3

Cities as Six-by-Six-Mile Squares
Zipf's Law?

Thomas J. Holmes and Sanghoon Lee

3.1 Introduction

Economists analyzing urban economics questions commonly use geographic units from the Census Bureau; for example, *metropolitan statistical areas* (MSAs). The Census Bureau, in turn, typically uses arbitrarily defined political boundaries to construct its reporting units. The Census Bureau must satisfy numerous constituents with its reporting. In its determination of reporting unit boundaries, the Census Bureau would not be likely to place a high priority on what would be best for research in urban economics. Put another way, there is a high probability of measurement error between the *economic units* that researchers want and the *reporting units* such as MSAs that the Census Bureau provides.

A question in urban economics that has attracted much attention is the extent to which the size distribution of cities obeys Zipf's law.[1] If Zipf's law holds perfectly, then when we rank cities and plot the log of the rank against the log of the city population, we get a straight line with a slope

Thomas J. Holmes is professor of economics at the University of Minnesota, a visiting scholar at the Federal Reserve Bank of Minneapolis, and a research associate of the National Bureau of Economic Research. Sanghoon Lee is an assistant professor in the Sauder School of Business, University of British Columbia.

We thank NSF Grant 0551062 for research support. We are grateful to conference volume editor Ed Glaeser for his detailed and helpful comments. We thank Joel Waldfogel for his comments as a discussant at the NBER Economics of Agglomeration conference. We received many helpful comments from the conference participants, including, in particular, a comment by Gilles Duranton. We thank Jan Eeckhout, Xavier Gabaix, and Erzo Luttmer for feedback. We thank Julia Thornton and Steve Schmeiser for research assistance. The views expressed herein are solely those of the authors and do not represent the views of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. See Gabaix and Ioannides (2004) for a literature survey.

of 1. Equivalently, the largest city is twice as big as the second largest, three times as big as the third largest, and so on (the *rank-size rule*). Researchers who have used MSAs to define cities, such as Gabaix (1999), have found that Zipf's law holds to a striking degree. But what does it mean to say that Zipf's law holds, when the boundaries are determined by bureaucrats and politicians?

We are concerned about how to interpret Zipf's law results with these data for three reasons. First, MSAs are aggregations of counties, and the county is a crude geographic unit for such a building block. In some parts of the country, counties cover an extremely large land area, and locations get wrapped together as an MSA that clearly does not comprise a coherent metropolitan area.[2] We note that even if measurement error is unsystematic, it causes potential problems for a study of the size distribution, because the distribution *with* measurement error is generally different from the one *without* it. Second, we are particularly concerned about how boundaries are drawn for the largest cities. These cities can often be found in densely populated parts of the country where MSAs form contiguous blocks, such as the Northeast Corridor extending from Washington, DC, to Boston. It is often a tough call determining whether a given area should be classified as one or two MSAs, and if the latter, where to delineate the boundary. If bureaucrats tend to use broad definitions of MSAs that subsume contiguous areas into single large MSAs, this process may itself contribute to the findings of Zipf's law. Third, with MSA data, we leave out approximately 20 percent of the population not living in MSAs. So, we do not see what is going on with small cities, the left tail of the size distribution.[3] Eeckhout (2004) has recently advocated looking at the left tail by using data on census places that include very small towns. But as argued next, census places are heavily dependent on arbitrary political decisions of where to draw boundaries.

Our chapter considers a new approach to looking at population distributions that sweeps out any decisions made by bureaucrats or politicians. When comparing populations of geographic units, we can think of differences as falling along two margins. First, one unit can have a larger population than another because it encompasses more land area, holding population density fixed. Second, a unit can have a larger population on a fixed amount of land; that is, higher population density. In our analysis of the size distribution, we completely eliminate the first margin and allow only the second. We cut the map of the continental United States into a uniform grid of *six-by-six-mile squares* (and some other size grids as well) and examine the distribution of population across the *squares*. We document several regularities that

2. This point about MSAs is well appreciated in the literature. See, for example, Bryan, Minton, and Sarte (2007) for a recent discussion.

3. The Census recently released data on what are called *micropolitan areas,* which are essentially moderate-sized counties that do not qualify as MSAs. So, our concern that the county is a crude geographic unit applies here.

are robust to various ways of cutting the data. We also examine the extent to which Zipf's law holds for squares.

Our first result is that the extreme left tail of the distribution looks approximately lognormal—roughly, a bell curve. With the Zipf distribution, there are always more smaller cities than bigger cities; there is never a bell curve with a modal point below which the density of log population decreases as size decreases. This works well on the right tail of the distribution (e.g., there are more squares with 50,000 people than with 100,000) but does not work well around the left tail. This point can be highlighted by a discussion of the extreme cases of squares with population one and two. There are 713 squares with exactly one person (a bachelor farmer, a forest ranger) living in them. A much larger number of squares (1,285) have exactly two people living in them. (Perhaps a forest ranger couple?) Given priors about scale economies and basic agglomeration benefits, it not surprising that squares with one lonely person in them are rarer than squares with two. The recent literature has not focused on scale economies and agglomeration benefits to try to understand the size distribution; instead, it has focused on the impacts of cumulative random productivity shocks (e.g., Gabaix [1999] and Eeckhout [2004]). We suspect that to understand the shape of the extreme left tail of the distribution of squares, issues of scale economies and agglomeration are of first-order importance.

Our second result throws out the extreme left tail and looks at the distribution of population across squares with population 1,000 or more. Approximately 24,000 squares meet this population threshold, and these squares account for 28 percent of the surface area of the continental United States. We construct a Zipf plot and find a striking pattern. To a remarkable degree, the plot is linear until it hits a kink at square population around 50,000. Below the kink, the slope is approximately 0.75; above the kink, the slope is approximately 2. This piecewise linear function fits the data extremely well. Moreover, when we split the data by region and make a Zipf's plot in each individual region, the same piecewise linear relationship shows up, with the kinks in approximately the same place. Our results are not like the standard Zipf's law findings, and the objects we are looking at—with no variation on the land-area margin—are different from the standard objects people look at. But we find our results intriguing in the same way that the usual Zipf's law findings are intriguing.

The third result concerns the extent that Gibrat's law for growth rates holds with squares. Under a typical statement of Gibrat's law, the mean and variance of growth is independent of initial size. Gibrat's law does not hold for squares. The relationship between growth and size is an inverted U, with the smallest and the largest population squares having the lowest growth rates. It is not surprising that the highest population squares have a low growth rate, since these areas typically are fully developed, and little vacant land is available for further growth.

Our fourth result links our findings to results in the previous literature about Zipf's law for MSAs. As mentioned, the main finding in the literature is that when we look at the upper tail of the MSA's size distribution, the regression coefficient of log rank on log population equals 1. Now, if we were to replace *MSA population* with *MSA average density* in the regression, we do not necessarily expect to get a coefficient of 1, because it depends on the elasticity of MSA surface area to MSA population. If this elasticity equals 0.05 (which is approximately what we find it to be), then the expected slope coefficient on density is actually 2 rather than 1. This is, in fact, our approximate result when we replace MSA population with MSA density. This is also our result when we use the maximum density square rather than the average density in the MSA. We find it interesting that the slope we are getting in the right tail of these *MSA-level* regressions is similar to the slope we get in the right tail of the *square-level* regressions (i.e., the slope to the right of the previously mentioned kink). We interpret this result as evidence of some kind of fractal structure, where the distribution of average density of the right tail of MSAs is similar to the distribution of the right tail of squares *within* MSAs, which in turn is similar to the distribution of the right tail of squares *across* all of the continental United States.

Given our wariness about using the MSA surface-area measure, we are somewhat surprised that when we use it to construct average MSA density, we get numerical results that we can connect to our results with squares. Perhaps the bureaucrats are doing a reasonably good job after all. Even if they are, our analysis of squares rather than MSAs is still interesting, because we are looking at something different from the previous literature with new insights. The fractal pattern of the right tails—across MSAs similar to squares within MSAs similar to squares across the continent—suggests an underlying common explanation. The dominant explanation in the recent literature of the size distribution of MSAs is the random growth explanation of Gabaix (1999),[4] but it certainly cannot explain the size distribution of squares within MSAs and squares across the continent. For one thing, Gibrat's law does not hold for squares, as already noted, and Gibrat's law is needed to get the random growth theory to work. For another, it is clear that the size distribution of squares within MSAs is better understood by economic theories like the Alonzo-Muth-Mills monocentric model of the city than by a random growth theory. We believe that a unified theory of the size distribution of squares within MSAs and across MSAs will have to incorporate economic factors like scale economies and include an explicit spatial structure. See Hsu (2008) for an attempt to do exactly this.

The closely related work of Eeckhout (2004) merits further discussion. He made a compelling case that the use of MSAs truncates out low population areas, and he suggested the use of the *census place* as a way to see what is

4. For related work on firms, see Luttmer (2007).

happening at the bottom tail of the distribution. Interestingly, Eeckhout found that the distribution of places is lognormal rather than Zipf. However, we are even more concerned about the use of census places to define geographic boundaries than we are about the use of MSAs. First, only 74 percent of the 2000 population actually lives in what the Census calls a place; the rest of the population are in unincorporated areas.

Next, consider table 3.1. To construct it, we take a list of all census places from the 2000 Census (Eeckhout's data) and tabulate all those places with population five or less. Two places in the census file have exactly one resident (including Lost Springs, Wyoming), and two places have population equal to two, including Hove Mobile Park City, North Dakota. The arbitrary decision that Lost Springs with its one resident is considered a place, while a farmhouse in an unincorporated area with a family of five living in it is not a place of five people is dependent on legal particulars that are not likely to be of interest in our analysis of city size distributions. These concerns arise at the top of the size distribution as well. Saint Paul and Minneapolis in the Twin Cities are adjacent to each other and are different census places, since they have never merged. Manhattan and Brooklyn are part of the same census place (New York City), because they merged in the nineteenth century. Our six-by-six-square analysis pulls in all of the land in the continental United States and treats it in a uniform way: the one resident of Hove Mobile Park City is on equal footing with a bachelor farmer in an unincorporated area, and New York City is treated the same way as the Twin Cities.

Many others have noted the inadequacies of MSA definitions for vari-

Table 3.1 **Census places with population five or less (2000 Census)**

Place	Population
New Amsterdam town, IN	1
Lost Springs town, WY	1
Hove Mobile Park city, ND	2
Monowi village, NE	2
Hobart Bay CDP, AK	3
East Blythe CDP, CA	3
Hillsview town, SD	3
Point of Rocks CDP, WY	3
Flat CDP, AK	4
Blacksville CDP, GA	4
Prudhoe Bay CDP, AK	5
Storrie CDP, CA	5
Baker village, MO	5
Maza city, ND	5
Gross village, NE	5

Note: CDP = Census designated place.

ous research questions and have used geographical techniques to improve on these boundaries. For example, Duranton and Turner (2008) use buffers around 1976 settlements within MSA boundaries to obtain more meaningful MSA definitions for their analysis of urban growth and transportation. Others have used rich geographic data to determine the location of employment subcenters. (See Anas, Arnott, and Small [1998] and McMillen and McDonald [1998].) In principle, rather than fix squares like we do, it might be possible to draw some kind of optimal city boundaries to let the land margin back in. We view this approach as fruitful and complementary. But once the economists take the job of drawing the metropolitan boundaries away from the bureaucrats, we need to worry about the mistakes the economists might make. For this reason, we think it is useful to nail down what happens when we completely eliminate the land margin across locations, as we do here.

While the focus of our work is the size distribution and Zipf's law, our work also makes a broader point that research in urban economics should not be constrained by standard geographic units handed to us by statistical agencies. The Census releases population data at an extremely high level of geographic precision—the *block level* (which in urban areas is a city block or an apartment building)—so there is great flexibility in choosing boundaries. Moreover, such analysis is facilitated by advances in geographic information system software. We therefore have great flexibility in defining the boundaries to be whatever we want them to be. In many applications in urban economics, researchers might be well served by defining their own boundaries rather than using the off-the-shelf boundaries. The construction of segregation indices is one example. Other papers highlighting the flexibility of continuous geographic data include Duranton and Overman (2005) and Burchfield et al. (2006). Another related work is the G-Econ database, which contains the worldwide geographic distribution of economic activity (gross domestic product; GDP) on a 1-degree-latitude-by-1-degree-longitude grid (Nordhaus et al. 2006).

3.2 Data

We draw a grid of six-by-six-mile squares across the map of the continental United States. A map is a two-dimensional projection of the three-dimensional globe, and the square grid may look different on maps using different projection methods. We use the USA Contiguous Albers Equal Area Conic projection method, which preserves area size: the size of an area on a map is equal to the real size of the area on the globe.[5]

5. This may not be true in maps using other projections. For example, maps using Mercator projections present Greenland as being roughly as large as Africa, but Africa is about fourteen times as big as Greenland.

We use six miles for our baseline, because in the first version of this chapter, we used the original township grid of six-by-six-mile squares. This grid was laid down in the early 1800s by the Public Land Survey System (PLSS) for the purpose of selling federal lands. (See Linklater [2003] and Holmes and Lee [2008].) That was a good place to start, but we eventually realized that drawing our own grid would be much cleaner. That way, we could cover states that were otherwise left out (e.g., the original thirteen states were not surveyed, because there were no federal lands to sell). Moreover, the original survey done with chains and landmarks was sloppy compared to what we can do now on a computer. We have to anchor the grid at some place, but as we show later, shifting the grid up or down or left or right is irrelevant. As discussed in section 3.7, a large enough change in the grid size can make a difference but not a small change.

The grid has 85,527 squares, each exactly thirty-six square miles, summing up to 3.1 million square miles of the continental United States. Figure 3.1 illustrates the grid in the vicinity of New York City. Note the six-by-six squares along the coast project into the water. We treat these areas as full six-by-six-mile squares and do not distinguish between dry land and water when delineating the surface area within the square. We make no distinc-

Fig. 3.1 Map of grid lines for six-by-six squares in the vicinity of New York City

tion, because people can live on the water (e.g., on houseboats) in some cases more easily than they can live on dry land, particularly in remote desert areas. We return to the water issue in section 3.7.

We use the population data from the 2000 and 1990 decennial Census reported at the level of the census block. In urban areas, a census block is a city block or an apartment building. For 2000, there are 7 million census blocks in the continental United States. Of those reporting any population, the area of the median census block for 2000 equaled 0.014 square miles, a tiny unit of land compared to a six-by-six square. The ninety-fifth percentile of block area equals 1.43 miles, still a small amount. The Census Bureau reports the longitude and latitude of a point within the boundaries of each census block, and we use this point to map each block into a six-by-six square. Figure 3.2 illustrates the location of census blocks in the vicinity of New York City. In this area, a thousand or more blocks can be assigned to a particular square.

We need to address the possibility of measurement error in the allocation of population to squares. A block boundary might cross the boundaries of a six-by-six square, and when this happens, someone living in the block on one side of the boundary can be mistakenly allocated to the six-by-six square on the other side. Because blocks are typically very small, this issue is negligible,

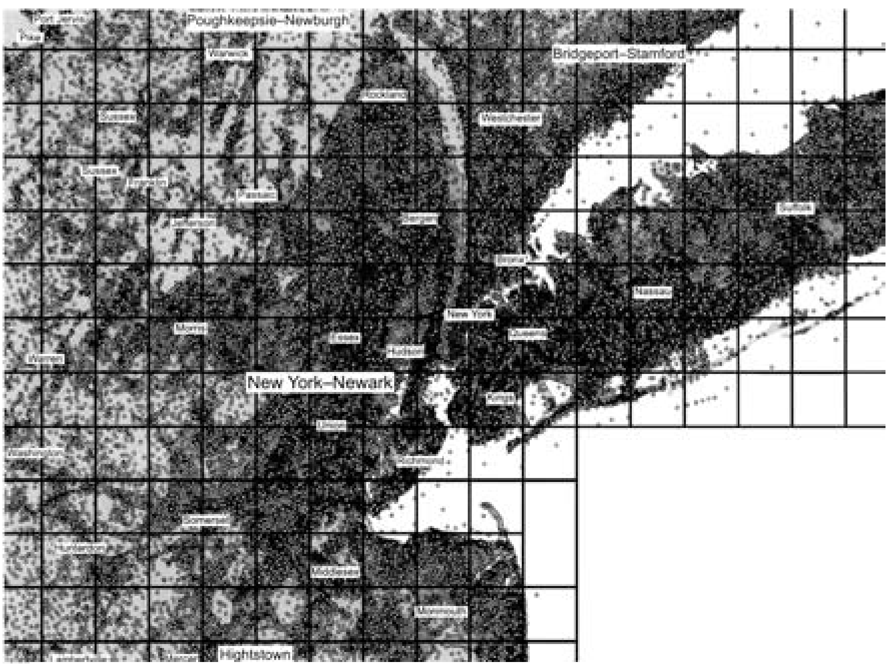

Fig. 3.2 Location of census blocks (2000 Census) in the vicinity of New York City

except in a few extreme cases. To get some sense of this issue, we determine for each of the 280 million people in the population what six-by-six square they are assigned to and the number of block groups assigned to the same six-by-six square. The first percentile of this statistic is thirty-five blocks. This means that all but 1 percent of the population live in six-by-six squares with at least thirty-five blocks assigned to them. Now, thirty-five blocks will trace out a fairly clean square. The fifth percentile is 74 blocks, the fiftieth is 719, and the seventy-fifth is 1609. We are confident that for 99 percent of the population, our assignment is very good. We note that even in remote rural areas, the Census typically defines blocks at a fine level of granularity.[6]

To compare our results with what comes out of the traditional approach with MSA-level data, it is useful to aggregate our squares to MSAs. We allocate squares to the MSAs as defined for the 2000 Census. In certain metropolitan areas, the Census offers a choice of consolidated areas (e.g., the New York CMSA) versus a breakdown into component areas. We use the consolidated definitions. There are 274 different such MSAs in the continental United States. We allocate squares to MSAs according to the following rule. A square gets assigned to an MSA if any block in the square is part of the MSA. In the event a square is at a boundary where MSAs overlap in the square, we assign the square to the MSA with the largest surface area based on blocks.

Table 3.2 presents summary statistics of how population from the 2000 Census varies across squares. Mean population across the 85,527 squares is 3,269. Population is highly skewed, with two squares in the New York MSA having 1.3 million in population. The area unit used in the analysis to calculate density is the six-by-six-mile square. So, each square has one unit of area, and the population density equals the population.

Table 3.2 also presents summary statistics for the 274 MSAs. Mean density is 7,881 per square, which is twice the density of squares overall. The mean number of squares across MSAs is 87, with the minimum being 14 and the maximum being 981 squares. So clearly, the square is a much smaller geographic unit than the MSA. The maximum land area is attained by the Las Vegas MSA, which is a good example of the limitations of Census MSA definitions. The surface areas of counties in Nevada are huge. Since the Census uses the county as a building block unit for MSAs, much of the surrounding area that is not actually part of the Las Vegas metropolitan area is folded into the MSA bearing its name.[7]

6. In a relatively small number of cases, a square has only one block group assigned to it. There are 592 such blocks, accounting for 20,000 people (out of 280 million). These look like unusual and exceptional cases rather than just simply rural cases. Of these 20,000 people, 5,677 are in the 29 Palms military base in California. The base is in a census block covering 272 square miles. Another block is in the Mohave Desert. Others are in national parks and national forests.

7. Another example of this problem with huge counties is the case of the Flagstaff MSA in Arizona. The city of Flagstaff is located in the geographically huge Coconino county (over

Table 3.2 Summary statistics: Squares and MSAs (population from 2000 Census)

Unit	Variable	Number	Mean	Standard deviation	Minimum	Maximum	Sum across units
Square	Population	85,527	3,269	18,181	0	1,317,207	279,583,434
	Log(population)	70,590	5.69	2.48	0	14.09	—
	Area (6 × 6 square)	85,527	1	0	1	1	85,527
MSA	Population	274	843,209	1,986,836	60,744	21,343,534	231,039,389
	Population density	274	7,881	7,073	215	55,151	—
	Log(population density)	274	8.67	.80	5.37	10.92	—
	Area (6 × 6 square)	274	87	103	14	981	23,798

3.3 Background Equations

Discussing some background equations on the size distribution is useful. Following the notation of Gabaix and Ioannides (2004), let S_i denote the population size of city i, and suppose the distribution of populations across cities is Pareto:

$$\text{Rank}_i = P(\text{Size} > S_i) = \frac{\alpha}{S_i^{\zeta}}. \tag{1}$$

Taking logs, we get

$$\ln \text{Rank}_i = \ln \alpha - \zeta \ln S_i. \tag{2}$$

The slope ζ is called the *tail coefficient.* Zipf's law is said to hold if $\zeta = 1$.

Let L_i be the land area of city i and the population density D_i be

$$D_i = \frac{S_i}{L_i}.$$

The analysis remains in a log-linear form if there is a constant elasticity η relationship between land and population,

$$L_i = \gamma S_i^{\eta}.$$

Taking logs yields

$$\ln L_i = \ln \gamma + \eta \ln S_i. \tag{3}$$

18,000 square miles). The Census classifies the whole county as the Flagstaff MSA. Flagstaff is the third largest MSA by land area. Cities quite distant from Flagstaff, including Tuba City (78 miles) and Page (119 miles), are folded into the Flagstaff MSA because they happen to be in this county. A large percentage of the Flagstaff MSA population reported by the Census comes from distant places like these that clearly are not part of the economic unit of Flagstaff city. Researchers might be tempted to use the city boundaries of Flagstaff rather than the MSA boundaries. But this raises the issue of the often arbitrary political decisions that determine municipal boundaries.

Solving the preceding for $\ln S_i$ and substituting into equation (2) yields

$$\text{(4)} \qquad \ln \text{Rank}_i = \left[\ln \alpha + \frac{\zeta}{\eta} \ln \gamma\right] - \frac{\zeta}{\eta} \ln L_i.$$

This is a Zipf's relationship using land instead of population. Note the slope is ζ/η, not ζ. In the special case where population density is constant across cities (e.g., each individual inelastically demands one unit of land), then $\eta = 1$, and the slope coefficient for the land regression in equation (4) is identical to the slope coefficient for the population regression in equation (2). But otherwise, in the empirically relevant case where $\eta < 1$, the slope is higher for the land regression than the population regression.

Analogously, using $\ln D_i = \ln S_i - \ln L_i$ and equation (3), we can solve for $\ln S_i$ in equation (2) in terms of $\ln D_i$ to get

$$\text{(5)} \qquad \ln \text{Rank}_i = \left[\ln \alpha - \frac{\zeta \ln \gamma}{(1 - \eta)}\right] - \frac{\zeta}{1 - \eta} \ln D_i.$$

This is a Zipf's plot for population density. The tail coefficient is $\zeta / (1 - \eta)$. If Zipf's law holds so that $\zeta = 1$, and if $\eta < 1$, then this slope will be greater than 1.

Next, consider squares. Let the squares be indexed by j, and let s_j be the population of square j. Let A_i be the set of squares that are in city i. Then, city population, land area, and density equal

$$S_i = \sum_{j \in A_i} s_j.$$

$$L_i = \text{Number of squares in } A_i,$$

$$D_i = \frac{S_i}{L_i} = \text{mean } s_j, j \in A_i.$$

In general, the relationship between the size distribution of the squares s_j and of the cities S_i is quite complicated, except for the special case where each square is a city. We leave to future research a theoretical analysis of this relationship and focus instead on a descriptive analysis of the distribution of the squares s_j and how it compares to the distribution of MSA-defined cities.

We are able to make one immediate observation. Let $s_i^{\max}$ be the highest population square in city i,

$$s_i^{\max} = \max_{j \in A_i} s_j.$$

If the maximum density square is proportionate to the overall city population density,

$$\text{(6)} \qquad s_i^{\max} = \lambda D_i,$$

and if we replace D_i in equation (5) with $s_i^{\max}$, then we obtain the same slope coefficient. This is interesting, because the maximum population square is more reliably measured than the average population density of an MSA. The latter heavily depends on where the boundaries are drawn. Typically, there is rural land at the boundary of an MSA, so the wider the boundaries are drawn, the lower the overall MSA population density. The s_i^{max} variable is determined in the interior of the MSA, the "central business district," far from the boundaries of the MSA. So, it will not be affected if the MSA boundary is arbitrarily increased twenty miles out or twenty miles in.[8] (The MSA boundaries still impact the s_i^{max} measure if the Census merges two MSAs into one.)

3.4 The Size Distribution of MSAs

As a benchmark, this section examines the size distribution of MSAs. Following Gabaix (1999), we focus on the 135 largest MSAs, treating this area as the upper tail of the distribution.

Figure 3.3 presents three Zipf plots. Panel A is the standard plot where we use population. Panel B replaces population with land area as in equation (4); panel C replaces population with density as in equation (5).[9] Table 3.3 reports estimated slope coefficients. As is common in the literature, we estimate the tail index two ways: standard ordinary least squares (OLS) and the Hill method (the maximum likelihood procedure under the null hypothesis that the distribution is Pareto). See Gabaix and Ioannides (2004) for a discussion of econometric practice in this literature. As recommended in this work, we use simulation methods to estimate the OLS standard errors, because the usual method yields biased estimates. Zipf's law for the population holds in a striking fashion. The OLS estimate of the slope coefficient for the population regression is 1.01. The fit is excellent, as can be seen by the straight line in figure 3.3 and by the R^2 of 0.988 in table 3.3.

The Hill estimate of the population coefficient is 0.94—a little less than 1. But the estimated standard error is 0.07, so we cannot reject that the slope equals one with a standard statistical test. Here and elsewhere in the chapter, the Hill estimates are a little smaller than the OLS estimates and have a higher estimated standard error but are otherwise similar. Since the OLS and Hill estimates are basically telling the same story, for the rest of the chapter, we will discuss just the OLS estimates in the text but report both in the tables.

The OLS slope coefficients on land and density are 1.70 and 1.90, respectively. Straight lines fit reasonably well. To relate this result to the equations in the previous section, we look at the relationship between land area and

8. One issue with s_i^{max} one could raise is that it might depend on where the grid is positioned. We show in the following text that we can shift around the grid and our results with s_i^{max} do not change.

9. Analogous to what we do for population, for land, we take the top 135 MSAs ranked by land, and for density, we take the top 135 MSAs ranked by density.

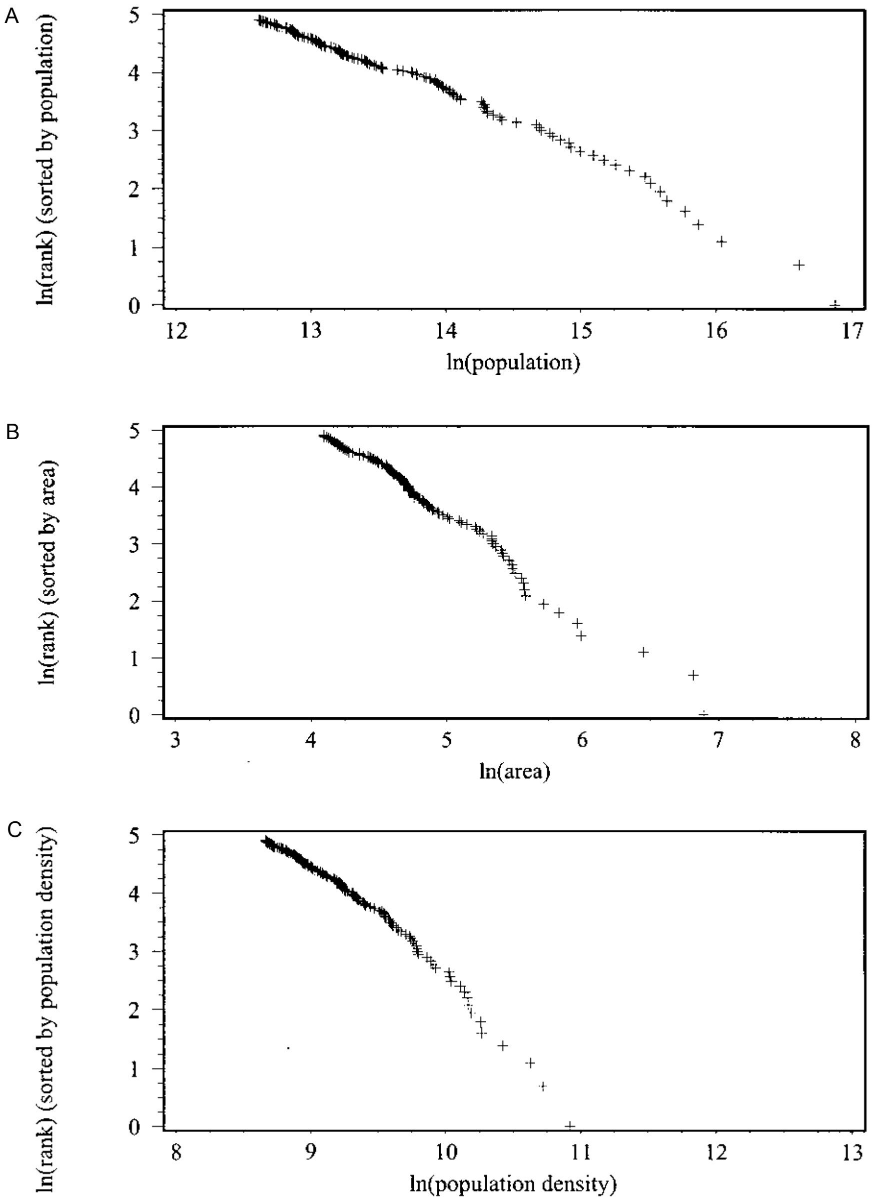

Fig. 3.3 MSA-level Zipf plots: *A*, top 135 MSAs by population; *B*, top 135 MSAs by area; *C*, top 135 MSAs by population density

Table 3.3 **MSA-level Zipf regression results: Alternative size measures**

	OLS		Hill method
Size measure	Slope (absolute value)	R^2	Slope (absolute value)
Population	1.013	.985	.944
	(.12)		(.078)
Land area	1.70	.984	1.569
	(.12)		(.176)
Density	1.896	.973	1.616
	(.12)		(.120)
s_i^{max} (maximum population square in MSA)	1.761	.988	1.546
	(.12)		(.125)

Note: Each regression uses top 135 MSAs ranked by given size measure.

population in the top 135 MSAs by population. A regression of the log of MSA area on log MSA population yields a slope coefficient of 0.52.[10] Let us take this as an estimate of η from the previous section. Equations (4) and (5) from the previous section suggest the slope coefficient on both land and density should approximately equal 2 if $\zeta = 1$ and $\eta = 0.5$ approximately hold. Our estimates of 1.70 and 1.90 are in the ballpark of 2.

Next, we bring in our information about squares into an MSA-level analysis. For each MSA *i*, we determine s_i^{max}, the maximum population square of all the squares in MSA *i*. We substitute s_i^{max} for the average density D_i, as discussed in the previous section. The results are reported in the bottom row of table 3.3. The estimated slope coefficient equals 1.76. The estimate is close to the 1.90 estimate obtained with average density, and the fit is little better: $R^2 = 0.988$ instead of $R^2 = 0.973$. Recall that the land measure for MSAs is crude, making the derived measure of average MSA density a relatively crude object. Yet, the results are similar with the two alternative measures of density. Suppose the population of the maximum density square is proportionate to average density as in equation (6) *and* that the average density measure is measured precisely. Then, these two regressions would yield similar slopes. We interpret this finding as encouraging for those wishing to use MSA-defined cities.

It is worth noting that even with the s_i^{max} regression, we are still dependent on Census decisions about whether two nearby metropolitan areas should be grouped into one or two MSAs. The Census groups San Francisco and Oakland into one MSA, so the observation of s_i^{max} is downtown San Francisco. If Oakland were separated into a distinct MSA, we would get another observation of s_i^{max} for downtown Oakland. In our exercise in the next section with squares, we do not depend on such Census classifications.

10. The standard error is 0.04, and the $R^2 = 0.52$.

So far, our focus has been on the upper tail of the MSA distribution. Next, we look at the entire distribution of MSAs. It is known in the literature that Zipf plots of MSAs tend to exhibit a concave shape when the lower tail of the distribution is included. (See, for example, Rossi-Hansberg and Wright [2007].) When a Zipf's plot is not a straight line, a standard density plot of the distribution can be more revealing than a Zipf's plot. As a segue into looking at the whole distribution, we first illustrate in panel A of figure 3.4 a density plot (histogram) of log population for just the upper tail, the 135 highest population MSAs. Also illustrated in the plot is the best-fitting

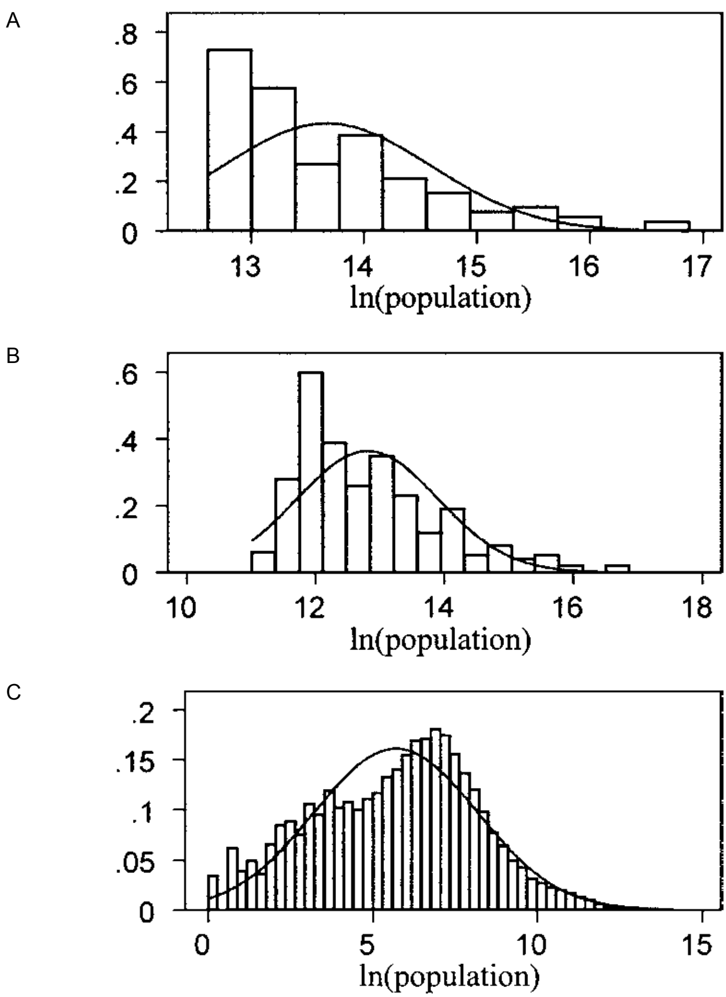

Fig. 3.4 Density plots: *A*, 135 largest MSAs; *B*, all 274 MSAs; *C*, populated squares

normal curve. Clearly, the bell-curved shape of the normal does not fit the distribution within the top 135 MSAs very well. Rather, a Pareto distribution is a good fit here. With the Pareto, the density is a straight line that is strictly decreasing; the smaller the units, the more units there are.

Panel B in figure 3.4 illustrates the distribution of log population for all 274 MSAs. Now, the tendency for monotone decline of the density is not as pronounced as it is with just the top 135, but this is still the clear pattern. Certainly, the bell curve of the normal does not fit the distribution of MSAs very well.

3.5 The Size Distribution of Six-by-Six Squares

We turn now to the size distribution of six-by-six squares. Table 3.4 provides cell counts for population size groupings. Approximately 15,000 of the 86,000 squares are unpopulated. There are 713 squares where only one person lives and 1,285 where two people live. Clearly, the Pareto in which the density is always decreasing cannot fit this distribution.

Panel C of figure 3.4 is a density plot of log population across all squares with at least one person. For the unpopulated squares, the log of population is minus infinity, so the figure leaves out a spike at minus infinity. For squares with one person, log population equals 0, so the plot begins here. The last column of table 3.4 provides a conversion from population to log population to aid in interpretation of the figure. When log population is less

Table 3.4 **Distribution of population across six-by-six squares (Census 2000 population in the contiguous United States)**

	Number of squares	Percent of population	Log(population) at bottom of grouping
All squares	85,527		
Population = 0	14,937	0.00	$-\infty$
Population > 0	70,590	100.00	0.00
By population size grouping			
Population = 1	713	.00	0.00
Population = 2	1,285	.00	0.69
$3 \leq$ population ≤ 5	2,564	.00	1.10
$6 \leq$ population < 10	2,532	.01	1.79
$10 \leq$ population < 100	16,233	.23	2.30
$100 \leq$ population $< 1{,}000$	23,289	3.59	4.61
$1{,}000 \leq$ population $< 10{,}000$	19,271	21.20	6.91
$10{,}000 \leq$ population $< 50{,}000$	3,521	27.40	9.21
$50{,}000 \leq$ population $< 1{,}000{,}000$	1,179	46.28	10.82
$1{,}000{,}000 \leq$ population	3	1.29	13.82
Size groupings of later interest			
$1{,}000 \leq$ population	23,974	96.17	6.91
$50{,}000 \leq$ population	1,182	47.57	10.82

than 4 (when population is less than about fifty), the best fit normal curve fits reasonably well, though the fit is choppy. Certainly, the lognormal fits the distribution better than the Pareto on the right tail.

Our finding that the lognormal is a rough approximation to the right tail of the distribution of squares is like Eeckhout's (2004) finding that the lognormal fits the right tail of the distribution of census places. But as argued in the introduction, the census place is a problematic geographic unit to use in examining the size distribution. Eeckhout presents a random growth model with shocks to location productivities that generates a lognormal distribution. We do not attempt any formal analysis in this chapter to try to explain why the size distribution has the shape that it has. But a look at the raw data makes us skeptical that random location-specific productivity shocks are the main driving factor, at least at the extreme left tail. That there are more squares with two people than with one person (1,285 instead of 713) seems to us more likely due to basic agglomeration benefits in the human condition rather than the variance of location-specific productivity shocks. It seems likely that as we move beyond the one- and two-person size classes, related agglomeration forces are also at work.

We now turn our attention away from the extreme left tail and consider what the distribution looks like with the extreme left tail truncated. If any part of the distribution is to look anything like Zipf, it has to be on the downward-sloping portion of the density. Inspection of figure 3.4 (panel C) reveals that the mode of the distribution is approximately at a log population of 7, which corresponds to approximately a population of 1,000. Henceforth, we truncate all squares with population less than 1,000. From table 3.4, we see that there are 23,974 squares with 1,000 people or more and that these account for about 28 percent of the U.S. land mass and 96 percent of the population. The coverage of the population is very significant here. Even with the truncation, we are including areas that are quite remote.

Figure 3.5 is a Zipf's plot of the population distribution of squares with 1,000 or more people. It exhibits a clear pattern. The relationship looks piecewise linear, with a kink around log population of 11 (which corresponds to a population of approximately 50,000). Above the kink, the relationship steepens. We use nonlinear least squares to fit a piecewise linear function to the plot in figure 3.5. The estimates are reported in table 3.5. Because of the large number of observations, the estimated standard errors are quite small, so they are not reported. The estimated kink is at a log population of 10.89. Below the kink, the (absolute value of) the slope is 0.75; above the kink, it is 1.94. The $R^2 = 0.998$ is extremely high, so the piecewise linear function fits very well. For comparison purposes, we also fit a linear function. The slope in the linear case is between the estimates for the piecewise linear case, and the fit is noticeably worse.

The Census groups states into nine different census divisions. Our next exercise is to examine the distribution of population across squares within

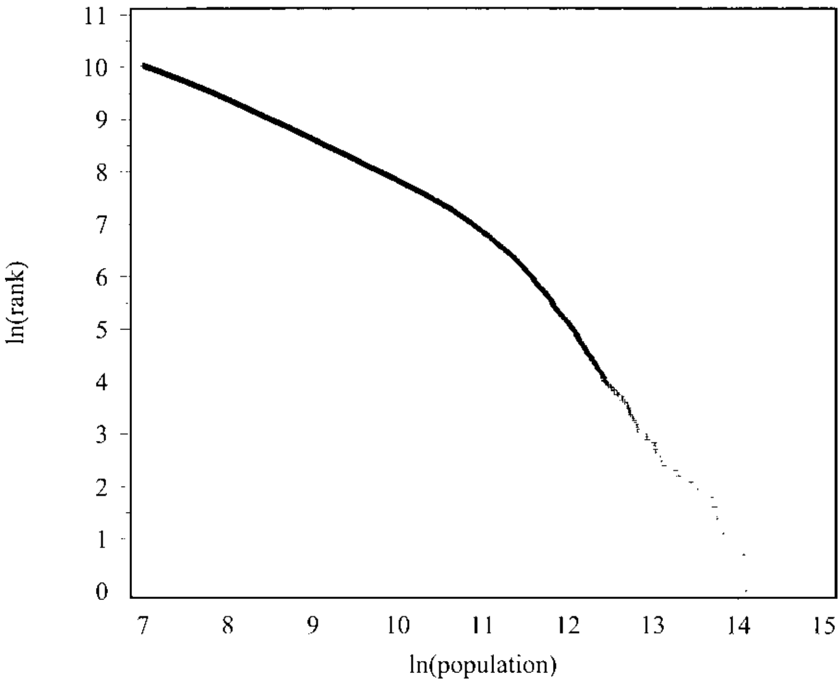

Fig. 3.5 Square-level Zipf plot for continental United States (all 23,974 squares with population at least 1,000)

Table 3.5 Six-by-six-square-level Zipf regression results (squares with population 1,000 and above)

		Piecewise linear				Linear	
Sample of squares	*N*	Kink	Slope1	Slope2	R^2	Slope	R^2
All squares with population ≥ 1,000	23,974	10.89	.747	1.937	.998	.833	.969
By Census division							
New England	1,027	9.96	.569	1.521	.996	.763	.930
Middle Atlantic	2,184	10.28	.669	1.249	.997	.759	.965
East North Central	4,313	10.92	.784	1.982	.999	.861	.975
West North Central	2,337	11.04	.886	2.607	.999	.941	.984
South Atlantic	4,977	10.72	.756	2.175	.995	.857	.959
East South Central	2,898	10.48	1.010	2.357	.997	1.072	.983
West South Central	3,078	11.17	.786	2.834	.997	.857	.969
Mountain	1,383	11.55	.723	3.662	.997	.791	.964
Pacific	1,777	11.21	.521	1.872	.992	.646	.922

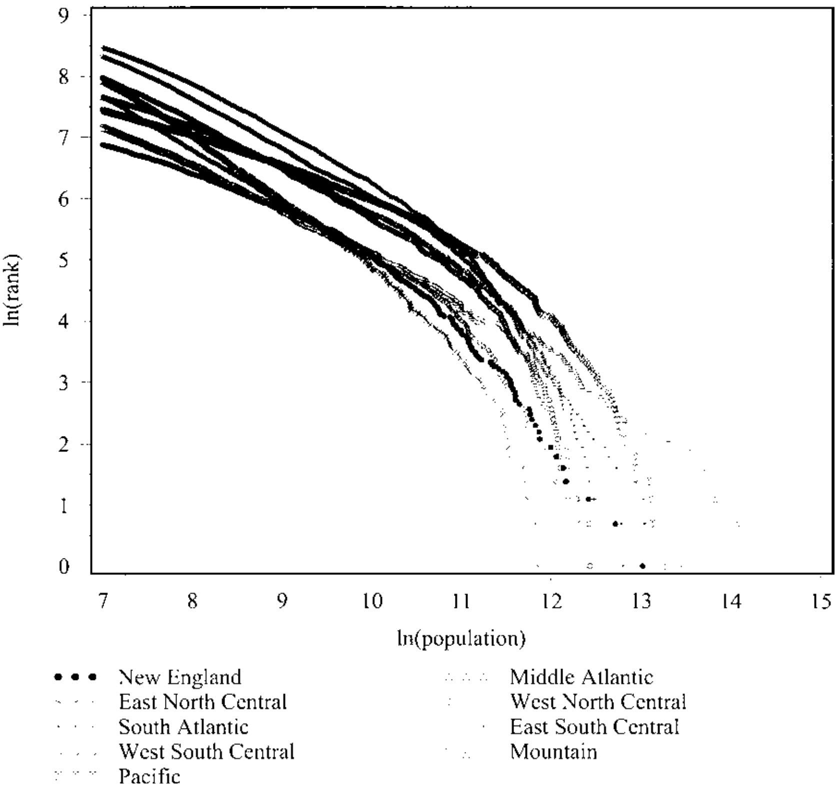

Fig. 3.6 Square-level Zipf plots for census divisions (square population at least 1,000)

each census division. Figure 3.6 contains Zipf plots for all nine divisions, and table 3.5 lists the estimates. To a remarkable degree, the pattern we have established for the country as a whole occurs in each division individually. Table 3.5 shows that the estimated location of the kink varies little across the divisions, roughly eleven for each. In figure 3.6, we see that the slope on the left side of the kink is approximately the same for each division. The plots look something like vertical shifts across the divisions. In all cases, the slope to the right of the kink is strictly greater than 1, and to left of the kink, the slope is less than 1 (with the exception that for the East South Central, the slope actually equals 1 to the left of the kink).

The kink at log population of 10.9 suggests we should explore this upper tail. This corresponds approximately to a population of 50,000. Now, truncate all squares with population less than 50,000. We are left with 1,182 squares, accounting for 48 percent of the population. Table 3.6 reports the

Table 3.6 **Six-by-six-square-level Zipf regression results (squares with population 50,000 and above)**

		OLS		Hill method
Sample of squares	*N*	Slope (absolute value)	R^2	Slope
All squares with population ≥ 50,000	1,182	1.889	.983	1.569
By Census division				
New England	58	1.865	.989	1.892
Middle Atlantic	154	1.318	.989	1.302
East North Central	193	1.929	.987	1.641
West North Central	74	2.389	.969	2.108
South Atlantic	218	2.271	.972	1.847
East South Central	44	2.763	.923	2.575
West South Central	138	2.286	.918	1.778
Mountain	85	1.951	.853	1.487
Pacific	218	1.597	.931	1.236
Mean across divisions	131.3	2.041	.948	1.763
By MSA (10 largest)				
Boston	26	1.462	.987	1.491
Chicago	54	1.412	.974	1.246
Dallas	35	2.208	.869	1.401
Detroit	35	1.718	.938	1.603
Houston	29	1.751	.894	1.469
Los Angeles	82	1.265	.870	0.986
New York	95	1.139	.981	1.173
Philadelphia	32	1.425	.982	1.612
San Francisco	43	1.451	.935	1.373
Washington	43	1.639	.955	1.336
Mean across top ten MSAs	47.4	1.547	.939	1.369
Mean across top twenty-five MSAs	29.2	1.776	.915	1.556

results of a linear Zipf's regression on this tail of the distribution. Taking the country as a whole, the slope is 1.889. Looking at each census division individually, the variation in the slope is relatively small, and the mean is 2.

We conclude this section by connecting our results from the square-level analysis to the previous section's results for the MSA-level analysis. The bottom of table 3.6 reports the results of Zipf regressions across squares *within* MSAs. For example, there are twenty-six squares with 50,000 people or more in the Boston MSA, and when we estimate the Zipf's regression on this sample, we get a slope of 1.46. The table reports the results of individual regressions for the top ten MSAs (by population), as well as the mean coefficients across these regressions for the top ten and top twenty-five MSAs. (We only do this for large MSAs, since small MSAs have few 50,000+ squares with which to run the regression.)

Recall from table 3.3 that in an *MSA-level* regression with the 135 top MSAs, when we use the maximum population square $s_i^{\max}$ as the size measure, we get a slope of 1.761. It is notable that when we take the MSA that is ranked 135 according to this measure, its value of $s_i^{\max}$ is 65,000, which approximately equals the 50,000 cutoff we are using here. The 1.761 slope approximately equals the slope of the *within-MSA, square-level* regressions we are doing here. The average slope across the top twenty-five MSAs is in fact 1.776.

The results here are interesting in two ways. First, there is an interesting fractal-like pattern among squares with 50,000 or more in population. Looking *within* a given MSA, the Zipf coefficient across squares is on the order of 1.7. This is approximately what we get when we take the maximum population square in each MSA and look *across* MSAs. It is also approximately what we get when we take all such squares across the whole country and look at them together (the 1.9 estimate in table 3.6), as well as when we look at squares in individual regions.

Second, this coefficient is also approximately the result we get when we do not use the squares and just use average MSA density (the 1.896 coefficient on density in table 3.3). We have raised concerns about the arbitrary way MSAs are defined, and there is certainly measurement error. Yet, our analysis in which MSA definitions play no role whatsoever (1.889 Zipf coefficient in table 3.6) is very close to our results in the MSA density analysis of table 3.3 (again, the 1.896 coefficient in table 3.3). Now, these are different objects that need not be the same, even if with perfect measurement. Yet, the suggestive fractal pattern here hints that they might very well be the same or very close if we did have perfect measurement. And even with the imperfect measurement of MSAs we have to work with, our analysis may not be very far off.

3.6 Growth Rates

The theoretical literature has emphasized the link between the size distribution of cities and their growth rates. In particular, Gabaix has shown a connection between Gibrat's law and Zipf's law. One version of Gibrat's law is that the mean and variance of the growth rate of a city are independent of the initial size of a city. Authors such as Ioannides and Overman (2003) have noted that Gibrat's law is a reasonable first-order approximation to the data. (See also Black and Henderson [2003] for an analysis.)

Table 3.7 shows that Gibrat's law is a reasonable first-order approximation for MSA growth in our data. The measure of growth rate used here is the difference in log population between 2000 and 1990. Mean growth over all MSAs during the period is 0.124. The mean growth varies relatively little over the four different MSA groupings in the table. It takes a low of 0.114 for cities with less than 250,000 people and has a peak of 0.141 for cities in

Table 3.7 Growth rates (change in log population), 1990 to 2000, by size (MSAs and squares)

	Number with positive population in 1990 and 2000	Change in log population	
		Mean	Standard deviation
MSAs	274	.124	.100
MSAs by 1990 population			
Population < 250,000	135	.114	.098
250,000 ≤ population < 500,000	66	.127	.093
500,000 ≤ population < 1,000,000	32	.141	.129
1,000,000 ≤ population	41	.139	.094
Squares	65,975	.081	.6186
Squares by 1990 population			
Population < 1,000	43,723	.054	.741
1,000 ≤ population < 2,000	8,057	.129	.228
2,000 ≤ population < 5,000	7,117	.139	.242
5,000 ≤ population < 10,000	2,953	.144	.223
10,000 ≤ population < 50,000	3,118	.149	.204
50,000 ≤ population < 100,000	616	.093	.128
100,000 ≤ population < 250,000	341	.056	.095
250,000 ≤ population < 500,000	39	.046	.071
500,000 ≤ population	11	.060	.061

the 0.5 to 1 million range. Moreover, the standard deviation does not vary much across the different groups.

Table 3.7 shows that Gibrat's law is not a good approximation for the growth of squares. The mean and variance of growth depend on size in a clear pattern. Mean growth in the smallest size category is 0.054—the lowest over all categories. Growth increases with size until it attains a maximum value of 0.149 for squares in the 10,000 to 50,000 range. Beyond this, mean growth decreases, falling to 0.093 in the 50,000 to 100,000 range and to around 0.05 beyond that. The standard deviation is not flat but decreases sharply with population.

These results for the growth rates of squares are not surprising, given what we know about the patterns of urban and rural growth. As is well known, remote rural areas have been declining in their share of population, so not surprisingly, mean growth is lowest in the smallest size category, under 1,000 people in the square. Also well understood is that in large urban areas, population expansions take place at the edges where new housing is constructed. For this reason, the most dense squares (those with more than 100,000 in 1990 population) have the lowest growth rate besides the under-1,000 category. These dense areas are already built up, and additional housing units are hard to squeeze in. Those squares that tend to be on the edge of metropolitan areas (in the range of 10,000 to 50,000 people) have the highest growth rate of 0.149.

It is also easy to see why the highest population squares have the lowest variance of growth. The absence of a large stock of vacant buildable land eliminates the possibility of upside growth, and the existence of a housing stock decreases the downside of population outflow (see Glaeser and Gyourko [2005]). It is easy to see why the smallest locations have the highest variance of growth. If the forest ranger living by himself or herself in a six-by-six square gets married, population in the square doubles.

3.7 Robustness

In setting our grid of squares, we had to determine: (a) what grid size to use (we picked six miles), and (b) where to start the grid. Let us begin by exploring this second decision, which is analogous to the decision of where to put the prime meridian for longitude, an arbitrary placement that by international convention passes through Greenwich. With the way we have placed the grid in figure 3.1, we can see that downtown Manhattan is in the same six-by-six square with Jersey City and other places across the river in New Jersey. If we had shifted the grid two miles to the east, downtown Manhattan would have been in a square with Queens.

One may wonder whether this arbitrary decision on our part impacts our results. Fortunately, the answer is no: where to start the grid has virtually no impact on our results. Table 3.8 shows what happens when we shift the grid two miles and four miles to the north. (Note that if we shift it north six miles, the grid remains the same.) Analogously, it shows what happens when we shift the grid two and four miles to the east. The top row contains the original baseline results. The rows below are the results with the shift and show that they are the same up to two-digit accuracy, and for some columns, up to three digits.

Next, we consider changing the size of the grid. Significant changes in the grid will impact the results. If we make the grid size 1,000 miles, there will be only three squares. If we make the grid one meter by one meter, then our first problem is the Census data are not fine enough for this size. Our second problem is that populations would typically be one if a person happened to be standing in the one-by-one-meter square at the time of the census and zero otherwise, so the size distribution would not be interesting.

Next, we focus on the robustness of our results to relatively small changes in the grid size. We consider two smaller grid sizes (two and four miles) and four larger ones (eight, ten, fifteen, and twenty miles). To a remarkable degree, our results are robust to these changes in grid size. Recall that in the original six-by-six analysis, we used a 1,000 population cutoff for the piecewise linear regression and a 50,000 cutoff in the linear regression. When we change the grid size, we also change the population cutoffs to keep population density at the cutoff the same. For example, the area of a two-by-two square is 1/9 times the area of a six-by-six square. So, for the two-by-two

Table 3.8 Robustness of results to alternative grids

	MSA-level regression on s_i^{max}		Square-level piecewise linear regression, population ≥ 1,000 per 6 × 6 square[a]				Square-level linear regression, population ≥ 50,000 per 6 × 6 square[a]	
	OLS slope	R^2	Kink	Slope1	Slope2	R^2	OLS slope	R^2
Baseline 6 × 6 grid	1.761	.988	10.89	.747	1.937	.998	1.889	.983
Shift of baseline grid								
2 miles north	1.790	.986	10.95	.751	1.984	.998	1.892	.980
4 miles north	1.838	.986	10.90	.750	1.923	.998	1.879	.981
2 miles east	1.715	.988	10.90	.745	1.957	.998	1.919	.987
4 miles east	1.774	.989	10.92	.747	1.979	.998	1.924	.983
Alternative grid size								
2 miles	1.981	.977	9.072	.680	2.097	.999	1.800	.968
4 miles	1.873	.992	10.262	.719	2.037	.999	1.886	.979
6 miles	1.761	.988	10.899	.747	1.937	.998	1.889	.983
8 miles	1.595	.981	11.433	.773	1.976	.998	1.959	.986
10 miles	1.483	.978	11.655	.786	1.819	.998	1.914	.987
15 miles	1.325	.979	12.328	.816	1.850	.998	1.959	.994
20 miles	1.246	.969	12.482	.822	1.630	.997	1.994	.983

[a]We adjust the population cutoffs for the squares to keep the population density the same across cutoffs for the different grid sizes. For example, in the two-by-two-square linear regression, we use all the squares with population sizes greater than or equal to 5,556 (= 50,000/9). The 50,000 comes from the base case of the six-by-six-mile square. The nine takes account of the fact that the area of a six-by-six square is nine times as large as a two-by-two square.

case, the linear regression cutoff is 5,556 = 50,000/9. The piecewise linear function fits extremely well throughout all the grid sizes (R^2 = 0.997 and above). The coefficient estimates do not vary much: 0.7 to 0.8 below the kink, and 1.8 to 2.0 above the kink. Moreover, the locations of the kink increase by the expected magnitude. For example, going from a two-by-two grid to a four-by-four grid increases the area by a factor of 4 (ln [4] = 1.39). If density at the kink stayed the same, then the kink should increase by 1.39 when moving from a two-by-two grid to a four-by-four grid. The actual increase of 1.19 = 10.26 – 9.07 is fairly close. We see an analogous pattern for the other grid sizes. We conclude that our results are not an artifact of an arbitrary choice of a six-mile grid length.

One notable pattern in table 3.8 is the decline of the MSA-level regression coefficient on s_i^{max} as the grid size is increased. As grid sizes increase, the squares begin to incorporate the entirety of the MSA. So, the population of the biggest square s_i^{max} begins to approximate the population of the MSA as a whole, and the coefficient gets close to 1 (Zipf's law), as it is in table 3.3.

One last issue concerns what is happening on the coasts with the squares. As can be seen in figure 3.1, some of the squares in the New York metro area are partly in the very dense island of Manhattan and partly in the water. Since the highest population density locations (New York, Chicago, etc.) tend to border bodies of water, one might wonder whether some systematic biases might be present. We think this is an interesting point but not one of much quantitative significance, because we are working with logs rather than levels. We make two distinct arguments. First, in these dense cities, the log population of the squares changes relatively slowly as we move away from the coasts (at least at a six-by-six grid size). The possibility of systematic biases at the coasts is not quantitatively a big problem, because many other squares nearby that are approximately equal in log population will average things out. Second, even at the coast, variations in density are not quantitatively significant. Suppose, for example, that a square at the coast is half in the water (ln [1/2] = –0.3). At the dense squares near or in Manhattan, log population is around 14. If we shifted such a square and put it half in the water, log population would fall to 13.7 = 14 – 0.3. This is a small difference compared to the vast differences in log population between squares close to Manhattan (regardless of whether in the water) and squares in less-dense places, such as upstate New York. Even if the square were 99 percent in the water, this would not matter either, because such a square at a six-by-six resolution would represent a negligible portion of the downtown area.

3.8 Conclusion

Our chapter studies the distribution of population across six-by-six-mile squares, examining the extent to which Zipf's law and Gibrat's law hold. The main results are as follows:

1. At the bottom tail of the distribution, the distribution is roughly lognormal, certainly not Zipf.
2. For squares above 1,000 in population, a Zipf's plot has a piecewise linear shape, with a kink at around a population of 50,000. Below the kink, the slope is 0.75; above the kink, it is around 2. The finding is robust across different regions in the country.
3. Gibrat's law does not hold with squares. Mean growth has an inverted U-shaped relationship with population size. The variance of growth declines with size.
4. The slope of 2 in the upper tail matches what we get with MSA-level data if we substitute population density for population in a Zipf's plot. This is consistent with the usual Zipf coefficient of 1 for the population regression if the land elasticity of population is 0.5. The slope of 2 also matches what we get if we use the maximum population square in the MSA instead of average density, as well as what we get in the upper tail when we look at squares *within* MSAs. All of this suggests some kind of fractal pattern in

the left tail in which the distribution of squares within MSAs looks like the distribution of MSAs across the country, which in turn looks like the distribution of squares across the country and within individual regions.

In our title, we put a question mark after "Zipf's Law." It is clear that the standard Zipf's law does not apply for squares in the upper tail, because the slope is around 2, not 1. Nevertheless, if we take the land elasticity of population to be 0.5 (which roughly fits the data for large MSAs), then a slope coefficient of 2 for squares (where the land margin is fixed) is consistent with a slope coefficient of 1 for regularly defined MSAs (where the land margin varies). In this sense, Zipf's law holds for squares in the right tail. But what about below the kink of a square population of 50,000? For relatively less-populated squares like these, an expansion of the population might not put much pressure on the land margin, as vacant rural land in the square can be converted to housing sites. If the land elasticity were zero, the coefficient on density in equation (5) would be the same as the coefficient on population in equation (2). In this extreme case, the relevant comparison is between the 0.75 slope for squares and the standard slope of 1, and Zipf's law does not hold. If the land elasticity is a little higher than zero, Zipf's law works better. Regardless of this matter, the fact that the Zipf's plot is straight as an arrow for population in the range between 1,000 and 50,000 is very intriguing. The presence of the kink is intriguing, as well.

We believe a joint analysis of the distribution of population of squares within and across metropolitan areas is a fruitful area for further research. We see opportunities for progress in theories that emphasize economic considerations and spatial factors, such as the work of Hsu (2008). In terms of directions for future empirical work, we believe it would be promising to examine the size distribution of squares in an international context.

References

Anas, A., R. Arnott, and K. A. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 36 (3): 1426–64.

Black, D., and V. Henderson. 2003. Urban evolution in the USA. *Journal of Economic Geography* 3 (4): 343–72.

Bryan, K. A., B. D. Minton, and P.-D. G. Sarte. 2007. The evolution of city population density in the United States. *Economic Quarterly* 93 (4): 341–60.

Burchfield, M., H. G. Overman, D. Puga, and M. A. Turner. 2006. Causes of sprawl: A portrait from space. *Quarterly Journal of Economics* 121 (2): 587–633.

Duranton, G., and H. Overman. 2005. Testing for localization using micro-geographic data. *Review of Economic Studies* 72 (4): 1077–106.

Duranton, G., and M. A. Turner. 2008. Urban growth and transportation. University of Toronto, Department of Economics. Manuscript, November.

Eeckhout, J. 2004. Gibrat's law for (all) cities. *American Economic Review* 94 (5): 1429–51.
Gabaix, X. 1999. Zipf's law for cities: An explanation. *Quarterly Journal of Economics* 114 (3): 739–67.
Gabaix, X., and Y. M. Ioannides. 2004. Evolution of city size distributions. In *Handbook of regional and urban economics,* vol. 4, ed. J. V. Henderson and J.-F. Thisse, 2341–78. Amsterdam: Elsevier.
Glaeser, E. L., and J. Gyourko. 2005. Urban decline and durable housing. *Journal of Political Economy* 113 (2): 345–75.
Holmes, T. J., and S. Lee. 2008. Economies of density versus natural advantage: Crop choice on the back forty. University of Minnesota, Department of Economics. Manuscript, October.
Hsu, W.-T. 2008. Central place theory and Zipf's law. University of Minnesota, Department of Economics. Manuscript, January.
Ioannides, Y. M., and H. G. Overman. 2003. Zipf's law for cities: An empirical examination. *Regional Science and Urban Economics* 33 (2): 127–37.
Linklater, A. 2003. *Measuring America: How the United States was shaped by the greatest land sale in history.* New York: Plume.
Luttmer, E. G. J. 2007. Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics* 122 (3): 1103–44.
McMillen, D. P., and J. F. McDonald. 1998. Suburban subcenters and employment density in metropolitan Chicago. *Journal of Urban Economics* 43 (2): 157–80.
Nordhaus, W., Q. Azam, D. Corderi, K. Hood, N. M. Victor, M. Mohammed, A. Miltner, and J. Weiss. 2006. The G-Econ Database on Gridded Output: Methods and data. Yale University, Geographically Based Economic Data (G-Econ). Available at: http://gecon.yale.edu.
Rossi-Hansberg, E., and M. L. J. Wright. 2007. Urban structure and growth. *Review of Economic Studies* 74 (2): 597–624.