*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. 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 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), 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 Maıtre 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 Fe ́lix 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 (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 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. 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. 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

(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.

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 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, 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 spillovers.” 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 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 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.

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.”

A substantial portion of research into the size distribution of the U.S. population has been done using the MA19 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.

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 􏲢ˆ 􏲒 7.28 (0.01) and the standard deviation is 􏲣ˆ 􏲒 1.75. The theoretical density function of the lognormal size distribution is normal in ln *S* and given by 􏲤 ( 􏲢ˆ , 􏲣ˆ ) :

A Kolmogorov-Smirnov (KS) test of goodness of fit of the empirical density function against the lognormal with sample mean 􏲢ˆ 􏲒 7.28 and sample standard deviation 􏲣ˆ 􏲒 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, 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.

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 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 for all *i* 􏲒 1, ... , 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 density24 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 *gi* 􏲒 *m*(*Si*) 􏲖 􏲦*i* , *i* 􏲒 1, ... , 19361, where *gi* is the normalized growth rate, i.e., the difference between growth and the sample mean divided by the sample standard deviation, and *Si* 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 *Si*. The estimate of *m*(*s*) will be denoted *mˆ* (*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 *Kh* denotes the dependence of *K* on the bandwidth *h*. With the kernel weights, we calculate the estimate of *m* using the NadarayaWatson method,26 where

In Figure 7 there is a plot of *mˆ* (*s*) calculated for a bandwidth of *h* 􏲒 0.5 (see Silverman, 1986). The Figure also shows the bootstrapped 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 seventyfifth and twenty-fifth percentiles (*Q*3 􏲕 *Q*1). 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 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 increases28 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 􏲣*ˆ*2(*s*) (see Ha ̈rdle, 1990) is calculated as

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 variance30 (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:

(*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 􏲖 *ga*)10 􏲒 1.103 or *ga* 􏲒 1 percent. The lack of significance of city growth on size is further confirmed when the dependent variable is the population size in 1990:

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

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 1⁄2 the size of New York. The tenth-ranked city, Detroit, should have a size 1⁄10 of New York. Above in Tables 1 and 2, *S*/*SNY* 􏲗 *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

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

where *N*􏲷 is the number of cities above the truncation point. Taking natural logs, we get that rank is inversely proportional to size

where *K* 􏲒 ln *N*􏲷 􏲖 *a* ln *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 *aˆ* equal to 1.354:

(*N*􏲷 􏲒 135, *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 *N*􏲷 .

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

(*N*􏲷 􏲒 135, *S*􏲷 􏲒 280,843, *R*2 􏲒 0.985). Observe that for MAs, the estimated coefficient *aˆ* is nearly exactly equal to 1, as originally described by Zipf (1949). Unfortunately, the fact that *aˆ* is equal to 1 is highly sensitive to the choice of the truncation point in either direction: for *N*􏲷 􏲒 276 (the entire MA sample and roughly double the original), *aˆ* 􏲒 0.850, and for *N*􏲷 􏲒 67 (half the original sample), *aˆ* 􏲒 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 *N*􏲷 􏲒 135, the sample ensures a perfect fit with Zipf’s original observation.34

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 􏲤􏲸 as given in equation (1). To simplify notation, let *x* 􏲒 ln *S*, and denote the normal cumulative density function by 􏲘(*x*). Now consider the truncated lognormal distribution at truncation point *x*􏲷 􏲒 ln *S*􏲷. Then the cdf of the truncated lognormal is

As before, let *N*􏲷 be the sample size of the truncated distribution. Then the rank can be written as

and taking logs

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 􏲕 􏲘(*x*)) is not linear in *x* 􏲒 ln *S*. Calculating the derivative of the term that depends on *x* in equation (2) gives

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 􏲢ˆ 􏲒 7.28, 􏲣ˆ 􏲒 1.75 is strictly increasing over the entire domain (and positive by definition). The plot of the hazard function *h*( *x*; 􏲢*ˆ* , 􏲣*ˆ* ) is given in Figure 11.

A strictly increasing hazard rate implies that the second derivative of the term ln(1 􏲕 􏲘(*x*)) is strictly concave, i.e., *d*2/*dx*2ln(1 􏲕 􏲘(*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* 􏲘( *x*; 􏲢*ˆ* , 􏲣*ˆ* ), *then the estimate of the parameter aˆ of the Pareto distribution is increasing in the truncation city size* (*daˆ* /*dS*􏲷 ) 􏲛 0 *and decreasing in the truncated sample population* (*daˆ*/*dN*􏲷) 􏲚 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

Not only are the estimates of *aˆ* 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 *S*􏲷 (decreasing *N*􏲷 ), *aˆ* 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.

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 lowerthan-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. 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* 􏲺 I 􏲒 {1, ... , I}. Each city has a continuum population of size *Si*,*t*, and the total, country-wide population size S 􏲒 ¥I *Si*,*t*. All individuals are infinitely lived and can perform exactly one job. Let *Ai*,*t* be the productivity parameter that reflects the technological advancement of city *i* at time *t*. The law of motion of *Ai*,*t* is *Ai*,*t* 􏲒 *Ai*,*t* 􏲕 1(1 􏲖 􏲣*i*,*t*). Each city experiences an exogenous technology shock 􏲣*i*,*t*. Let 􏲣*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 􏲖 􏲣*i*,*t* 􏲛 0.38 On aggregate, there is no growth in productivity.39

Firms are identical, consist of one worker, and are infinitesimally small. The marginal product *yi*,*t* of a worker is composed of the city’s productivity parameter and the positive local externality *a*􏲖(*Si*,*t*) from being in a city of size S:

where *a*􏲖􏲙 (*Si*,*t*) 􏲛 0 is the positive external effect. It is increasing in *Si*,*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 *wi*,*t* received by a worker is equal to the marginal product *yi*,*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 *li*,*t* 􏲺 [0, 1] as the amount of labor employed, and 1 􏲕 *li*,*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 *Li*,*t* 􏲒 *a*􏲕(*Si*,*t*)*li*,*t*, where *a*􏲕(*Si*,*t*) 􏲺 [0, 1] denotes the negative external effect and *a*􏲕􏲙 (*Si*,*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* . An individual citizen’s consumption of *i*,*t*

land is denoted by *hi*,*t*.  
Citizens have preferences over consumption

*ci*,*t*, the amount of land (or housing) *hi*,*t*, and the amount of leisure 1 􏲕 *li*,*t*. The representative consumer’s preferences in city *i* at period *t* are represented by

where 􏲪, 􏲫, 􏲪 􏲖 􏲫 􏲺 (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 􏲣*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 *Si*,*t* equates utilities across cities. In what follows, given a population size *Si*,*t* in city *i*, agents choose consumption bundles {*ci*,*t*, *hi*,*t*, *li*,*t*} in a Walrasian economy with local externalities. The “population market” clears if all *Si* imply that the equilibrium utilities are the same across cities.

Given *Si*, any individual maximizes utility *u*(*ci*,*t*, *hi*,*t*, *li*,*t*) subject to the budget constraint (where the tradeable consumption good is the numeraire, i.e., with price unity)

where *wi*,*t* 􏲒 *Ai*,*ta*􏲖(*Si*,*t*) and *Li*,*t* 􏲒 *a*􏲕(*Si*,*t*)*li*,*t*. A competitive equilibrium allocation for this problem satisfies the first-order conditions (where 􏲭 is the Lagrange multiplier)

which, after substituting for the market clearing condition of the housing market (*hi*,*tSi*,*t* 􏲒 *H*) and for the budget constraint, give the following equilibrium prices

and the equilibrium allocation

Observe that wages are higher in cities with positive productivity shocks (higher *Ai*,*t*) and they are also higher in cities with a larger population (due to the externality *a*􏲖(*Si*,*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 *hi*,*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

for all cities *i* and *j* and where This implies that

is equal to a constant for all cities. Denote

the net local effect, so *Ai*,*t* 􏲹 􏲞(*Si*,*t*) is constant. Then provided the inverse exists and 􏲞􏲕1 is a positive power function, we get

where *K* is a positive constant. After substituting for the law of motion of technology *Ai*,*t* 􏲒 *Ai*,*t* 􏲕 1(1 􏲖 􏲣*i*,*t*), we obtain

This expression now helps establish the follow

ing result:

􏲕1  
PROPOSITION 2: *Let* 􏲞 *be a positive power function. If* 􏲞(*Si*,*t*) *is decreasing, i.e.,* 􏲞􏲙 􏲚 0, *then* (*ex ante identical*) *cities with larger shocks will have larger populations:* (*dSi*,*t* /*d*􏲣*i*,*t*) 􏲛 0.

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

Since is positive provided. This establishes the proof.

Consider the following example. Let *a* (*S* )􏲒*S*􏲮,and*a* (*S* )􏲒*S*􏲕􏲯 then where note that is a positive power function. As a result, we write

From Proposition 2, bigger shocks will lead to larger cities, provided. Observe that this condition requires  
that the positive knowledge spillover in production not be too large. 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

For all t. Since and provided is a power function, it follows that and therefore

After substituting equation (5) evaluated at t – 1. We now define to get

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 {*Si*,*t*}indexedbyplace*i*􏲒1,...,*I*andtime*t*􏲒 0, 1, ... , where *Si*,*t* is the population size of a place *i* at time *t*. Let 􏲦*i*,*t* be an identically and independently distributed random variable44 denoting the growth rate between period *t* 􏲕 1 and *t* for place *i*. If growth is proportionate, then

Or

Rewriting and taking the summation, we get

and since for small intervals

or equivalently between any two periods

As a result, it follows that ln *Si*,*T* 􏲒 ln *Si*,0 􏲖 􏲦*i*,1 􏲖 ... 􏲖 􏲦*i*,*T*. From the central limit theorem, ln *Si*,*T* is asymptotically normally distributed, and hence *Si*,*T* is asymptotically lognormally distributed, provided the shocks are independently distributed and small (thus justifying ln(1 􏲖 􏲦*i*,*t*) 􏲗 􏲦*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* C *be an economy with local externalities, let* 􏲞􏲕1 *be a positive power function, and let* 􏲞(*Si*,*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 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.

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. 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 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.