PROPOSITION 1: *If the underlying distribution is the lognormal distribution* $\Phi(x; \hat{\mu}, \hat{\sigma})$,

[35] While the analysis below for cities as opposed to MAs may be suggestive for the relation between the estimated Pareto coefficient and the truncation point, other explanations have been suggested in the literature. In particular, see the review on MAs by Gabaix and Ioannides (2003).

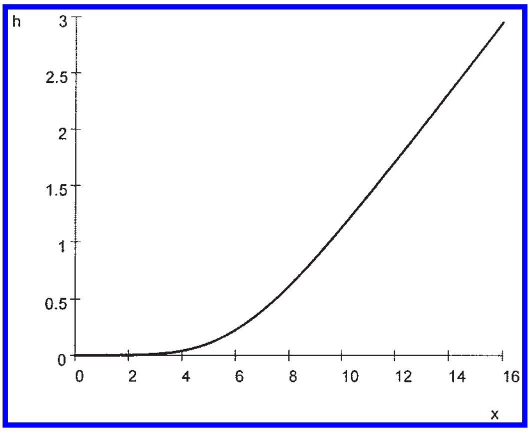

FIGURE 11. HAZARD RATE FOR THE LOGNORMAL $h(x;\ \hat{\mu},\ \hat{\sigma})$

then the estimate of the parameter $\hat{a}$ *of the Pareto distribution is increasing in the truncation city size* $(d\hat{a}/d\underline{S}) > 0$ *and decreasing in the truncated sample population* $(d\hat{a}/d\underline{N}) < 0$.

Given this theoretical prediction, Table 3 is consistent with the fact that the underlying empirical distribution function of city sizes (as established in the former section) is indeed lognormal. Estimated parameters are reported for the regression

$$\ln r = \hat{K} - \hat{a} \ln S, \qquad \underline{N}, \underline{S}.$$

Not only are the estimates of $\hat{a}$ highly sensitive to the choice of the truncation point, they are so in a systematic fashion, consistent with the fact that the underlying distribution is lognormal. For increasing $\underline{S}$ (decreasing $\underline{N}$), $\hat{a}$ is systematically increasing.[36]

Finally, Figure 12 provides a plot of the data for the entire size distribution ($\ln r$ against $\ln S$), and the regression lines as obtained from the linear regressions reported in Table 3.

III. A Theory of Local Externalities

The empirical analysis above supports the hypothesis that the underlying mechanism that governs the evolution of the size distribution of cities satisfies Gibrat's proposition. Growth rates of cities are observed to be proportionate to city size, and the limiting size distribution is lognormal. From an economics viewpoint, the question remains how economic forces can lead to such population dynamics. While there may be many idiosyncratic reasons why individuals decide to live in one city over another, or choose to move between cities, it is hard to deny that economic forces are a major determinant in population mobility. Cities like Detroit and Philadelphia have seen a significant drop in population, while at the same time experiencing a serious decline in their manufacturing industries. In Silicon Valley on the contrary, cities have seen higher-than-average population growth rates over the 1990s (and often equally lower-than-average rates since 2000). Cupertino City, home to technology companies like Apple and HP, saw a population growth rate of 25 percent between 1990 and 2000, two and a half times the national average. There is no doubt that the economic impact of the technology boom in those cities has contributed to attracting citizens.

We therefore propose a general equilibrium theory that incorporates those differences in technological change across cities. In addition, the main reason for the existence of cities and the determination of population boundaries is the presence of local externalities within cities. Firms and workers locate in cities because there are positive spillovers in production from workers,[37] consumers, suppliers, and even competitors. Without those external benefits, firms would locate in rural areas where property prices are much lower. At the same time though, land and space are in limited supply. All firms and workers ideally want to locate as closely together as possible, but that tendency is slowed by a counteracting force. Not only does a higher population lead to higher property prices (which has been experienced extensively in cities like Cupertino), the presence of more inhabitants causes congestion. There is a negative external cost due to increased commuting time.

[36] The GI s.e. is the Gabaix-Ioannides (2003) corrected standard error. They show that the nominal OLS s.e. underestimates the true standard error due to the positive correlations between residuals caused by the ranking procedure. They derive that for large N, the approximate true standard error is $\hat{a}(2/N)^{1/2}$.

[37] Guy Dumais et al. (1997) provide evidence that sharing a common pool of workers is the main reason why industries locate together.

TABLE 3—PARETO COEFFICIENT REGRESSIONS

Truncation point			Estimates		
$\underline{N}$	$\underline{S}$	City	$\hat{K}$ (s.e.)	$\hat{a}$ (s.e.) (GI s.e.)	R^2
135	155,554	Chattanooga (city), TN	21.099 (0.144)	1.354 (0.011) (0.165)	0.99
2,000	19,383	Lyndhurst (CDP), NJ	20.648 (0.017)	1.314 (0.002) (0.042)	0.997
5,000	6,592	Attalla (city), AL	18.588 (0.019)	1.125 (0.002) (0.023)	0.985
12,500	1,378	Fullerton (city), NE	15.944 (0.014)	0.863 (0.002) (0.011)	0.961
25,000	42	Paoli (town), CO	13.029 (0.010)	0.534 (0.001) (0.005)	0.860

Notes: Dependent variable: Rank (ln). s.e. standard error; GI s.e. Gabaix-Ioannides (2003) corrected standard error ($\hat{a}(2/N)^{1/2}$).
Source: Census Bureau, 2000.

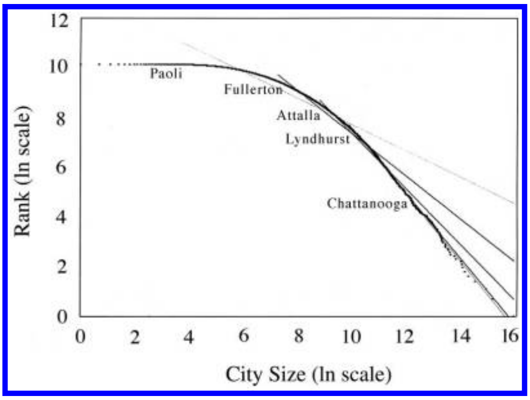

FIGURE 12. ENTIRE SIZE DISTRIBUTION AGAINST RANK AND LINEAR REGRESSION LINES FOR DIFFERENT TRUNCATION POINTS

Citizens in large cities must devote part of their leisure time to nonproductive but work-related commuting.

The model, like Lucas and Rossi-Hansberg's (2002) theory of the internal structure of cities, incorporates those two counteracting external forces. The current model does not explicitly model internal geographic heterogeneity of the city. Because in Lucas and Rossi-Hansberg (2002) citizens obtain the same utility over different locations, it is without loss of generality that citizens within a given city are considered identical. The main objective is to understand economic and population differences between cities, rather than within cities. The city is therefore not considered in isolation, but rather experiences population mobility from and to different cities. The main aim is to extend the work in this literature on the internal structure of cities and allow for competition between cities of different sizes. The space in which heterogeneous cities are considered is therefore the size space rather than a given geographical space.

Define an economy with local externalities C. Time is discrete and indexed by t. Let there be a set of locations (cities) $i \in \mathcal{I} = \{1, \ldots, I\}$. Each city has a continuum population of size $S_{i,t}$, and the total, country-wide population size $S = \sum_{\mathcal{I}} S_{i,t}$. All individuals are infinitely lived and can perform exactly one job. Let $A_{i,t}$ be the productivity parameter that reflects the technological advancement of city i at time t. The law of motion of $A_{i,t}$ is $A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t})$. Each city experiences an exogenous technology shock $\sigma_{i,t}$. Let σ_t denote the vector of shocks of all cities. The city-specific shock is symmetric and is identically and independently distributed with mean zero, and $1 + \sigma_{i,t} > 0$.[38] On aggregate, there is no growth in productivity.[39]

[38] This law of motion implies that $\ln(A_{i,t})$ follows a unit root process. In empirical applications, the presence of a unit root often cannot be rejected. In the real business cycle literature, for example, using the Solow residual to measure TFP, the point estimates found on the persistence parameter ρ in $A_{i,t} = (A_{i,t-1})^{\rho}(1 + \sigma_{i,t})$ cannot be rejected to be different from 1 (see, for example, Robert G. King and Sergio T. Rebelo, 1999).

[39] Recent work by Rossi-Hansberg and Wright (2004) and Gilles Duranton (2002) has proposed different growth models that can explain Zipf's law. Rossi-Hansberg and Wright (2004), for example, have shocks at the industry level. The implication is that while industry size is persistent over time, the size of a given city is not related to that of industries, as industries and workers can relocate each pe-

Firms are identical, consist of one worker, and are infinitesimally small. The marginal product $y_{i,t}$ of a worker is composed of the city's productivity parameter and the positive local externality $a_+(S_{i,t})$ from being in a city of size $S_{i,t}$:

$$y_{i,t} = A_{i,t}a_+(S_{i,t})$$

where $a'_+(S_{i,t}) > 0$ is the positive external effect. It is increasing in $S_{i,t}$ which reflects the fact that larger cities generate bigger externalities in production. The city's labor market is considered perfectly competitive. Identical firms compete for labor of a representative worker, so the wage rate $w_{i,t}$ received by a worker is equal to the marginal product $y_{i,t}$. As a result of the fact that larger cities have a higher marginal product, they also have higher wages.

Workers are endowed with one unit of leisure, which can be employed as labor. Denote $l_{i,t} \in [0, 1]$ as the amount of labor employed, and $1 - l_{i,t}$ as the amount of leisure. Unfortunately, not all labor employed is productive. Because of the negative commuting externality, out of the total amount of labor employed, a fraction needs to be devoted to commuting. As a result, productive labor $L_{i,t} = a_-(S_{i,t})l_{i,t}$, where $a_-(S_{i,t}) \in [0, 1]$ denotes the negative external effect and $a'_-(S_{i,t}) < 0$. The larger the population, the lower the fraction of time that remains to be devoted to productive labor.

The amount of land in a city is fixed and denoted by H. Land is a scarce resource, and it is assumed that the total stock of land available is for residential use. The price of land is given by $p_{i,t}$. An individual citizen's consumption of land is denoted by $h_{i,t}$.

Citizens have preferences over consumption $c_{i,t}$, the amount of land (or housing) $h_{i,t}$, and the amount of leisure $1 - l_{i,t}$. The representative consumer's preferences in city i at period t are represented by

$$u(c_{i,t}, h_{i,t}, l_{i,t}) = c_{i,t}^{\alpha}h_{i,t}^{\beta}(1 - l_{i,t})^{1-\alpha-\beta}$$

where α, β, $\alpha + \beta \in (0, 1)$. Workers and firms are perfectly mobile, so they can relocate to another city instantaneously and at no cost.[40] After observing the realization of the vector of technology shocks σ_t in each period t, citizens choose location i to maximize the discounted stream of utilities. Because all citizens are identical, each of them should obtain the same utility level. Moreover, because there is no aggregate uncertainty over different locations, and because capital markets are perfect, the location decision in each period depends only on the current period utility. The problem is therefore a static problem of maximizing current utility for a given population distribution, and the population distribution must be such that in all cities, the population $S_{i,t}$ equates utilities across cities. In what follows, given a population size $S_{i,t}$ in city i, agents choose consumption bundles $\{c_{i,t}, h_{i,t}, l_{i,t}\}$ in a Walrasian economy with local externalities. The "population market" clears if all S_i imply that the equilibrium utilities are the same across cities.

Given S_i, any individual maximizes utility $u(c_{i,t}, h_{i,t}, l_{i,t})$ subject to the budget constraint (where the tradeable consumption good is the numeraire, i.e., with price unity)

$$\max_{\{c_{i,t}, h_{i,t}, l_{i,t}\}} u(c_{i,t}, h_{i,t}, l_{i,t}; S_i) = c_{i,t}^{\alpha}h_{i,t}^{\beta}(1 - l_{i,t})^{1-\alpha-\beta}$$

$$\text{s.t. } c_{i,t} + p_{i,t}h_{i,t} \leq w_{i,t}L_{i,t}$$

where $w_{i,t} = A_{i,t}a_+(S_{i,t})$ and $L_{i,t} = a_-(S_{i,t})l_{i,t}$. A competitive equilibrium allocation for this problem satisfies the first-order conditions (where λ is the Lagrange multiplier)

$$\alpha c_{i,t}^{\alpha-1}h_{i,t}^{\beta}(1 - l_{i,t})^{1-\alpha-\beta} + \lambda = 0$$

$$\beta c_{i,t}^{\alpha}h_{i,t}^{\beta-1}(1 - l_{i,t})^{1-\alpha-\beta} + \lambda p_{i,t} = 0$$

$$(1 - \alpha - \beta)c_{i,t}^{\alpha}h_{i,t}^{\beta}(1 - l_{i,t})^{-\alpha-\beta} + \lambda w_{i,t}a_-(S_{i,t}) = 0$$

which, after substituting for the market clearing condition of the housing market ($h_{i,t}S_{i,t} = H$)

riod. Their theory therefore also predicts a particular size distribution of industries, in addition to that of cities.

[40] In the real world, there are obviously transportation and relocation costs. Dumais et al. (1997) find, however, that transportation costs have become far less important.

and for the budget constraint, give the following equilibrium prices

$$p^*_{i,t} = \frac{\beta A_{i,t} a_+(S_{i,t}) a_-(S_{i,t}) S_{i,t}}{H}$$

$$w^*_{i,t} = A_{i,t} a_+(S_{i,t})$$

and the equilibrium allocation

$$c^*_{i,t} = \alpha A_{i,t} a_+(S_{i,t}) a_-(S_{i,t})$$

$$h^*_{i,t} = \frac{H}{S_{i,t}}$$

$$l^*_{i,t} = \alpha + \beta.$$

Observe that wages are higher in cities with positive productivity shocks (higher $A_{i,t}$) and they are also higher in cities with a larger population (due to the externality $a_+(S_{i,t})$). This is consistent with the empirical fact that there is an urban wage premium (for an overview, see Glaeser, 1998). Higher wages are in part offset by higher property prices, which in equilibrium implies that less $h_{i,t}$ is consumed, and in part by the fact that more time must be devoted to commuting in larger cities.[41]

Perfect mobility implies that upon realization of the shocks, citizens must be indifferent across different locations.[42] As a result, in equilibrium, city populations will be such that citizens will obtain the same equilibrium utility U[43]

$$u^*(S_{i,t}) = u^*(S_{j,t}) = U$$

for all cities i and j and where $u^*(S_{i,t}) = u(c^*_{i,t}, h^*_{i,t}, l^*_{i,t}; S_{i,t})$. This implies that

$$A_{i,t} \cdot a_+(S_{i,t}) a_-(S_{i,t}) S_{i,t}^{-\beta/\alpha} = A_{j,t} \cdot a_+(S_{j,t}) a_-(S_{j,t}) S_{j,t}^{-\beta/\alpha}$$

is equal to a constant for all cities. Denote $\Lambda(S_{i,t}) = a_+(S_{i,t}) a_-(S_{i,t}) S_{i,t}^{-\beta/\alpha}$ the net local size effect, so $A_{i,t} \cdot \Lambda(S_{i,t})$ is constant. Then provided the inverse exists and Λ^{-1} is a positive power function, we get

$$\Lambda^{-1}(A_{i,t}) S_{i,t} = K \tag{3}$$

where K is a positive constant. After substituting for the law of motion of technology $A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t})$, we obtain

$$\Lambda^{-1}(A_{i,t-1}(1 + \sigma_{i,t})) S_{i,t} = K. \tag{4}$$

This expression now helps establish the following result:

PROPOSITION 2: *Let Λ^{-1} be a positive power function. If $\Lambda(S_{i,t})$ is decreasing, i.e., $\Lambda' < 0$, then (ex ante identical) cities with larger shocks will have larger populations: $(dS_{i,t}/d\sigma_{i,t}) > 0$.*

PROOF:

Apply the implicit function theorem to equation (4), then we get that

$$\frac{dS_{i,t}}{d\sigma_{i,t}} = -\frac{[\Lambda^{-1}]'(\cdot) A_{i,t-1}}{\Lambda(A_{i,t-1}(1 + \sigma_{i,t}))}.$$

Since $[\Lambda^{-1}]'(\cdot) = 1/\Lambda'(\cdot)$, $dS_{i,t}/d\sigma_{i,t}$ is positive provided $\Lambda' < 0$. This establishes the proof.

Consider the following example. Let $a_+(S_{i,t}) = S_{i,t}^{\theta}$, and $a_-(S_{i,t}) = S_{i,t}^{-\gamma}$ then

$$\Lambda(S_{i,t}) = S_{i,t}^{\theta - \gamma - \beta/\alpha} = S_{i,t}^{\Theta}$$

where $\Theta = \theta - \gamma - \beta/\alpha$. Note that $\Lambda^{-1}(S_{i,t}) = S_{i,t}^{1/\Theta}$ is a positive power function. As a result, we write

[41] The amount of time devoted to productive and nonproductive labor is the same across city sizes. An argument could be made that in larger cities the total labor time is larger than in small cities. Unfortunately, our simple model with homothetic preferences cannot account for this. A more sophisticated model with nonhomothetic preferences, or even with heterogeneous agents, may provide a way to introduce it.

[42] Interestingly, Gibrat (1931) himself discusses wage heterogeneity of a given profession (*terrassiers*) across different cities (Saint-Etienne and Lyon) in the presence of random shocks. Unlike the perfect mobility economy considered here, he assumes the complete absence of mobility of workers.

[43] Like in Lucas and Rossi-Hansberg (2001), the location choice of an atomless agent does not change market equilibrium. For an analysis of the impact of individual location choices on market equilibrium, see an interesting model by Ellison and Drew Fudenberg (2003).

$$(A_{i,t-1}(1+\sigma_{i,t}))^{1/\Theta} \cdot S_{i,t} = K.$$

From Proposition 2, bigger shocks will lead to larger cities, provided $\Theta < 0$, i.e. $\theta - \gamma - \beta/\alpha < 0$. Observe that this condition requires that the positive knowledge spillover in production not be too large: $\theta < \gamma + \beta/\alpha$. If the positive spillover is very large, an equilibrium will involve a degenerate distribution of cities where all citizens live in the city with the largest productivity shock.

From equation (3), it is immediate that

$$S_{i,t} = 1/\Lambda^{-1}(A_{i,t}) \cdot K \tag{5}$$

for all t. Since $A_{i,t} = A_{i,t-1}(1 + \sigma_{i,t})$ and provided Λ^{-1} is a power function, it follows that $\Lambda^{-1}(A_{i,t}) = \Lambda^{-1}(A_{i,t-1}) \cdot \Lambda^{-1}(1 + \sigma_{i,t})$ and therefore

$$\begin{aligned} S_{i,t} &= 1/\Lambda^{-1}(A_{i,t-1}) \cdot 1/\Lambda^{-1}(1+\sigma_{i,t}) \cdot K \\ &= 1/\Lambda^{-1}(1+\sigma_{i,t}) \cdot S_{i,t-1} \end{aligned}$$

after substituting equation (5) evaluated at $t - 1$. We now redefine $1/\Lambda^{-1}(1 + \sigma_{i,t}) = (1 + \varepsilon_{i,t})$ to get

$$S_{i,t} = (1 + \varepsilon_{i,t}) \cdot S_{i,t-1}. \tag{6}$$

The latter equation is exactly what gives rise to Gibrat's law provided shocks are small.

Gibrat's Law of Proportionate Growth.—Gibrat (1931) (following the discovery by Kapteyn, 1903) establishes the law of proportionate effect. Consider a stochastic process $\{S_{i,t}\}$ indexed by place $i = 1, \ldots, I$ and time $t = 0, 1, \ldots$, where $S_{i,t}$ is the population size of a place i at time t. Let $\varepsilon_{i,t}$ be an identically and independently distributed random variable[44] denoting the growth rate between period $t - 1$ and t for place i. If growth is proportionate, then

$$S_{i,t} - S_{i,t-1} = \varepsilon_{i,t} S_{i,t-1}$$

or

$$S_{i,t} = (1 + \varepsilon_{i,t}) \cdot S_{i,t-1}.$$

Rewriting and taking the summation, we get

$$\sum_{t=1}^{T} \frac{S_{i,t} - S_{i,t-1}}{S_{i,t-1}} = \sum_{t=1}^{T} \varepsilon_{i,t}$$

and since for small intervals

$$\sum_{t=1}^{T} \frac{S_{i,t} - S_{i,t-1}}{S_{i,t-1}} \approx \int_{S_{i,0}}^{S_{i,T}} \frac{dS_i}{S_i} = \ln S_{i,T} - \ln S_{i,0}$$

or equivalently between any two periods

$$\ln S_{i,t} = \ln S_{i,t-1} + \varepsilon_{i,t}.$$

As a result, it follows that $\ln S_{i,T} = \ln S_{i,0} + \varepsilon_{i,1} + \cdots + \varepsilon_{i,T}$. From the central limit theorem, $\ln S_{i,T}$ is asymptotically normally distributed, and hence $S_{i,T}$ is asymptotically lognormally distributed, provided the shocks are independently distributed and small (thus justifying $\ln(1 + \varepsilon_{i,t}) \approx \varepsilon_{i,t}$). In other words, in line with Gibrat's proposition, a proportionate stochastic growth process leads to the lognormal distribution.

As a result, the above establishes the main proposition of the theory of local externalities:

PROPOSITION 3: *Let $\mathcal{C}$ be an economy with local externalities, let Λ^{-1} be a positive power function, and let $\Lambda(S_{i,t})$ be decreasing. Then city size satisfies Gibrat's law: the population growth process is proportionate and the asymptotic size distribution is lognormal.*

It is important to note that Gibrat's law will still hold for economies with local externalities that in addition have economy-wide externalities. In fact, by introducing a technological parameter A, common to all cities, economy-wide technological progress can be captured which results from external effects. This typically denotes an aggregate measure—most often the mean or the max—of the economy-wide technological progress. For Gibrat's law to hold, it

[44] Donald R. Davis and David E. Weinstein (2002) provide some evidence for persistence in the population shocks, in particular when those shocks are extremely big. This was the case, for example, in the Japanese cities of Hiroshima and Nagasaki after they were destroyed by the atom bomb in August 1945. For small variance, i.e., "normal" shocks, nonpersistence is justified.

does need to be satisfied that this country-wide technology parameter is *independent* of city size. For the case of the size distribution of firms and not of cities, Eeckhout and Jovanovic (2002) provide evidence that spillovers across firms are dependent on the size of firms. In fact, spillovers between firms are larger for smaller firms. We do not find such evidence across cities. If the true model of the economy is the one proposed in this section, then the proportionate population growth process is consistent with the fact that there are no net local spillovers across cities of different sizes. That does not, of course, rule out the possibility that there are local spillovers between cities of different sizes that are geographically close, but the net effect over the entire distribution cancels out.[45] Our results provide no evidence in that direction. Recent work on MAs by Linda H. Dobkins and Ioannides (2001), however, establishes that distance from the nearest higher-tier city (i.e., the nearest larger city in a higher tier) is not significant as a determinant of size and growth.

Ideally, further analysis of the data should be done. In particular, one would like to analyze the entire size distribution over time. This would provide an exact description of the moments of the distribution at different points in time which would allow for further verification of the underlying statistical process. It would answer questions concerning the limit variance, whether Gibrat's law satisfies exactly a Geometric Brownian motion, thus pinning down the detailed process that generates a limit lognormal size distribution.[46] Unfortunately, due to the lack of available data covering the entire size distribution, those further analyses are not possible at this time.

IV. Concluding Remarks

In this paper, a simple but robust underlying mechanism of population dynamics of all cities in the United States has been uncovered. Cities grow proportionately, i.e., at a stochastic rate that is independent of city size, and this gives rise to a lognormal distribution of cities. This property of the stochastic process has been known at least since Gibrat (1931). At the same time, this result can account for what for over half a century has been the benchmark stylized fact of economic geography, that the upper tail of the city size distribution satisfies Zipf's law. It has been shown that the results confirming Zipf's law and the corresponding estimates of the power coefficient can be obtained even if the true underlying distribution is not the Pareto (or Zipf) distribution. Estimated power coefficients are sensitive to the choice of the truncation point and are consistently increasing in the truncation. Given a lognormal distribution, we have proposed a simple resolution of one of the major puzzles related to the size distribution of cities based on Gibrat's law.

This breakthrough can be made only now because it hinges on the availability of new data in Census 2000 for the entire size distribution. The change in conclusion following the availability of different data does not seem to be an isolated occurrence in science. A similar phenomenon has occurred in material sciences, in particular in the measurement of the atmospheric aerosol size distribution.[47] Atmospheric aerosols are particles of different components floating in the air. When the measurement of particles is restricted to those with the largest size (often due to the absence of measurement technology that can capture the distribution of the smaller ones), the resulting observed distribution is in fact the truncated distribution and is often fit to a power law. With the advent of advanced measurement technology, however, smaller particles and hence the total size distribution can be measured. Knowledge of the entire atmospheric aerosol distribution is important mainly because, for humans, inhalation of small aerosol is much more harmful than large. The latter get stopped in the nostrils and throat and never enter the lungs. For the entire size distribution of many aerosol types, the distribu-

[45] Consider, for example, an economy with many pairs of cities, each pair with one large and one small city. If the shocks between the large and the small city are correlated, but shocks across the many pairs of cities are not, growth will still be proportionate.

[46] Kalecki (1945) extends the class of stochastic processes that lead to the lognormal distribution. This is motivated by his observation that Gibrat's process leads to a lognormal distribution with linearly and unbounded increasing variance.

[47] I am grateful to Samuel Pessoa for pointing this out. See John H. Seinfeld and Spyros N. Pandis (1997), Amin Haaf and Rainer Jaenicke (1980) and William Hinds (1982).

tion is actually lognormal, or a convolution of different lognormals.

The fact that Gibrat's proposition is established concerning the population mobility of cities is a necessary requirement for an empirically consistent theory of the underlying economic activity. The second main purpose of this paper is to propose and solve an equilibrium model of local externalities where wages and prices guide citizens in their location decision. Consistent with proportionate growth and a lognormal size distribution, the model establishes a mechanism of local productivity shocks in the presence of local externalities and their effect, through worker mobility, on the population size distribution of cities.

Appendix A: The Size Distribution of Counties

We investigate the size distribution of counties. While counties may not necessarily be the right geographical unit that an economist is interested in, they do have the major advantage that the size distribution of counties comprises 100 percent of the U.S. population, i.e., 281 million in 2000. According to the Census, counties are described as the primary legal divisions of most states. For example, voting for most elections is organized at the county level. Most counties are functioning governmental units, whose powers and functions vary from state to state. Legal changes to county boundaries or names are typically infrequent.

TABLE A-1—TEN LARGEST COUNTIES IN THE UNITED STATES

Rank	City	Population S	S_{LA}/S
1	Los Angeles County, CA	9,519,338	1.000
2	Cook County, IL	5,376,741	1.770
3	Harris County, TX	3,400,578	2.799
4	Maricopa County, AZ	3,072,149	3.099
5	Orange County, CA	2,846,289	3.344
6	San Diego County, CA	2,813,833	3.383
7	Kings County, NY	2,465,326	3.861
8	Miami-Dade County, FL	2,253,362	4.225
9	Queens County, NY	2,229,379	4.269
10	Dallas County, TX	2,218,899	4.290

Note: S_{LA}/S denotes the ratio of population size relative to Los Angeles.
Source: Census Bureau, 2000.

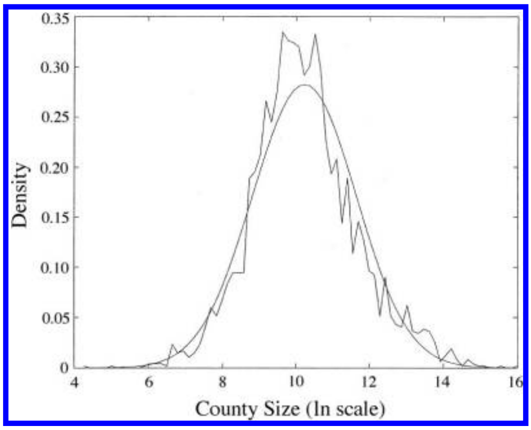

FIGURE A-1. EMPIRICAL AND THEORETICAL DENSITY FUNCTIONS OF ALL COUNTIES

In 2000, there were 3,141 counties in the United States covering the entire population. The ten largest are listed in Table A-1. The largest, Los Angeles County, California, had 9.5 million inhabitants and the smallest, Loving County, Texas, 67 inhabitants. The sample mean (in ln, standard error in brackets) is $\hat{\mu}$ = 10.22 (0.02) and the standard deviation is $\hat{\sigma}$ = 1.41.

In Figure A-1 we plot the size distribution, together with the normal density $\phi(\hat{\mu}, \hat{\sigma})$.

The size empirical density is remarkably similar to the normal. There is somewhat more mass near the mode, and the distribution may be slightly skewed, but the distribution of county size is nonetheless surprisingly close to lognormal.

Appendix B: Additional Statistics of City Growth

TABLE B-1—DESCRIPTIVE STATISTICS OF CITY GROWTH

Range of cities		Growth rate (non-normalized)		
	N	mean	stdev	IQR ($Q3 - Q1$)
All	19,361	0.103	0.729	0.199
Top 100	100	0.108	0.158	0.154
Bottom 100	100	−0.127	0.671	0.493
P_{10} to P_{90}	15,488	0.106	0.786	0.191

Source: Census Bureau, 1990–2000.

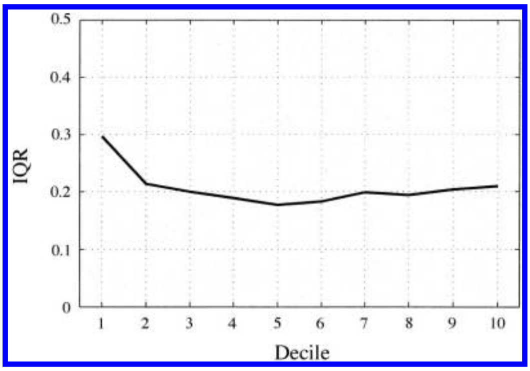

FIGURE B-1. INTERQUARTILE RANGE BY DECILE

REFERENCES

Auerbach, Felix. "Das Gesetz Der Bevolkerungskonzentration." *Petermanns Geographische Mitteilungen*, 1913, *49*, pp. 73–76.

Blank, Aharon and Solomon, Sorin. "Power Laws in Cities' Population, Financial Markets and Internet Sites (Scaling in Systems with a Variable Number of Components)." *Physica A*, 2000, *287*(1–2), pp. 279–88.

Champernowne, D. G. "A Model of Income Distribution." *Economic Journal*, 1953, *63*(250), pp. 318–475.

Davis, Donald R. and Weinstein, David E. "Bones, Bombs, and Break Points: The Geography of Economic Activity." *American Economic Review*, 2002, *92*(5), pp. 1269–89.

Dobkins, Linda H. and Ioannides, Yannis M. "Spatial Interaction among US Cities: 1900–1990." *Regional Science and Urban Economics*, 2001, *31*(6), pp. 701–31.

Dumais, Guy; Ellison, Glenn and Glaeser, Edward L. "Geographic Concentration as a Dynamic Process." *Review of Economics and Statistics*, 2002, *84*(2), pp. 193–204.

Duranton, Gilles. "City Size Distributions as a Consequence of the Growth Process." Centre for Economic Performance, LSE, CEP Discussion Papers, 2002.

Eaton, Jonathan and Eckstein, Zvi. "Cities and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics*, 1997, *27*(4–5), pp. 443–74.

Eeckhout, Jan and Jovanovic, Boyan. "Knowledge Spillovers and Inequality." *American Economic Review*, 2002, *92*(5), pp. 1290–307.

Ellison, Glenn and Fudenberg, Drew. "Knife Edge or Plateau: When Do Market Models Tip?" National Bureau of Economic Research, Inc., NBER Working Papers: 9528, 2003.

Gabaix, Xavier. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 1999, *114*(3), pp. 739–67.

Gabaix, Xavier and Ioannides, Yannis M. "The Evolution of City Size Distributions," in J. V. Henderson and J. F. Thisse, eds., *Handbook of urban and regional economics, Vol. 4*. Amsterdam: Elsevier Science, North-Holland, forthcoming 2003.

Gibrat, Robert. *Les inégalités économiques; applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel.* Paris: Librairie du Recueil Sirey, 1931.

Glaeser, Edward L. "Are Cities Dying?" *Journal of Economic Perspectives*, 1998, *12*(2), pp. 139–60.

Glaeser, Edward L.; Scheinkman, Jose A. and Shleifer, Andrei. "Economic Growth in a Cross-Section of Cities." *Journal of Monetary Economics*, 1995, *36*(1), pp. 117–43.

Haaf, Amin and Jaenicke, Rainer. "Results of Improved Size Distribution Measurement in the Aitken Range of Atmospheric Aerosols." *Journal of Aerosol Science*, 1980, *11*(3), pp. 321–30.

Hardle, Wolfgang. *Applied nonparametric regression*. Econometric Society Monographs. Cambridge, New York and Melbourne: Cambridge University Press, 1990.

Henderson, J. Vernon. "The Sizes and Types of Cities." *American Economic Review*, 1974, *64*(4), pp. 640–56.

Hinds, William C. *Aerosol technology*. New York: Wiley, 1982.

Ioannides, Yannis M. and Overman, Henry G. "Zipf's Law for Cities: An Empirical Examination." *Regional Science and Urban Economics*, 2003, *33*(2), pp. 127–37.

Jovanovic, Boyan. "Selection and the Evolution of Industry." *Econometrica*, 1982, *50*(3), pp. 649–70.

Kalecki, Michael. "On the Gibrat Distribution." *Econometrica*, 1945, *13*(2), pp. 161–70.

Kapteyn, Jacobus C. *Skew frequency curves in biology and statistics*. Astronomical Laboratory, Groningen: Noordhoff, 1903.

Kesten, Harry. "Random Difference Equations as Renewal Theory for Products of Random Matrices." *Acta Mathematica*, 1973, *131*, pp. 207–48.

King, Robert G. and Rebelo, Sergio T. "Resuscitating Real Business Cycles," in J. B. Taylor and M. Woodford, eds., *Handbook of macroeconomics, Volume 1B. Handbooks in Economics, Vol. 15.* Amsterdam, New York and Oxford: Elsevier Science, North-Holland, 1999, pp. 927–1007.

Krugman, Paul. *Development, geography, and economic theory.* Cambridge, MA: MIT Press, 1995.

Krugman, Paul. *The Self-organizing economy*. Cambridge: Blackwell, 1996.

Le Maitre, Alexandre B. *La metropolitée, ou de l'établissement des villes capitales, de leur utilité passive & active, de l'union de leurs parties & de leur anatomie, de leur commerce. [Publisher: Amsterdam, Ches B. Boekhold, Pour J. Van Grop, 1682]: Editions d'histoire sociale*. Paris: 1973, Vol. 17, pp. 192.

Lucas, Robert E., Jr. and Rossi-Hansberg, Esteban. "The Internal Structure of Cities." *Econometrica*, 2002, *70*(4), pp. 1445–76.

Pareto, Vilfredo. *Cours d'économie politique [1896].* New Edition, G.-H. Bousquet and G. Busino, eds. Geneva: Librairie Droz, 1964.

Rossi-Hansberg, Esteban and Wright, Mark. "Urban Structure and Growth." Unpublished Paper, 2004.

Seinfeld, John H. and Pandis, Spyros N. *Atmospheric chemistry and physics: From air pollution to climate change*. New York: John Wiley & Sons, 1997.

Silverman, Bernard W. *Density estimation for statistics and data analysis*. New York: Chapman and Hall, 1986.

Sutton, John. "Gibrat's Legacy." *Journal of Economic Literature*, 1997, *35*(1), pp. 40–59.

Zipf, George K. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley Press, 1949.

This article has been cited by:

1. Yannick Malevergne, Vladilen Pisarenko, Didier Sornette. 2011. Testing the Pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *Physical Review E* **83**:3. . [CrossRef]
2. Xavier Gabaix, Rustam Ibragimov. 2011. Rank − 1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents. *Journal of Business and Economic Statistics* **29**:1, 24-39. [CrossRef]
3. Rafael González-Val. 2010. THE EVOLUTION OF U.S. CITY SIZE DISTRIBUTION FROM A LONG-TERM PERSPECTIVE (1900-2000)*. *Journal of Regional Science* **50**:5, 952-972. [CrossRef]
4. AURÉLIE LALANNE, RICHARD SHEARMUR. 2010. La métropolisation à l'aune de la 'loi' taille/rang : Le cas Canadien, 1971-2001. *Canadian Geographer / Le Géographe canadien* **54**:4, 476-491. [CrossRef]
5. Michael Batty. 2010. Space, Scale, and Scaling in Entropy Maximizing. *Geographical Analysis* **42**:4, 395-421. [CrossRef]
6. Peter Huber, Michael Pfaffermayr. 2010. Testing for Conditional Convergence in Variance and Skewness: The Firm Size Distribution Revisited*. *Oxford Bulletin of Economics and Statistics* **72**:5, 648-668. [CrossRef]
7. STIJN VAN NIEUWERBURGH, PIERRE-OLIVIER WEILL. 2010. Why Has House Price Dispersion Gone Up?. *Review of Economic Studies* **77**:4, 1567-1606. [CrossRef]
8. Boris A. Portnov, Ben Reiser, Moshe Schwartz. 2010. Does Gibrat's law for cities hold when location counts?. *The Annals of Regional Science* . [CrossRef]
9. Rafael González-Val, Marcos Sanso-Navarro. 2010. Gibrat's law for countries. *Journal of Population Economics* **23**:4, 1371-1389. [CrossRef]
10. I. S. Gill, C.-C. Goh. 2010. Scale Economies and Cities. *The World Bank Research Observer* **25**:2, 235-262. [CrossRef]
11. K. Giesen, J. Sudekum. 2010. Zipf's law for cities in the regions and the country. *Journal of Economic Geography* . [CrossRef]
12. Jesús Clemente, Rafael González-Val, Irene Olloqui. 2010. Zipf's and Gibrat's laws for migrations. *The Annals of Regional Science* . [CrossRef]
13. Marcus Berliant. 2010. MISBEHAVIORAL URBAN ECONOMICS. *Journal of Regional Science* **50**:1, 93-101. [CrossRef]
14. Klaus Desmet, Esteban Rossi-Hansberg. 2010. ON SPATIAL DYNAMICS. *Journal of Regional Science* **50**:1, 43-63. [CrossRef]
15. Stephen J. Redding. 2010. THE EMPIRICS OF NEW ECONOMIC GEOGRAPHY. *Journal of Regional Science* **50**:1, 297-311. [CrossRef]
16. Xavier Gabaix. 2009. Power Laws in Economics and Finance. *Annual Review of Economics* **1**:1, 255-294. [CrossRef]
17. Moshe Levy. 2009. Gibrat's Law for (All) Cities: CommentGibrat's Law for (All) Cities: Comment. *American Economic Review* **99**:4, 1672-1675. [Abstract] [View PDF article] [PDF with links]
18. Jan Eeckhout. 2009. Gibrat's Law for (All) Cities: ReplyGibrat's Law for (All) Cities: Reply. *American Economic Review* **99**:4, 1676-1683. [Abstract] [View PDF article] [PDF with links]
19. Gautier Krings, Francesco Calabrese, Carlo Ratti, Vincent D Blondel. 2009. Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment* **2009**:07, L07003. [CrossRef]

20. Mark D. Partridge, Dan S. Rickman, Kamar Ali, M. Rose Olfert. 2009. Do New Economic Geography agglomeration shadows underlie current population dynamics across the urban hierarchy?. *Papers in Regional Science* **88**:2, 445-466. [CrossRef]

21. H. D. Rozenfeld, D. Rybski, J. S. Andrade, M. Batty, H. E. Stanley, H. A. Makse. 2008. Laws of population growth. *Proceedings of the National Academy of Sciences* **105**:48, 18702-18707. [CrossRef]

22. ANNA BOGOMOLNAIA, MICHEL LE BRETON, ALEXEI SAVVATEEV, SHLOMO WEBER. 2008. Heterogeneity Gap in Stable Jurisdiction Structures. *Journal of Public Economic Theory* **10**:3, 455-473. [CrossRef]

23. M. D. Partridge, D. S. Rickman, K. Ali, M. R. Olfert. 2008. Lost in space: population growth in the American hinterlands and small cities. *Journal of Economic Geography* **8**:6, 727-757. [CrossRef]

24. Jan Loecker, Jozef Konings, Patrick Van Cayseele. 2008. Merger Review: How Much of Industry is Affected in an International Perspective?. *Journal of Industry, Competition and Trade* **8**:1, 1-19. [CrossRef]

25. M. Batty. 2008. The Size, Scale, and Shape of Cities. *Science* **319**:5864, 769-771. [CrossRef]

26. Tomoya Mori, Koji Nishikimi, Tony E. Smith. 2008. THE NUMBER-AVERAGE SIZE RULE: A NEW EMPIRICAL RELATIONSHIP BETWEEN INDUSTRIAL LOCATION AND CITY SIZE*. *Journal of Regional Science* **48**:1, 165-211. [CrossRef]

27. Alan Collins. 2007. MAKING TRULY COMPETITIVE CITIES ? ON THE APPROPRIATE ROLE FOR LOCAL GOVERNMENT. *Economic Affairs* **27**:3, 75-80. [CrossRef]

28. Yo Sasaki, Hiroto Kuninaka, Naoki Kobayashi, Mitsugu Matsushita. 2007. Characteristics of Population Distributions in Municipalities. *Journal of the Physical Society of Japan* **76**:7, 074801. [CrossRef]

29. Ali Ataullah, Andrew Higson, Mark Tippett. 2007. The Distributional Properties of the Debt to Equity Ratio: Some Implications for Empirical Research. *Abacus* **43**:2, 111-135. [CrossRef]

30. Matthew O. Jackson, Brian W. Rogers. 2007. Meeting Strangers and Friends of Friends: How Random Are Social Networks?Meeting Strangers and Friends of Friends: How Random Are Social Networks?. *American Economic Review* **97**:3, 890-915. [Abstract] [View PDF article] [PDF with links]

31. ESTEBAN ROSSI-HANSBERG, MARK L. J. WRIGHT. 2007. Urban Structure and Growth. *Review of Economic Studies* **74**:2, 597-624. [CrossRef]

32. Gilles Duranton. 2007. Urban Evolutions: The Fast, the Slow, and the StillUrban Evolutions: The Fast, the Slow, and the Still. *American Economic Review* **97**:1, 197-221. [Abstract] [View PDF article] [PDF with links]

33. Mark Partridge, M. Rose Olfert, Alessandro Alasia. 2007. Canadian cities as regional engines of growth: agglomeration and amenities. *Canadian Journal of Economics/Revue canadienne d'économique* **40**:1, 39-68. [CrossRef]

34. MARIO POLÈSE, RICHARD SHEARMUR. 2006. Growth and Location of Economic Activity: The Spatial Dynamics of Industries in Canada 1971-2001. *Growth and Change* **37**:3, 362-395. [CrossRef]

35. G DURANTON. 2006. Some foundations for Zipf's law: Product proliferation and local spillovers. *Regional Science and Urban Economics* **36**:4, 542-563. [CrossRef]

36. K DESMET, M FAFCHAMPS. 2006. Employment concentration across U.S. counties. *Regional Science and Urban Economics* **36**:4, 482-509. [CrossRef]

37. Robert W. Wassmer. 2006. The Influence of Local Urban Containment Policies and Statewide Growth Management on the Size of United States Urban Areas*. *Journal of Regional Science* **46**:1, 25-65. [CrossRef]