Negative Exponential Model Parameters and 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
February 2017

Abstract

The change in the negative exponential density gradient has often been used as a measure of urban decentralization, with the gradient itself also being taken as a measure of the degree of centralization for comparing urban areas. Changes and levels of estimated negative exponential model parameters for 43 areas from 1950 to 2010 are compared with a “pure” measure of centralization, the centralization ratio. The proportional change in the density gradient is indeed significantly related to the change in the centralization ratio, while the relationship of the change in central density to centralization is inconsistent. However, the level of the density gradient is not significantly related to the level of centralization, except that controlling for the size of the urban area produces a significant relationship. The central density is related to the level of centralization.

Introduction

The previous paper in this pair (Ottensmann 2017) examined levels and changes in the centralization of housing units in large urban areas in the United States from 1950 to 2010. That paper used a “pure” measure of centralization—the centralization ratio—that does not depend on any assumptions regarding the distribution of housing units and housing unit densities in the urban area. In a contrasting approach, numerous researchers have studied decentralization using the density gradient estimated assuming a negative exponential decline of density with distance from the center. This raises the issue of the appropriateness the density gradient as a measure of centralization and decentralization in urban areas.

This paper compares the centralization ratio with the estimated parameters of the negative exponential model, the density gradient and the central density. The next section discusses the relationship between the density gradient and the centralization of housing units and population. The following part describes the data and methods used in this study. The results of the comparison of both changes in the parameters of the
negative exponential model and centralization and the magnitudes of those parameters and centralization are presented.

Changes in the Negative Exponential Model Density Gradient and Centralization

Numerous studies have shown that population and housing unit densities tend to decline as a negative exponential function of distance from the centers of urban areas. The monocentric economic model of urban structure developed by Muth (1969) and Mills (1972) predicts this pattern of density decline with some simple but reasonable assumptions. The negative exponential function requires that two parameters be estimated, the density at the center of the urban area and the density gradient, the rate of decline of density with distance from the center. Because the density gradient reflects the key aspect of the monocentric model, the tradeoff people make between accessibility to the center (reducing transportation costs) and more space for their residences (lower densities), its value has typically received greater attention.

Since the middle of the twentieth century (at least), increasing suburbanization and the decentralization of urban population and housing has been occurring and has received extensive attention from urban researchers. Many of those looking at urban decentralization have chosen to focus on the decline of the density gradient estimated for the negative exponential model. This has been especially true for economists considering these issues, which is not surprising given the prediction of the economic monocentric model. However, others have also chosen to focus on the density gradient. In the following paragraphs, a small number of such studies of decentralization using the density gradient are considered.

The primary objective of these studies examining decentralization has been to consider changes in levels of centralization over time. They have therefore naturally focused primarily on decreases in density gradients over time as measures of decentralization. The assumption has been that greater decentralization would be associated with larger declines in density gradients. Even if this assumption were valid, however, it does not logically follow that the level of the density gradient will be a good measure of the level of centralization in an urban area. Yet numbers of researchers have gone on to make this further claim.

One of the first to consider the decline in the density gradient as an indicator of decentralization was Mills (1972), one of the developers of the monocentric model. He looked at the declines of density gradients for a sample of urban areas, for a few going back as far as 1880, showing a long-term trend of density gradient decline and decentralization.

Edmonston (1975, p.5) defined suburbanization “as the diminishing of the slope or the gradient.” This is explicitly referring to the change in the density gradient over time. But just a few paragraphs later, he suggested that “the density gradient offers a
measure of population concentration superior to other indices.” That is, the level of the
gradient at a single point in time can be used to measure concentration or centralization,
the term I am using here. And in his research, he compared the levels of density
gradients at specific points in time in cross-sectional analyses.

Edmonston, Goldberg, and Mercer (1985, p. 211) described the density
gradient as “a readily grasped measure of population concentration.” They did go on to
say that it can also be used to consider population decentralization over time. Their
objective in this paper was to compare the estimated parameters of the negative
exponential model—both the density gradient and the central density—between urban
areas in Canada and the United States. They had estimates for multiple years and
observed the decline in both the gradients and the central densities over time for urban
areas in the two countries. Unlike many authors, they referred to the decreases in the
central densities as well as the gradients as indicators of urban decentralization.

Mieszkowski and Mills (1993) explicitly described the density gradient as a
measure of decentralization. They noted that “many writers have used estimates of the
density gradient... as a measure of the degree of decentralization.” But then they
proceeded to say that decreases in the gradients have been used as indicators of
increased decentralization or suburbanization. So references were being made to both
the level of the density gradient and its changes over time. They reviewed many studies
that have used the density gradient, looking first at the changes and trends in density
gradients over time. They also considered studies that compared the magnitudes of
density gradients between countries.

Finally, Anas, Arnott, and Small (1998) described the density gradient as “a
useful index of population centralization.” In considering evidence from studies using
the density gradient to examine urban decentralization, they focused on changes in the
gradient over time. They did note a strong negative correlation between density
gradients and total population, concluding from this that larger cities are more
decentralized. This is despite their prior expectation from their version of the
monocentric model that the gradient would be positively related to the size of the urban
area.

It is easy to see where the notion that the density gradient can be used as a
measure of centralization arises. For an area experiencing decentralization with no
change in the population or number of housing units, population or housing units must
be decreasing near the center and increasing farther out. If the distributions conform to
the negative exponential model, then the central density must be lower and the density
curve is flattened, implying a reduction in the density gradient. This is illustrated in
Figure 1 by the two curves showing the decline of density with distance from the center
before (A) and after (B) the decentralization.
But from the fact that decentralization may be associated with the decline in the density gradient it does not logically follow that the density gradient is necessarily a good measure of the degree of centralization of population or housing units in an urban area. Consider the density curves in Figure 2 for 2 urban areas with the same central density (or for a single urban area at 2 points in time). The higher curve A is for an area with a larger population and number of housing units, given the area under the curve is greater and extends farther from the center. It is flatter and therefore has a smaller density gradient than the smaller city B. And yes, more people and housing units are located farther from the center of the city in the larger city A. The average distance people and housing units are located from the center is greater. This is a larger urban area, extending out much farther from the center. But is the area less centralized? One would obviously expect in a larger area, the average distance to the center would be greater. It is not clear from the density curves in Figure 2 that the larger urban area with the lower density gradient is less centralized.
To reinforce this point and set the stage for the analysis that follows, using the measure of centralization employed in this paper, the New York area has by far the highest level of centralization of any large urban area. Most people would hardly find this to be surprising. But the density gradient for New York in 2010 was similar to (actually slightly less than) Albuquerque. And it was much less than half the density gradient for Dayton.

This paper examines the relationship of a measure of centralization to the parameters estimated for the negative exponential model for 43 large urban areas in the United States from 1950 to 2010. The following section describes the data used, the measure of centralization, and the estimation of the negative exponential model parameters. This is followed by the results of the comparisons of both the changes in centralization with the changes in the parameters and the levels of centralization with the levels of the parameters.

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). Because the estimation of the negative exponential model requires distances to a single center (for the monocentric model), only the 43 urban areas with a single center are included in the analysis in this paper.

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

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 conversions 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 CBDs 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 from the centroids of each of the census tracts in the urban area.¹

The earlier paper on the centralization of urban areas (Ottensmann 2017) examined alternative measures of centralization and presented the arguments for using the centralization ratio that was developed for this research. The centralization ratio is the proportional reduction in the mean distance of housing units or population to the center compared with the mean distance if they were uniformly distributed. The formula for the centralization ratio CR is:

¹ More detail on the construction of the dataset and the delineation of the urban areas is provided in Ottensmann (2014).
where \( u_i \) is the number of housing units in census tract \( i \), \( a_i \) is the land area of the tract, and \( s_i \) is the distance from the tract centroid to the center of the urban area. The numerator in the ratio is the mean distance housing units in the area are located from the center. The denominator is the mean distance if housing units were uniformly distributed throughout the urban area, if densities were everywhere the same. If all housing units were actually uniformly distributed, the value of the ratio would be 1. As centralization increases, the mean distance units are located from the center decreases, so the ratio declines to 0 if all housing units were located at the center. Subtracting the ratio from 1 produces an index measuring centralization, varying from 0 for a uniform distribution to 1 with complete centralization. (The value could be negative if units were actually more decentralized than a uniform distribution, not likely for a realistic urban area.)

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.

The monocentric model posits the negative exponential decline of density with distance from the center:

\[
D_i = D_0 e^{-\beta s_i}
\]

where \( D_i \) is the density in tract \( i \), \( s_i \) is the distance from the center to tract \( i \), \( D_0 \) is the density at the center, \( \beta \) is the density gradient, and \( e \) is the base of the natural logarithms. For estimation, this is transformed by taking the natural logarithm of both sides,

\[
\ln(D_i) = \ln(D_0) - \beta s_i
\]

producing a linear relationship that can be estimated using ordinary least squares regression. Log of density is regressed on distance to estimate the central density and density gradient parameters.
Relationships between Negative Exponential Model Parameters and Centralization

Values for the centralization ratio have been calculated and the parameters of the negative exponential model estimated for the 43 large urban areas for each census year from 1950 to 2010. The analysis first examines the relationships between the changes in the exponential model parameters and the changes in centralization from decade to decade. Following is the analysis of the relationships between the parameter estimates themselves and the values of the centralization ratios for each year.

Changes in Parameters and Centralization

To estimate the relationship between the changes in the estimated negative exponential model parameters and the changes in centralization, the proportional changes in the parameters were regressed on the proportional changes in the centralization ratios from one decade to the next. (Using proportional changes yielded slightly more robust results than the absolute changes.) Table 1 presents the results for both the density gradient and the central density, with the regression coefficients for the exponential model parameter and the coefficient of determination $R^2$. Regressions are reported for the changes for each decade and for the pooled regression including changes for all areas and all decades.

Table 1. Regressions of Proportional Change in Density Gradient and Central Density on Proportional Change in Centralization Ratio.

<table>
<thead>
<tr>
<th>Period</th>
<th>Proportional Change in Density Gradient</th>
<th>Proportional Change in Central Density</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regression Coefficient</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1950-1960</td>
<td>0.608 ***</td>
<td>0.298</td>
</tr>
<tr>
<td>1960-1970</td>
<td>0.955 ***</td>
<td>0.454</td>
</tr>
<tr>
<td>1970-1980</td>
<td>0.465 ***</td>
<td>0.304</td>
</tr>
<tr>
<td>1980-1990</td>
<td>0.846 ***</td>
<td>0.797</td>
</tr>
<tr>
<td>1990-2000</td>
<td>0.194</td>
<td>0.041</td>
</tr>
<tr>
<td>2000-2010</td>
<td>1.120 ***</td>
<td>0.326</td>
</tr>
<tr>
<td>All Cases (Pooled)</td>
<td>0.695 ***</td>
<td>0.434</td>
</tr>
</tbody>
</table>

*p < 0.05,  **p < 0.01,  ***p < 0.001
The proportional change in the density gradient is definitely related to the proportional change in the centralization ratio. All but one of the regression coefficients are significant at the 0.001 level. $R^2$ values fall between 0.3 and 0.8 for all but one of the decades. And since these are proportional changes, the regression coefficients are not dependent on units, so a coefficient of 1.0 would be associated with perfect correspondence between change in the gradient and change in centralization. Numbers of the coefficients are quite close to 1.0 and all but one have values of at least 0.47. So it is reasonable to conclude that the change in the density gradient can be considered to be a reasonable measure of change in centralization.

Proportional change in central density, on the other hand, is a far less reasonable measure of change in centralization. Only 3 of the 6 decade-change regressions are statistically significant (and the pooled regression has 6 times the number of cases, making significance far easier to achieve). While 2 of the $R^2$ values exceed 0.3, 3 are exceedingly small, less than 0.1. So while decreases in the central density may be associated with decentralization, this does not consistently seem to be the case.

Levels of Parameters and Centralization

The analysis now shifts to the examination of the relationships between the levels of the exponential model parameters, the density gradient and the central density, and the levels of centralization as measured by the centralization ratio. The question being addressed is this: Can the density gradient (or perhaps the central density) be used as a measure of the centralization of housing units or population in an urban area at a single point in time? Does it make sense to compare density gradients across urban areas as a comparison of relative levels of centralization?

Following the same procedure as before, the parameters estimated for the negative exponential model are regressed on the centralization ratio, for each year and for all of the observations taken together. Table 2 presents the results for the density gradient, with the regression coefficients and $R^2$ values using only the gradient to predict centralization in the left 2 data columns. The results are quite unambiguous: There is virtually no relationship between the density gradient and the centralization ratio. The regression coefficients are not statistically significant except for 1 year with significance at the 0.05 level. At most, the gradient accounts for about 10 percent of the variation in centralization, with $R^2$ values around 0.1 or less. Assuming the centralization ratio to be a reasonable measure of centralization, the density gradient indicates little about the level of centralization at a given point in time. (See the appendix to this paper for some additional evidence supporting this conclusion.)
Table 2. Regressions of Density Gradient on Centralization Ratio without and with Log of Number of Housing Units in Urban Area.

<table>
<thead>
<tr>
<th>Year</th>
<th>Density Gradient Only</th>
<th></th>
<th></th>
<th>Density Gradient Plus Log Number of Housing Units</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Density Gradient Regression Coefficient</td>
<td>$R^2$</td>
<td></td>
<td>Density Gradient Regression Coefficient</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1950</td>
<td>0.065</td>
<td>0.015</td>
<td></td>
<td>0.301 ***</td>
<td>0.331</td>
</tr>
<tr>
<td>1960</td>
<td>0.172</td>
<td>0.075</td>
<td></td>
<td>0.475 ***</td>
<td>0.524</td>
</tr>
<tr>
<td>1970</td>
<td>0.048</td>
<td>0.004</td>
<td></td>
<td>0.465 ***</td>
<td>0.608</td>
</tr>
<tr>
<td>1980</td>
<td>0.183</td>
<td>0.023</td>
<td></td>
<td>0.638 ***</td>
<td>0.592</td>
</tr>
<tr>
<td>1990</td>
<td>0.297</td>
<td>0.040</td>
<td></td>
<td>0.780 ***</td>
<td>0.538</td>
</tr>
<tr>
<td>2000</td>
<td>0.487 *</td>
<td>0.104</td>
<td></td>
<td>0.938 ***</td>
<td>0.658</td>
</tr>
<tr>
<td>2010</td>
<td>0.395</td>
<td>0.059</td>
<td></td>
<td>0.924 ***</td>
<td>0.597</td>
</tr>
<tr>
<td>All Cases (Pooled)</td>
<td>0.199</td>
<td>0.106</td>
<td></td>
<td>0.555 ***</td>
<td>0.498</td>
</tr>
</tbody>
</table>

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

The clue to what is happening comes from the observation made repeatedly that the density gradient is inversely related to the size of the urban area (see, e.g., Mills 1972; Edmonston 1975; Edmonston, Goldberg, and Mercer 1985; Anas, Arnott, and Small 1998). This is especially noteworthy given that the monocentric model suggests that the relationship of the gradient to population might be positive (Mills 1972; Anas, Arnott, and Small 1998). For the 43 urban areas considered here, the density gradient is negatively correlated with the log of the number of housing units in each year and for all observations pooled, with all coefficients significant at least at the 0.01 level.

The density gradient is affected by the size of the urban area, decreasing as areas become larger. Then it is possible that controlling for the effect of size when using the density gradient to predict centralization may allow a relationship of the gradient to centralization to emerge. The right 2 columns of Table 2 show the regression coefficients for the density gradient and the $R^2$ values for regressions in which the log of the number of housing units in the urban area has been included as a second independent variable. The regression coefficients for the density gradient are all much larger than when not controlling for size. All are significant at the 0.001 level. $R^2$ values range from a low of 0.33 to a high of 0.66, with all but one of the year values over 0.5. After the effect of the size of the urban area is removed, the density gradient is related to the centralization
ratio. This also explains why a significant relationship exists between the change in the density gradient and the change in the centralization ratio. By taking the difference, most of the effect of the size of the area is eliminated (aside from the change in size over the period of the change).

Most have focused on the density gradient as a measure of centralization. Far less consideration has been given to the central density, despite the obvious fact that higher central densities imply greater numbers of housing units or persons concentrated near the center. Table 3 presents the results of the regressions of the central densities on the centralization ratios for each year and for all of the observations pooled. As before, the regression coefficients and $R^2$ values using only the central density to predict centralization are in the left 2 data columns. These results are also unambiguous, but in the opposite way from those for the density gradients. All of the regression coefficients are highly significant, with $p$-values less than 0.001. Except for 1950, the $R^2$ values are 0.49 or above, with 3 over 0.70. Central density thus generally accounts for about half or more of the variation in the centralization ratio. So the central density can be considered a reasonable measure of centralization, while the density gradient, which has often been used for that, is not.

Table 3. Regressions of Central Density on Centralization Ratio without and with Log of Number of Housing Units in Urban Area.

<table>
<thead>
<tr>
<th>Year</th>
<th>Central Density Only</th>
<th>Central Density Plus Log Number of Housing Units</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Central Density Regression Coefficient</td>
<td>$R^2$</td>
</tr>
<tr>
<td>1950</td>
<td>0.0148 ***</td>
<td>0.298</td>
</tr>
<tr>
<td>1960</td>
<td>0.0219 ***</td>
<td>0.487</td>
</tr>
<tr>
<td>1970</td>
<td>0.0217 ***</td>
<td>0.708</td>
</tr>
<tr>
<td>1980</td>
<td>0.0259 ***</td>
<td>0.589</td>
</tr>
<tr>
<td>1990</td>
<td>0.0296 ***</td>
<td>0.621</td>
</tr>
<tr>
<td>2000</td>
<td>0.0299 ***</td>
<td>0.701</td>
</tr>
<tr>
<td>2010</td>
<td>0.0297 ***</td>
<td>0.723</td>
</tr>
<tr>
<td>All Cases (Pooled)</td>
<td>0.0240 ***</td>
<td>0.551</td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01, ***p < 0.001
The central density is also related to the size of the area, but in a positive direction. This has been noted, for example, by Mills (1972), Edmonston (1975), and Edmonston, Goldberg, and Mercer (1985). This is also the case for the 43 areas considered here, for all years. Given the findings for the the relationship of the density gradient to the centralization ratio when controlling for the log of the number of housing units, the question arises as to the effect with the central densities. The right 2 columns of Table 3 show the regression coefficients for the central density and the $R^2$ values for regressions in which the log of the number of housing units in the urban area has been included as a second independent variable. The regression coefficients showed little change and the $R^2$ values indicated little or no increase in the fit of the model. So unlike the density gradient, the relationship of central density to the centralization ratio is essentially independent of the size of the urban area.

**Conclusions**

As expected, the change in the density gradient was clearly positively related to the centralization ratio. The widespread use of the decline of the density gradient as a measure of urban decentralization is therefore supported. On the other hand, the change in the central density was not consistently associated with the measure of centralization. This is perhaps not surprising, given the magnitude of some of the decreases in central densities over the period considered (Ottensmann 2016).

However the level of the density gradient was not clearly related to the centralization ratio. The density gradient is inversely related to the size of the urban area. When controlling for the size of the area, the density gradient is then significantly related to centralization. But it is clearly inappropriate to use the magnitude of the density gradient as a measure of the degree of centralization in an urban area. Interestingly, the central density is clearly related to the centralization ratio and would not be an unreasonable indicator of the level of urban centralization.

A skeptic—especially one committed to the use of the monocentric model and the negative exponential density gradient as a measure of centralization—might argue that results showing the density gradient not clearly related to the centralization ratio do not show that the density gradient is not a reasonable measure of centralization. Instead, they only show that two measures offered as measures of centralization yield different result, not which one is the better measure of centralization. This argument fails to address the case made for the centralization ratio as a good measure of centralization in the previous paper (see Ottensmann 2017). And it fails to address the rather inconvenient fact that the three urban areas with the highest centralization ratios in 2010—New York, Chicago, and Philadelphia—areas that most observers would expect to be among the most centralized urban areas in the country—have density
gradients less than the mean density gradient for all 43 large urban areas included in this study.

In developing the dataset and the centralization ratio, multiple alternative measures of centralization were calculated for 2010 for a sample of 12 of the 43 urban areas. The appendix to this paper presents the results of analysis using these alternative measures. The conclusions are clear: The centralization ratio is a good measure of centralization. The density gradient is not significantly related to any of the measures of centralization. The conclusion stands: The density gradient is not a reasonable measure of the level of centralization in an urban area.

References

Appendix. Alternative Measures of Centralization

In developing the dataset used for this study, including the centralization ratio, I chose to compare this measure to alternative measures of centralization as a check on my work. For doing this, I selected a sample of 12 of the urban areas, the 6 with the highest centralization ratios in 2010 and the 6 with the lowest. I calculated a number of
alternative measures of centralization for this sample to observe how they track with the centralization ratio.

This first alternative was the most sophisticated, the absolute centralization index, developed by Massey and Denton (1988) and also used by Lee (2007). For this index, the census tracts are ordered by distance from the center and the cumulative proportions of housing units and land areas are calculated. The greater the degree of centralization, the faster will the cumulative proportion of housing units rise compared with the cumulative proportion of area. The measure of centralization is the difference between these distributions. The measure is analogous to the Gini index commonly used for the measurement of income inequality.

The remaining alternatives are all variations on the very simple idea of centralization as the proportion of all housing units located close to the center. Glaeser and Kahn (2001, 2004) used this for calculating the centralization of both population and employment. However, they used proportions within the same fixed radii for urban all areas, which does not make a lot of sense for urban areas having very different sizes. The determination of the radii of the circles for determining the number of housing units close to the center should be dependent upon the extent of the urban area.

Two approaches were taken to establishing these thresholds. The first starts with the total land area for the urban area. The census tracts closest to the center are selected that include some fraction of the total area. The proportion of housing units within these tracts is taken as the measure of centralization. Different fractions of the total area can be considered. For this exercise, the sets of closest tracts including 25 percent and 10 percent of the total area were used to provide alternative measures of centralization.

The second method starts with the calculation of the mean distance housing units would be located from the center if they were uniformly distributed in the urban area, with the same density in each tract. (This is the denominator in the centralization ratio.) Tracts were identified as being close to the center if their distance to the center were some fraction of that mean distance for the uniform distribution. Values of 75 percent and 50 percent of the mean distance were used and the proportions of housing units in these tracts were the centralization measures. Note that these distances were selected to provide approximately the same areas as the 25 and 10 percent thresholds for total area used in the prior measures, at least for a perfectly circular urban area.

These alternative measures were first compared with the centralization ratio by regressing each of these measures on the ratio in the manner used for the comparisons with the parameters of the negative exponential model. The $R^2$ values indicate the level of correspondence. These values are presented in the first column of data in the following table:
As can be seen, the $R^2$ values are very high, with 4 of the 5 exceeding 0.9 and the lowest still at 0.85. All of the proportions are lower than the absolute centralization index, perhaps not surprising because they are very simple measures based on dichotomizing the set of housing units using an arbitrary fraction. The absolute centralization index makes more effective use of the data, as does the centralization ratio, and the value of $R^2$ for this comparison is an extremely high 0.97.

The centralization ratio and each of the alternatives all seem to be measuring pretty much the same thing. Given the logic of each as a measure of centralization, it would be difficult to conclude that the centralization ratio is not a reasonable measure of centralization.

Given the alternative measures of centralization, it is of course possible to repeat the analysis of the relationship of the density gradient to each of these as was done for the centralization ratio. The density gradient is once again regressed on each measure of centralization, including the centralization ratio, and the $R^2$ values for each regression are reported in the right-hand column of the table. (Remember that these results are for the sample of high- and low-centralization urban areas, not the entire set, so the results are not directly comparable to those reported in Table 2.) The results were very unambiguous: The density gradient is not significantly related to any of the 6 measures of centralization. The greatest $R^2$ value was 0.16.

<table>
<thead>
<tr>
<th>Alternative Centralization Measures</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With Centralization Ratio</td>
</tr>
<tr>
<td>Centralization ratio</td>
<td>—</td>
</tr>
<tr>
<td>Absolute centralization index</td>
<td>0.970</td>
</tr>
<tr>
<td>Proportion of units in tracts with 25 percent of the total area that are closest to the center</td>
<td>0.942</td>
</tr>
<tr>
<td>Proportion of units in tracts with 10 percent of the total area that are closest to the center</td>
<td>0.950</td>
</tr>
<tr>
<td>Proportion of units in tracts within 75 percent of the mean distance for a uniform distribution</td>
<td>0.850</td>
</tr>
<tr>
<td>Proportion of units in tracts within 50 percent of the mean distance for a uniform distribution</td>
<td>0.938</td>
</tr>
</tbody>
</table>
To me, the conclusion is obvious and reinforces the conclusion from the previous analysis: The density gradient is not a good measure of the level of centralization in an urban area.