The Degree of Centralization in Large Urban Areas in the U.S., 1950-2010

John R. Ottensmann
Indiana University-Purdue University Indianapolis
john.ottensmann@gmail.com
urbanpatternsblog.wordpress.com
January 2017

Abstract

Levels of centralization of population and housing and the phenomenon of decentralization are an important aspect of large urban areas. A measure of centralization called the centralization ratio is developed that reflects the proportional reduction in the mean distance housing units are located from the center compared to a uniform distribution in the urban area. Values for this measure are computed using data on numbers of housing units by census tracts for 59 large urban areas defined for each census from 1950 to 2010. The results show that mean levels of centralization have declined steadily and significantly over the period. This decline was not universal, however, with 14 areas showing increases. Centralization varied by region of the country, highest in the Northeast and lowest in the South. Mean levels of centralization were also higher for the largest urban areas.

Introduction

Since at least the middle of the twentieth century, a great deal of attention has been focused on the decentralization of people, residences, and jobs within America's urban areas. Perhaps the most basic method for reporting on these changes has been the examination of the percentages inside and outside of the central cities of metropolitan areas. Indeed, Massey and Denton (1988) present and use percentage in the central city as one measure of urban centralization.

Central city percentage is a crude measure that suffers from the problem that the boundaries of those cities depend on historical and political factors that vary widely. The proportion of a metropolitan area's population or housing units included within the central city boundary varies widely, from a fairly small percentage of the total to a substantial majority. This is even subject to sudden, major change resulting from changes in local government structure, such as city-county consolidations.

A more general approach has been to address the question of decentralization by employing the widely recognized regularity that densities of population and housing units tend to decline as a negative exponential function of distance from the center. This

pattern is predicted by economists' monocentric model (Muth 1969; Mills 1972). It is therefore not surprising that using measures of exponential decline to examine decentralization has been especially common among economists (e.g., Mieszkowski and Mills 1993; Anas, Arnott, and Small 1998).

The attraction of using the negative exponential model parameters to examine decentralization is clear. For any metropolitan area, the decentralization of population or housing units will result in a decrease in the negative exponential density gradient (all other things being equal). However, the truth of this statement does not necessarily imply the converse, that the density gradient can therefore be used as a reasonable measure of centralization. The relationship between the parameters for the negative exponential decline of density and the level of centralization is more complex.

Questions involving the levels of centralization of urban areas over time and the relationship to the parameters of the exponential model are being addressed in two papers. This paper examines centralization in large urban areas over time using a "pure" measure of centralization with no consideration of the negative exponential model. The following paper will then examine the relationship between this measure of centralization and the parameters of the negative exponential model.

The next section discusses the measurement of centralization and presents the measure used in this research. This is followed by the description of the data being used and the details of the method of measurement of centralization. Results are presented for levels of centralization for 59 large urban areas in the United States from 1950 to 2010, along with the changes in centralization over time.

The Measurement of Centralization

This section describes the measure of centralization—the centralization ratio—that was developed for and is used in this research. The centralization of people and housing units in urban area has been measured in a variety of ways. Massey and Denton (1988) and Lee (2007) review and use numbers of indices of centralization. Several of the more important approaches that have been used are discussed here, serving as the background for explaining the decision to develop the new measure.

A straightforward, easily understood measure of centralization is the proportion of the objects of concern in the urban area within some distance from the center. Glaeser and Kahn (2001, 2004) use this in examining the centralization of employment and population. The problem with their approach is that they use several fixed distances of 3, 5, and 10 miles for calculating the proportions for all of the urban areas being considered. But obviously the proportion within some fixed distance has very different implications for the degree of centralization for urban areas of different sizes. The area within 10 miles of the center of New York or Los Angeles is still a small fraction of the total urban area. On the other hand, for much smaller urban areas, this area within the

same 10-mile radius can encompass a significant portion of those areas. This becomes even more of an issue going back in time, when urban areas encompassed much smaller areas. For the urban areas defined for 1950 in the urban patterns dataset, for 35 of the 59 areas the maximum distance from the center was less than 10 miles. Indeed, for 7 of the areas, the maximum was less than 5 miles and for 2 of those it was less than 3 miles. So it is obvious that proportions within such fixed distances cannot be a generally useful measure of centralization.

The proportion of housing units or population within some distance of the center can be a meaningful measure of centralization if the threshold distance is made dependent upon the size of the urban area. This could be based on the total land area or on the mean area weighted distances of tracts from the center. (The latter would be preferable as the use of the former is unduly affected by departures from a circular urban area.) However, such a measure makes very inefficient use of the data on the distribution of units or persons in calculating centralization. The continuous measure of distance to the center for each census tract is reduced to a dichotomy—less than or greater than the threshold distance. Variations within these areas are not reflected in the measure.

The absolute centralization index presented by Massey and Denton (1988) constitutes a major improvement over the proportion within some distance measures. To create the index, the census tracts or other spatial units are ordered by distance from the center and the cumulative proportions of housing units or population and of area are calculated. The index is best understood by considering the plot of the cumulative proportion of the units or persons versus the cumulative proportion of area. If the units are uniformly distributed, the values will fall on the diagonal from (0, 0) to (1, 1). With centralization, the cumulative proportion of units will increase more rapidly than the area, resulting in a curve from (0, 0) to (1, 1) that is above the diagonal and concave downwards. The greater the degree of centralization, the farther the curve will be from the diagonal. The index is then the area between the curve and the diagonal as a proportion of the total area above the diagonal, 0.5. This is analogous to the Lorenz curve and the Gini coefficient used for the measurement of income inequality, though with the curve below the diagonal in that case.

The absolute centralization index is a reasonable measure of centralization, but it has two shortcomings. First, it utilizes the distances from the tracts to the center as an ordinal variable, using the ordering to calculate the cumulative units and areas. While far superior to the binary division of tracts into those with distances less than or greater than some value, this still makes inefficient use of the actual distances to the center. Small differences have the same effect as large.

The second limitation is that the value of the absolute centralization index has no interpretable meaning beyond the maximum value of 1 meaning all units are located at the center, 0 for a uniform distribution, and -1 for all units at the edge of the area

(complete decentralization). Massey and Denton (1988) erroneously state that the index can be interpreted as the proportion of units that would have to be relocated to produce a uniform distribution. This is not only false for this index but cannot be true for any reasonable measure of centralization. Start with any nonuniform distribution of units with some degree of centralization. Any relocation of units from one tract to a tract that is closer to the center or farther away should result in a change in the measure of centralization. But if the relocation is such that the number of units in both the giving and receiving tract both before and after relocation is less than or equal to the numbers for those tracts for a uniform distribution, then that relocation would have no effect on the number of units that would need to be relocated to produce a uniform distribution.

The centralization index developed for this research, called the centralization ratio, addresses the issues of making maximum use of the data and provides a meaningful interpretation. Calculation of the index begins with the mean distance housing units are located from the center:

$$\overline{s} = \frac{\sum u_i s_i}{\sum u_i}$$

where u_i is the number of housing units in tract i and s_i is the distance from the centroid of tract i to the center. This will be compared to the mean distance to the center for housing units uniformly distributed across the urban area, with the same density in each tract:

$$\overline{s}_U = \frac{\sum D a_i s_i}{\sum D a_i} = \frac{\sum a_i s_i}{\sum a_i}$$

where D is the overall density of the urban area. The ratio of the actual mean distance to the center to the mean distance if units are uniformly distributed is $\overline{s}/\overline{s}_U$. If the units are actually distributed uniformly, the value of the ratio will be 1. As centralization increases, the actual mean distance to the center decreases and the ratio will decline to a minimum value of 0 if all units are located at the center. This ratio is thus a measure of decentralization, increasing as units are located farther from the center. To create a measure of centralization, this ratio is subtracted from 1, giving the centralization ratio CR:

$$CR = 1 - \frac{\overline{s}}{\overline{s}_U} = \frac{\overline{s}_U - \overline{s}}{\overline{s}_U}$$

The centralization ratio ranges from 0 for a uniform distribution of units to 1 if all units are located in the center. If units are actually more decentralized than a uniform distribution (not likely for an urban area), the centralization ratio can be negative. As can be seen from the final term above, the centralization ratio can be interpreted as the proportional reduction of the mean distance of housing units from the center as compared with the mean distance if they were uniformly distributed.

Galster and his colleagues employed measures of centralization using the mean distance of housing units from the center in two papers presenting multiple measures of urban sprawl. In Galster, *et al.* (2001), they used the inverse of the mean distance standardized by the square root of the area of the Urbanized Area. This would be appropriate for normalizing the distance for circular urban areas but is less satisfactory for irregularly shaped areas. A subsequent paper, Cutsinger, *et al.* (2005), measured centralization using the ratio of the mean distance of their cell centroids from center to the mean distance of housing units from the center. Since the cells had a uniform size, the mean distance of cell centroids is the distance if housing units were uniformly distributed. This ratio is the inverse of the ratio used here and increases with greater centralization. Its value is 1 for a uniform distribution and is unbounded as centralization increases. The value lacks a readily interpretable meaning, unlike the measure used here, described above.

As a practical matter, each of the measures discussed is a reasonable measure of the centralization of housing units. When creating the dataset and the various measures of urban patterns, I calculated for a sample of the urban areas in 2010 multiple measures of centralization. These included the proportion of units within one-half the mean distance to the center for a uniform distribution, the Massey and Denton absolute centralization index, and the centralization ratio. The values were very highly correlated, with the highest correlation coefficient, over 0.98, between the absolute centralization index and the centralization ratio.

Data and Methods

This research uses a dataset that was developed with data on numbers of housing units in census tracts for large urban areas in the United States from 1950 to 2010. The tracts for urban portions of metropolitan areas were identified within the Combined Statistical Areas (CSAs) as delineated by the Office of Management and Budget for 2013 (U.S. Bureau of the Census 2013). CSAs were used rather than the more commonly employed Metropolitan Statistical Areas (MSAs) as it was felt they more properly represented the full extent of the metropolitan areas, including those instances in which 2 or 3 MSAs should more properly be considered to be parts of a single area. For those MSAs which were not incorporated into a CSA, the MSA was used.

The 59 CSAs and MSAs with 2010 populations over one million were selected for the creation of the dataset. A number of these areas had multiple large centers associated with separate urban areas than had grown together. This posed the issue of identifying those cases in which a second or third urban area could be considered sufficiently large in relation to the largest area to be considered as an additional center. The decision was made by comparing the population of census Urbanized Areas (either from the current census or the last census in which the areas were separate) with the largest area. A center was considered to be an additional center if its population were greater than 28 percent of the population of the largest area. The three areas included with the lowest percentages were Akron (with Cleveland), Tacoma (with Seattle), and Providence (with Boston).

The primary data source for this research was the Neighborhood Change Database developed by the Urban Institute and Geolytics (2003). This unique dataset provides census tract data from the 1970 through 2000 censuses, with the data for 1970 through 1990 normalized to the 2000 census tract boundaries. Population and housing unit data from the 2010 census were added by aggregating the counts from the 2010 census block data (U.S. Bureau of the Census 2012).

Housing unit densities—the numbers of housing units divided by the land areas of the tracts in square miles—are used in this research rather than the more commonly employed population density measure for two reasons. Housing units better represent the physical pattern of urban development as they are relatively fixed, while the population of an area can change without any changes in the stock of housing. Other studies of urban patterns have made similar arguments for choosing housing units over population, for example Galster, *et al.* (2001); Theobald (2001); Radeloff, Hammer, and Stewart (2005); and Paulsen (2014).

Using housing units also allows the extension of the analysis to census years prior to 1970. The census includes data on housing units classified by the year in which the structure was built, and these data are included in the Neighborhood Change Database. The 1970 year-built data can be used to estimate the numbers of housing units present in the census tracts for 1940, 1950, and 1960. Several prior studies have used the housing units by year-built data to make estimates for prior years in this manner, though they have used more recent census data to make the estimates, not the earlier 1970 census data (Radeloff, *et al.* 2001; Theobald 2001; Hammer, *et al.* 2004; Radeloff, Hammer, and Stewart 2005).

Sources of error in these housing unit estimates for earlier years from the year-built data arise from imperfect knowledge of the year in which the structure was built and from changes to the housing stock due to demolitions, subdivisions, and conversion to or from nonresidential uses. These errors increase for estimates farther back in time. Numbers of housing units for 1970 to 1990 were estimated from the 2000 year-built data and compared with the census counts in the Neighborhood Change

Database. The judgment was made that estimates 2 decades back involved acceptable levels of error, but this was not the case for 3 decades back. As a result, the decision was made to use the housing unit estimates for 1950 and 1960 but not for 1940.

Urban areas were defined for each census year from 1950 to 2010 consisting of those contiguous tracts meeting a minimum housing unit density threshold. For the definition of Urbanized Areas for the 2000 and 2010 censuses, a minimum population density of 500 persons per square mile was required for a block or larger area to be added to an Urbanized Area (U.S. Bureau of the Census 2002, 2011). Using the ratio of population to housing units for the nation in 2000 of 2.34 persons per unit, a density of 500 persons per square mile is almost exactly equivalent to 1 housing unit per 3 acres or 213.33 units per square mile. This was used as the minimum urban density threshold. Note that this is a measure of gross density, not lot size, as the areas of roads, nonresidential uses, and vacant land are included.

The location of the CBD must be specified to measure distances. One of the only efforts by the Census to do so came in a report for the 1982 economic censuses (U.S. Bureau of the Census 1983). This lists the census tracts comprising the CBD for many larger cities. This information was used to identify the CBD tracts for those urban areas included and for which the tract numbering and boundaries were the same for 2000. For the other urban centers, the tract or tracts for the CBD were identified by determining the location of the city hall or other major government buildings and examining the pattern of major roads, which generally converge on the CBD. The centroid of the CBD tract or tracts was taken as the center. Distances to the center were measured in miles to the centroids of each of the census tracts in the urban area.[1]

The centralization ratios were produced by first calculating the mean distance of housing units from the center by summing the distance times the number of units in each tract and dividing by the total number of units. This was then divided by the mean distance for a uniform distribution, which is the summation of the distance times the land area of each tract, divided by the total land area of the urban area.

For the 16 areas with 2 or 3 centers, the centralization ratios were separately calculated for the portions of the urban area associated with each center. These were then combined, taking the mean of those ratios weighted by the number of housing units in the tracts assigned to each center. This, of course, required the partitioning of those urban areas and the assignment of each tract to one of the multiple centers. For those areas for which separate Urbanized Areas were delineated for the census in 2010, the tracts were assigned using the Urbanized Area boundaries. For the urban areas with a single Urbanized Area encompassing the entire urban area, tracts were assigned to the nearest center. In a few instances, modifications were made based on the particular

[1] More detail on the construction of the dataset and the delineation of the urban areas is provided in Ottensmann (2014).

geography of the area. For example, for Tampa-St. Petersburg, tracts on either side of Tampa Bay were assigned to the center on that side, even though in a few instances they were actually closer to the other center.

Centralization Over Time

Values for the centralization ratio were computed for each of the 59 urban areas as defined for each census year from 1950 to 2010. Table 1 presents the mean, minimum, and maximum centralization ratios across the areas for each year.

Table 1. Centralization Ratios for 59 Urban Areas, 1950-2010.

Year	Mean Centralization Ratio	Minimum Centralization Ratio	Maximum Centralization Ratio
1950	0.246	0.102	0.515
1960	0.243	0.091	0.442
1970	0.236	0.110	0.463
1980	0.206	0.053	0.450
1990	0.188	0.048	0.434
2000	0.184	0.045	0.435
2010	0.184	0.078	0.458

The mean centralization ratio for 1950 was about 0.25, so that on average the mean distance housing units were located from the center was about a quarter less than if the units had been uniformly distributed. However, the ratios varied widely across the areas, from a low of 0.10 to a high value of 0.52. The mean ratio dropped steady from 1950 to 2010, down to a value of 0.18 for the last two years. This is consistent with the widely observed suburbanization and decentralization occurring in American urban areas over this period.

Two tests showed the decline of the centralization ratios to be statistically significant. When all of the values from 1950 to 2010 were regressed on the year, the regression coefficient for the year was significant at the 0.001 level. Likewise, the simple test of regressing the mean ratios in Table 1 on time yielded the same significance.

Not surprisingly, the maximum centralization ratio from 1960 through 2010 was for the New York area. The values were clear outliers in the distributions. (For 1950, the

El Paso area had the highest level of centralization. The El Paso area was very small at that time, with only about 29,000 housing units, which may have contributed it having this unusual value.)

The notion of centralization is obviously more complex for those areas with 2 or 3 centers as compared with single-center areas. As noted above, centralization ratios were calculated separately for the portions of the urban area assigned to each center. These values were combined by calculating a weighted average to produce a value for the overall centralization ratio for the entire urban area. This raises the question as to whether values of the centralization ratios might differ for the multi-centered areas as compared with the areas with only 1 center. To examine this, *t*-tests for the difference of means were performed for, each year, comparing the mean for the 16 areas with 2 or 3 centers with the mean for the 43 single-centered areas. In every case, the differences were not statistically significant. Indeed, for 6 of the 7 tests, the *t*-values were actually less than 1, indicating the complete absence of any statistical significance.

The next question involves the changes in the centralization ratios from 1950 to 2010. The mean change across the 59 areas was -0.062, a change of -18.1 percent. But there was wide variation in the changes among the urban areas. The greatest decline in the centralization ratio was a change of -0.42, a drop of 81.6 percent from 1950 to 2010. On the other hand, in the other direction, one area experienced an increase in the centralization ratio over this period of +0.113, which was a doubling of +103.1 percent. So it is clear that while decentralization was occurring on average from 1950 to 2010, this was certainly not the case for all urban areas.

With the wide variation, it is useful to look at a histogram of the distribution of change over the period. Figure 1 shows the numbers of areas by the change in the centralization ratio from 1950 to 2010. The overwhelming majority of the areas experienced a change in the ratio between -0.15 and +0.10. Only 9 have the areas had changes outside this range. Of the 59 urban areas, 14 experienced positive change, that is, increases in the level of centralization. So decentralization was clearly the dominant experience in urban areas over the period, though it was not universal. Looking at the distribution of percentage changes (histogram not shown), most areas had changes ranging from a drop of 60 percent to an increase of 40 percent. Only 6 areas had changes that were more extreme.

Which urban areas—and which types of areas—had the highest and lowest levels of centralization? The five areas with the highest centralization ratios in 2010, in descending order, were New York, Chicago, Philadelphia, Milwaukee, and Rochester. These are all, of course, older urban areas in the Northeast and Midwest and are among the larger areas, except for Rochester. The next 5 are similar areas with the exception of Los Angeles, which had the ninth highest centralization ratio. Most of the areas with the high levels of centralization in 2010 also had relatively high levels of centralization in 1950. The list of the areas with the highest levels in 1950 did include several smaller

areas that experienced the greatest drops in their centralization ratios to be among the least centralized in 2010.

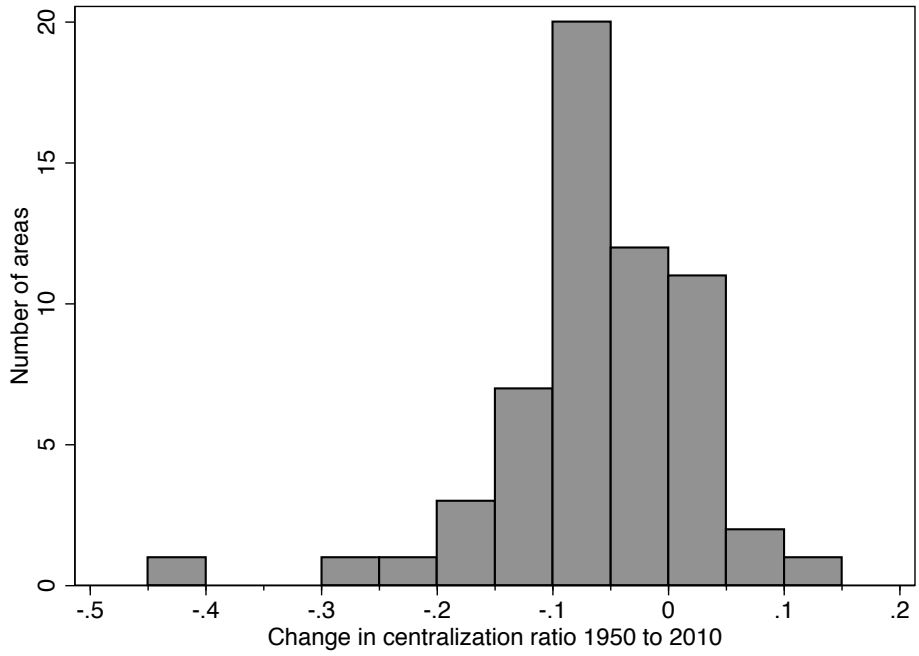

Figure 1. Histogram of Change in Centralization from 1950 to 2010.

The 5 areas with the lowest levels of centralization in 2010, in increasing order from the most decentralized, were Tampa-St. Petersburg, El Paso (which was highest in 1950), Jacksonville, Las Vegas, and Albuquerque. All were areas in the South or the Southwest and were not among the larger urban areas. This continued to be the pattern for the next most decentralized with the exception of Kansas City. Areas with low levels of centralization in 1950 were similar areas.

These results suggest that the region of the country and perhaps the size of the urban area may be associated with current levels of centralization. Table 2 presents the mean, minimum, and maximum centralization ratios in 2010 for the urban areas in each of the four census regions The urban areas in those regions have distinctly different levels of centralization, with the mean being highest for the Northeast, followed by the Midwest and the West, with the South having the lowest mean centralization ratios. This same ordering holds for the minimum values in each region and almost holds for the maximum values, with a small reversal for the South and the West. (The maximum for the South would be lower if Washington-Baltimore were excluded. That area is in

Table 2. Centralization Ratios by Region for 59 Large Urban Areas in 2010.

Year	Mean Centralization Ratio	Minimum Centralization Ratio	Maximum Centralization Ratio
Northeast	0.257	0.152	0.458
Midwest	0.203	0.118	0.307
South	0.157	0.078	0.253
West	0.169	0.109	0.237

the South as defined by the census, but is obviously one of the major areas comprising the Northeast urban corridor.) A one-way analysis of variance confirms that the differences in the means across the regions is statistically significant at the 0.001 level. The dramatic differences in levels of centralization by region is emphasized by the box plots in Figure 2 showing little or no overlap between the interquartile ranges of the Northeast, Midwest, and South.

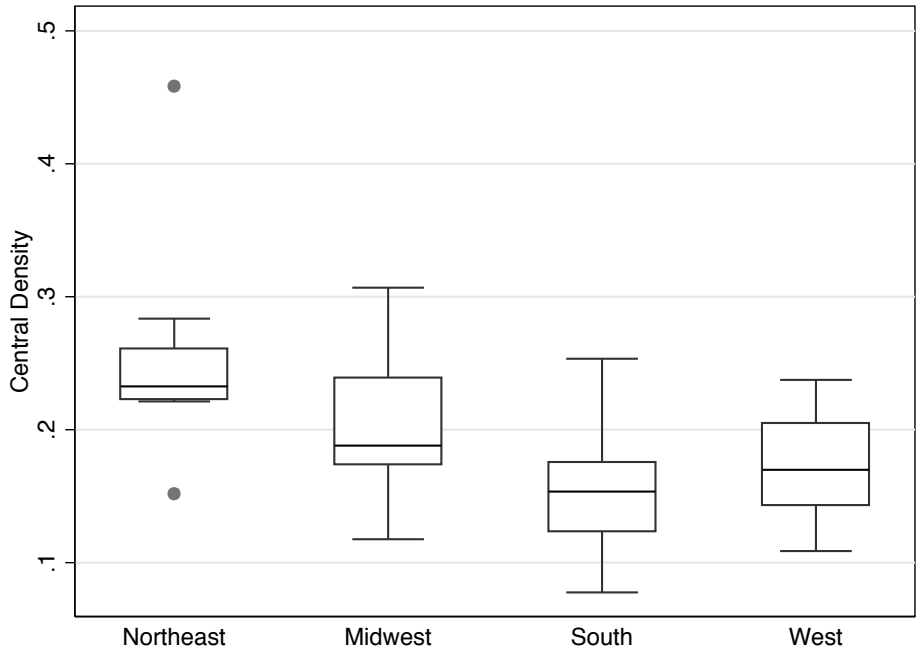

Figure 2. Box Plots of Centralization Ratios in 2010 by Region.

Now for the relationship between centralization and the size of the urban area. The correlation between the centralization ratio in 2010 and the population in the urban area is 0.62, which is statistically significant at the 0.001 level. New York, however, is an outlier with respect to both centralization and population. When New York is dropped, the correlation coefficient declines to 0.39, still statistically significant but now just at the 0.01 level. Centralization and population are directly related, but the strength of that relationship may not be as great as the initial correlation suggested.

To look at how centralization varies across areas of different sizes in a manner comparable to that used for looking at regions, the urban areas were divided into quartiles by population. Summary values for the centralization ratio were calculated for the urban areas in each population quartile. These are shown in Table 3. Centralization is greatest for the top quartile, those urban areas with populations over 2.8 million. There was little difference and no pattern to the values for the first three quartiles. Perhaps it is reasonable to find greater centralization in the largest areas where distance to the center becomes more of an issue. The differences in means across the population quartiles are statistically significant, but only at the 0.05 level.

Table 3. Centralization Ratios by Population Quartiles for 59 Large Urban Areas in 2010.

Quartile, Urban Area Population in 2010	Mean Centralization Ratio	Minimum Centralization Ratio	Maximum Centralization Ratio
First quartile 412,453-736,494	0.171	0.095	0.261
Second quartile 805,637-1,337,417	0.161	0.098	0.230
Third quartile 1,382,783-2,614,441	0.181	0.078	0.275
Fourth quartile 2,852,230-19,889,896	0.223	0.115	0.458

Conclusions

Levels of centralization vary widely. For the New York area, the most centralized, the centralization of housing units reduces their mean distance to the center by nearly half compared with a uniform distribution. At the other extreme, the reduction for the least centralized areas can be less than 10 percent.

Mean levels of centralization across the 59 urban areas declined significantly from 1950 to 2010. The increased decentralization over this period was far from universal, however, with 14 areas actually experiencing increases in levels of centralization.

The urban areas with the highest levels of centralization tend to be the larger areas located in the Northeast and Midwest, though there are some exceptions. The least centralized areas, on the other hand, were smaller and generally located in the South and Southwest. Differences in levels of centralization by region were very clear and highly significant. Average levels of centralization were highest for urban areas in the Northeast and lowest in the South. Areas in the Midwest and West fell between, in that order. The relationship of centralization to the size of the area was more ambiguous. The correlation of the centralization ratio with population was statistically significant, as were the differences in means across urban areas arrayed by population quartiles. But there were only small differences in the means for the smallest 3 quartiles. Only the largest urban areas showed higher mean centralization.

The examination of the performance of the negative exponential model of density decline from 1950 to 2010 also showed some patterns related to region and the size of the area (Ottensmann 2016). The following paper will directly address the relationship between the estimated parameters for the negative exponential model and the levels the centralization ratio examined in this paper.

References

Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998. Urban spatial structure. *Journal of Economic Literature* 36, 1 (September): 1426-1464.

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 (August): 235-259.

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 Kahn. 2001. *Job Sprawl: Employment Location in U.S. Metropolitan Areas*. Washington, DC: Brookings Institution.

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. Amsterdam: Elsevier North-Holland, pp. 2481-2527.

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

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

Mieszkowski, Peter, and Edwin S. Mills. 1993. The causes of metropolitan suburbanization. *Journal of Economic Perspectives* 7, 3 (Summer): 135-147.

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

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

Ottensmann, John R. 2014. Urban patterns dataset description. Available at https://urbanpatternsblog.files.wordpress.com/2016/12/urban-patterns-dataset-description.pdf.

Ottensmann, John R. 2016. The negative exponential decline of density in large urban areas in the U.S., 1950-2010. Available at https://urbanpatternsblog.files.wordpress.com/2016/12/negative-exponential-density-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 (September): 2629-2645.

Radeloff, Volker C., Roger B. Hammer, Paul R. Voss, Alice E. Hagen, Donald R. Field, and David J. Mladenoff. 2001. Human demographic trends and landscape level forest management in the northwest Wisconsin Pine Barrens. *Forest Science* 47, 2 (May): 229-241.

Radeloff, Volker C., Roger B. Hammer, and Susan L. Stewart. 2005. Rural and suburban sprawl in the U.S. midwest from 1940 to 2000 and its relation to forest fragmentation. *Conservation Biology* 19, 3 (June): 793-805.

Theobald, David M. 2001. Land-use dynamics beyond the urban fringe. *Geographical Review* 91, 3 (July): 544-564.

U.S. Bureau of the Census. 1983. *1982 Economic Censuses. Reference Series. Geographic Reference Manual.* EC82-R-1. Washington, DC: U.S. Government Printing Office, Table 9, pp. 309-322.

U.S. Bureau of the Census. 2002. Urban area criteria for the 2000 Census. *Federal Register* 67, 51 (Friday, March 15): 11663-11670.

U.S. Bureau of the Census. 2011. Urban area criteria for the 2010 Census. *Federal Register* 76, 164 (Wednesday, August 24): 53030-53043.

U.S. Bureau of the Census. 2012. Census block shapefiles with 2010 Census population and housing unit counts. Downloaded from http://www.census.gov/geo/www/tiger/tgrshp2010/pophu.html on September 29, 2012.

U.S. Bureau of the Census. 2013. Core based statistical areas (CBSAs) and combined statistical areas (CSAs) 2013 delineation file. February 2013. Downloaded from https://www.census.gov/population/metro/data/def.html on February 13, 2014.

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