Negative Exponential Model Parameters and the Size of Large Urban Areas in the U.S., 1950-2010

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

Abstract

The density of urban development tends toward a negative exponential decline with distance from the center. The density gradient has been observed to be inversely related to the size of the urban area, while the central density is directly related. If the negative exponential density decline holds, the size of the urban area, the density gradient, and the central density are necessarily mathematically related. This relationship is derived, and a simplified approximation is compared with the sizes of large urban areas and the estimated density gradients and central densities. The results generally confirm the expectations, although the coefficient for the central density is larger than expected, possibly because of the use of gross rather than net residential density in estimating the parameters for the negative exponential model.

Introduction

Researchers have long observed the negative exponential decline of density with distance from the center of a city. They have used data to estimate the parameters of this relationship, the density gradient—the rate of decline of density with distance—and the central density. Researchers have further have looked at the relationships of these parameters of the negative exponential model with a wide range of characteristics of urban areas, most notably the size of urban areas. But for the most part, researchers have failed to look simultaneously at the relationships of these parameters, the density gradient and the central density, with each other and with urban area size.

Numerous studies have come to consistent conclusions regarding the relationships of the exponential model parameters and the size of urban areas (for example, Mills 1972; Edmonston 1975; Edmonston, Goldberg and Mercer 1984; Anas, Arnott, and Small 1998). The density gradient is inversely related to urban area population. Larger cities tend to have lower rates of decline of density with distance from the center, a flatter density curve. Central densities, on the other hand, are directly related to urban area size. Larger cities tend to have higher densities near their centers, with New York the obvious extreme example in the United States.

The general consensus is that the positive relationship of central density with size is to be expected, given the greater demand for access to the center associated with more population (Mills 1972; Anas, Arnott, and Small 1998). Expectations regarding the density gradient have varied. Mills (1972) argued that the gradient should indeed be lower for larger cities, as they could better support outlying employment centers and would be less dependent on the CBD. (Note that this, of course, would be somewhat at variance with the assumptions underlying the basic monocentric model predicting the negative exponential decline of density.) On the other hand, Anas, Arnott, and Small (1998) were surprised at the inverse relationship, stating that the monocentric model should suggest either no relationship or, in their version of the model, a mild positive correlation.

The lack of attention given to the relationships of the density gradient and central density to each other and the size of the city is surprising given that, if the negative exponential model really describes the distribution of density in an urban area, a mathematical relationship necessarily exists among these three values. This paper proceeds by describing this mathematical relationship. Then, after describing the data and methods, results are presented for the empirical analysis of these relationships for 42 large urban areas in the United States from 1950 to 2010. This paper is an extension of a prior paper examining the pattern of the negative exponential decline of density is large urban areas from 1950 to 2010 (Ottensmann 2016).

Relationship of Urban Area Size with the Density Gradient and Central Density

This section derives the relationship among the density gradient, central density, and the size of the urban area under the assumption that densities are distributed as predicted by the negative exponential model. The argument is then made for an approximate form of this relationships that can be estimated empirically using data on density for urban areas.

Begin with the assumption that densities in an urban area are distributed as a negative exponential function of distance from the center:

$$D(s) = D_0 e^{-\beta s}$$

where $D(s)$ is the density at distance s from the center, D_0 is the density at the center, β is density gradient, and e is the base of the natural logarithms. Note that the density could be either the density of population or housing units. The latter is used in the empirical portion of this research.

The urban area is assumed to be circular, with a radius of *R*. The total population or number of housing units in the urban area (corresponding to how density is being measured), is *U*.

Given the density at every distance *s* from the center, the total number of housing units *U* in the urban area is the integral over distance *s* of density times the circumference at *s*:

$$U = \int_0^R 2\pi s D_0 e^{-\beta s}\, ds$$

Integrating gives this expression for the total number of housing units:

$$U = \frac{2\pi D_0}{\beta^2} - \frac{2\pi D_0 e^{-\beta R}\left(\beta R + 1\right)}{\beta^2}$$

The first term of this expression is the total number of housing units under the curve from the center out to infinity. The second term is the number of housing units under the curve from the edge of the urban area at *R* out to infinity. This is subtracted from the first term to get the total number of housing units in the urban area out to *R*.

The first term is proportional to D_0/β^2, the ratio of the central density to the density gradient squared. The second term includes this ratio, modified by a further function of the density gradient. As an approximation, the total number of units in the urban area can be taken to be proportional to this ratio:

$$U \approx k \frac{D_0}{\beta^2}$$

where *k* is simply a constant of proportionality. As an alternative motivation for this approximation, Mills (1972) derived his method for estimating the parameters of the negative exponential model using just two values by assuming that the total for the urban area was the value from the center out to infinity, ignoring the second term of the expression above.

Now the issue becomes how to arrange the above expression involving the size of the urban area (total number of housing units), the central density, and the density gradient for examining the relationship empirically. The size of the urban area can reasonably be assumed to be determined primarily by factors other than the exponential model parameters. So this can be considered to be exogenous.

Now for the density gradient and the central density. Housing is very durable and long lasting. Additions to the housing stock can be more readily made closer to the periphery of the urban area where there is vacant land available. Much of the housing near the center will have been developed decades ago, so it may not be unreasonable to consider current central densities to have been exogenously determined. Some empirical justification for this assumption can be seen by looking at the relationship of estimated central densities for the urban areas considered here with metropolitan area populations in 1910. Correlations are presented in Table 1. (Detail on the data and methods is provided in the following section.) Populations for metropolitan areas in 1910 were chosen because this represents the period before the start of widespread use of the automobile. Larger populations in 1910 should be associated with greater amounts of higher density development.

Table 1. Correlations of Estimated Central Density for Housing Units with 1910 Metropolitan Area Population.

Year	Correlation of Central Density with 1910 Population
1950	0.771 ***
1960	0.716 ***
1970	0.782 ***
1980	0.849 ***
1990	0.865 ***
2000	0.874 ***
2010	0.898 ***
All Cases (Pooled)	0.780 ***

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

The correlations of central density with 1910 metropolitan population are consistently high for all years, varying from a low 0f 0.72 to a high of 0.90 in 2010. The correlations are all highly statistically significant. This would support the idea that central densities have been determined, to a considerable extent, by earlier urban area growth and densities.

If both urban area size and central density can be assumed to be exogenous, the density gradient can be moved to the left side, giving this equation for the relationship:

$$\beta = \sqrt{\frac{kD_0}{U}} = k^{1/2} D_0^{\ 1/2} U^{-1/2}$$

(The form with the fractional exponents is included to make clear the next step.) Taking the logarithm of both sides and allowing the values of the exponents to vary gives the following expression:

$$\ln(\beta) = b_0 + b_1 \ln(D_0) + b_2 \ln(U)$$

This is linear. With the assumption that central density and urban area size are exogenous, this can be estimated using ordinary least squares regression. The assumption of the exogeneity of central density can be tested as well.

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 that 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 populations 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 was 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). Sixteen of the areas had multiple centers, which made them unsuitable for estimating the monocentric negative exponential model, leaving the 43 areas with a single center.

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 allowed 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 1983 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]

In 1910, the Census for the first time attempted to capture the larger urban areas that extended beyond the major cities by defining Metropolitan Districts as aggregations of civil divisions. The populations of these Metropolitan Districts in 1910 were used to estimate of the amount of older development. For areas that were not designated as Metropolitan Districts or the related city plus adjacent area designation in 1910, the populations of the cities themselves were used (U.S. Bureau of the Census 1913: pp. 73-74 and Tables 50, 51, and 59). Since these were smaller cities with populations less than 100,000, the populations in surrounding suburban areas was likely rather small. Las Vegas was not incorporated as a city until 1911, so the 1910 census does not report a population for Las Vegas. Since the 1910 population is used in some of the analyses, Las Vegas has been excluded from the set of urban areas considered, leaving 42 urban areas with single center for the analyses in this paper.

The monocentric model (Muth 1969; Mills 1972) 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, β 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,

$$\ln(D_i) = \ln(D_0) - \beta s_i$$

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

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. This was done using the tracts within the urban areas for each census year for each of the 42 urban areas.

Relationship of Estimated Density Gradients to Estimated Central Density and Urban Area Size

The parameters of the negative exponential model have been estimated from housing unit densities by census tracts for 42 large urban areas for each census year from 1950 to 2010. The first step is the examination of the correlations of the density gradient and the central density with the size of the urban area, the total number of housing units. These correlations are shown in Table 2. For the density gradient, the correlations with the log of the number of housing units is reported, as the relationship was stronger than with the natural form. The density gradient was negatively correlated with the size of the area, as many others have found. The correlations are moderately large, ranging from -0.37 to -0.66 for the pooled results. All are statistically significant. Also as expected, the central density is positively related to the number of housing units (this time in natural form), with correlations ranging from 0.64 for the pooled regression to as high as 0.85. Of course these are also statistically significant.

Table 2. Correlations of the Density Gradient and the Central Density with the Total Number of Housing Units in the Urban Area.

Year	Correlation of Density Gradient with Log of Housing Units	Correlation of Central Density with Number of Housing Units
1950	-0.547 ***	0.764 ***
1960	-0.541 ***	0.689 ***
1970	-0.539 ***	0.746 ***
1980	-0.436 **	0.845 ***
1990	-0.411 **	0.817 ***
2000	-0.372 *	0.819 ***
2010	-0.397 **	0.841 ***
All Cases (Pooled)	-0.656 ***	0.635 ***

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

The objective is the examination of the relationship among the negative exponential model parameters and the size of the urban area. As explained above, the log of the density gradient is assumed to be the dependent variable, with the log of central density and the log of the number of housing units as independent variables. Assuming that the latter two variables are exogenous, ordinary least squares is used to estimate the model for each year and for all cases pooled. The regression coefficients for the logs of central density and number of units and the R^2 values are presented in Table 3.

Table 3. Ordinary Least Squares Regressions of the Log of the Density Gradient on the Log of Central Density and the Log of the Number of Housing Units.

Year	Regression Coefficient for Log of Central Density	Regression Coefficient for Log of Housing Units	R^2
1950	0.894 ***	-0.619 ***	0.879
1960	0.977 ***	-0.635 ***	0.834
1970	0.950 ***	-0.628 ***	0.808
1980	1.040 ***	-0.615 ***	0.589
1990	0.922 ***	-0.576 ***	0.640
2000	0.943 ***	-0.573 ***	0.680
2010	0.931 ***	-0.556 ***	0.656
All Cases (Pooled)	1.004 ***	-0.637 ***	0.869

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

The models account for a substantial proportion of the variation in the log of the density gradient, with R^2 values ranging from 0.59 to 0.88. The regression coefficients are reasonably consistent from year to year and are all highly statistically significant. The density gradient is positively related to the central density and inversely related to the size of the area, as was anticipated.

Looking at the magnitudes of the regression coefficients, first for the size of the area, the estimates ranged from -0.56 to -0.64. The approximation for an area conforming to the negative exponential model had the density gradient varying with the inverse of the square root of area size, implying a coefficient in the model using logs of -0.5. So the actual values are reasonably close.

For the central density, the model for the ideal area had the density gradient varying with the square root of the central density, implying a coefficient of 0.5 in the logged model. The actual estimates are much higher, varying from 0.89 ti 1.04, about twice as large. One possible explanation for this difference arises from the way density is measured here (and most other negative exponential analyses). The census tract density is a measure of gross density, the number of housing units in the tract divided by the total land area. This includes not only residential land but areas of other land uses, land in streets, and vacant land, varying widely across tracts. Including this nonresidential land decreases the densities compared to net residential densities. This would result in lower estimates of the central density, which could then account for the larger regression coefficients.

The assumption was made that the two independent variables in the model, central density and urban area size, were both exogenous. This seems very reasonable for area size. However, the exogeneity of central density is less clear and might be questioned. If central density were endogenous to the model, the ordinary least squares estimates would be biased. However, the model could be estimated using instrumental variable regression. Fortunately, a suitable instrument that is clearly exogenous and is strongly related to central density is already available, the population of the metropolitan area in 1910 (refer back to Table 1). This variable was used to estimate the model using two-stage least-squares instrumental variable regression. The regression results for the instrumental variable regression using all cases pooled is presented in Table 4 along with the ordinary least squares regression results.

The regression coefficients are quite similar for the two models. The magnitudes of the regression coefficients are slightly larger for the instrumental variable regression, but would not lead to a significantly different interpretation. The standard errors are, of course, larger for the instrumental variable regression, as would be expected. The Durban and Wu-Hausmann tests for endogeneity as implemented in Stata reject the null hypothesis that central density is exogenous. This suggests that the instrumental variable regression should be providing better estimates of the regression coefficients than the ordinary least squares regression (though, of course, the estimates are not that different).

Turning to the instrumental variable regression for each year (results not shown) provided a decidedly more mixed picture. Of the 7 regressions, the tests for endogeneity were clearly significant for 3, completely not significant for 3, and only barely significant at the 0.05 level for 1 other. For 6 of the 7 years, the magnitudes of the regression coefficients were somewhat larger for the instrumental variable regression, as observed for the pooled results in Table 4. However, for 1 of the years, both regression coefficients had magnitudes several times larger than any of the other regression

Table 4. Ordinary Least Squares and Instrumental Variable Regressions of the Log of the Density Gradient on the Log of Central Density and the Log of the Number of Housing Units for All Cases Pooled.

Variable	Ordinary Least Square Regression		2SLS Instrumental Variable Regression	
Log of Central Density	1.004	***	1.180	***
	(0.033)		(0.059)	
Log of Housing Units	-0.637	***	-0.672	***
	(0.015)		(0.018)	
Constant	-2.175	***	-3.142	***
	(0.253)		(0.371)	
R^2	0.869	***	0.856	***

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

coefficients for both the ordinary least squares and instrumental variable regressions. Given the similarities of the other regression coefficients, the inconsistencies in the tests for endogeneity, and the great inconsistency of the instrumental variable regression coefficient estimates for one of the years, the decision was made to rely on the more straightforward ordinary least squares estimates as presented in Table 3.

Conclusions

A great deal of the interrelationship of the density gradient, the central density, and the size of the urban area can be explained by the necessary mathematical relationship that must exist among these values if the distribution of density is a negative exponential function of distance from the center. The density gradient must be approximately proportional to the square root of the central density and inversely proportional to the square root of the size of the urban area.

Regression estimates of a model based on these relationships for 42 urban areas for the census years from 1950 to 2010 showed that the log of central density and the log of the number of housing units in the urban area accounted for well over half of the variation in the log of the density gradient. The directions of the relationship was as expected, positive for the central density and negative for the size of the urban area. The estimated regression coefficients for the number of housing units was reasonably consistent with the expectation of an inverse relationship to the square root of area size. The regression coefficient for the central density was higher than would be expected based on proportionality to the square root. It is possible that this difference results from the necessary use of gross densities for the estimation of the negative exponential model parameters, producing estimates of the central density that are lower than if net residential density involving only residential land could have been used.

This analysis provides a more comprehensive account of the relationships of the exponential model parameters to the size of the urban area than just the simple bivariate observations that density gradients decline and central densities increase with size. The relationship of the density gradient to central density is an important addition to understanding urban density patterns. In his analysis of the determinants of the exponential model parameters, Mills (1972) reached the conclusion that "the cause of the historical flattening of population density functions has been the growth of urban population and income rather than the passage of time or whatever it stands for." This analysis helps explain the role of the growth of urban population in this.

References

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

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

Edmonston, Barry, Michael A. Goldberg, and John Mercer. 1985. Urban form in Canada and the United States: An examination of urban density gradients. *Urban Studies* 22, 3 (June): 209-217.

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.

Hammer, Roger B., Susan I. Stewart, Richelle L. Winkler, Volker C. Radeloff, and Paul R. Voss. 2004. Characterizing dynamic spatial and temporal residential density patterns from 1940-1990 across the North Central United States. *Landscape and Urban Planning* 69, 2-3 (August): 183-199.

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. 1913. *Thirteenth Census of the United States Taken in the Year 1910*. Vol. 1. *Population 1910. General Report and Analysis.* Washington DC: Government Printing Office.

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.