The Extent of Centralization of Housing Units in Large American Cities, 1970-2020

John R. Ottensmann
Indiana University Indianapolis
john.ottensmann@gmail.com
urbanpatternsblog.wordpress.com
September 2023

Abstract

The centralization of housing units in 56 large urban areas from 1970 to 2020 is measured using an index based on the ratio of the mean distance housing units are located from the center to the mean distance if the housing units were uniformly distributed across the urban area. Mean centralization declines from 1970 to 2010 and then increases in 2020. Highest levels of centralization tend to occur in old large urban areas, especially in the Northeast. Centralization increases with urban area size and declines as areas grow more rapidly. Measures based on units in the urban core and suburban periphery, and the central density from the negative exponential model are reasonable measures of centralization. The negative exponential density gradient is not. Several of these measures show patterns over time similar to the centralization index, first declining and then increasing in recent decades. This raises the possibility that long-standing trends in urban decentralization may be ending.

Introduction

Researchers were studying the decentralization of the population in urban areas by the middle of the last century. Early research compared rates of growth in the central city to the remainder of the urban area. Thompson (1948) studied the growth of Metropolitan Districts from 1900 to 1940. Bogue (1953) did the same for Standard Metropolitan Areas up through 1950.

Mills (1972) used the density gradient from the negative exponential model of the decline of density with distance from the center as a measure of decentralization of both population and employment. His work popularized that approach, presenting a simple method for estimating the parameters of the model using only totals for the urban area and the central city. Edmonston (1975) extended the work to more urban areas over longer periods of time. An extensive literature using the negative exponential model followed, summarized in the review by McDonald (1989). This became something of a standard method for the study of the subject, with Anas, Arnott, and Small (1998) describing the density gradient as "a useful index of population centralization."

Schnore (1959) identified a problem with such studies, asking "at what point does `normal' city growth—which almost inevitably occurs at the periphery of existing settlement—become `decentralization'?" Guterbock (2021) also faults researchers for failing to make the distinction between decentralization and the general expansion of urban areas, citing Duncan (1959).

It is easy to see why researchers addressing decentralization selected the measures they used. Decentralization of the population within an urban area will result in the rate of population increase in the suburbs being greater than in the central city. Likewise, decentralization will produce a decline in the density gradient. The problem is that the converse is not necessarily true. The growth of an urban area without the population redistribution that one would consider decentralization can also be associated with more rapid suburban growth and density gradient decline.

One needs to start with a clear understanding of what one means by centralization and decentralization. Then a measure of centralization can be developed that directly implements that understanding. An index of decentralization is created in this manner. The index is described in the following section, which also discusses the measurement of centralization more generally.[1]

This paper considers the centralization of housing units from 1970 to 2020 using a dataset with housing units in census tracts for 56 large urban areas, described next. Housing units are used rather than the more frequently employed population for several reasons. Housing units are more fixed than population so they better characterize the urban pattern, an argument made by others as well (e.g., Galster, *et al*. 2001; Paulsen 2014). The Census Bureau (2022) changed from using population density to housing unit density for delineating urban areas saying that "housing unit density provides a more direct measure of the densely developed landscape than population density." The standard monocentric model in urban economics (Muth 1969; Mills 1972) predicts the decline of population density with distance from the center. But the model assumes identical households and addresses the choice of location for residences, making housing units and housing unit density actually more appropriate.

Results for the centralization of housing units in the large urban areas from 1970 to 2020 follow, including basic statistics, examples for selected areas, and simple exploratory models of the relationship of centralization to basic characteristics of the urban areas. Since measures related to counts in urban cores and suburban peripheries and to parameters estimated for the negative exponential model have been used for research on decentralization, the relationships of several of these measures to the index

---

[1] This index was first employed in an earlier paper to look at centralization using an different dataset (Ottensmann 2017a).

of centralization are explored. Finally, the mean index of centralization declined until until 2010, with an uptick to 2020. Such reversals have also been observed for some of the core-periphery and negative exponential measures. These potentially important changes in the trends are compared.

Measurement of centralization

Developing a measure of centralization and decentralization first requires a clear statement of what this concept means. A reasonable place to start would be the dictionary. Most useful is this definition of decentralization:

> decentralization… 2. sociology : the redistribution of population and industry from urban centers to outlying areas (Merriam-Webster 2022)[2]

Most would agree that if the population, numbers of housing units, or employment near the center decrease and numbers near the edge of the urban area increase, decentralization is occurring. But both need not occur to produce decentralization. Either the loss of population or units near the center or the gain of units at the periphery can reasonably be considered to be decentralization, resulting in an decrease in the level of centralization in an urban area having an unchanged size.

All of these forms of decentralization share a common characteristic. Each of the changes increases the mean distance of the population, housing units, or employment from the center. Thus for an urban area within fixed boundaries, mean distance to the center can be taken as a measure of the level of centralization.

The qualification of fixed boundaries is important. It is obvious that larger urban areas will tend to have greater mean distances to the center. Therefore, if mean distance is to be used in a measure of centralization, it must be standardized to account for the size of the urban area.

An urban area can be considered to be neither centralized nor decentralized if the number of housing units or other entities being considered are uniformly distributed across the urban area. The uniform distribution means that the density of housing units would be everywhere the same across the urban area. The mean distance for a uniform distribution can be used for the standardization of the actual mean distance to produce a measure comparable across urban areas of different sizes.

[2] The first definition of decentralization in Merriam-Webster and the only one in many dictionary definitions of centralization and decentralization refers to these in the context of governments and other organizations. The Oxford English Dictionary does give one definition of centralize as, "To position in or at the physical centre; to place in a central position; to make central. Also: to position or adjust with reference to a pre-existing central point." This is consistent with the definition quoted without specifically referring to population, industry, and urban centers.

This provides the basis for the development of an index of centralization for an urban area. Begin with data for the numbers of housing units for small areas such as census tracts within the urban area. The actual mean distance of housing units to the center, $\bar{d}$ is

$$\bar{d} = \frac{\sum h_i d_i}{\sum h_i}$$

where h_i is the number of housing units in census tract i, and d_i is the distance to the center. The mean distance of uniformly distributed housing units to the center, $\bar{d}_u$ is

$$\bar{d}_u = \frac{\sum H a_i d_i}{\sum H a_i} = \frac{\sum a_i d_i}{\sum a_i}$$

where H is the overall housing unit density for the urban area, assumed to be the same for each tract, and a_i is land area of the tract.

The ratio of the actual mean distance to the mean distance for a uniform distribution will have a value of one if housing units are neither centralized nor decentralized. The value will be zero when all housing units are located at the center. So as the ratio decreases from 1 to zero, the level of centralization increases to the maximum possible level. (The ratio can be greater than one if the distribution is more decentralized than a uniform distribution, but this is less likely for urban areas.)

To create an index of centralization which increases as centralization increases, the ratio can be subtracted from one, ranging from one for maximum centralization to zero for a uniform distribution (and negative for even more decentralized patterns). Presentation of the index is enhanced by scaling it to have a value of one hundred for maximum centralization. The centralization index C is then

$$C = 100\left(1 - \frac{\bar{d}}{\bar{d}_u}\right) = 100\left(\frac{\bar{d}_u - \bar{d}}{\bar{d}_u}\right)$$

This is the measure of centralization used in this paper. It ranges from one hundred for maximum centralization to zero for a uniform distribution, and negative for greater decentralization. The second expression for the index shows that it can be interpreted as the percent reduction in mean distance to the center compared to the mean distance for a uniform distribution.

One note on computation. Some of the urban areas in the dataset have two or three centers with each census tract assigned to one of the centers. For these areas, values of the centralization index are first computed for each center, for the tracts assigned to the center using the distances to the center. The final centralization index for the urban area is the mean of these values weighted by the number of housing units in the tracts assigned to each center.

A variety of other measures of centralization have been proposed and used. Massey and Denton (1988) and Lee (2007) describe various alternatives. Perhaps the simplest, both conceptually and in terms of data requirements, is the percentage of the population of the urban area within the central city. Schnore (1962) provides one example. The varying extent of central cities, which are political entities, makes this measure problematic.

Glaeser and Kahn (2001, 2004) use the percentages of employment and population within different distances from the center of the city as measures of centralization. The problem with their application is that they used the same fixed set of distances—3 miles, 5 miles, and 10 miles—for all of the metropolitan areas considered regardless of size. For the New York urban area in 2020 as defined for this research, the area of a 3-mile circle would be less than one percent of the area of the whole area. For Albuquerque, on the other hand, the area of the 3-mile circle would be 12 percent of the total area. For the 56 large urban areas as delineated for 1970, eight would be entirely within the 10-mile radius. A measure that was the percentage within some distance would be reasonable if the distance were to vary with the size of the urban area. Even then, this would make inefficient use of the information on the distances from small areas to the center, collapsing that to two categories.

The absolute centralization index described by Massey and Denton (1988) makes better use of the information by ordering the small areas by distance. The cumulative proportion of population is plotted against the cumulative proportion of area. A uniform distribution would fall on the diagonal. The measure is based on the area between the curve and the diagonal, much like the Gini coefficient.

Mean distance uses the full interval-level information. Lee (2007) uses this to compare levels of centralization over time for a single-metropolitan area. Obviously comparing mean distances across areas of different sizes does not make sense (which is also the case for comparisons of median distances, done by Glaeser and Kahn (2001, 2004)).

Two papers by Galster, Cutsinger, and others (Galster, *et al.* 2001; Cutsinger, *et al.* 2005) measure centralization using mean distance to the center adjusted using the size of the urban area. The first standardizes the inverse of the mean distance by the square root of the area of the urban area. The second (Cutsinger, *et al.* 2005) comes closest to the centralization index presented here, using the ratio of the mean distance for a uniform distribution to the center to the mean distance for housing units, the inverse of the ratio employed in the current index. It has a value one for a uniform distribution and increases without limit as centralization increases. The index proposed here has several advantages. The distribution of the Cutsinger index is much more highly skewed. The open-ended range makes judgements concerning what is a high level of centralization more difficult. And the proposed index can be interpreted as the percentage reduction in mean distance compared to the uniform distribution.

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 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 estimates of the numbers present in those years. These estimates include error resulting from changes to the housing stock over time, especially the loss of units, but analyses suggest that the estimates for urban area totals are reasonable for two decades back in time. Census tract boundaries for 2020 are used for the dataset. The census tract relationship files are used to estimate values for the 2020 tracts from data for earlier years. Detailed documentation of the dataset and listings of all data sources are included in Ottensmann (2023a).

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, Combined Statistical Areas (CSAs) are used (and MSAs that are not included in a CSA). CSAs are used rather than the more commonly used Metropolitan Statistical Areas (MSAs) as 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 have 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 is considered to be urban and is included in an urban area if it has a housing unit density greater than 200 housing units per square mile and is contiguous to the urban area. To include urban territory that is nonresidential, a tract is also included if over one-third of its area has impervious surface of 20 percent or more. An additional condition is that a tract is only considered to be urban if it has been designated as urban for the following census year. This is to provide a pattern of cumulative expansion of the urban areas. This direction has been chosen rather than the reverse (if urban, then urban later) because the more recent data are considered to be more accurate.

Urban areas include multiple areas of urban territory 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 area with tracts assigned to one of the urban centers. The Dallas-Fort Worth area 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 an urban area 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 all initially separate areas deemed large enough to potentially be considered urban centers and then setting the threshold. The smallest urban centers in relation to the total size of the urban area 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 all urban centers that have been included.

The centralization index requires distances from the census tracts to the center. This is taken to be the location of the Central Business District (CBD). The last time the Census identified CBDs was for the 1982 economic censuses. Many researchers have continued to use this information as the best available for designating CBDs. The Census report lists the census tract or tracts constituting the CBD for a large number of cities (U.S. Census Bureau 1983). The CBD tracts for each urban center are identified on a map, combined into a single feature, and the centroid of the feature is determined. This is taken as the point location of the CBD for each urban center. Distance to the CBD in miles from the centroid of each census tract to the CBD is calculated. For urban areas having two or three urban centers, distance is to the CBD for the center to which the tract has been assigned in the delineation of the urban areas.

As the dataset extends back to 1950, values for the centralization index can be calculated starting with that year. Two earlier papers examined the distribution of housing units within the urban areas. The first looked at numbers in the urban core and suburban periphery and the second the negative exponential decline of density with distance from the center (Ottensmann 2023b, c). In both cases, results for 1950 and 1960 were inconsistent with those obtained for the later years. The very small sizes of some of the urban areas in those earlier years combined with error associated with the estimation of housing unit counts from the year-built data are potential causes of these problems. Clear evidence points to the role of small size for the negative exponential analysis. Similar inconsistencies exist for the centralization index. As a result, analysis is once again limited to the period from 1970 to 2020.

Housing unit centralization, 1970-2020

This section presents the centralization index for the 56 large urban areas from 1970 to 2020. Table 1 gives the basic statistics. Mean centralization declines steadily from 24.2 in 1970 to 18.8 in 2010, followed by a small uptick to 19.3 in 2020 to nearly the level in 2000. The urban areas show a great deal of variation in the centralization index, ranging from 8.4 to 47.0 in 2020 with similarly large ranges in the earlier years. The minimum centralization dropped a little in an uneven fashion over the period while the maximum remained fairly steady.

Table 1. Basic statistics for centralization from 1970 to 2020 for the 56 large urban areas.

Year	Mean	Minimum	Median	Maximum
1970	24.2	11.1	24.1	46.1
1980	21.7	8.9	21.0	44.7
1990	19.9	8.1	18.6	44.3
2000	19.4	8.6	18.4	45.6
2010	18.8	9.7	17.9	46.2
2020	19.3	8.4	18.0	47.0

It is useful to look at the urban areas at the extremes with respect to levels of centralization and change. Table 2 lists the areas with the five highest and five lowest levels of centralization in 2020 and change in centralization from 1970 to 2020. Not surprisingly, housing units are most centralized in the New York urban area. New Orleans in second might be something of a surprise. Chicago, Milwaukee, and Philadelphia follow. All are older urban areas that saw significant shares of their development in the nineteenth century. All but New Orleans are larger areas in the Northeast and Midwest. Given earlier dependence on walking and public transportation, high levels of centralization make sense.

Leading the list of the least centralized urban areas are Oklahoma City, Las Vegas and Cape Coral-Fort Myers-Naples. These are areas in the Sunbelt that have seen much more of their growth in recent decades, the era dominated by the automobile, reducing the benefits of centralization. Kansas City and Memphis head a list of nine more areas with centralization of 15 or less, nearly all in the South and West.

Table 2. Urban areas with highest and lowest levels of centralization in 2020 and change in centralization from 1970 to 2020.

Area	Centralization 2020
New York	47.0
New Orleans	38.7
Chicago	30.3
Milwaukee	29.0
Philadelphia	28.1
...	...
Memphis	12.1
Kansas City	11.3
Cape Coral-Fort Myers-Naples	8.8
Las Vegas	8.6
Oklahoma City	8.4

Area	Change centralization 1970-2020
Los Angeles	6.9
Tucson	5.7
New Orleans	4.3
Sarasota-Bradenton	4.2
Albuquerque	3.5
...	...
Salt Lake City-Provo	-11.3
Nashville	-11.7
St Louis	-12.8
Las Vegas	-15.9
Honolulu	-19.0

Change in centralization from 1970 to 2020 presents a very different picture. The largest increase in centralization of 6.9 came in Los Angeles, not where it might have been expected. But the Los Angeles urban area expanded to add large amounts of much less dense territory over the years, making it relatively more centralized. Tucson, Sarasota-Bradenton, and Albuquerque are all rapidly growing urban areas that likely saw increased development near what had started out as small CBDs combined with low density development farther out. For some reason New Orleans had a much lower level of centralization in 1970 and increased dramatically to 1980, producing the large jump from 1970 to 2020.

Making sense of the list of urban areas showing the greatest declines in the level of centralization is more difficult. Four are very different types of areas in the South and West. And then there is St. Louis, presumably dropping because of large losses in the central city.

While the patterns of centralization and change in centralization show wide variation, regional differences were noted. Table 3 gives means by region for centralization in 2020 and for the changes in centralization from 1970 to 2020 and in the final decade from 2010 to 2020. (While Washington-Baltimore is in the South according

Table 3. Mean centralization by region and changes in centralization from 1970 to 2020 and 2010 to 2020 by region.

Region	Centralization 2020**	Change in centralization 1970-2020	Change in centralization 2010-2020
Northeast	26.8	-5.11	0.36
Midwest	19.9	-6.29	-0.06
South	16.9	-4.28	0.36
West	18.5	-4.81	1.28

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

to the census, it is part of the large expanse of urbanization extending up past Boston, so for this research it is considered to be in the Northeast.) Mean centralization in 2020 varies significantly for the urban areas in the four census regions. Levels of centralization are by far the highest for the areas in the Northeast. Areas in the South have on average the lowest centralization. The areas in the Midwest and West have similar levels, somewhat higher than in the South.

Mean changes do not vary significantly by region but the differences are still interesting. The greatest average drop in centralization from 1970 to 2010 takes place for the urban areas in the Midwest. Perhaps a result of the decline in manufacturing? The smallest decrease is in the South, with the other two regions between.

Change in centralization from 2010 to 2020 is considered because the uptick in mean centralization for all areas in this decade after steady decline suggests something may have changed. Urban areas in the Northeast, Midwest, and South have minimal changes in their average levels of centralization, ranging from -0.06 to 0.36. The major increase in mean centralization of 1.28 takes place in urban areas in the West. This is one piece of information that may be valuable in attempting to understand what may have happened in this decade.

The urban areas varied widely not only in their levels of centralization but in the patterns of change from 1970 to 2020. As might be expected, centralization in many of the areas went up and down over the period. But 27 of the areas, nearly half, had simpler, more distinct and consistent patterns. Figure 1 shows the trends for five of the areas exhibiting these patterns with the trend in the mean for all areas in light gray. This varies far less than any of the individual areas since it is the mean, affected by the ups and downs of all 56 areas.

By far the most common of these patterns, for 17 areas, was one of first declining centralization and then increasing centralization. This, of course, is the trend for the

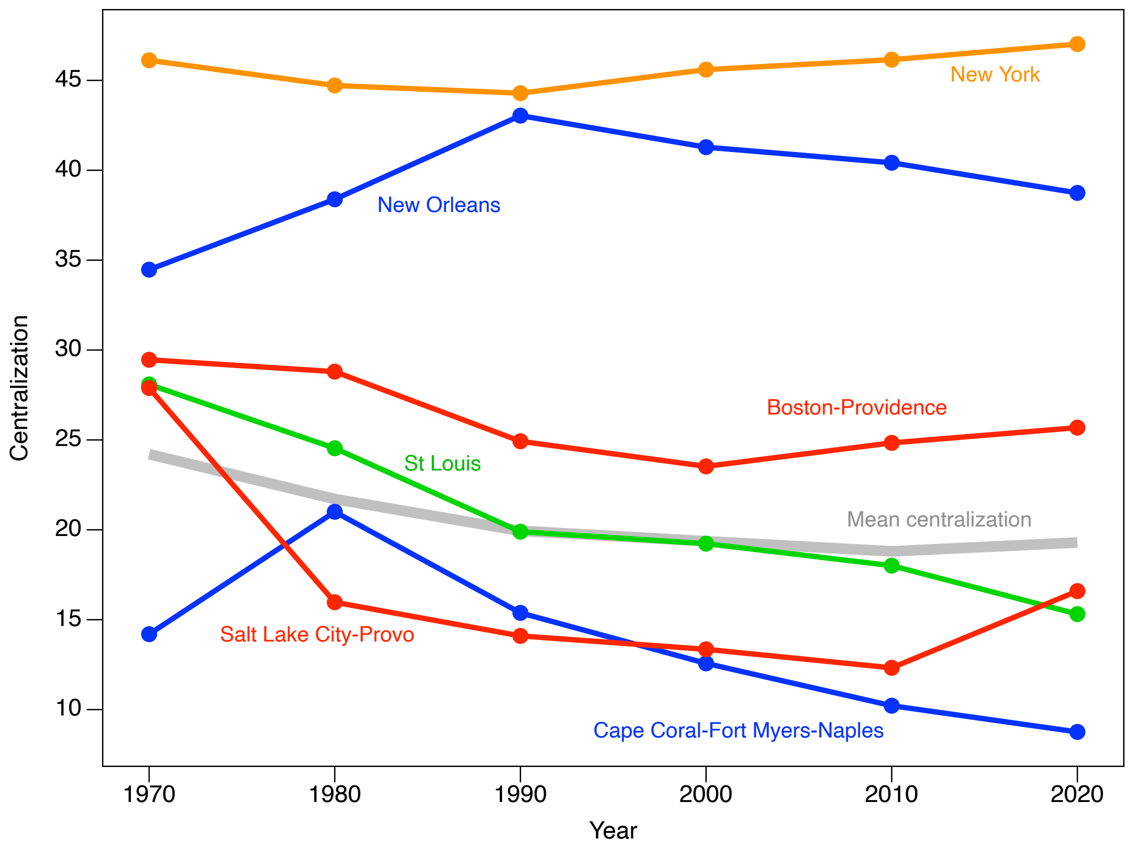

Figure 1. Centralization from 1970 to 2020 for selected urban areas.

mean for all areas, though the year of reversal varies for the areas. The examples of Boston-Providence and Salt Lake City-Provo, in red are illustrative. The former reverses direction and starts increasing in 2010, the latter in 2000. Some urban areas turn up even earlier.

But five the areas exhibit exactly the opposite pattern, with centralization first increasing and then falling. New Orleans and Cape Coral-Fort Myers-Naples are shown, in blue. Consistent decline in centralization in every decade occurs in four urban areas, including St. Louis, green. Finally, the level of centralization in New York varies within a limited range of less than three, smaller than any other area. It seems reasonable to consider the New York pattern, in orange, an example of a steady level of centralization.

Regression is used to examine the relationships of centralization and change in centralization to a few basic characteristics of the urban areas. This is exploratory analysis, with the number of housing units and housing unit density and measures of their change used to predict centralization and change. The models presented include those variables that are statistically significant. By minimizing the number of potential

Table 4. Exploratory regression model predicting centralization in 2020 (standard errors in parentheses).

Independent variable	Centralization 2020	
Housing units 2020 (millions)	6.13	***
	(1.01)	
Change housing units 1970-2020 (millions)	-7.75	**
	(2.23)	
Constant	17.14	***
	(1.02)	
R^2	0.499	***

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

predictors considered, the possibility of obtaining significant results by chance are reduced.

Results for the regression model predicting centralization in 2020 are presented in Table 4. Two urban area characteristics are significant predictors of centralization in 2020, the number of housing units in that year and the change in the number of housing units from 1970 to 2020. These account for half of the variation in centralization with R^2 at 0.50. Centralization increases with the number of housing units. Larger urban areas are more centralized, reflecting the greater benefit of locating relatively closer to the center as the area expands. But centralization is inversely related to the change in the number of housing units. More rapidly growing areas are more decentralized. Since much of the growth of an urban area occurs closer to the periphery, more growth would tend to increase the mean distance to the center and decrease the level of centralization.

Table 5 shows the results for regression models predicting centralization change to 2020, from 1970 and from 2010. Three urban area characteristics are significant predictors of the change in centralization from 1970 to 2020. R^2 is only 0.19, much lower than for the model predicting centralization in 2020. The number of housing units in 1970, in log form, is inversely related to the change in centralization. Growth in the number of housing units over the period is associated increased levels of centralization. This latter result is consistent with the observation previously that larger cities tend to be more centralized. And higher housing unit density at the start of the period is associated with increased centralization as well.

As discussed when considering mean change by region, change in centralization from 2010 to 2020 was considered along with change over the entire period because of

Table 5. Exploratory regression models predicting change in centralization from 1970 to 2000 and from 2010 to 2020 (standard errors in parentheses).

Independent variable	Change centralization 1970-2020	
Log housing units 1970	-2.49	*
	(1.17)	
Change housing units 1970-2020 (millions)	3.44	*
	(1.30)	
Housing unit density 1970 (thousands)	7.41	*
	(2.95)	
Constant	15.32	
	(11.74)	
R^2	0.192	*

Independent variable	Change centralization 2010-2020	
Change housing units 2010-2020 (millions)	5.86	**
	(1.73)	
Change housing unit density 2010-2020 (thousands)	-11.02	
	(4.13)	
Constant	-0.011	*
	(0.325)	
R^2	0.257	***

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

the reversal in the direction of change in mean centralization in the final decade. The second model in Table 5 shows the prediction of change in centralization from 2010 to 2020. The prediction is slightly better than for change over the longer interval with R^2 at 0.26. Change in housing units in that final decade is a significant predictor of change in centralization, just as change in units since 1970 was for change in centralization since 1970. The change in housing density from 2010 to 2020 is inversely related to the change in centralization. It is not clear why this might be the case.

Relationship to other measures of centralization

As described in the introduction, variables related to the central city-suburban distribution of population and housing units and the negative exponential model of the decline of density have been used as measures of centralization and decentralization. This section examines the relationship of measures associated with these to the index of centralization. This is accomplished by using the measures to predict the value of the index of centralization using regression, with the R^2 from the regression indicating the proportion of the variation in centralization accounted for by the different measures.

The early approach to the study of decentralization used the distribution of the population in the central city and in the remainder of the urban area, however defined.

The latter would be referred to as the "ring" or more often the "suburbs." The basic notion was that the central city encompassed the older portion of the urban area while the suburban ring was the more recently developed area. The problem is that cities are political entities and their boundaries encompass widely varying portions of their urban areas. In some urban areas, territory that had been developed and became a part of the urban area as early as the nineteenth or early twentieth century are outside the central city, in the area considered to be the suburbs. Conversely, central cities in other urban areas have been able to continue expanding their boundaries such that they include a large share of more recent development.

The dataset used here allows for the designation of the older and the more recently developed portions of the urban area in a consistent manner. The extent of the urban area in 1940 is delineated using estimates of housing units by census tracts. This is referred to as the urban core. Starting with 1950, that portion of the urban area outside the urban core is considered to be the suburban periphery. Thus the suburban periphery is approximately that part of the urban area developed after World War II, the area often considered to be the suburbs. This urban core-suburban periphery distinction has been used to examine change in the large urban areas (Ottensmann 2023b).

Early studies using central city-suburb data looked at decentralization over time by comparing rates of growth in the suburban ring and the central city. The centralization index is a measure of the level of centralization at one point in time, with change in that measure used to examine decentralization over time. For comparison with the centralization index, measures are needed for the distribution of housing units across the urban core and suburban periphery at specific times. The most obvious and frequently used measure is the percentage of all housing units in the urban area in the urban core. All other things being equal, a higher proportion in the core should be associated with greater centralization.

Percent in the core fails to take into account the relative portions of the urban area in the core and periphery. Some of the more rapidly growing urban areas were very small in 1940 and thus have a small urban core. Housing units close to the center but outside the core would not be seen as contributing to centralization. An alternative measure is the ratio of the housing unit density in the urban core to the density in the urban periphery. The underlying counts are still affected by the size of the core relative to the periphery, but the densities at least standardize those counts for the sizes of the areas.

Table 6 presents the results for the regressions of the centralization index for 2020 on the percent of housing units in the urban core and the ratio of density in the urban core to the suburban periphery. R^2 for the percent in the core is 0.48 and for the core-periphery ratio is 0.50, nearly identical. Each is accounting for approximately half of the variation in the index of centralization. This similarity does not result from the two

Table 6. Regression models for percent urban area housing units in the urban core and ratio of density in the urban core to the suburban periphery predicting centralization index in 2020 (standard errors in parentheses).

Independent variable	Centralization 2020	Centralization 2020
Percent urban area housing units in urban core 2020	0.330 *** (0.047)	- -
Ratio density urban core to suburban periphery 2020	- -	4.539 *** (0.620)
Constant	11.904 *** (1.248)	5.245 * (2.029)
R^2	0.477 ***	0.498 ***

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

measures being that closely related. The R^2 for the regression of one on the other is only 0.42.

To see if these relationships hold over time, the same regressions are repeated for the values for 1970. R^2 for percent in core was slightly lower while the value for the core-periphery ratio was larger, 0.70. Further comparisons regressed the changes in percent in core and the core-periphery ratio from 1970 to 2020 on the change in the centralization index. Here performance diverged. R^2 for the change in the percent in the urban core was only 0.053, not statistically significant. For the core-periphery ratio it was 0.44, highly significant and accounting for nearly half the variation in the change in centralization. Given these results, the ratio of the density in the urban core to the density in the suburban periphery would seem to be a more consistent and effective measure of centralization.

The model for the negative exponential decline of density with distance from the center can be expressed as

$$D_i = D_0 e^{-\beta d_i}$$

where D_i is the density (housing unit density in the current research) in census tract i, d_i is the distance from the census tract to the center of the urban area. D_0 is the central density and β is the density gradient, both parameters to be estimated from the data, and e is the base of the natural logarithms.

Winsborough (1963) suggests that both parameters of the negative exponential model are measures of the distribution of population or housing units. He proposes that the density gradient can be considered to be a measure of concentration and the central density a measure of congestion, without further explaining what either of these terms means. However, the density gradient has been nearly universally used as a measure of centralization, equating the decline of the density gradient with decentralization. That it might not be a very good measure should have been obvious early, as some of the oldest studies found a strong inverse relationship between the density gradient and the population of the urban area. As urban areas grow, density gradients tend to decline. Whether this decline represents actual decentralization or is simply a product of normal growth was not considered.

Guterbock (2021) and Ottensmann (2017b, 2022, 2023c) confront the problem with using the density gradient as a measure of decentralization. They demonstrate the empirical relationship between the gradient and the size of the urban area. They further show the mathematical relationship that exists between the density gradient, the central density, and the size of the urban area that underlies the problem.

The parameters of the negative exponential decline of housing unit density have been estimated for the large urban areas from 1970 to 2020 using nonlinear regression (Ottensmann 2023). Because the model assumes the decline of density from a single center, the model was estimated only for the 40 large urban areas having a single center. The relationships of both model parameters, the density gradient β and the central density D_0 to the index of decentralization for 2020 are shown in the first two sets of regression results in Table 7. While the density gradient is statistically significantly related to the centralization index, that relationship is not very strong. Across the 40 urban areas, the gradient accounts for only 18 percent of the variation in the centralization index. It is the central density that is far more strongly related to the centralization index, with the R^2 of 0.51 showing it accounts for half of the variation in the index. So the central density is actually a much better measure of centralization than the density gradient.

Similar results were obtained when regressing the centralization index for 1970 and the density gradient and the central density. In this case the value of R^2 for the density gradient was only 0.045, not statistically significant. The central density, on the other hand, was even more strongly related to the centralization index, accounting for 64 percent of the variation. But as measures of the change in centralization and decentralization over time, both fail. Regressions of the change in the in the density gradient and central density on the change in the centralization index from 1970 to 2020 both yield R^2 values less than 0.1 and the relationships are not statistically significant.

Guterbock (2021) proceeds further to develop a measure combining the central density and the density gradient in such a way that it is independent of the size of the urban area. This measure is called the Density Distribution Index (DDI) and is

Table 7. Regression models for negative exponential model density gradient and central density and Density Distribution Index predicting centralization index in 2020 for 40 urban areas having a single center (standard errors in parentheses).

Independent variable	Centralization 2020	Centralization 2020	Centralization 2020
Density gradient 2020	44.235 ** (15.208)	- -	- -
Central density 2020 (thousands)	- -	0.537 *** (0.086)	- -
Density Distribution Index 2020	- -	- -	5.051 *** 0.856
Constant	14.476 *** (2.016)	15.404 *** (1.058)	-17.588 ** 6.327
R^2	0.182 **	0.509 ***	0.478 ***

** p < 0.05   ** p < 0.01   *** p < 0.001*

expressed as a function of the negative exponential model parameters with this formula:

$$DDI = \ln\left(D_0\right) + \frac{1}{2}\ln\left(\beta\right)$$

The idea is that a measure not related to the size of the urban area will then be a superior measure of centralization. This does not consider the possibility that levels of centralization may in some way be related to size such that a size-independent measure may not be the best measure of centralization.

Values of the DDI are calculated from the central density and density gradient for each urban area for 2020 and are used to predict the centralization. The regression results are also presented in Table 7. With an R^2 of 0.48, the relationship to the index is nearly the same as for the central density. That this is the case is hardly surprising. While the DDI is a function of both the central density and the density gradient, most of the variation in the DDI comes from the central density. Calculating the ratio of the two terms for DDI, log of central density and 0.5 times log of density gradient for each urban area, the mean of the the ratios is 7.87. In other words, the magnitude of the first term with the central density is on average nearly eight times the magnitude of the second term with the density gradient. Regressing the log of central density on the DDI produces an R^2 of 0.94. In other words, the DDI is not that different from central density, explaining the similarity of the relationships to the centralization index.

Going back to 1970, like central density, DDI did better in predicting the centralization index with R^2 at 0.76, even higher than central density in 1970. Also following central density, the change in DDI from 1970 to 2020 was unrelated to the index, accounting for less than 10 percent of the variation in the change in centralization.

Given their close relationship and their similar performance, the central density and the DDI would be the most reasonable measures of centralization derived from the negative exponential model. Both provide estimates of centralization at a point in time comparable to the core-periphery measures, accounting for approximately half or more of the variation in the centralization index. But these measures are inferior to the core-periphery density ratio in failing to show a significant relationship to the change in centralization from 1970 to 2020.

The trend analysis in the following section uses central density rather than DDI as the centralization measure associated with the negative exponential model for two reasons: The central density has the intuitive meaning of number of housing units per square mile at the point where the density curve intersections the y-axis, the center. DDI is a combination of the logs of the two negative exponential parameters expressed in different units. And central density avoids the assumption made in constructing the DDI that it is independent of the size of the urban area, which seems less than appropriate given that the results in the last section showed the centralization index to be related to the number of housing units in the urban area.

Trends in measures related to centralization

The mean centralization for the 56 large urban areas declined consistently from 1970 to 2010. Then from 2010 to 2020 it increased slightly. This is not the only measure for which such a reversal has been observed. The paper looking at the negative exponential decline of density from 1970 to to 2020 found the mean central density declined through 2000 and then increased to 2020 (Ottensmann 2023c). The paper describes this as "an interesting and potentially very important development." This section compares the trends since 1970 in centralization, the ratio of density in the urban core to the suburban periphery, and central density, the latter two being measures related to centralization. Housing unit density in the urban areas is also included.

Table 8 gives the mean values of the centralization index, the core-periphery ratio, central density, and the housing unit density of the urban areas. These results are for the 56 large urban areas with the exception of central density, which is for the 40 urban areas with a single urban center for which the negative exponential model could

Table 8. Mean centralization, core-periphery ratio, central density, and overall urban area density from 1970 to 2020.

Year	Centralization	Core-periphery ratio	Central density	Overall density
1970	24.2	3.51	6,989	1,154
1980	21.7	3.09	6,475	1,141
1990	19.9	2.88	6,206	1,158
2000	19.4	2.90	6,031	1,093
2010	18.8	2.93	6,339	1,098
2020	19.3	3.09	7,464	1,121

be estimated. As shown earlier, centralization declines steadily to 2010 and then increases somewhat by 2020. The core-periphery ratio declines to 2000, increases very slightly to 2010, and then somewhat more by 2020. Central density also moves downward through 2000 and then increases significantly to 2010 and by a much larger amount by 2020, above the 1970 value. The pattern for the overall density of the urban areas is less consistent, dropping by a small amount, then increasing by a small amount, before more substantial drop to 2000. This is followed by a small increase to 2010 and a jump to 2020 to nearly the 1970 starting density. So all of the measures are showing increases in their mean values in the final one or two decades after previous declines.

It should be emphasized that this pattern in the means over time does not imply that individual urban areas necessarily followed this trajectory. As seen for centralization, urban areas vary greatly in the patterns of their trends over time. The other measures show similar variation. Over time an increasing percentage of the urban areas experience increases from one decade to the next. Table 9 gives the percent of the urban areas for which the value at the end of each decade was greater than the value at the start, the areas in which the measure was increasing. Centralization and central density increased steady, only exceeding fifty percent in 2020. The core-periphery ratio jumps to that level by 2000, drops slightly, and shoots up again to 2020. Once again, the overall density was an outlier, with high levels of increase early, a big drop from 1990 to 2000, and then large increases. The increases observed in the means of these measures was at least in part a function of higher percentages of the urban areas showing increases over time.

The four measures have very different magnitudes and are measured in different units. Directly comparing the trends requires that they be transformed. For each the the

Table 9. Percent of of urban areas with increases in centralization, core-periphery ratio, central density, and overall urban area density in each decade from 1970 to 2020.

Decade	Centralization	Core-periphery ratio	Central density	Overall density
1970-1980	27	13	14	41
1980-1990	27	21	18	54
1990-2000	36	55	23	14
2000-2010	43	50	34	52
2010-2020	59	71	57	68

values, the mean for the measure over time is subtracted and those values are divided by the standard deviation to give standardized z-scores. These are plotted to compare the relative changes in the measures from 1970 to 2020, shown in Figure 2. As observed, centralization declines steady to 2010, followed by a small increase to 2020. Despite having a different inflection point, the core-periphery ratio and the central density

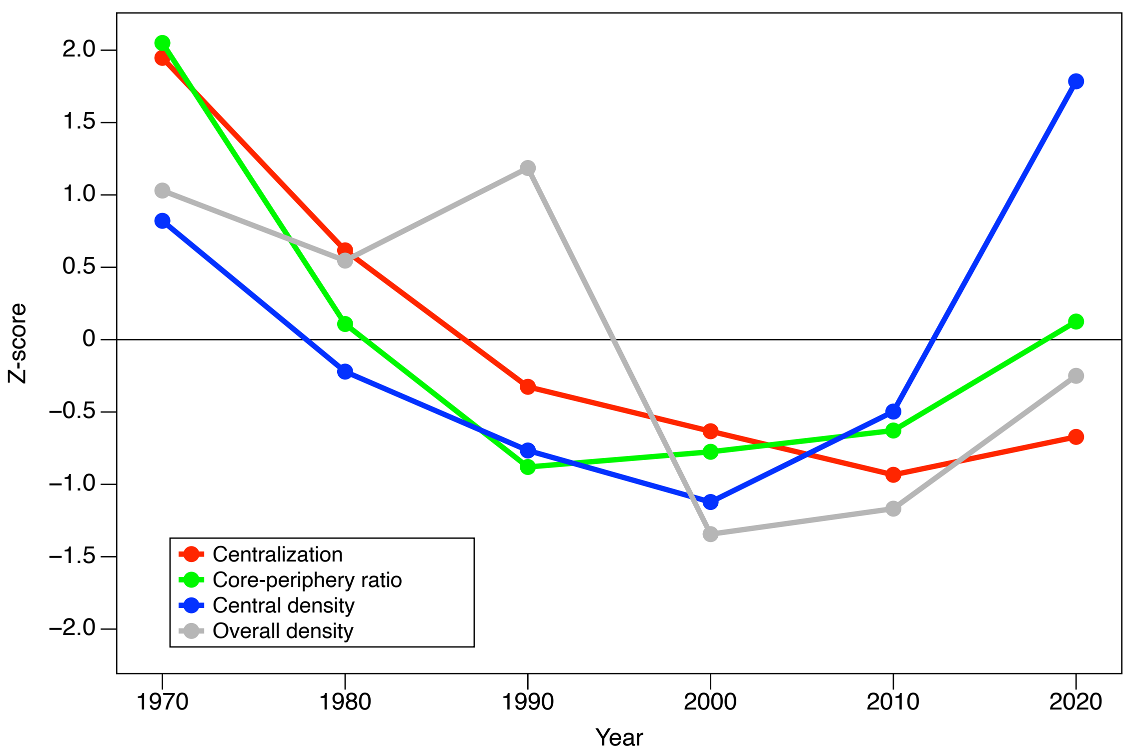

Figure 2. Z-scores for centralization, core-periphery ratio, central density, and overall urban area density from 1970 to 2020.

follow a similar pattern over much of the period. Central density declines to 2000 and then increases. The core-periphery ratio actually has the minimum z-score in 1990 but then increases only slightly before a larger increase from 2010 to 2020. And while overall density has the modest down and up over the first two decades, a large drop from 1990 to 2000 is followed by increases in the following two decades. So while the year that the increase starts varies, all four measures have declines followed by increases some time in the final three decades.

Conclusions

Measures of centralization based on traditional approaches do a fairly decent job, accounting for around half of the variation in the centralization index. Two variables related to the numbers of housing units in the urban core and the suburban periphery performed equally well—percent of housing units in the core and the ratio of housing unit density in the core to density in the periphery. Of course comparable measures using central city-suburb data will likely do less well due to the arbitrary nature of central city boundaries affecting the proportion of the urban area included in the central city.

The density gradient estimated for the negative exponential model of density decline has been a widely used measure of centralization. It is not a very good measure, accounting for only about a fifth of the variation in the centralization index. But while the gradient performed poorly, the other negative exponential model parameter, the central density is a more acceptable measure of centralization, accounting for about half of the variation in the centralization index.

The trend in the average levels of centralization raises interesting questions. Mean centralization declined in each decade from 1970 to 2010 but this consistent pattern was broken with an uptick in centralization for 2020. That this is not an aberration is confirmed by the trends in two of the other measures of centralization, the ratio of core to periphery density and the central density from the negative exponential model. Both show declines followed by a reversal, increasing after 2000. And while the mean of overall densities in the urban areas varied somewhat more, a large drop before 2000 also precedes increases in the final two decades.

So what to conclude from these observations? Has an extremely long period decentralization within urban areas come to an end? Has the trend reversed and are urban areas now becoming more centralized? Guterbock (2021) observed the deconcentration/decentralization of American urban areas though most of the twentieth century, becoming especially rapid after 1950. Then starting in 1990, levels of centralization remain fairly constant. This leads to the proclamation of "the end of population deconcentration in the United States." (He does wisely caution that the aftermath of COVID-19 may see other changes to urban patterns.)

I am more cautious about assuming that decentralization has come to an end and that levels of centralization will either remain steady or increase. I have seen too many examples where an observation of a change in direction is initially touted as a fundamental shift away from long established patterns but is later acknowledged to have been a brief anomaly. For example, after 1970, observations that nonmetropolitan areas had started growing more rapidly than metropolitan areas prompted observers to declare the end of centuries of urbanization and the start of counter urbanization (e.g., Berry 1980). But it was not long before metropolitan areas were again growing more rapidly and the idea of counter urbanization was forgotten (Beaile and Fugitt 1986; Long and DeArea 1988).

The increases in average levels of centralization in recent decades is certainly significant and provocative. Note however that it is not as if something had changed around the turn of the century. Increases in mean centralization resulted at least in part from greater numbers of urban areas experiencing increases in each decade. Some forces may be acting that will be important to identify and understand.

References

Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. Urban spatial structure. J*ournal of Economic Literature* 36, 3: 1426-1464.

Beale, Calvin L., and Glenn V. Fuguitt. 1986. Metropolitan and nonmetropolitan population growth in the United States since 1980. In *New Dimensions in Rural Policy: Building Upon Our Heritage*, Joint Economic Committee, U.S. Congress. U.S. Government Printing Office, pp. 46-62.

Berry, Brian J. L. 1980. Urbanization and counter urbanization in the United States. *Annals of the American Academy of Political and Social Science* 451, *Changing Cities: A Challenge to Planning*: 13-20.

Bogue, Donald J. 1953. *Population Growth in Standard Metropolitan Areas, 1900-1950*. U.S. housing and Home Finance Agency.

Cutsinger, Jackie, George Galster, Harold Wolman, Royce Hanson, and Douglas Towns. 2005. Verifying the multi-dimensional nature of metropolitan land use: advancing the understanding and measurement of sprawl. *Journal of Urban Affairs* 27, 3: 235-259.

Duncan, Otis Dudley. 1959. Human ecology and population studies. In *The Study of Population*, Philip M. Hauser and Otis Dudley Duncan, eds. University of Chicago Press, pp. 678-716.

Edmonston, Barry. 1975. *Population Distribution in American Cities*. Lexington Books.

Galster, George, Royce Hansen, Michael R. Ratcliffe, Harold Wolman, Stephen Coleman, and Jason Freihage. 2001. Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing Policy Debate* 12, 4: 681-717.

Glaeser, Edward L., and Matthew E. Kahn. 2001. Decentralized employment and the transformation of the American city. In *Brookings-Wharton Papers on Urban Affairs, 2000*, William G. Gale and Janet Rothenberg Pack, eds. Brookings Institution Press, pp. 1-47.

Glaeser, Edward L., and Matthew Kahn. 2004. Sprawl and urban growth. In *Handbook of Regional and Urban Economics*, vol. 4, J.V Henderson and J.E Thisse, eds. Elsevier North-Holland, pp. 2481-2527.

Guterbock, Thomas M. 2021. The beginning and the end of urban population deconcentration in the United States: new insights from application of the Density Distribution Index. *Cities* 118: 103349.

Lee, Bumsoo. 2007. "Edge" or "edgeless" cities? Urban spatial structure in U.S. metropolitan areas, 1980-2000. *Journal of Regional Science* 47, 3: 479-515.

Long, Larry, and Diana DeAre. 1988. A perspective on the nonmetropolitan turnaround. *Population and Development Review* 14, 3: 433-450.

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.

McDonald, John R. 1989. Econometric studies of urban population density: a survey. *Journal of Urban Economics* 26, 3: 361-385.

Massey, Douglas S., and Nancy A. Denton. 1988. The dimensions of residential segregation. *Social Forces* 67, 2: 281-315.

Merriam-Webster. 2002. "decentralization." *Merriam-Webster Dictionary*. Accessed at *https://www.merriam-webster.com/dictionary/decentralization* on June 10, 2022.

Mills, Edwin S. 1972. *Studies in the Structure of the Urban Economy*. Johns Hopkins Press for Resources for the Future.

Muth, Richard F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press.

Ottensmann, John R. 2017a. The degree of centralization in large urban areas in the U.S., 1950-2010. Available at *https://urbanpatternsblog.files.wordpress.com/2017/01/urban-area-centralization.pdf*.

Ottensmann, John R. 2017b. Negative exponential model parameters and centralization in large urban areas in the U.S., 1950-2020. Available at *https://urbanpatternsblog.files.wordpress.com/2017/02/negative-exponential-centralization.pdf*.

Ottensmann, John R. 2022. The decline of the density gradient is not a good measure of decentralization. Available at *https://urbanpatternsblog.files.wordpress.com/2022/06/density-gradient-centralization.pages_.pdf*.

Ottensmann, John R. 2023a. 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*.

Ottensmann, John R. 2023b. Density and sprawl in the older and newer parts of large urban areas in the U.S., 1970-2020. Available at *https://urbanpatternsblog.files.wordpress.com/2023/06/density-and-sprawl-in-older-and-newer.pdf*.

Ottensmann, John R. 2023c. The negative exponential decline of density in large urban areas in the U.S., 1970-2020. Available at *https://urbanpatternsblog.files.wordpress.com/2023/07/negative-exponential-decline.pdf*.

Paulsen, Kurt. 2014. Geography, policy or market? new evidence on the measurement and causes of sprawl (and infill) in US metropolitan regions. *Urban Studies* 41, 12: 2629-2645.

Schnore, Leo F. 1959. The timing of metropolitan decentralization: a contribution to the debate. *Journal of the American Institute of Planners* 25, 4: 200-206.

Schnore, Leo F. 1962. Municipal annexation and the growth of metropolitan suburbs, 1950-60. *American Journal of Sociology* 67, 4: 406-417.
Thompson, Warren S. 1948. *The Growth of Metropolitan Districts in the United States, 1900-1940*. U.S. Government Printing Office.
U.S. Census Bureau. 1983. *1982 Economic Censuses. Reference Series. Geographic Reference Manual.* EC82-R-1. U.S. Government Printing Office.
U.S. Census Bureau. 2022. Urban Area criteria for the 2020 Census—final criteria. *Federal Register* 97, 57 (March 24): 16706-16715.
Urban Institute and Geolytics. 2003. *Census CD Neighborhood Change Database (NCDB): 1970-2000 Tract Data*. Geolytics, Somerville, NJ.
Winsborough, Hal H. 1963. An ecological approach to the theory of suburbanization. *American Journal of Sociology* 68, 5: 565-570.