Change in the Size and Distribution of Large Urban Areas in the U.S., 1950-2020

John R. Ottensmann
Indiana University-Purdue University Indianapolis
john.ottensmann@gmail.com
urbanpatternsblog.wordpress.com
March 2023

Abstract

A dataset with the numbers of housing units for the 56 largest urban areas in 2020, going back to 1950, is used to examine the changes in their sizes. While all of the areas more than doubled in size, the growth of some was astounding, adding over 99 percent of their 2020 housing units since 1950. The differences in rates of growth produced significant changes to the list of the largest 15 areas, with six new areas added compared only one change in the previous four decades. The areas in the South grew most rapidly. In 1950, the large urban areas in that region accounted for by far the smallest share of all of the housing units in the large urban areas but had the largest share in 2020. In the final decade, nearly half of all growth in housing units for the large urban areas was in the South. The distribution of the sizes of urban areas is often thought to conform to Zipf's law, postulating that the size and rank are related by a power law with an exponent of one. In 1950 the exponent was slightly greater than one. The exponent dropped by 2020 such that the the smaller urban areas were larger than expected.

Introduction

Perhaps the most fundamental characteristic of an urban area is its size. The New York and Orlando urban areas differ in many ways, but the most obvious difference is size. The New York area is nearly seven times the size of Orlando. The growth of an urban area determines its size, of course. But the time period when an urban area has grown and how rapidly also shapes the character of the area. Over 98 percent of the housing units in the Orlando area have been added since 1950. Only 60 percent of the units in New York were added during the same period, with about 40 percent being older. And urban areas vary in size in a rather regular fashion. Only a small number of urban areas are among the very largest, while more areas have somewhat smaller sizes.

Many studies have looked at the growth and size over urban areas over numbers of decades. These necessarily confront the issue of the areas used for determining the sizes. One option is to use the populations of cities, the legally incorporated jurisdictions. For studies extending back to the eighteenth and nineteenth centuries (e.g., Lampard 1968; Kim 2000), this is the only data available from the Census. But using cities has several problems. Cites include only a portion of the urban areas, especially in the twentieth and twenty-first centuries. Urban areas vary widely in the proportion included in the major city. And city boundaries change over time.

Since 1950 the Census has used Urbanized Areas and Metropolitan Statistical Areas (MSAs, with varying names) to delineate in different ways and provide statistics for the more inclusive areas. A difficulty here is that these definitions have changed over time, making units at one census not strictly comparable with units from another. Most studies of urban size and change have used the metropolitan areas. Many, including Kim (2000) and Overman and Ioannides (2001) (for a portion of the period studied) used the MSAs as defined for each census. Others (e.g., Frey 1993) have taken advantage of the fact that MSAs are defined using entire counties and have assembled data for each year for those counties included in the MSA in the final year. Using data for these constant areas avoids the issue of changing definitions. However, this raises the question of whether these areas are overbounding the urban area in the earlier years before the urban area grew to its current size. This becomes more of an issue when using the method many decades earlier. Black and Henderson (2003) used such constant areas equivalent to the 1990 MSAs going back to 1900 but then addressed the issue of growth over time by using the urban population as reported by the Census rather than the total population. This brings in the problems of changing city boundaries, changing definitions of urban, and the fact that some of the urban population will not be associated with the major urban area. Frey and Speare (1988) used both constant boundaries and the metropolitan areas as defined for each census from 1950 to 1980 but chose to shift some counties in New York and New Jersey around for 1980. And this for a Census Monograph, no less!

Beginning in 1983, subdivisions of some of the largest MSAs were made. Confusingly, the subdivided MSAs were then called Consolidated Metropolitan Statistical Areas (CMSAs), even though they were not formed by consolidating anything. The subdivisions were then called Primary Metropolitan Statistical Areas (PMSAs), though these were not first, having been established by splitting the CMSAs. Some have used the PMSAs in analyzing change over time (Mills and Lubeuele 1995; Glaeser and Shapiro 2003). I would simply comment that I have difficulty considering the suburban counties of Nassau and Suffolk on Long Island (a PMSA) as constituting a metropolitan area.

The Census significantly changed the criteria for delineating Urbanized Areas for the 2000 census. Far more significant changes were made for the 2020 census, including replacing Urbanized Areas with Urban Areas, which now refers to both the larger and smaller urban agglomerations (U.S. Census Bureau 2002, 2022). Major changes were also made to the MSA definition starting in 2003 (U.S. Office of Management and

Budget 2000). As a result, making comparisons over time including these more recent censuses becomes even more problematic.

The issues of urban area size, growth, and distribution are addressed in this paper using urban areas defined in a consistent manner for each census from 1950 to 2020. The areas have been delineated using data on housing unit densities for census tracts, paralleling the new Census criteria for Urban Areas for 2020. Housing units are then used as the measure of the size of an urban area due to data availability in the earlier years and because they better represent the physical pattern of urban development.

The next section describes the *Urban patterns 2* dataset for the 56 largest urban areas in 2020 covering the period from 1950 to 2020 that is used for the analysis. The next section begins the analysis, focusing on the size and growth of these urban areas. The large differences by region motivate the following analysis of the total numbers of housing units in the large urban areas in each of the regions over the period. Finally, it has been observed over and over that the distribution of urban area sizes tends to follow a patterns often called the rank-size rule or Zipf's law, and this is examined using these data.

The *Urban patterns 2* data

The *Urban patterns* 2 dataset includes housing unit counts for census tracts from 1950 to 2020 that have been used to delineate 56 large urban areas in the United States for each census year. Data for 2010 and 2020 are from the Census and from the National Historical Geographic Information System (Manson, *et al.* 2022). Data from the censuses from 1970 to 2000 are from a unique dataset from the Urban Institute and Geolytics (2003) with the data normalized to 2000 census tract boundaries. Housing units for 1950 and 1960 are estimated from the data on housing units by year built from later years, taking the numbers built before 1950 and 1960 as the estimate of the numbers present at those times. This will include error because of changes to the housing stock over time, especially the loss of some units, but analyses suggest that the estimates are reasonable for two decades back in time. Census tract boundaries for 2020 are used for the dataset, with estimates from data for earlier year tract boundaries transformed using the census tract relationship files. Detailed documentation of the dataset and listings of all data sources are included in Ottensmann (2023).

Urban areas consist of contiguous census tracts that meet urban criteria. Some large areas of continuous urban tracts include what should reasonably be considered two or more urban areas. Areas in the northeastern United States are a major example. To distinguish separate urban areas, the Combined Statistical Areas (CSAs) are used (and MSAs that are not included in a CSA). CSAs are used rather than the more commonly used MSAs as it is believed that they better represent the full extent of urban areas. The CSAs are only used to identify the urban areas, such as Philadelphia, New York, and Hartford. The boundaries are established at the locations where the urban areas have become contiguous as they expanded. The urban areas included in the dataset are the 56 areas containing more than 300,000 housing units in 2020.

The criteria defining the urban areas are as close as possible to those being used for delineating the 2020 census Urban Areas, which include what were formerly called Urbanized Areas (U.S. Census Bureau 2022). A census tract can be considered urban and be added to the urban area if it has a housing unit density greater than 200 housing units per square mile. To include urban territory that is nonresidential, a tract will also be included if over one-third of its area has impervious surface of 20 percent or more. An additional condition is that a tract could only be considered urban if it had been designated as urban for the following census year to provide a pattern of cumulative urban development. This direction is chosen rather than the reverse (if urban, then urban later) because the more recent data are considered to be more accurate.

Some of the urban areas include two or more areas that were originally separate but that have since growth together. Areas that are sufficiently large are considered to be urban centers and are included in an urban areas. Dallas and Fort Worth is an example. As the areas become contiguous, tracts are assigned to the center growing more rapidly toward the other and to provide more continuous, less irregular boundaries. Areas are considered separate urban centers and are included in the urban areas if the number of housing units in 2020 exceeds 16 percent of the total units in the urban area. This cutoff was established by identifying as candidates any separate area deemed large enough to potentially be considered an urban center and then setting the threshold. The smallest areas in relation to the total size of the urban area included are Providence, with Boston; Tacoma, with Seattle; and High Point, with Greensboro and Winston-Salem. Next highest, at 11 percent are Port Charlotte in the Sarasota-Bradenton area and Winter Haven in the Orlando area. The names given to the urban areas include the names of the additional urban centers that have been added.

The size and growth of large urban areas

This section looks at the sizes and changes in size of the 56 large urban areas from 1950 to 2020. As discussed, numbers of housing units are the measure of size. It concludes examining the relationship of change to size with a test of what is called Gibrat's law.

Lists of the very largest urban areas are always of interest. Table 1 presents the lists of the 15 largest urban areas in 1950 and 2020. Twelve of the 15 largest areas in 1950 are in the Northeast and Midwest. (This assumes one counts Washington-Baltimore in the Northeast, which is reasonable in this context as it is part of the string of large urban areas in the Northeast extending up to Boston, sometimes referred to as the Northeast

Corridor or megalopolis.) The other three areas are in the West—Los Angeles, San Francisco-Oakland-San Jose, and Seattle-Tacoma.

Numbers of authors (e.g., Lampard 1968; Monkkonen 1990) have commented on the stability of the rankings of cities in the United States over time. Looking back from 1950, the Census first identified urban areas consisting of large cities and surrounding territory in 1910, called Metropolitan Districts (U.S. Census Bureau 1913). This list of the 15 largest Metropolitan Districts in 1910 by population (the Census had not yet started collecting data on housing units) is remarkably similar. Fourteen of the areas on that list are also on the list of the 15 largest urban areas in 1950. The one change had Cincinnati dropping off the list and Seattle-Tacoma being added. And most of the rankings within the list did not change that much. Nine of the 14 on both lists either did not change their

Table 1. Largest urban areas in terms of housing units in 1950 and 2020.

Largest urban areas in 1950

Urban area	Housing units 1950
New York	3,353,121
Chicago	1,242,271
Los Angeles	1,076,921
Philadelphia	793,855
Boston-Providence	779,201
Detroit	618,033
Washington-Baltimore	596,615
San Francisco-Oakland-San Jose	507,310
Cleveland-Akron	428,936
St Louis	316,665
Pittsburgh	310,238
Minneapolis-St Paul	234,307
Seattle-Tacoma	214,488
Buffalo	205,966
Milwaukee	205,103

Largest urban areas in 2020

Urban area	Housing units 2020
New York	8,286,942
Los Angeles	5,620,280
Chicago	3,667,811
Washington-Baltimore	3,084,652
Boston-Providence	2,871,820
Miami-Fort Lauderdale-West Palm Beach	2,794,627
Dallas-Fort Worth	2,558,011
San Francisco-Oakland-San Jose	2,474,088
Houston	2,453,609
Philadelphia	2,432,138
Atlanta	2,062,894
Detroit	1,815,429
Phoenix	1,788,759
Seattle-Tacoma	1,616,560
Tampa-St Petersburg	1,366,665

rank at all or moved up or down by a single position in the rankings. The greatest change in rank was for Los Angeles, up 11 positions from 14th to third. The next larger changes were for Detroit, up 6, and Pittsburgh, down 6.

That stability among the largest urban areas did not persist over the next seven decades. The list of the largest urban areas in 2020 shows a significant number of changes since 1950. Nine of the largest urban areas in 1950 remained on the 2020 list. Ranks changed by one position or less for six of the nine. The six areas that dropped off the list are in the area territory from Buffalo to St. Louis and Minneapolis-St. Paul. The additions include Phoenix plus five urban areas in the South, two in Florida, two in Texas, and Atlanta. This is in contrast to the set of the largest areas earlier, which include none in the South.

To show how the sizes of these areas changed over time, Figure 1 plots the lines showing the number of housing units at each census from 1950 to 2020 for each of the urban areas on the lists of the 15 largest for either 1950 or 2020. Note that because of the widely varying sizes of the areas, which include some areas making the list in 2020 that

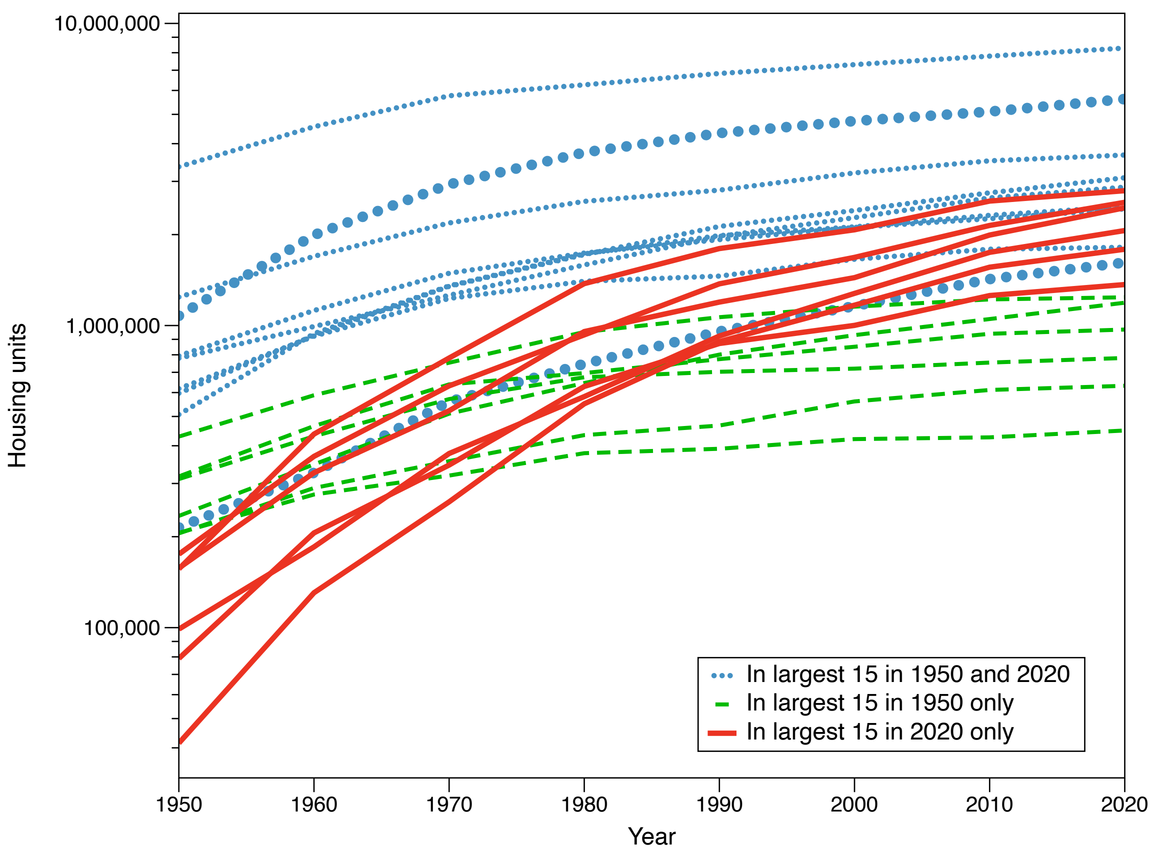

Figure 1. Number of housing units from 1950 to 2020 for 15 largest urban areas in either 1950 or 2020.

were very much smaller in 1950, the number of housing units are plotted using a logarithmic axis, with the major divisions at 100,000, 1 million, and 10 million housing units. Three groups of areas are plotted with distinct colors and line styles. The areas that were on the list in both years are in blue using a dotted line. The areas that were on the list in 1950 but then dropped off are in green using a dashed line. And a solid red line is used for the areas that were among the largest in 2020 but not 1950.

Not surprisingly, the dropouts with the green dashed line were among the areas growing more slowly, with some showing very limited increases in the later decades. These areas started out as six of the seven smallest areas on the list in 1950 and their limited growth was not sufficient to keep them on the list. The areas making the list in both years, displayed with the blue dotted lines, showed differing growth trajectories. Most common was a moderate level of increase sufficient to keep them on the list at about the same position. Two areas, Philadelphia and Detroit, grew more slowly and dropped six positions on the list but managed to remain among to 15 largest because they began with much larger numbers of housing units. And Los Angeles and Seattle-Tacoma, with the heavier blue dotted lines grew more rapidly, more like the new additions in red. This increased Los Angeles' rank to second, surpassing Chicago by a large amount. The growth of Seattle-Tacoma allowed it to remain among the 15 largest in spite of being close to the bottom of the list in 1950. Finally, the red lines show the very rapid growth among the six areas on the 2020 list as they exceeded not only those that dropped off but a number of areas that remained as well.

Widening the picture, Table 2 presents the basic statistics of the mean, minimum, and maximum number of housing units for the 56 urban areas in each of the census years from 1950 to 2020. On average the urban areas got a lot larger, with the mean number of housing units increasing fivefold from about 250,000 to 1,250,000. The largest area, New York, grew from something over three million to over eight million, becoming two and one-half times as large.

But the sizes of the smaller urban areas and their growth were the most striking. In 1950, the smallest urban area, Cape Coral-Fort Myers-Naples, had fewer than 3,000 housing units, the size of a fairly small town. Las Vegas and Sarasota-Bradenton were two other areas below 10,000, and Orlando was had fewer than 20,000 housing units. With urban areas starting out this small and being among the 56 largest areas by 2020, very great growth in size was required. The smallest urban areas in 2020 were Birmingham, Rochester, and El Paso, all with just over 300,000 housing units.

Because most people are more used to thinking about the sizes of urban areas in terms of population, the final column of Table 2 presents the mean populations from 1970 to 2020. (Census tract data for each census from the Neighborhood Change Database goes back to 1970. Housing units for the two earlier years were estimated using the year-built data and no tract population data are available.) From 1970 to 2010,

Table 2. Mean, minimum, and maximum number of housing units and mean population for urban areas from 1950 to 2020.

Year	Housing units			Mean population
	Mean	Minimum	Maximum	
1950	251,626	2,720	3,353,121	—
1960	386,665	9,378	4,549,968	—
1970	543,533	29,805	5,760,634	1,605,349
1980	701,843	98,176	6,262,219	1,802,544
1990	842,859	139,444	6,828,490	2,077,173
2000	971,580	187,883	7,306,187	2,438,235
2010	1,137,993	245,732	7,794,399	2,755,264
2020	1,267,376	302,234	8,286,942	3,101,683

the average size of the urban areas in the dataset just about doubled from a little over one and one-half million to over three million.

Table 3 presents the mean, minimum, and maximum for the percent rate of change of housing units in each decade. The mean percent change from 1950 to 1960 was very high, 86 percent, nearly a doubling of the number of housing units. The urban areas continued to grow over the period but at steadily decreasing rates. By the most recent decade, the average percent rate of growth had dropped to 14 percent.

The urban areas did grow in each decade with one exception. The minimum growth rate of -2.6 percent in the period from 2000 to 2010 was for the New Orleans area, obviously a result of Hurricane Katrina. The lowest growth rates dropped from a substantial 27 percent in 1950 to 1960 to minimal growth of two percent or less in the more recent decades.

The differences in the rates of growth for the slowest and fastest growing areas were very large, with the maximum percent growth rate at least 12 or more times as large as the minimum in each decade, often far more. Especially in the earlier decades, the maximum rate of growth was especially striking, indicating a tripling or even quadrupling of the number of housing units in the fastest growing area. The 330 percent maximum increase from 1950 to 1960 was for Las Vegas, which increased from 6,222 to 26,768 housing units. Also growing by more than 200 percent during that decade were Sarasota-Bradenton, Cape Coral-Fort Myers-Naples, and Phoenix. These four areas remained the most rapidly growing growing through 1980.

Table 3. Mean, minimum, and maximum percent change in housing units by decade from 1950 to 2020.

Period of change	Percent change in housing units		
	Mean	Minimum	Maximum
1950-1960	86.4	27.4	330.2
1960-1970	58.6	15.5	217.8
1970-1980	48.5	8.7	229.4
1980-1990	28.7	3.4	117.5
1990-2000	21.5	2.1	106.6
2000-2010	22.5	-2.6	78.0
2010-2020	13.9	1.5	45.2
1950-2020	1,614.6	118.1	21,388.8

It is instructive to consider lists of the urban areas that grew most rapidly and most slowly over the entire period from 1950 to 2020. As is obvious from the very small sizes of some of the urban areas at the beginning, some areas necessarily experienced astounding rates of growth to place them among the 56 largest urban areas in 2020. This makes using the standard percent rate of change—change divided by the number of housing units at the start of the period—unworkable. The maximum rate of change from 1950 to 2020 was literally 21,389 percent for Cape Coral-Fort Myers-Naples, followed by 13,859 percent for Las Vegas. With values this extreme, it is not surprising that the mean change was an similarly outsized 1,625 percent. These values are not particularly understandable or useful. For this reason, an alternative measure of change in relation to size is employed. This is the change over the period from 1950 to 2020 as a percent of the number of housing units at the end of the period, in 2020. This is readily interpretable as the percent of all of the housing units in 2020 that were added to the urban area since 1950.

Table 4 gives the lists of the areas in which the percent of the 2020 housing units added during the period 1950 to 2020 were the largest and smallest. At the top of the list with over 99 percent were, again, Cape Coral-Fort Myers-Naples and Las Vegas. These are truly very new urban areas in which less than 1 percent of the housing units date from before 1950. And while these were the highest, all of the 15 fastest growing areas from 1950 to 2020 had about 93 percent or more of their housing units added during that period. These urban areas are located in the South and West regions, in the area often referred to as the Sunbelt.

Table 4. Urban areas with largest and smallest housing unit change from 1950 to 2020 as percent of housing units in 2020.

Largest 1950-2020 change as percent of 2020

Urban area	Percent
Cape Coral-Fort Myers-Naples	99.5
Las Vegas	99.3
Orlando	98.4
Sarasota-Bradenton	98.2
Phoenix	97.7
Charlotte	96.9
Austin	96.6
Raleigh-Durham	95.6
Atlanta	95.2
Miami-Fort Lauderdale-West Palm Beach	94.4
Tampa-St Petersburg	94.2
Houston	93.6
Dallas-Fort Worth	93.2
Tucson	92.9
Sacramento	92.9

Smallest 1950-2020 change as percent of 2020

Urban area	Percent
Buffalo	54.2
New York	59.5
Pittsburgh	60.3
Rochester	62.6
Cleveland-Akron	65.4
New Orleans	65.8
Detroit	66.0
Chicago	66.1
St Louis	67.3
Philadelphia	67.4
Milwaukee	67.5
Cincinnati	71.4
Boston-Providence	72.9
Birmingham	72.9
Louisville	75.2

All of the urban areas experienced substantial growth over the 70-year period. Buffalo had the smallest change as a percent of the 2020 housing units but that was still over 50 percent. Over half of the urban area is post-1950. The number of housing units more than doubled. And by the fifteenth slowest growing area, Louisville, three quarters of the housing units were added during the period. The slow growing areas are located predominantly in the Northeast and Midwest. The three areas in the South—New Orleans, Louisville, and Birmingham—are all among the older urban areas in the region.

Comparing these lists to the lists of the largest urban areas in 1950 and 2020: All six of the urban areas that were new to the list in 2020 are among the fastest growing areas from 1950 to 2020. Five of the six areas that dropped off the list from 1950 to 2020 are on the list of the areas growing more slowly. Among the nine areas continuing on the lists of the largest, five are among the areas growing slowly as well. All of these are located in the Northeast and Midwest.

A final question to be addressed in this section is whether the rates of growth of their urban areas are related to their sizes. The idea that the distribution of growth rates do not vary with size is referred to in economics as Gibrat's law of proportionate effect (Rosen and Resnick 1980; Gabaix and Ioannides 2004; Berry and Okulicz-Koraryn 2011). This is relevant not only to the general question of how urban areas increase in size over time but for the consideration of the rank-size rule or Zipf's law, discussed below.

A simple test of whether the rate of growth varies with the size of the urban area and whether Gibrat's law holds is to regress the log of the size of the urban area at one time on the log of the size at the earlier time, the lagged log of the size (Berry and Okulicz-Koraryn 2011). If the estimated regression coefficient is 1.0, Gibrat's law holds and the rate of growth is independent of the size of the area.

Table 5 presents the results of regressions of the log of housing units in each year on the value for the preceding year, the test of the extent to which the rate of growth is independent of the size of the urban area. The regression coefficients are all less than one, which means that there is at least some tendency for rates of growth to be greater for smaller urban areas. They increase over time, becoming progressively closer to one in the later decades. In addition, the fit of the model as measured by R^2 also increases over the 70-year period.

The standard errors are all quite small. Most of the regression coefficients are statistically significantly different from one at the 0.05 level. This is not the case for the

Table 5. Regressions of log of number of housing units on log of lagged housing units in preceding decade, 1950-2020 (test of Gibrat's law).

Period	Regression coefficient on log of lagged housing units	Standard error	R^2
1960 on 1950	0.862	0.021	0.968
1970 on 1960	0.897	0.017	0.981
1980 on 1970	0.880	0.023	0.966
1990 on 1980	0.932	0.019	0.978
2000 on 1990	0.945	0.019	0.978
2010 on 2000	0.960	0.020	0.978
2020 on 2010	0.975	0.014	0.990

coefficient of 0.975 for the final decade from 2010 to 2020 where the null hypothesis that the coefficient is one cannot be rejected. The coefficient for the previous decade is right on the borderline of statistical significance.

An alternative way of looking at the relationship between the rate of growth and the size of the urban areas is to simply calculate the correlation coefficients between housing units at the beginning of each decade with the percent rate of growth for the period. These are all negative, again showing a tendency for smaller areas to grow more rapidly. The correlations for the first four decades range from -0.27 to -0.32 and are all statistically significant at the 0.05 level. The correlations for the final three decades are slightly lower, from -0.23 to -0.26 and are not statistically significant. So these results parallel those from the regressions in Table 5: Some modest tendency for the rate of growth to be negatively related to size in the earlier years, decreasing to a lack of any statistically significant relationship by the end of the period.

The pattern may be explained by the nature of the sample of urban areas and the length of the period considered. Gibrat's law has been found to hold especially in the upper-tail of the distribution of city sizes. The set of urban areas considered here are the largest areas in 2020, so they represent the upper tail in that year. But has been shown, many of these urban areas were much smaller in 1950, far from being in the upper tail of the distribution in the earlier years. And they were experiencing very rapid growth that eventually led to their being among the largest areas in 2020. This could by why a somewhat stronger negative relationship between rate of growth and size is observed in the earlier periods.

Large urban area growth by region

The previous results make it very clear that there are major differences by region in the sizes and growth of the large urban areas. These differences are addressed in this section. Rather than focusing on the sizes of the individual urban areas as in the last section, an alternative approach is taken here. The analysis compares the total numbers of housing units in all of the large urban areas in each of the regions. This is done because of the very different sets of the large urban areas in the regions. The Northeast has only eight large urban areas having an average of 2.3 million housing units in 2020 with the median being almost the same. At the other extreme, the 24 large urban areas in the South have a mean of 900,000 housing units with a median of less than 600,000. The region totals give the overall magnitudes of the large urban areas within the regions that are not affected by such differences. The regions are the four census regions—Northeast, Midwest, South, and West—with the exception discussed above that the Washington-Baltimore urban area is included in the Northeast since it is part of the larger urban expanse often referred to as the Northeast Corridor.

Starting with the total number of housing units in the large urban areas in Table 6, the tremendous growth from 1950 to 2020 is obvious. The differences among the regions are as well. In 1950, the urban areas in the Northeast had over six million housing units while those in the South had fewer than two million, with the West not that much higher. Numbers were much larger in 2020, 18 million, nearly three times as large in the Northeast. But this was eclipsed by the urban areas in the South with nearly 22 million housing units, greater than the Northeast and highest among all regions. Starting from a much lower number of units in 1950, the West urban areas had grown to include nearly as many as the Northeast. Meanwhile, large urban areas in the Midwest started with nearly 4 million units, solidly in second place among the regions. By growing to only something over 12 million by 2020, the Midwest had become by far the smallest of the four regions with respect to the number of housing units in the large urban areas.

Another way to look at these sizes over time is to examine the shares of total large urban area housing units at each census within the urban areas of the four regions. These shares are displayed in Figure 2, showing the trends in shares over time for the regions. Starting in 1950, the gap between the share in the Northeast, 44 percent, and the other regions is apparent. Likewise, the areas in the South came in last with only a 12 percent share.

Looking at the trends from 1950 to 2020, the shares of large urban area housing units in both the Northeast and Midwest declined steadily. Percent share in the

Table 6. Total number of housing units in the large urban areas in each region, 1950-2020.

Year	Total number of housing units in all large urban areas			
	Northeast	Midwest	South	West
1950	6,242,547	3,721,570	1,714,598	2,412,323
1960	8,574,812	5,291,187	3,311,269	4,475,972
1970	11,136,383	7,073,091	5,395,725	6,832,640
1980	12,843,681	8,433,138	8,421,518	9,604,846
1990	14,515,782	9,396,045	11,438,270	11,850,017
2000	15,838,619	10,702,566	14,047,257	13,820,026
2010	17,274,774	11,888,873	18,401,895	16,162,081
2020	18,585,077	12,619,190	21,797,871	17,970,934

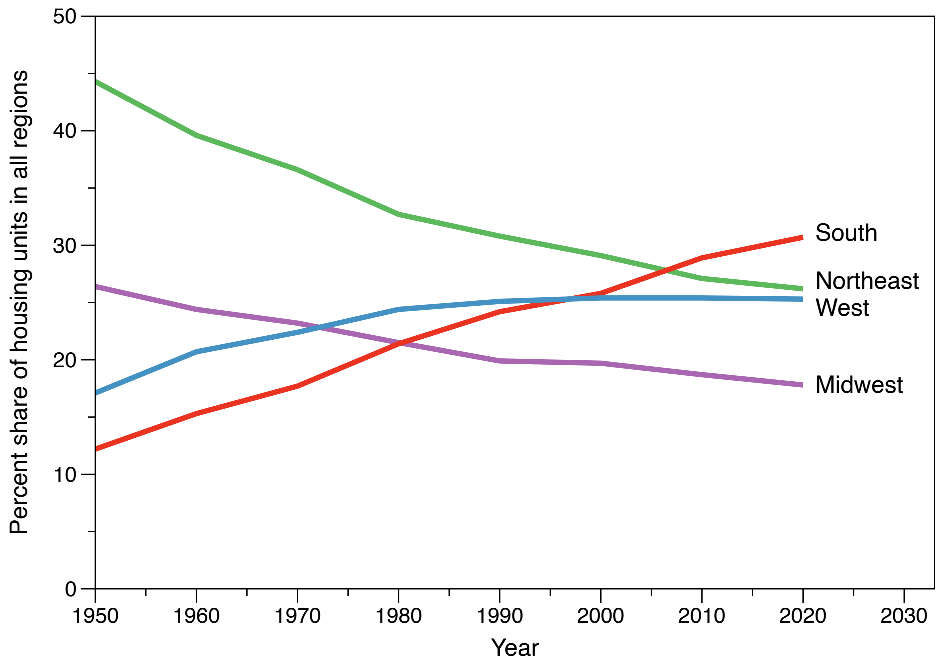

Figure 2. Percent share of housing units in the large urban areas in each region, 1950-2020.

Northeast went from 44 to 26 and the Midwest fell from 26 to 18 percent. The West's share increased during the first half of the period and then remained remarkably constant at 25 percent. The share of total housing units in the areas in the South shot up to 31 percent in 2020, with that region coming to have the greatest proportion of housing units among the four regions.

Table 7 gives the rates of change of housing units in each region for the seven decades from 1950 to 2020. Percent change generally declines dramatically over time in each region with a few reversals. The rate of change was greater in the South than for any other region in each decade. The West was always second, close to the South in the earlier periods, lower later. The totals for large urban areas in the Northeast and Midwest grew more slowly, with the Northeast lowest in most decades. Percent increase in these regions was less than half the percent increase in the South except for the Midwest in two decades This more rapid growth in the South and to a lesser extent in the West resulted in the greater sizes for these areas by 2020 and for their larger shares of total housing units over time. Housing units increased in the Northeast and Midwest but at a much slower rate, resulting in the declines in share.

One again, another view of differences in growth by region comes from looking at the region shares of the total growth in the large urban areas during each decade. These are illustrated in Figure 3. Growth during the decade from 1950 to 1960 was

Table 7. Percent change in housing units in the large urban areas in each region by decade, 1950-2020.

Year	Percent change in total housing units in all large urban areas			
	Northeast	Midwest	South	West
1950-1960	37.4	42.2	93.1	85.5
1960-1970	29.9	33.7	63.0	52.7
1970-1980	15.3	19.2	56.1	40.6
1980-1990	13.0	11.4	35.8	23.4
1990-2000	9.1	13.9	22.8	16.6
2000-2010	9.1	11.1	31.0	16.9
2010-2020	7.6	6.1	18.5	11.2

relatively evenly distributed among the regions, with shares ranging from 21 to 31 percent. Though growing at the slowest rate, the urban areas in the Northeast still

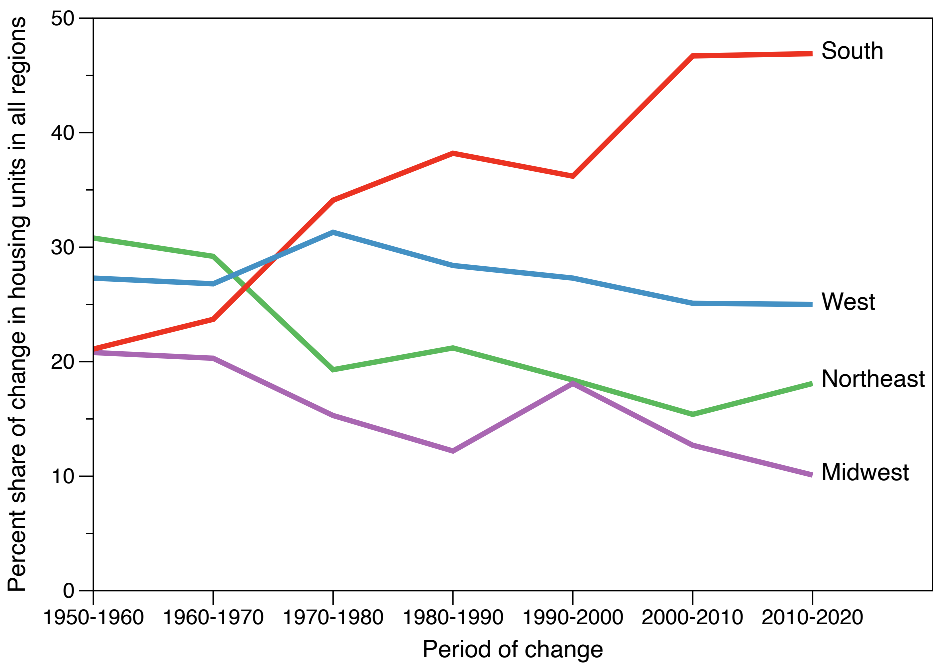

Figure 3. Percent share of total change in housing units in each region by decade, 1950-2020.

received the greatest share of the increase during that decade as this region began with a far larger number of housing units. The West received nearly as large a share, 27 percent, resulting from a very high rates of growth for a region that was not the smallest. The Midwest and South had the lowest shares of 21 percent for opposite reasons. The Midwest had a low rate of growth but a fairly large number of units, while the very high rate of growth in the South came on a base that was very small.

Urban areas in each region showed fairly consistent trends over time, though with some ups and downs. The South's urban areas increased their share of all of the growth in the large urban areas from 21 percent for the first decade to 47 percent for both decades from 2000 to 2020. In other words, nearly half of all of the increase in housing units during this latter period was in those urban areas. The West garnered a fairly consistent share of about a quarter of large urban area growth over the entire period, with shares between 25 and 28 percent. Progressively less of all growth was taking place in the Northeast and Midwest. In the final decade, the Northeast got 18 percent of the growth, compared to 31 percent in 1950. The Midwest was down to just 10 percent.

Distribution of urban areas by size: the rank-size rule (Zipf's law)

The rank-size rule, also called Zipf's law, has long been investigated as a description of the distribution the the sizes of urban areas. Auerbach in 1913 first recognized that city populations were related to their ranks by a power law (Carroll 1982). Zipf (1949) proposed a more specific form for the relationship. Major reviews of the work relating to this topic have been done by Carroll (1982), Cheshire (1999) and Gabaix and Ioannides (2004) and will not be reprised here. This section examines the extent to which the large urban areas conform to the rank-size rule, comparing these results to sets of areas defined in other ways.

According to Zipf (1947) the population of a city times its rank should be a constant value for all cities. This implies that the second-largest city would have a population one-half that of the largest city, the population of the third city would be one-third, and so forth. This can be expressed as

$$S_i R_i = c \quad or \quad S_i = c R_i^{-1} \tag{1}$$

where S_i is the size of city or urban area i, R_i is the rank of that area, and c is the constant, which is necessarily the population of the largest city if the proposition is correct. Because this will never hold exactly, the relationship is generalized, substituting a variable for the exponent

$$S_i = c R_i^{-a} \tag{2}$$

where a and c are parameters to be estimated using data on size and rank for a set of urban areas. The usual approach is to take the log of both sides of the equation giving

$$\ln\left(S_i\right) = \ln(c) - a \ln\left(R_i\right) \tag{3}$$

which is linear in the logs and can be estimated using Ordinary Least Squares to determine values for a and $\ln(c)$. The distribution of urban area sizes can then be considered to follow the Zipf's law to the extent the estimated value of the exponent a is close to one.

At least one author has made the distinction between the restrictive version of the rank-rule with the exponent having a value of one and a general rank-size rule with the exponent taking on any value (Carroll 1982). In this paper, the term rank-size rule will be used to refer to any distribution generally conforming the the relationships in equations (2) and (3) in which the exponent can have any value. The term Zipf's law will be restricted to those cases which at least approximately follow the equations in (1), that is, where the exponent is approximately one. In addition to the rank-size rule and Zipf's law, the relationship between urban area size and rank is often called a Pareto distribution (e.g., Rosen and Resnick 1980) and sometimes a power law (e.g., Krugman 1996; Gabaix and Ioannides 2004). Early research estimated what is sometimes called the Zipf form of the relationship as stated above, with urban area size a function of rank raised to a power. Later work, especially referring to the Pareto distribution, estimated the reverse formulation with rank as a function of size to a power. These are, of course, mathematically equivalent and the exponent for one form is the inverse of the exponent for the other (Berry and Okulicz-Koraryn 2011; Arshad, Hu, and Arshad 2018). The Zipf form given above is used in this paper because its development is considered to be more intuitive and the plotting of (log) size versus (log) rank provides a more natural ordering starting with the largest area highest on the y-axis at the left moving towards lower sizes as the rank increases moving toward the right on the x-axis.

Krugman (1996) considers it to be "an astonishing empirical regularity" that the size distribution of metropolitan areas is described by a power law with an exponent of about one. He is not alone in assuming the universality of Zipf's law, with others saying that this "may well be the most accurate regularity in economics" (Gabaix 1999b), "one of the most well-documented facts in social science"(Glaeser 2000), and "a remarkable empirical regularity…so widely accepted among social scientists that it has gained the status of a law" (Córodoba 2008). Krugman says that "the classic study by Rosen and Resnick (1980) suggests that most national metropolitan size distributions are well described by a power law with an exponent not too far from 1." Rosen and Resknick looked at the population size distributions for cities in 44 countries. The exponents ranged from 0.81 to 1.96 with a mean of 1.14. Soo (2005) performed similar analyses for 73 countries finding a obtaining a range of 0.73 to 1.72, mean 1.11 for the Ordinary Least

Squares estimation and lower values from 0.59 to 1.23, mean 0.87 using the Hill estimator. Whether these results are "not too far from 1" would be a judgment for each reader to make. In addition, studies have examined the change in the exponent over time (e.g., Black and Henderson 2003) raising the question as to whether one can consider Zipf's law holding with an exponent of one when the exponent changes.

Numerous authors have suggested that the rank-size rule hold in the upper tail of the distribution of city or urban area sizes, something emphasized in reviews of the literature (Carroll 1982; Gabaix and Ioannides 2004). But there seems to be little discussion of the extent of the city-size distribution that constitutes the upper tail. Rosen and Resnick (1980) include the 50 largest cities in each country, though the choice seemed to have been influenced by data availability and city sizes. Krugman (1996) and Gabaix (1999a) estimate the relationship for the 130 or 135 largest Metropolitan Statistical Areas (MSAs) using 1991 populations. Their data source was the *Statistical Abstract* which happened to present only the data for the 135 areas having populations greater than 250,000 so it is not clear as to whether this was assumed to represent the upper-tail.[1] Black and Duncan (2003) perform analysis from 1900 to 1990 for a constant set of areas constructed using counties and use the urban population at each census as a measure of population for the rank-size analysis. They use a minimum cutoff for including an area in each year that varies with the overall growth of urban areas over the period. Their selection includes all MSAs in 1990 and progressively fewer areas going back to 1900.

The log form of the rank-size equation presented above (3) is estimated using the counts of housing units as the measure of size for the 56 large urban areas for 1950 and 2020. For 2020, the process is straightforward. These are the 56 largest urban areas in that year, so they can simply be ranked by decreasing numbers of housing units from 1 to 56. But these are not the 56 largest areas in 1950 so the ranks cannot be obtained in that way, and other urban areas that were among the largest have not been delineated and the number of housing units are not available. An alternate procedure for obtaining ranks was followed. The census Urbanized Areas are the areas most similar to the urban areas as defined here. So the list of the Urbanized Areas and the housing units in those areas for 1950 is the starting point. But numbers of the urban areas being used here encompass two or more of the 1950 Urbanized Areas. The Dallas and Fort Worth Urbanized areas, separate in 1950, are an example. In such cases the housing units in the separate ares were combined and this replaced the original Urbanized Areas on the list. The resulting list of the Urbanized Areas and combined areas was then ordered by housing units and their ranks were assigned to the corresponding urban areas. Only 49

[1] Both cite *Statistical Abstract* as the source, with Gabaix specifically referencing the 1993 edition (U.S. Census Bureau 1993), and this is the only year giving the 1991 populations. This lists data for 135 MSAs, which is what Gabaix reported as including. Krugman refers to 130 areas and obtains estimates for the equation that a very slightly different from Gabaix.

of the 56 large urban areas received rankings, as the remaining areas were too small in 1950 to have qualified as Urbanized Areas, which required that the largest city have a population exceeding 50,000. Of the 49 ranked areas, 43 were among the 56 largest Urbanized Areas. The smallest of the remaining six was ranked at number 122.

For both 1950 and 2020, the log of the number of housing units for an urban area is regressed on its rank. The results are presented in Table 8. Starting with 1950, the regression coefficient, the exponent in the untransformed equation, is 1.09. This is not too much greater than the value of one suggested by Zipf's law (though the very small standard error indicates that this is statistically significantly different from one). The R^2 value of 0.98 shows a very close fit to the model. And while the value of the constant term is seldom considered, a perfect fit to the rank-size equation with an exponent of one implies that this should be the log of the size of the largest area. The estimated constant of 15.162 is the log of about 3.8 million, not that different from the 3.4 million housing units in the New York area in 1950.

The relation between size and rank in 1950 is shown in Figure 4. The number of housing units in each urban area is plotted against its rank, with logarithmic scales used for both axes. The best-fit line from the regression estimate is shown with the dashed green line. The individual urban areas cluster close to this line. The dotted blue line gives the values expected if the Zipf's law held exactly, with the second 'largest area having half the number of housing units of the largest and so forth. This line is not that far from the estimated best-fit line or from the points for the individual urban areas. The figure provides visual support for the proposition that Zipf's law approximately holds in this instance.

Table 8 also gives the results for the rank-size regression of log of housing units on log of rank for 2020. The regression coefficient, the exponent, is 0.88, not that much

Table 8. Results for regressions of log of housing units on log of rank, 1950 and 2020 (standard errors in parentheses).

	1950 log housing units	**2020 log housing units**
Exponent: regression coefficient on log rank	1.091 (0.021)	0.882 (0.025)
Constant	15.162 (0.068)	16.395 (0.081)
R^2	0.982	0.957
N	49	56

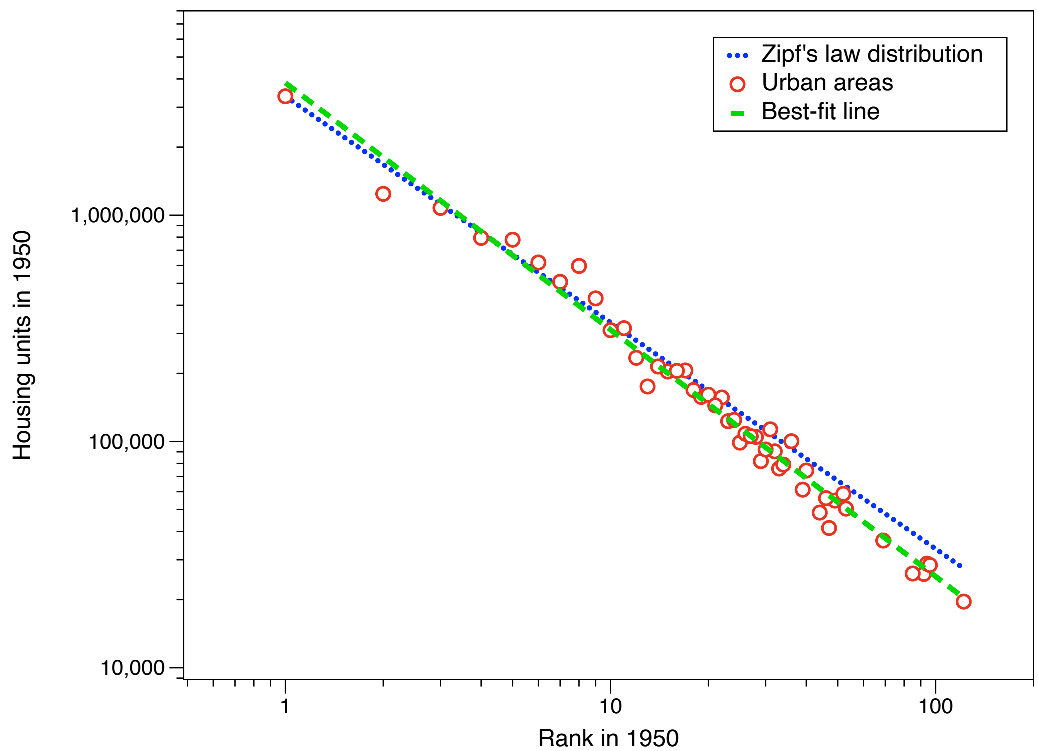

Figure 4. Plot of housing units versus rank of urban areas in 1950 using logarithmic scales on axes, with best-fit line from regression and line showing Zipf's law distribution based on size of largest area.

further below one than the 1950 coefficient is above. But that does represent a drop of 0.2 of the exponent over the 70-year period. At the beginning, the distribution was weighted somewhat towards the larger urban areas, with those being somewhat larger than predicted by Zipf's law. This shifted so that by 2020 the larger areas were relatively smaller in size. At 0.96, R^2 is slightly lower than in 1950. The constant term of 16.395, equivalent to about 13.2 million housing units while the New York area had only 8.3 million. This emphasizes that the larger areas are smaller than expected given Zipf's law.

Figure 5 is the plot of housing units versus rank for 2020. The urban areas do not seems to be clustered quite as close to the dashed green best-fit line from the regression. The very largest urban area are well below the line, with lower numbers of housing units. They are causing the slope of the line and the exponent to be lower. The smaller urban areas in the right portion of the distribution show somewhat more rapid decline than the estimated line. The entire distribution is well above the dotted blue line associated with Zipf's law holding exactly starting with the size of the largest urban area. Comparing the 2020 results with those for 1950, the exponent is significantly

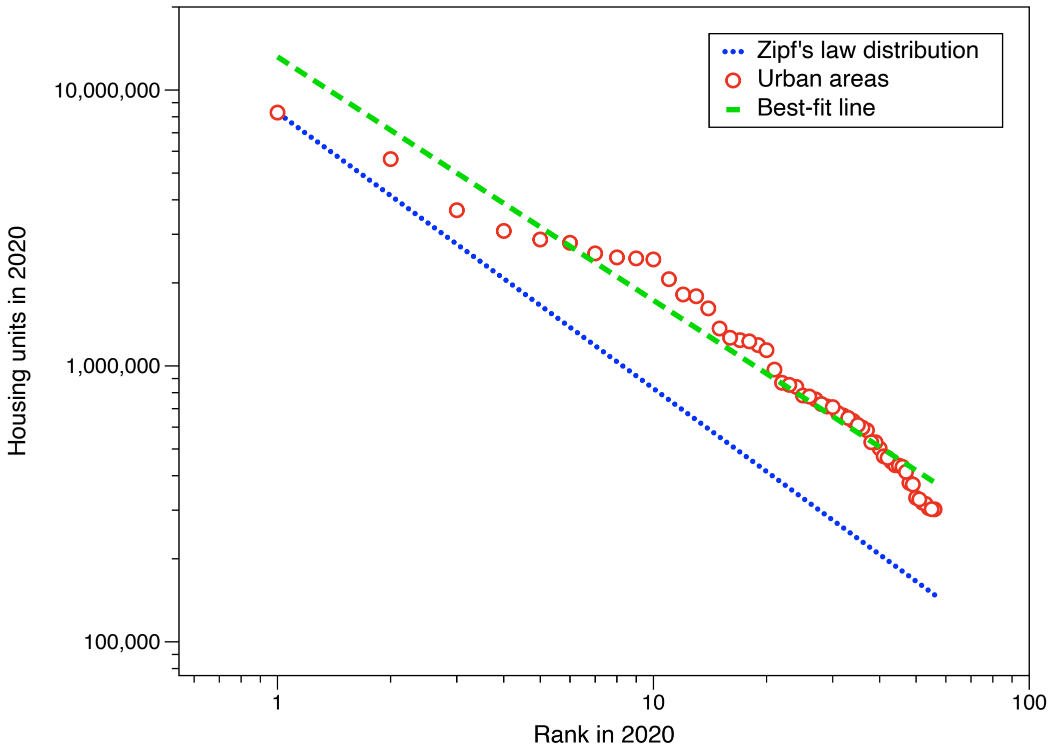

Figure 5. Plot of housing units versus rank of urban areas in 2020 using logarithmic scales on axes, with best-fit line from regression and line showing Zipf's law distribution based on size of largest area.

smaller, going from above one to below. And the distribution for 2020 does not seem to conform to the rank-size rule or Zipf's law quite as closely as 1950.

Before attempting to draw any conclusions from these results, several cautions are in order. Research on the rank-size rule starting with Rosen and Resnick (1980) has shown that the areas used for establishing the sizes of the urban areas has a significant effect on the results. This issue has arisen with the use of cities as the units when this is the only data available versus more inclusive urban agglomerations. The extent of the urban areas used here, established using Combined Statistical Areas, would likewise be expected to have some effect when compared with other alternatives. In addition, with the rank-size rule expected to better describe the upper-tail of the distribution of urban area sizes, the question of the numbers of urban areas to be included in the analysis is also an issue. To address these issue, comparable analyses are conducted using the primary urban agglomerations for which the Census reports data, Urbanized Areas and Metropolitan Statistical Areas.

Starting with 1950, Table 9 presents the exponent obtained for the 49 urban areas for which ranks were available and the exponents from comparable regressions of log

Table 9. Rank-size exponent, the regression coefficient for regressions of log housing units on log rank, for the urban areas and for Urbanized Areas and Standard Metropolitan Areas in 1950 (results for upper-tails of distributions highlighted).

1950 areas	N	Exponent
Urban patterns urban areas	49	1.091
Urban patterns urban areas among 56 largest	43	1.084
All Urbanized Areas	157	1.031
56 largest Urbanized Areas	56	0.992
Smaller Urbanized Areas	101	1.172
All Standard Metropolitan Areas	168	0.982
56 largest Standard Metropolitan Areas	56	0.924
Smaller Standard Metropolitan Areas	112	1.307

housing units on log rank for Urbanized Areas and Standard Metropolitan Areas in 1950. (The latter is the name for the areas that evolved to become the current Metropolitan Statistical Areas.) First the comparisons of the exponents for all of the areas of each kind: The exponent for the urban areas was 1.09 and the exponents for all of the Urbanized Areas and Standard Metropolitan Areas were lower and even closer to one at 1.03 and 0.98 respectively.

Next is the comparison of the exponents for just the areas in the upper-tail of the distribution, defined here as the largest 56 areas (as this will be the number of the largest urban areas in 2020). The urban areas saw only a small decline in the exponent from 1.09 to 1.08, not surprising since 43 of the 49 areas were among the 56 largest. For the Census areas, the exponents for the largest 56 areas were somewhat lower than for all of the areas: 0.99 versus 1.03 for the Urbanized Areas and 0.92 versus 0.98 for the Standard Metropolitan Areas. And the exponents for the smaller areas are much larger at 1.17 and 1.30. So the upper-tail does show a distinctly different distribution of sizes than the other areas. As to whether one could conclude that the upper-tail better conforms to Zipf's law with an exponent of one, the evidence is mixed. For the Urbanized Areas, the exponent for the upper-tail is almost exactly one, but the exponent for all Urbanized Areas was not that far away. But for the Standard Metropolitan Areas, while the exponent for all areas was very close to one, the upper-tail exponent was somewhat lower.

The exponents for the urban areas were larger than those for the Urbanized Areas and Standard Metropolitan Areas. The differences were greatest when comparing the results including just the largest areas. That the exponent might be slightly larger for the urban areas is perhaps not surprising. Some of the urban areas included multiple Urbanized Areas creating larger areas, especially near the top of the distribution. Also, the Combined Statistical Areas used for establishing the general extent of the urban areas and the areas of urban development included are in some instances larger than the Metropolitan Statistical Areas and their predecessors. The greatest effect again is for the very largest areas. This weights the size distribution for the urban areas somewhat more toward the larger areas, resulting in the larger exponent.

The Census has not yet reported results for the Urban Areas (the new name for Urbanized Areas and Urban Clusters) for 2020 so it is impossible to currently do a comparison for that year. Instead, the rank-size model was estimated for the urban areas for 2010, when the data for Urbanized Areas are available. As with 1950, the 56 urban areas were not necessarily the largest areas in 2010. Therefore, ranks were obtained using the Urbanized Area housing unit counts in the same manner as was done for 1950. The exponent of 0.897 is reported in the first row of Table 8 and is not the different from the 2020 exponent of 0.882. It is a little higher, consistent with the observed decline in the exponent from 1950 to 2020.

In addition to Urbanized Areas (areas with populations of 50,000 or more), the Census reported information for what they call Urban Clusters, smaller areas of urban settlement defined in the same way with populations down to 2,500. Including all of the Urbanized Areas and Urban Clusters—thousands—the exponent is high, 1.39, indicating the distribution is more weighted toward the larger areas. Considering only the Urbanized Areas, the exponent drops to 1.12, not that far from the value of one for Zipf's law.

Table 10. Rank-size exponent, the regression coefficient for regressions of log housing units on log rank, for the urban areas and for Urban Clusters and Urbanized Areas in 2010 (results for upper-tails of distributions highlighted).

2010 areas	N	Exponent
Urban patterns urban areas	56	0.897
Urban patterns urban areas among 56 largest	54	0.885
All Urbanized Areas and Urban Clusters	3592	1.389
All Urbanized Areas	497	1.116
56 largest Urbanized Areas	56	0.785
Smaller Urbanized Areas	441	1.243

For comparison with the 56 largest Urbanized Areas, the exponent was determined for the 54 urban areas among the 56 largest by rank. Given the small change in the sample size, the exponent of 0.89 was not very different from that for all of the areas. This is compared to the distribution for the 56 largest Urbanized Areas, which has an exponent of 0.79. Both exponents are less than one, indicating lower numbers of housing units in the largest areas. But this is much more the case for the Urbanized Areas with the smaller exponent. Again, this is reasonable because the some of the urban areas contain multiple Urbanized Areas, making them larger.

The distribution for the smaller Urbanized Areas having fewer housing units than the top 56 was very different, with an exponent of 1.243. So the upper tail, at least as representing by the largest 56 Urbanized Areas, has a very different size distribution from the remainder of the Urbanized Areas. This also raises an issue in interpreting some of the results. All Urbanized Areas taken together had an exponent of 1.12, not that far from one, suggesting that Zipf's law approximately described the size distribution of the Urbanized Areas. But separately considering the upper tail of the largest 56 areas and the other suggests that exponent was the result of very different patterns in the two portions of the overall distribution.

Results from the 2020 Census are available for the Metropolitan Statistical Areas and the smaller Micropolitan Statistical Areas having urban populations down to 10,000. Together, these are called Core-Based Statistical Areas (one of the least felicitous names used by the Census). This makes possible the comparison of the original results for the 56 largest urban areas in 2020 with results using these areas. The exponents for the various analyses are presented in Table 11. For the 56 urban areas, the exponent is 0.88, somewhat less than one.

Starting with all of the Core-Based Statistical Areas, the Metropolitan and Micropolitan Areas, the exponent is 1.24. This is well above one, though not as large as

Table 11. Rank-size exponents, the regression coefficients for regressions of log housing units on log rank for the urban areas and for Core-Based Statistical Areas in 2020 (results for upper-tails of distributions highlighted).

2020 areas	N	Exponent
Urban patterns urban areas	56	0.882
All Core-Based Statistical Areas	939	1.244
All Metropolitan Statistical Areas	392	1.057
56 largest Metropolitan Statistical Areas	56	0.726
Smaller Metropolitan Statistical Areas	336	1.223

the 1.39 for all Urban Clusters and Urbanized Areas in 2010. The values of the exponent are tending to decline as fewer areas are included. The 2020 Census had fewer than a thousand CBSAs compared to the over 3,500 Urban Clusters and Urbanized Areas in 2010. Restricting the analysis to the Metropolitan Statistical Areas (MSAs) reduces the exponent to 1.06, suggesting a very good fit to Zipf's law. But as with the Urbanized Areas, the pattern is not consistent within this distribution with the exponents for the upper tail and lower portion of the distribution being very different.

Once again, considering the upper tail of the 56 largest MSAs produces a much smaller exponent of 0.73, even lower than the value for the 56 Urbanized Areas in 2010 and much below the 0.88 for the urban areas. The extent of the urban areas was determined using the Combined Statistical Areas, which include two ore more Metropolitan or Micropolitan Areas. The larger sizes of especially some of the very largest urban areas contributes the the significantly larger exponent.

The results show very different distributions for the upper tail of the distribution of the MSAs, with an exponent of 0.73, and for the remainder of the distribution, with

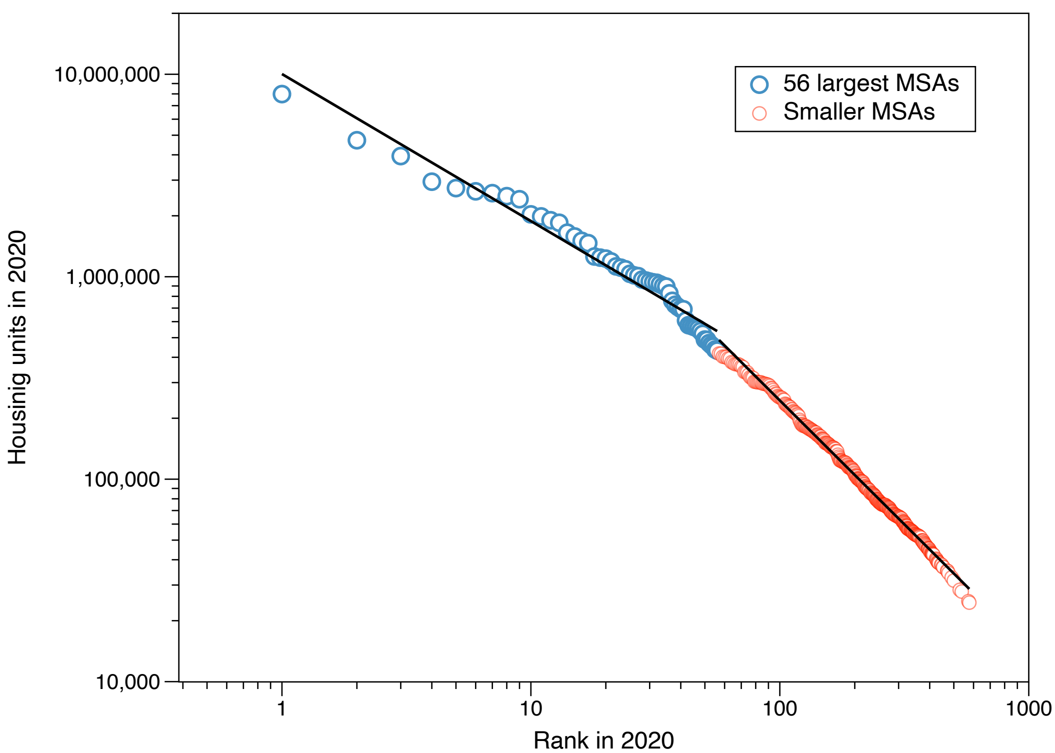

Figure 6. Plot of housing units versus rank of Metropolitan Statistical Areas (MSAs) in 2020 using logarithmic scales on axes, for 56 largest MSAs and the smaller MSAs, with best-fit line from regression for each group.

an exponent of 1.22. The sizes of the MSAs are plotted against their ranks using logarithmic scales on the axes in Figure 6, similar to the previous figures. Two distributions are shown, the 56 largest areas in the upper tail in blue and the remaining 336 smaller areas in red. The separate estimated regression lines, reflecting the different exponents and therefore slopes are shown for each segment. This clearly shows the distinct patterns for the size distributions, with both sets of areas clustered around their different lines.

The exponents for the two portions of the distribution are obviously very different. For reasons of space, only the exponents, which are the values of primary interest, have been reported in the last three tables. Standard errors are uniformly very small and the R^2 value are very large, all exceeding 0.95. However for drawing conclusions about the differences in the exponents for distributions of the upper tail and remainder of the MSAs, the standard errors should also be considered. For the 56 largest areas, the regression coefficient/exponent is 0.726 with a standard error of 0.0200. For the smaller areas, these values are 1.223 and 0.0039. This makes it immediately obvious that any confidence intervals will not overlap and that the exponents can be considered to be significantly different at any standard level of statistical significance.

Conclusion

After being quite stable for the first half of the twentieth century, the set of the 15 largest urban areas changed significantly from 1950 to 2020. Six of the areas on the list in the earlier year were replaced by new entrants by the latter year. The assignment of ranks to the urban areas in 1950 for the rank-size analysis showed 13 of the urban areas were not among the 56 largest in that year, and seven of these were not sufficiently large to even qualify as Urbanized Areas.

These changes in the relative sizes of the urban areas resulted from very different rates of growth. The numbers of housing units more than doubled for all of the urban areas over the seventy-year period. But the growth of many of the areas was astounding. For 24 of the 56 urban areas, over 90 percent of the housing units present in 2020 had been added to the areas since 1950. Cape Coral-Fort-Myers-Naples and Las Vegas had over 99 percent of their housing units added over the period. This constituted an almost ridiculous percentage increase of over 21,000 percent for the former.

Listings of the areas having the fastest and slowest growth showed a clear regional pattern. In 1950 the Northeast region contained close to half, 44 percent, of all housing units in the large urban areas followed by over a quarter in the Midwest. Only one-eight of all of the large-urban-area housing units were in the South. By 2020, the South had come to include the largest share of units, over 30 percent. The share in the

Northeast plummeted while the West's share rose so both had about a quarter of all units. The Midwest had fallen to just 18 percent.

Differences in the amounts and rates of housing unit growth necessarily produced these changes. During the first decade from 1950 to 1960, each of the regions received substantial proportions of the new housing units added to the large urban area. The Northeast and West got the most, and 31 and 27 percent, while the Midwest and South received a somewhat lower 21 percent. By the final decade from 2010 to 2020, the distribution of new housing units had shifted dramatically. Nearly half of all new housing units, 47 percent, were added to the urban areas in the South. The West held fairly steady at a quarter of new units. Shares in the Northeast and Midwest fell to fractions of their earlier values.

These large changes in the sizes and rankings of the urban areas raises the question as to the effect this may have had on the distribution of urban area sizes. The rank-size rule postulates that the sizes of urban areas should be proportional to their ranks raised to some power (and vice versa). Zipf's law further asserts that this exponent should have the value of one. Many have suggested that this holds, at least approximately, for metropolitan areas in the United States, especially for the larger areas, the upper-tail of the distribution. The size distributions at the beginning and end of the seventy-year period were examined not only for the urban areas but for the Urbanized Areas and metropolitan areas defined by the Census as well, because the definition of the areas affects the estimates of the exponents.

In 1950, Zipf's law might be considered to approximately hold for the urban areas, the Urbanized Areas, and the Standard Metropolitan Areas. Also, given that it is often suggested that this works better for the upper-tail, the exponents were also estimated for the areas that ranked among the 56 largest for each set of areas. All of the exponents fell between 1.09 and 0.92, so not that different from one. Estimates made using just the smaller Urbanized Areas and Standard Metropolitan Areas were considerably larger, supporting the idea of a difference between the upper tails and the remaining portions of the distributions.

The results for the recent years show a significant change from 1950. The exponent for the 56 urban areas in 2020 is lower, 0.88. This might be considered still reasonably close to one, supporting Zipf's law. However, that the exponent had declined by about 0.21 is inconsistent with the proposition that it would have a constant value of one. The decline in the exponent is consistent with the findings by Black and Henderson (2003) for the top third of their metropolitan areas from 1900 to 1990. (They reported an increase, having regressed rank on size, opposite the direction used here.)

For the 56 largest Urbanized Areas in 2010 and Metropolitan Statistical Areas in 2020, the exponents for these upper-tails of the distributions are much lower, 0.78 and 0.73 respectively. The values are definitely not close to one and Zipf's law cannot be said to hold for these areas. The contrast between these values and the much larger exponent for the urban areas shows the importance of how the areas are defined. Numbers of the urban areas are substantially larger, including two or more Urbanized Areas and Metropolitan Statistical Areas.

The exponent for all Metropolitan Statistical Areas in 2020 is 1.06, quite close to one. But this is an artifact of the combination of the distribution in the upper tail with an exponent of 0.73 and the distribution of the remaining areas having an exponent of 1.22. Given the two very different distributions, it is unreasonable to claim that this exponent confirms that the set of all Metropolitan Statistical Areas in 2020 conforms to Zipf's law.[2]

One curious aspect of the results relates to the findings that the rates of growth of the urban areas are only weakly related to their sizes, with the relationship becoming progressively less and statistically insignificant by the final decade. This indicates increasing conformance to Gibrat's law. Gabaix (1999a) suggests that if Gibrat's law holds, the size distribution of urban areas will conform to Zipf's law. However the results obtained here seem to suggest conformance to Zipf's law may be declining.

The large urban areas experienced a great deal of change from 1950 to 2020. And that change includes the conformity to the rank-size rule and Zipf's law, raising the question as to whether these propositions continue to be useful in describing the distribution of the sizes of large urban areas.

References

Arshad, Sidra, Shouting Hu, and Badar Nadeem Ashraf. 2018. Zipf's law and city size distribution: a survey of the literature and future research agenda. *Physica A: Statistical Mechanics and Its Applications* 492: 75-92.

Berry, Brian J.L., and Adam Okulicz-Kozaryn. 2011. The city size distribution debate: resolution for US urban regions and megalopolitan areas. *Cities* 29, Supplement 1: S17-S23.

Black, Duncan, and Vernon Henderson. 2003. Urban evolution in the USA. *Journal of Economic Geography* 3, 4 (October): 343-372.

Carroll, Glenn R. 1982. National city-size distributions: What do we know after 67 years of research? *Progress in Human Geography* 6, 1 (March): 1-43.

Cheshire, Paul. 1999. Trends in sizes and structures of urban areas. In *Handbook of Regional and Urban Economics: Applied Urban Economics*, vol. 3, Paul Cheshire and Edwin S. Mills, eds. Elsevier, pp. 1339-1373.

Córdoba, Juan-Carlos. 2008. On the distribution of city sizes. *Journal of Urban Economics* 63, 1: 177–197.

Frey, William H. 1993. The new urban revival in the United States. *Urban Studies* 30, 4/5: 741-774.

[2] Those who would compare these results for Metropolitan Statistical Areas in 2020 with results for such areas from 2000 and earlier are cautioned that major changes were made to the definition of these areas after 2000 that resulted in some of the earlier areas being split into multiple Metropolitan Statistical Areas starting in 2003.

Frey, William H., and Alden Speare, Jr. 1988. *Regional and Metropolitan Growth and Decline in the US*. Russell Sage Foundation.
Gabaix, Xavier. 1999a. Zipf's law for cities: an explanation. *Quarterly Journal of Economics* 114, 3: 739-767.
Gabaix, Xavier. 1999b. Zipf's law and the growth of cities. *American Economic Review* 89, 2: 129-132.
Gabaix, Xavier, and Yannis M. Ioannides. 2004. The evolution of city size distributions. In *Handbook of Regional and Urban Economics: Cities and Geography*, vol. 4, J. Vernon Henderson and Jacquard-François Thisse, eds. Elsevier, pp. 2341-2378.
Glaeser, Edward L. 2000. Urban and regional growth. In *The Oxford Handbook of Economic Geography*, Gordon L. Clark, Maryann P. Feldman, and Meric S. Gertler, eds. Oxford University Press, pp. 83-98.
Glaeser, Edward L., and Jesse Shapiro. 2003. Urban growth in the 1990s: Is city living back? *Journal of Regional Science* 43, 1: 139-165.
Kim, Sukkoo. 2000. Urban development in the United States, 1690-1990. *Southern Economic Journal* 66, 4 : 855-880.
Krugman, Paul. 1996. Confronting the mystery of urban hierarchy. *Journal of the Japanese and International Economies* 10, 4: 399-418.
Lampard, Eric E. 1968. The evolving system of cities in the United States: urbanization and economic development. In *Issues in Urban Economics*, Harvey S. Perloff and Lowdon Wingo, Jr., eds. Johns Hopkins Press for Resources for the Future, pp. 81-139.
Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. *IPUMS National Historical Geographic Information System: Version 17.0* [dataset]. Minneapolis MN: IPUMS. *http://doi.org/10.18128/D050.V17.0.* Data downloaded from *https://www.nhgis.org* on various dates, August to October 2022.
Mills, Edwin S. and Luan' Sende Lubuele. 1995. Projecting growth of metropolitan areas. *Journal of Urban Economics* 37, 3: 344-360.
Monkkonen, Eric H. 1990. *America Becomes Urban: The Development of U.S. Cities and Towns, 1790-1980.* University of California Press.
Ottensmann, John R. 2023. Urban patterns of the largest urban areas in the United States, 1950-2020: a new dataset. Available at *https://urbanpatternsblog.files.wordpress.com/2023/02/new-urban-patterns-dataset.pdf*.
Rosen, Kenneth T. and Mitchel Resnick. 1980. The size distribution of cities: an examination of the Pareto law and primacy. *Journal of Urban Economics* 8, 2: 165-186.
Soo, Kwok Tong. 2005. Zipf's law for cities: a cross country investigation. *Regional Science and Urban Economics* 35, 3: 239-263.
U.S. Census Bureau. 1913. *Thirteenth Census of the United States Taken in the Year 1910*. Vol. 1. *Population 1910. General Report and Analysis*. Government Printing Office.
U.S. Census Bureau. 1993. *Statistical Abstract of the United States: 1993.* U.S. Government Printing Office.
U.S. Census Bureau. 2002. Urban area criteria for the 2000 Census. *Federal Register* 67, 51 (Friday, March 15): 11663-11670.
U.S. Census Bureau. 2022. Urban Area criteria for the 2020 Census—final criteria. *Federal Register* 97, 57 (March 24): 16706-16715.
U.S. Office of Management and Budget. 2000. Standards for defining Metropolitan and Micropolitan Statistical Areas. *Federal Register* 65, 249 (December 27): 82228-82238.

Urban Institute and Geolytics. 2003. *Census CD Neighborhood Change Database (NCDB): 1970-2000 Tract Data*. Geolytics.

Zipf, G.K. 1949. *Human Behavior and the Principle of Least Effort*. Addison Wesley.