The Negative Exponential Decline of Density in Large Urban Areas in the U.S., 1970-2020

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

Abstract

The central density, density gradient, and goodness-of-fit for the negative exponential model are estimated using housing unit densities for census tracts for 40 large urban areas from 1970 to 2020. Nonlinear regression is used to estimate the model parameters. The values of the estimates in 2020 show extremely wide variation across the urban areas. Larger and older urban area, especially in the Northeast, tend to have the highest central densities, gradients and R^2 values, while areas in the Sunbelt are among the lowest. These values have generally declined from 1970 to 2020, again with wide variation across areas. Mean R^2 drops from 0.41 to 0.26, suggesting declining importance of the Central Business District associated with the decentralization of employment. The R^2 values for some areas in 2020 are less than 0.10; the negative exponential model is accounting for a very small proportion of the variation in density. But despite an overall decline in central density over the fifty-year period, the mean increases in the most recent decades, with an especially large jump from 1970 to 2020, raising questions regarding possible shifts in the patterns of urban change.

Introduction

Contemporary consideration of the negative exponential decline of population density goes back to an article by Clark (1951), though several authors had made similar observations earlier. Clark examines the decline of population densities for numbers of cities at different times in the nineteenth and twentieth centuries and shows the patterns generally conform to a negative exponential decline with distance from the center. Large numbers of subsequent studies have confirmed this. Reviews of this extensive literature include Thrall (1988), McDonald (1989), and Smith (1997). Anas, Arnott, and Small (1998) cite results from numerous studies in their broader review of urban structure. In a comprehensive study of urban patterns around the world, Angel (2012) discusses the negative exponential decline of densities.

Muth (1969) and Mills (1972) provide an economic basis for understanding this phenomenon, the monocentric model in urban economics. They assume employment is concentrated in the central business district (CBD), to which residents commute. In selecting residential locations, people make a tradeoff between the costs of commuting and the desire for more space. Commuting costs are reduced for locations closer to the CBD, so residents were willing to pay more for those locations. Land rents then decline with distance, allowing households to obtain more space and live at lower densities farther from the center. Some simple but reasonable assumptions regarding functional forms and parameters yield the negative exponential decline of density.

During the nineteenth century and the first part of the twentieth century, employment was concentrated in and near the CBD, which was the major commuting destination. But employment has become progressively more decentralized and the journey-to-work less oriented towards the center. In the same work cited presenting the monocentric model, Mills (1972) also examines the decentralization of employment over time. The basic assumption of the monocentric model of the concentration of employment in a single center is less consistent with the evolving distribution of employment. This raises the question of the applicability of the negative exponential model over time. To what extent might the ability of the model to account for population and housing unit densities be declining?

Addressing what the extensive literature documenting the negative exponential model can say in response to this requires a brief detour on how the parameters of the model can be estimated. The original approach involved using data on densities for small areas (such as census tracts) and looking at those densities in relation to distance to the CBD, either graphically as Clark did, or by using regression. This approach not only provides for the estimation of the model parameters but also provides a basis for assessing how closely the densities conform to a negative exponential decline. This requires large amounts of data, which until the relatively recent availability of the necessary data in digital form was a laborious process.

Mills (1972, again!) devised an an ingenious solution. Since significant evidence had shown that the negative exponential model describes the decline of densities, he simply assume this to be the case. Then one can estimate the parameters of the model using just two data pairs, the population and estimated radius of the central city and of the entire urban area. A key limitation of the two-point method arises from the initial assumption that the density in the urban area declines as a negative exponential function of distance. This allows the estimation using the limited data but provides no basis for assessing the extent to which the distribution conforms to the negative exponential model, as that has been assumed.

Using the two-point method is of course is much easier and faster than using small-area data. This became a popular, widely used method for looking at the negative exponential decline of density. Studies looking at densities in significant numbers of urban areas for multiple years have especially taken this approach. But that has meant that such studies could not address questions about the degree to which distributions of density conform to the negative exponential model and whether this may have declined over time.

This paper examines the negative exponential decline of housing unit densities for forty large urban areas in the United States using data for census tracts from 1970 to 2020. The estimated parameters provide measures of the density patterns and their changes up through the most recent census. But especially important is that measures of the goodness-of-fit of the models provide evidence of the degree to which densities conform to a negative exponential pattern. This allows examination of changes over time that may be associated with the decreased importance of the CBD.

The following section describes the data used and the delineation of consistent urban areas using the census tract housing unit data. Next is the description of the procedures for model estimation, arguing for the use of nonlinear regression. (A brief appendix compares these results with those obtained using log linear regression, which is typically employed.) Interpretation of the parameters of the negative exponential model have emphasized the use of the density gradient and its decline as a measure of decentralization. The case is made that the density gradient is not a very good measure for this. Then come the results. One section addresses the negative exponential decline of density in 2020, describing the results and presenting some very simple exploratory models of factors associated with the estimated values. This is followed by a similar treatment of the changes in the values from 1970 to 2020, providing evidence addressing the question posed about the performance of the model over time.

The *Urban Patterns 2* data

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

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

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

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

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

Nonlinear estimation of the negative exponential model

This section addresses the procedures followed in the estimation of the parameters of the negative exponential model using housing unit densities for the urban areas. The major focus is on the use of nonlinear regression for the estimation.[1] Also addressed is the measurement of the goodness-of-fit of the model, and the urban areas and years of data included in the analysis.

The usual method used to estimate the parameters of the negative exponential model from small area data for density is to use standard linear regression with the log-transformed version of the model—log linear regression. The following paragraphs argue for the use of nonlinear regression procedures instead. The possibility of using nonlinear regression in this application has received very limited attention. Kemper and Schmenner (1974) conclude that nonlinear regression performs better for analysis of the negative exponential decline of manufacturing employment density. White (1977) uses both log linear and nonlinear methods, but his primary objective is the assessment of the accuracy of Mills' (1972) two-point method of estimation. Greene and Barnbrock (1978) compare the performance of the two methods by examining the residuals. (Curiously they do not present and compare parameter estimates.) They ultimately conclude that the residuals do not show a pattern indicating an underestimate of densities near the center as Kemper and Schmenner have found. Finding this for the one urban area they examined, they consider the log linear model "adequate."

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

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

where D_i is the density (housing unit density in the current research) in census tract i, d_i is the distance from the census tract to the center of the urban area. D_0 is the central

[1] The choice of nonlinear regression for the negative exponential model is addressed more fully in an earlier paper (Ottensmann 2022b).

density and β is the density gradient, both parameters to be estimated from the data, and e is the base of the natural logarithms.

The virtually universal method used to estimate the parameters from data for census tracts involves transforming the equation by taking the natural logarithm of both sides, giving

$$\ln D_i = \ln D_0 - \beta d_i$$

This is linear in log density and distance, so the model can be estimated using ordinary least squares regression. This is referred to as log linear estimation.

As the two equations above are mathematically equivalent, it might appear to make no difference which is used as the starting point for the estimation. This is not true. The data will not fit the model perfectly. Estimation involves selecting values for the parameters that minimize the error in prediction. So the equation for the estimation necessarily includes an error term which is generally the difference between the actual and predicted densities and is added to the equation:

$$\ln D_i = \ln D_0 - \beta d_i + e_i$$

where e_i is the error in prediction for tract i. Now note that the dependent variable is the log of density, so the model predictions and the the error terms are likewise measured in terms of the log of density. This means that the error associated with the difference in densities between 20,000 and 30,000 is identical to the error associated with the difference between 200 and 300, 0.4055. By taking the logs, the errors in predicting densities for higher density tracts are made relatively smaller and are given less weight in the estimation of the model parameters.

Nonlinear estimation procedures can add the error term to the original, nonlinear equation:

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

The error terms are then in units of density, not log density, preserving the larger magnitudes of error at higher densities. Those working with the negative exponential model seem to be more interested in density than the log of density. After doing the log linear estimation to obtain the estimate of the log of central density, it is nearly always transformed back to density terms when reporting the results. This suggests that using nonlinear estimation to minimize the error in prediction in terms of density rather than log density would be more appropriate.

The magnitudes of errors and the effect of the difference between nonlinear and log linear estimation of the negative exponential model will be greater for urban areas having more census tracts having higher densities. The extreme example is then the New York urban area, which has by far the highest densities. Figure 1 plots the housing unit densities of the thousands of census tracts in the New York area versus distance to the CBD in miles. The inset in the upper right expands the lower-left quarter of the original plot to display more detail. Nonlinear and log linear estimation are used to estimate the parameters for negative exponential models. The predicted densities are shown on the graphs, red for the nonlinear estimates and blue for the log linear estimates. The difference in the central densities—where the lines intersect the y-axis—is striking. The nonlinear estimate of the central density is 58,533 units per square mile while the log linear estimate is only 13,584. The very high density tracts extending upward closer to the center have little effect on the estimate of the log linear central density. The red, nonlinear prediction line goes more closely through the masses of

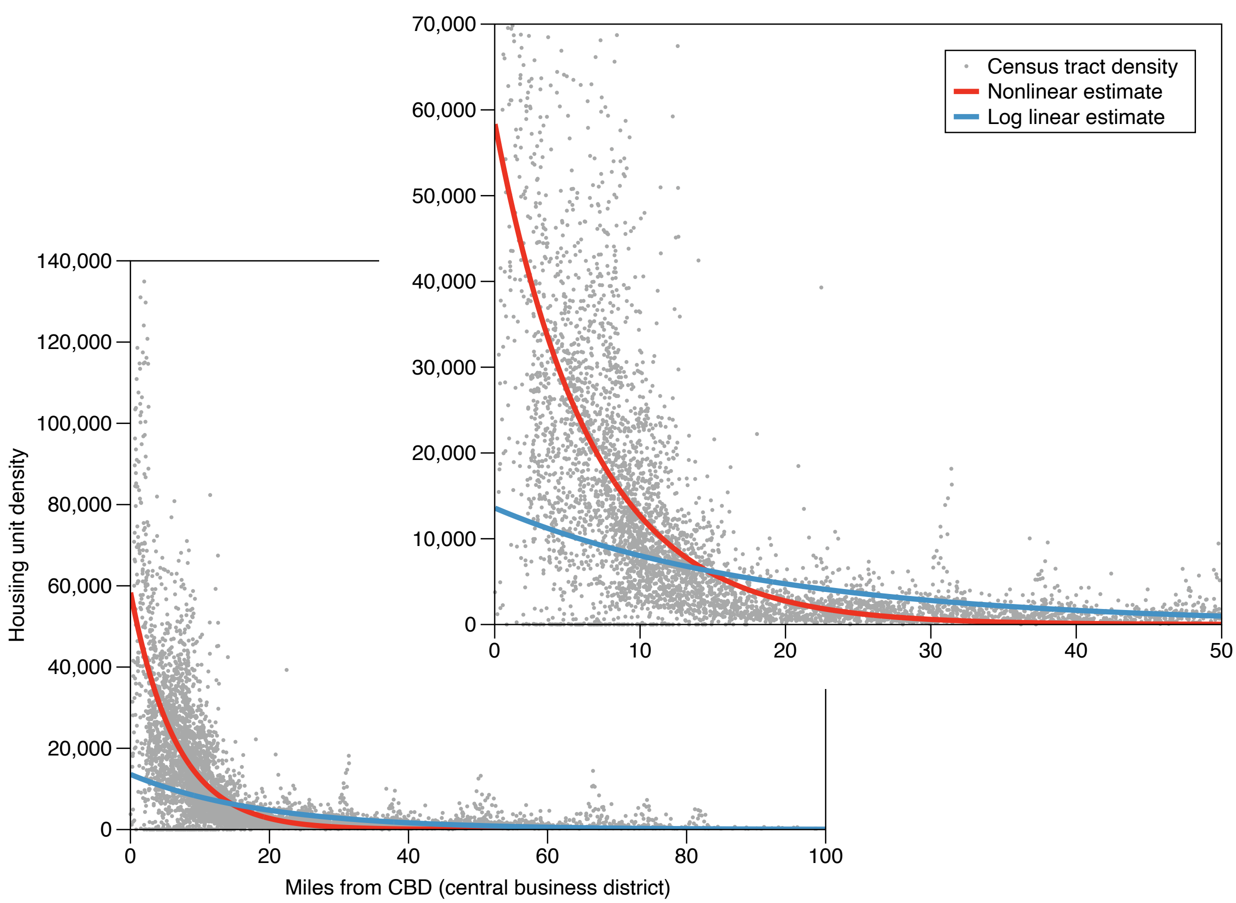

Figure 1. Housing unit densities versus miles from the CBD for census tracts in the New York urban area in 2020 with predicted densities using negative exponential models with nonlinear and log linear estimates of model parameters.

census tracts moving away from the center. It is not perfect—the pattern does not exactly conform to a negative exponential decline. But the nonlinear estimate certainly appears to provide a more reasonable prediction of densities than the log linear estimate.

The nonlinear regression procedure in Stata is used to estimate the parameters of the negative exponential model for each of the urban areas and for the different years of data. The procedure works by minimizing the sum of squared deviations of the errors in prediction, the error term added to the negative exponential model.

The regression reports the sum of squared errors analogously to linear regression. But nonlinear regression does not necessarily include a constant term like linear regression. Stata calculates a value for the total sum of squares that does not use squared deviations from the mean but rather squared deviations from zero. This value is used along with the error sum of squares to calculate a value for R^2 following the procedures used in linear regression. Because the total sum of squares is larger than that used for linear regression, those R^2 values will be larger and do not represent the proportion of the variation in the dependent variable accounted for by the model.

To provide a more meaningful measure of the goodness-of-fit of the model, an alternative approach is taken. The predicted densities produced by the estimated model are regressed on the actual densities using linear regression, producing a value called here the *prediction* R^2. This will be the proportion of the variation in density accounted for by the estimated negative exponential model. All of the R^2 values reported in this paper are these prediction R^2 measures.

Two aspects of the data being used restrict the estimation of the negative exponential model and the results to be considered. The model predicts the decline of density with distance from the center, the CBD. It assumes a single center. Urban areas having two or three urban centers would be expected to show declines in density with distance from the CBDs for each of the centers, with the effects of the CBDs interacting, and producing multiple sets of parameters. While there have been some attempts to examine density decline in such contexts (e.g., Griffith 1981; Gordon, Richardson, and Wong 1986), this is beyond the scope of what is being done here. As a result, the analysis is restricted to the 40 urban areas in the dataset having a single urban center.

The data on housing units by census tract in the dataset extend back to 1950 and the negative exponential models were estimated for each census year from 1950 to 2020. Inspection of the results suggests inconsistencies for the first two years of estimates before 1970. Prediction R^2 values generally increase in the first two decades and then decline in a steady fashion. Standard deviations of changes from 1950 to 2020 were significantly higher than changes from 1970. Some of the percent changes from 1950 were ridiculously high.

In appears that the problems may result at least in part from the small sizes of some of the urban areas in the earlier years. In 1950, the Las Vegas urban area had only about 6,000 housing units, and Orlando had fewer then 20,000. Ten of the urban areas, one quarter of the total, had fewer than 30,000 housing units in 1950. By 1970, only one area had fewer than 70,000 units. Las Vegas only had 19 census tracts in the urban area in 1950. By 1970, only two areas had fewer than 100 tracts.

For very small areas, distances to the center will be smaller and will not vary as much as in larger areas. These smaller distances may be of lesser importance for households' choice of location and density. Examination of the results for the smaller urban areas show most exhibiting inconsistencies in the pattern of estimates in the early years.

If the patterns of change over time were perfectly consistent over the entire period, then prediction of the change from 1950 to 2020 using the change from 1970 to 2020 should result in a perfect prediction. The magnitude of the residual from the regression is then a measure of the departure from consistency for each urban area. The correlation of the absolute value of the residuals with the size of the urban area, the number of housing units in 1950, is -0.58. In other words, the smaller the area, the greater the inconsistency. This provides some support to the notion that the small sizes present an issue for the negative exponential model and contribute to the inconsistency in the estimates for the early years. Given this, the decision has been made to include only the results from 1970 forward in the analysis that follows.

The density gradient is not a good measure of decentralization

The estimated central density and density gradient provide useful measures describing the general overall distribution of housing unit densities in an urban area. The predicted R^2, the measure of goodness-of-fit, indicates how closely the distribution conforms to the predictions of the negative exponential model. These values can be compared across urban areas and the changes in the values over time can be examined.

Mills (1972) gives prominence to the notion that the decline of the density gradient over time could be used as a measure of the decentralization of population and employment in an urban area. It further follows that the density gradient is therefore a good measure of the level of centralization or decentralization in an urban area. This has come to be the common interpretation of the density gradient.

It is understandable how one might have arrived at this conclusion. If decentralization takes place over time in an urban area, the density gradient will indeed fall. But the converse is not true. This is the logical problem of affirming the consequent. The decline in the density gradient is not necessarily associated with decentralization. It may be but it needn't be.[2]

[2] More extensive discussion of this issue is in an earlier paper (Ottensmann 2022a).

Demonstrating that the density gradient can decline without decentralization of population or housing units is quite simple. Begin with a circular urban area with densities declining according to the negative exponential model. Densities at time 1 decline from the central density D_0 at the center to the small minimum density for territory to be considered part of the urban area at distance R_1. This is illustrated by the red line in Figure 2. Now assume the urban area grows, increasing the number of housing units while continuing to conform to the negative exponential model with the central density unchanged. The urban area necessarily has to expand outwards to accommodate at least some of the new growth beyond the original edge of the area. The density curve at time 2, in blue, extends from the same central density as before but goes farther out to R_2 where the density declines to the minimum urban density.

The density gradient at time 2 is lower than the density gradient at time 1. But at no location has the density declined. It has actually increased everywhere except at the center, where it has remained constant. No net outmovement has taken place. This would seem to make it difficult to conclude that decentralization has taken place. To be sure, some people and housing units are located farther from the center than any at time 1. The urban area has expanded in terms of area. But that cannot be sufficient to conclude that decentralization has taken place. If it were, then growth of an urban area

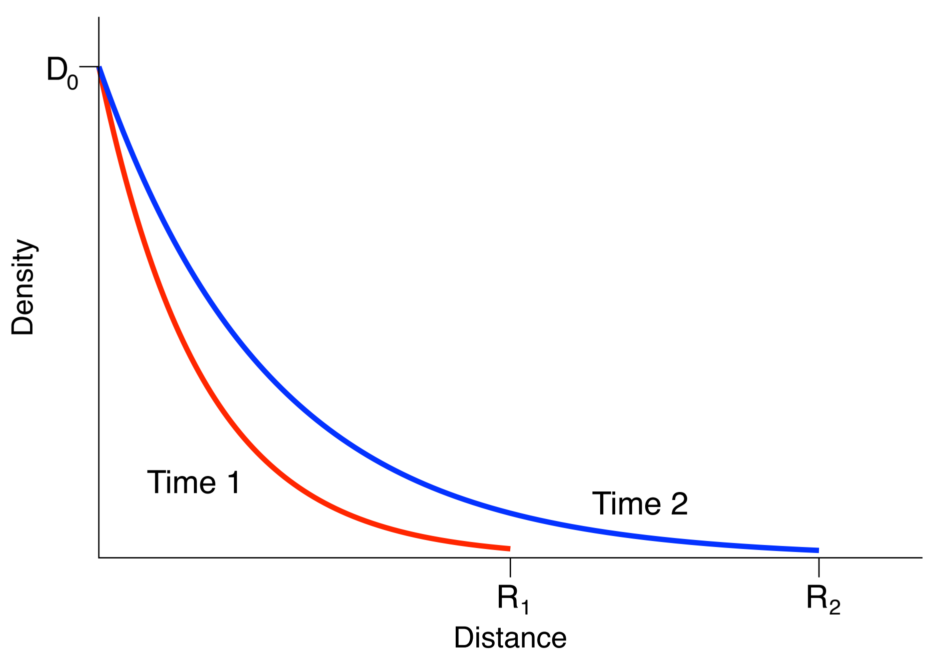

Figure 2. Negative exponential decline of density from the same central density to the minimum urban area density at two points in time.

that involves an increase in the central density and the density gradient while simultaneously expanding outwards would have to be considered decentralization, which would be nonsense.

The early observations (e.g., Muth 1969; Mills 1972) that density gradients tend to be negatively related to the size of an urban area should have led to questioning the extent to which the density gradient can be seen as a measure of centralization and decentralization. For example, can one realistically believe that the New York urban area, with a density gradient of 0.153 in 2020 is much more decentralized than Portland or Richmond with gradients of 0.255 and 0.242? Simply solving the negative exponential model for the density gradient shows that quantity to be positively related to the central density and inversely related to the radius for a circular area. Focusing on the the radius may not be quite sufficient, as that necessarily depends on the nature of the density decline. However, it is more difficult to argue that the total population and housing units in an urban area are determined by the density gradient.

For an urban area that is circular with densities conforming to the negative exponential model, the number of housing units can be calculated by integrating over area times the negative exponential function of distance. Following Mills (1972), this is a simplified approximation for housing units as a function of the central density and the density gradient:

$$H \approx \frac{kD_0}{\beta^2}$$

where H is the total number of housing units in the urban area and k is a constant. Solving for the density gradient β gives

$$\beta^2 \approx \frac{kD_0}{H} \quad \text{and} \quad \beta \approx \sqrt{\frac{kD_0}{H}}$$

so the density gradient should be positively related to the central density D_0 and inversely related to the number of housing units. Regressing the estimated density gradient for the urban areas in 2020 on the square root of central density and the square root of the number of housing units gives

$$\beta = -0.062 + 0.063\sqrt{D_0} - 0.107\sqrt{H} \quad \text{(density in thousands, housing units in millions)}$$

$$R^2 = 0.595$$

This demonstrates that the density gradient is highly dependent on the central density and the size of the urban area and shows that if the central density remains constant as an urban area grows, the density gradient would be expected to decline.

Negative exponential model estimates for 2020

The nonlinear regression estimates for the negative exponential parameters central density and the density gradient and the prediction R^2 for the 40 large urban areas in 2020 are considered. The section begins with the basic description of the results. This is followed by some simply exploratory models investigating characteristics associated with those values.

Basic descriptive statistics are shown in Table 1 for the central density, density gradient, and prediction R^2. The mean values are not inconsistent with those obtained in other studies, taking into account that the current work uses housing unit densities as opposed to population densities (e.g., Edmonston 1975; Guest 1975; Anas, Arnott, and Small, 1998; Angel 2012). The minima and maxima show the wide variation in the values across the urban areas. This is especially true for the central density with the extremely high value over 58,000 housing units per square mile for New York, as noted above. The distribution for this is highly skewed, with the median value of about 4,700 far less than the mean of nearly 7,500.

Table 1. Basic statistics for central density, density gradient, and prediction R^2 in 2020.

	Mean	Minimum	Median	Maximum
Central density	7,464	1,938	4,713	58,533
Density gradient	0.112	0.029	0.090	0.284
R^2	0.257	0.050	0.232	0.593

A clearer picture of the variation is seen when looking at the five highest and five lowest areas for each of the three measures. Beginning with central density, New York is clearly the high outlier. Three urban areas follow with central densities above 20,000, with Philadelphia and Chicago not unexpected, though Honolulu being somewhat of a surprise. Portland, the fifth highest, has a much lower central density not much over half these values. At the other extreme, central densities in some smaller areas in the South are very low, with four under 2,200 units per square mile.

Table 2. Urban areas with the highest and lowest values for central density, density gradient, and prediction R^2 in 2020.

Area	Central density
New York	58,533
Philadelphia	26,364
Chicago	22,912
Honolulu	20,279
Portland	12,421
...	...
El Paso	2,350
Memphis	2,176
Jacksonville	2,082
Birmingham	1,953
Oklahoma City	1,938

Area	Density gradient
Philadelphia	0.284
Rochester	0.265
Portland	0.255
Richmond	0.242
Hartford	0.219
...	...
Las Vegas	0.038
Orlando	0.036
San Antonio	0.035
Phoenix	0.030
Oklahoma City	0.029

Area	R^2
Rochester	0.593
Philadelphia	0.567
New York	0.527
Cincinnati	0.470
Milwaukee	0.464
...	...
Memphis	0.096
Phoenix	0.095
Albuquerque	0.065
Las Vegas	0.051
Oklahoma City	0.050

The highest density gradients are greater than 0.2 for a mix of areas of varying sizes and in different parts of the country. Four of the five are older urban areas. Portland is the exception. Its appearance on the lists for both highest central density and density gradient leads one to suspect that the area's urban growth boundary may be playing a role in its development and the distribution of density. Very different are the urban areas with density gradients less than 0.04. These a very small gradients, indicating little decline of density with distance, a very flat pattern. These are newer urban areas of varying size in the South and Southwest.

Urban areas with the highest prediction R^2 are all older areas in the Northeast and Midwest. They were developed and grew to significant size during the era where the CBDs were dominant and retain a pattern conforming to the prediction of the negative exponential decline of density with distance. Again, the urban areas having the lowest prediction R^2 values, less than 0.10, are newer areas that have experienced much of their development in recent decades. It is not surprising that three of these areas are also among the five areas having the smallest density gradients. With very little decline in density, these areas have less variation to be explained, so random variation in densities becomes relatively more important.

The lists of the areas with the highest and lowest values suggest clear regional variation in these values. Considering the means for the urban areas in the four regions, central densities are highest in the Northeast and lowest in the South, with the other regions in the middle. The same is true for the density gradients. While the urban areas in the Northeast also have the highest mean prediction R^2, the areas in both the South and West have the lowest values, with the areas in the Midwest in the middle.

The locations of the urban areas based on their prediction R^2 values is shown on the map in Figure 3. The areas have been divided into four groups of ten areas, with the areas having the highest R^2 shown in darker red, those with the lowest in light yellow, and the intermediate groups in shades of orange. Most of the areas with the highest prediction R^2 are located in the northeastern quarter of the country. Charlotte is a little farther south, New Orleans is one of the older cities in the country, and then once again is Portland. Most of the next group, in the darker orange, are in the Northeast and Midwest census regions with Denver and Los Angeles the outliers. For the group with the second lowest R^2 values, Detroit and Kansas City are the outliers with the rest

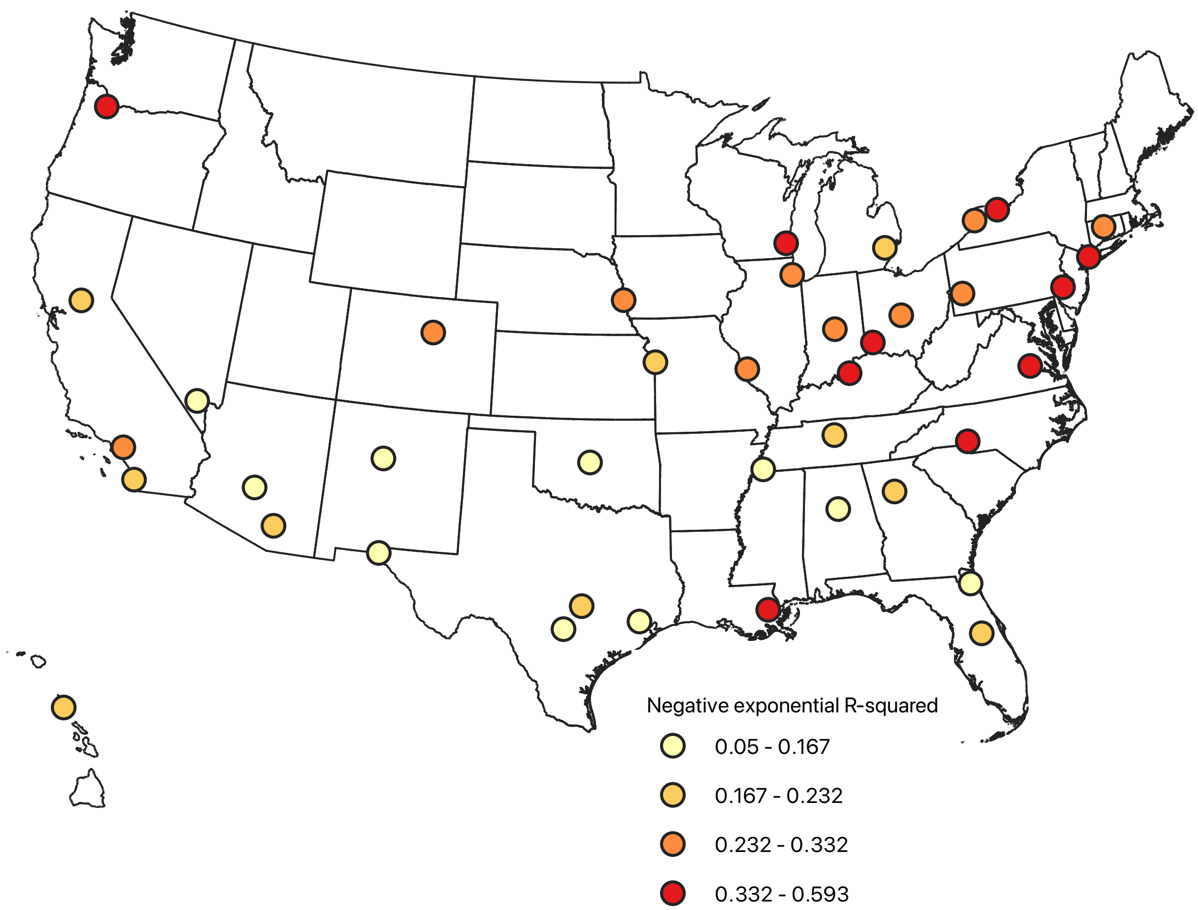

Figure 3. Urban areas classified by negative exponential model prediction R^2.

spread across the South and West. Finally, those urban areas in the lowest are clearly in the sunbelt, in the South and Southwest.

An exploration is undertaken of the association of some characteristics of the urban areas with central density, the density gradient, and the prediction R^2 in 2020. The approach takes a small set of urban area characteristics and uses regression to identify those that are statistically significant predictors of the negative exponential model results. The variables include the size of the urban area (number of housing units) and measures of change from 1970 to 2020, variables capturing the extent to which water and mountains serve as barriers to expansion, and when appropriate urban area density and change and other negative exponential results. The barrier variables are the proportion of the area of a five-mile ring surrounding the 2020 urban area covered by major bodies of water (oceans or the Great Lakes) and by mountainous lands. More detail on these measures is provided in an earlier paper on density and barriers to urban expansion (Ottensmann. 2023b).

Two variables are statistically significant in the regression predicting central density in a model that accounts for a large share of the variation, with an R^2 of 0.70. The results are given in Table 3. The larger the urban area, the higher the central density. This is reasonable as more households mean greater demand for locations closer to the

Table 3. Exploratory regression models predicting central density and density gradient in 2020 (standard errors in parentheses).

Independent variable	Central density 2020	
Housing units 2020 (millions)	4,397	***
	(622)	
Proportion ring area water	19,077	**
	(6,192)	
Constant	351	
	(1,172)	
***R*²**	0.701	***

Independent variable	Density gradient 2020	
Log central density 2020	0.1200	***
	(0.0103)	
Housing units 2020 (millions)	-0.0323	***
	(0.0045)	
Proportion ring area water	-0.1281	**
	(0.0416)	
Proportion ring area mountains	-0.0504	*
	(0.0198)	
Constant	-0.8522	***
	(0.0829)	
***R*²**	0.810	***

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

center. The water barrier to expansion—the proportion of the area of the ring water—is also positively related to central density. Most cities with significant water barriers were initially established on the coastline as ports with the CBD located near the water. The water reduces the supply of land near the CBD, increasing its price and thus the central density.

The prior section showed the density gradient to be related to the central density and the number of housing units in the urban area. These are two of the four significant variables in the model predicting the density gradient and have the expected signs, log of density positively related and housing units negatively related. Table 3 also gives the results for this model, which has an R^2 of 0.81. The two barrier variables, water and mountains, are both significant and negatively related to the density gradient. The presence of the barriers forces an urban area to expand farther outwards where there are no barriers compared to an urban are not so restricted. This more distant development then produces the lower density gradient.

The lists of the areas with the highest and lowest density gradients and prediction R^2 values suggests that the two could be related, with areas having lower density gradients having less variation in density giving relatively greater importance to random variation, producing a poorer fit for the negative exponential model. The results for the model for prediction R^2 in Table 4 show the gradient to be a highly significant predictor. Prediction R^2 also tends to be higher for larger urban areas. The third variable is the mountains barrier, negatively associated with prediction R^2. One

Table 4. Exploratory regression model predicting prediction R^2 in 2020 (standard errors in parentheses).

Independent variable	R^2 2020	
Density gradient 2020	1.5421	***
	(0.1324)	
Housing units 2020 (millions)	0.0198	***
	(0.0061)	
Proportion ring area mountains	-0.0986	***
	(0.0304)	
Constant	0.0539	*
	(0.0186)	
R^2	0.838	***

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

possibility is that development on more uneven terrain has greater variation in density, producing a poorer fit for the negative exponential model.

Change in negative exponential estimates 1970-2020

The previous section examined the current performance of the negative exponential model. This considers the changes in the performance of the model over time, looking at values and their changes from 1970 to 2020. Table 5 lists the basic statistics for central density, the density gradient, and the prediction R^2 for each census year from 1970 to 2020. The three measures exhibit different trends.

Central density presents a complex pattern. The mean drops from about 7,000 units per square mile in 1970 to 6,000 over the next three decades but then rises even higher to nearly 7,500 in 2020. This is pattern is found through much of the distribution, with both the minimum and median showing a similar pattern of decline and then rise, though the median does not quite reach the 1970 level in 2020. The maximum, which is New York, increases steady over the fifty-year period. Underlying these changes is much greater variation across the urban areas. The mean of the changes from 1970 to 2020 is nearly 500. But the median is -766. Over half of all the urban areas saw a substantial decline in the central density while at the same time the mean of the changes is positive, meaning larger gains outweighed smaller declines. This was obviously affected by the extremely large maximum increase of 15,000 for New York.

The density gradient experienced major declines, with the mean gradient dropping by half from 0.20 to 0.10 from 1970 to 2010. This was followed by a modest uptick in the final decade. Similar patterns occur for the median and the maximum gradients while the minimum fell and they stayed constant for from 2010 to 2020. The mean and median changes were similar and consistent with these trends, declines of about 0.09. But one urban area increased by that amount, while the urban area with the largest decline fell by a huge 0.31.

Prediction R^2 showed the most consistent trends over time, with the mean, median, and maximum declining in each decade. The minimum values did not show a steady trend, but the values were so small, all less than 0.10, that they reflect very little fit to the negative exponential model. The average declines in the R^2 values were substantial. The mean dropped from 0.41 to 0.26 and the median from 0.42 to 0.23. So the decline in the goodness-of-fit of the negative exponential model expected with the decentralization of employment is substantial. The mean and median of the changes from 1970 to 2020 are both negative and consistent with the year-to-year observations. But as with the other measures, changes for some areas are very large in both directions. The greatest decline of 0.52 meant the area had to have been among the highest in 1970 (it was Jacksonville in second place).

Table 5. Basic statistics for central density, density gradient, and prediction R^2 from 1970 to 2020 and changes from 1970 to 2020.

Year	Mean	Minimum	Median	Maximum
		Central density		
1970	6,989	1,828	4,965	43,120
1980	6,475	1,683	4,108	44,547
1990	6,206	1,709	3,782	47,909
2000	6,031	1,645	3,636	48,006
2010	6,339	1,706	3,979	52,297
2020	7,464	1,938	4,713	58,533
Change 1970-2020	475	-5,738	-766	15,413
		Density gradient		
1970	0.204	0.070	0.201	0.391
1980	0.150	0.032	0.149	0.289
1990	0.120	0.033	0.108	0.274
2000	0.107	0.031	0.095	0.272
2010	0.100	0.029	0.088	0.256
2020	0.112	0.029	0.090	0.284
Change 1970-2020	-0.092	-0.308	-0.089	0.091
		R^2		
1970	0.410	0.099	0.422	0.644
1980	0.348	0.016	0.354	0.633
1990	0.288	0.041	0.263	0.626
2000	0.283	0.050	0.263	0.642
2010	0.265	0.052	0.246	0.630
2020	0.257	0.050	0.232	0.593
Change 1970-2020	-0.154	-0.518	-0.139	0.105

Especially given that each measure saw large changes from 1970 to 2020 in both directions, it is again helpful to examine the areas at the extremes and their changes. Table 6 shows the areas having the greatest increases and greatest declines for central density, the density gradient, and the prediction R^2. Considerable consistency exists for the areas having the highest increases for all three of the measures. New York, Philadelphia, and Honolulu are on all three lists. Chicago and Portland are in the top 5 for both central density and the density gradient. And these were the areas that had the highest central density in 2020. New York, Philadelphia, and Chicago are older, very large urban areas. Honolulu is constrained by both water and mountains. And as noted earlier, Portland's urban growth boundary could be playing a role. These are among the areas that can be considered most consistent with the negative exponential model, with increasing parameter estimates and greater goodness-of-fit over time.

The lists of areas with the largest declines are not so consistent. Four of the urban areas in which the central density fell the most, by over 3,000 units per square mile, are older rustbelt areas in the Midwest and Northeast. The greatest decline occurred in New Orleans which in addition to being an older areas experienced the devastation of hurricane Katrina. Note that this decline also placed New Orleans on the list of the areas with the greatest decline in the density gradient. The other four areas with the largest drops in the gradient also had the largest drops in the prediction R^2 and were located in the sunbelt, in the South and Southwest.

Table 6. Urban areas with the highest and lowest changes in central density, density gradient, and prediction R^2 from 1970 to 2020.

Area	Change in central density
New York	15,413
Honolulu	10,855
Philadelphia	9,181
Chicago	6,072
Portland	6,012
...	...
St Louis	-3,390
Buffalo	-3,546
Detroit	-4,274
Cincinnati	-4,637
New Orleans	-5,738

Area	Change in density gradient
Philadelphia	0.091
Chicago	0.033
New York	0.020
Honolulu	0.014
Portland	0.000
...	...
San Antonio	-0.174
New Orleans	-0.184
Jacksonville	-0.225
Orlando	-0.300
Las Vegas	-0.308

Area	Change in R^2
New York	0.105
Honolulu	0.080
Richmond	0.066
Philadelphia	0.053
Milwaukee	0.033
...	...
Oklahoma City	-0.343
Las Vegas	-0.359
Orlando	-0.362
San Antonio	-0.465
Jacksonville	-0.518

The lists of areas having high and low change suggests the possibility of differences by region. While there are differences in the means for the central density and density gradient across the census regions, they are not statistically significant. Prediction R^2 does have significant differences across the region. The mean decline for the urban areas in the South is 0.25. Next lowest is in the Midwest dropping by 0.14, with the Northeast and West having even smaller mean declines. The differences in the prediction R^2 means across the regions are statistically significant.

The urban areas experience varying changes in the negative exponential estimates from 1970 to 2020 suggesting very different trends over the period. Some areas show steady increases, some consistent declines, with others remaining fairly constant. And some areas go down for a number of decades and then increase, some ending up below where they started in 1970 and others above. And in one case an area increases and then drops somewhat. To illustrate the variety of trajectories, the values for the central density, the density gradient, and the prediction R^2 for each year are plotted for a few selected urban areas.

Trends for central densities from 1970 to 2020 are shown in Figure 4. Two urban areas show consistent change over time, Austin (in red) increasing, and Cincinnati (purple) decreasing. Two more areas are examples of the decline and then rise seen for the means but with different final outcomes. Milwaukee (green) drops for three decades and then increases, but never gets back to the starting level in 1970. Central density in

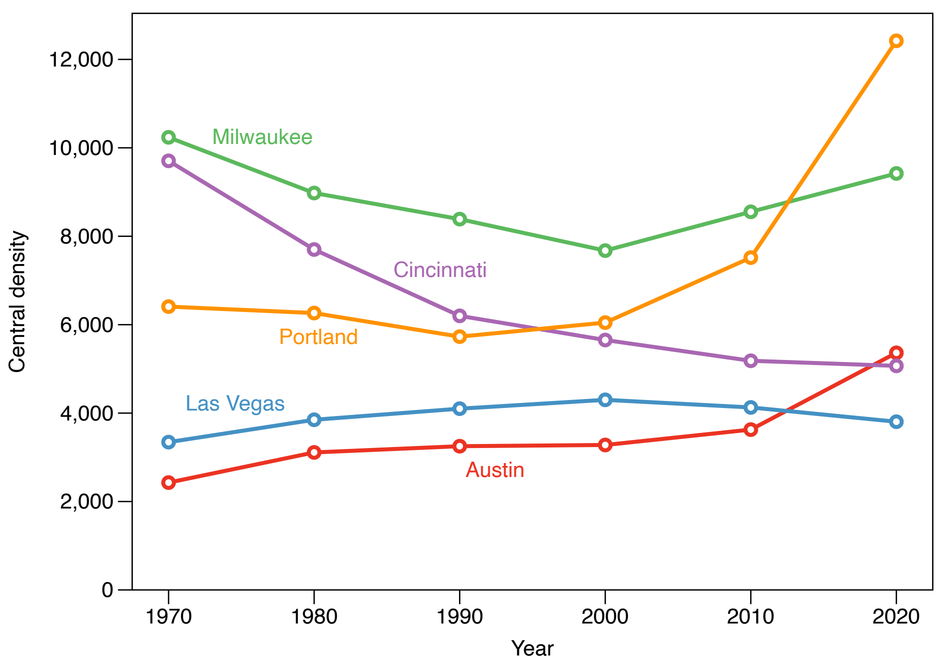

Figure 4. Estimated central densities for selected urban areas from 1970 to 2020.

Portland (orange) first declines but then shoots up to about twice its initial level by 2020. Finally, Las Vegas (blue) does the opposite, first rising and then falling. The changes are not large. But this case is unique in that this is the only instance of this among all of the urban areas for all three of the negative exponential measures.

Figure 5 displays the density gradients over time for four different urban areas. The gradient for Philadelphia (green) increases over time. Orlando's gradient (red) plummets, dropping to only 0.04 in 2020 from its initial level of 0.34. Charlotte (orange) is the example here of that pattern of decline followed by an upswing. If one looks closely, Los Angeles (blue) displays a similar pattern but the changes are so small that it is better seen as an example of an area for which the gradient has remained relatively stable over time.

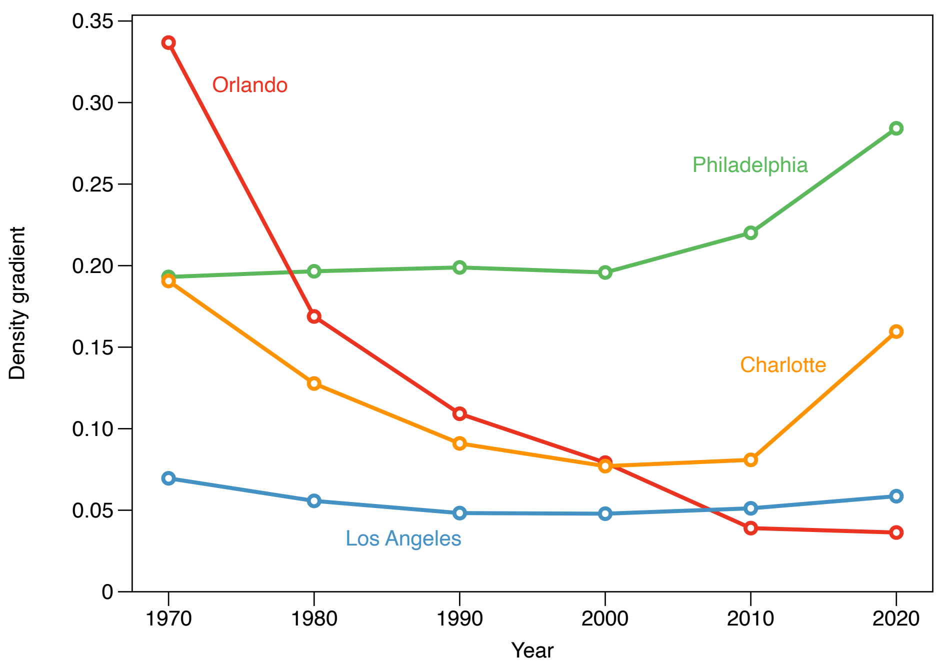

Figure 5. Estimated density gradients for selected urban areas from 1970 to 2020.

New York (orange) exhibits a moderate increase in the prediction R^2 over the fifty-year period as illustrated in Figure 6. Jacksonville (red) is the example for declining R^2, but the change is anything but modest. Its value drops from 0.62 in 1970, a very good fit to the negative exponential model, to 0.10 in the final year. The pattern of first decrease and then increase is exhibited here by Richmond (purple). Rochester (green) and Albuquerque (blue) are offered has examples of urban areas in which the R^2 values

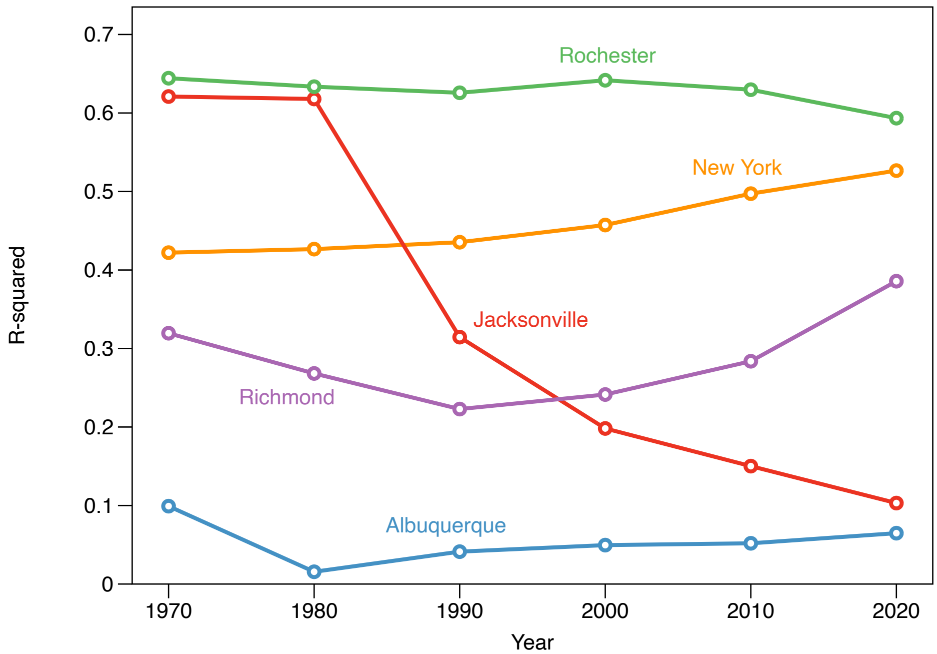

Figure 6. Estimated prediction R^2 values for selected urban areas from 1970 to 2020.

stay fairly constant over time but at very different levels. Rochester has the highest prediction R^2 among all urban areas in each of the years. Albuquerque was either the lowest or close to the lowest in every year.

Exploratory models are developed to consider the association of urban area characteristics with the changes in central density, the density gradient, and the prediction R^2 from 1970 to 2020. The procedure used for the previous analysis focused on the 2020 values of these measure is followed here: Regression models are created with statistically significant predictors from the same limited list of variables. The results are exactly what one would expect given the earlier findings for predicting the values in 2020. These models are completely consistent with the earlier models.

Table 7 gives the regression model results for the models predicting the change in central density, the density gradient, and the prediction R^2. Starting with the change in central density: Two variables are significant in the prediction of central density in 2020, the number of housing units and the water barrier. For the change in central density from 1970 to 2020, these two are again significant, along with the change in the number of housing units from 1970 to 2020. So for the change in central density, both the size of the urban area and the change in size are significant predictors.

In predicting the density gradient in 2020, the log of central density and the number of housing units are two of the four significant predictors. Here, for the change in the density gradient, the percent change of central density from 1970 to 2020 is

Table 7. Exploratory regression models predicting change in central density, density gradient, and prediction R^2 from 1970 to 2020 (standard errors in parentheses).

Independent variable	Change in central density 1970-2020	
Housing units 1970 (millions)	2,556	***
	(536)	
Change in log housing units 1970-2020	2,761	**
	(883)	
Proportion ring area water	7,660	*
	(3,461)	
Constant	-4,548	***
	(1,205)	
R2	0.530	***

Independent variable	Change in density gradient 1970-2020	
Percent change in central density 1970-2020	0.00128	***
	(0.00011)	
Log housing units 1970	0.01979	**
	(0.00634)	
Percent change in housing units 1970-2020	-0.00020	***
	(0.00003)	
Constant	-0.29521	**
	(0.08314)	
R^2	0.846	***

Independent variable	Change in R^2 1970-2020	
Density gradient 1970	1.059	***
	(0.206)	
Change in density gradient 1970-2020	1.998	***
	(0.192)	
Proportion ring area mountains	0.128	*
	(0.048)	
Constant	-0.206	**
	(0.039)	
R^2	0.760	***

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

significant along with both the log of housing units in 1970 and the percent change in the number of housing units over the period. The model predicting the density gradient in 2020 also includes both barrier variables, for water and mountains. These measures are not significant for predicting change in the density gradient.

For predicting R^2 values for 2020, the density gradient, the size of the urban area, and the mountains barrier are the three significant variables. Table 8 shows both the density gradient in 1970 and the change in the density gradient to be significant for the change in R^2, as is the mountains barrier variable. However, neither the size of the urban area nor the change in the size of the area are significant.

Conclusion

The urban areas exhibit tremendous variation in their patterns of density decline with distance from the CBD and how those patterns change over time. This means that simple conclusions about the results are not always possible. Making general statements about the findings from this study will require caveats. And even then it is important to recognize that it will be impossible to summarize the full range of the variation.

Starting with the negative exponential decline of density in 2020, central densities range from under 2,000 to nearly 60,000 units per square mile. A few older, very large urban areas have the highest central densities, and the mean is greatest for the urban areas in the Northeast. The number of housing units in the urban area and the extent to which water is a barrier to urban expansion are significantly related to the level of the central density.

The discussion of the reason that the density gradient is not a good measure of decentralization suggests that the gradient should be negatively related to the size of the urban area and directly related to the central density. Results confirm this, with the gradient significantly related to both of these variables and to the water and mountain barriers to expansion.

The predicted R^2 for the model in 2020 ranges from a high of 0.59 (over half the variation in density accounted for) down to 0.05 (the negative exponential model hardly fitting the density pattern at all). The map of the urban areas by their R^2 values shows a clear pattern. Most of the ten areas with R^2 highest are located in or near the Northeast. Those in the bottom quarter fall in a line stretching across the Sunbelt. This begins to account for the variation in the performance of the negative exponential model. Urban areas in the Northeast grew to substantial size in the nineteenth and early twentieth century, an era in which the CBD emerged and was dominant, with employment concentrated near the center as assumed in the monocentric model. The Sunbelt cities experienced most of their growth much more recently. They were small and did not have significant CBDs in those earlier years.

The changes in the negative exponential model values since 1970 gives some insight into what gave rise to these patterns. However it is important to stress that the urban areas followed very different trajectories with those values over the fifty years as illustrated above. The mean central density declines from 1970 to 2000 but then jumps significantly by 2020. This is an interesting and potentially very important development.

It will take more data and likely more time to understand the reason for this and to assess the significance. The small values for the mean and median changes, one positive and one negative, tell little about how central densities changed in the urban areas. The variation is astounding, with a drop of nearly 6,000 for one area (New Orleans) and an increase of 15,000 in another (New York). Change is positively related to the size of the area in 1970 and the change in the log of housing units over the period.

Just as the gradient is related to the central density and the size of the urban area, the change in the gradient is positively related to the change in the central density and to the size of the urban area in 1970. But it is negatively related to the change in the size of the urban area. More rapidly growing areas have greater declines in their density gradients. As the edge of the urban area extends farther out, the negative exponential curve would have to do so as well, decreasing the gradient.

Finally the change in predicted R^2 from 1970 to 2020 addresses the question as to whether the performance of the negative exponential model has declined over time. The mean change was -0.15 with urban areas ranging from -0.52 to +0.09. So the average urban area saw the drop in predicted R^2 from 0.41 in 1950 to 0.26 in 2020, a not inconsiderable decline of nearly forty percent. As the positive changes in R^2 for some urban areas shows, decline is not universal, but only a small number of urban areas had increases. So it is reasonable to conclude that the predicted R^2 values generally have been declining and the negative exponential model is doing more poorly in accounting for the variation in density within large urban areas.

References

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

Angel, Shlomo. 2012. *Planet of Cities*. Lincoln Institute of Land Policy.

Clark, Colin. 1951. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)* 114, 4: 490-496.

Gordon, P., H. W. Richardson, and H. L. Wong. 1986.The distribution of population and employment in a polycentric city: the case of Los Angeles. *Environment and Planning A* 18, 2: 161-173.

Greene, David L., and Joern Barnbrock. 1978. A note on problems in estimating exponential urban density models. *Journal of Urban Economics* 5, 2: 285-290.

Griffith, D. 1981, Modelling urban population density in a multi-centered city. *Journal of Urban Economics* 9, 3: 298-310.

Kemper, Peter, and Roger Schmenner. 1974. The density gradient for manufacturing industry. *Journal of Urban Economics* 1, 4 (October): 410-427.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. *IPUMS National Historical Geographic Information System: Version 17.0* [dataset].

Minneapolis MN: IPUMS. *http://doi.org/10.18128/D050.V17.0*. Data downloaded from *https://www.nhgis.org* on various dates, August to October 2022.
McDonald, John R. 1989. Econometric studies of urban population density: a survey. *Journal of Urban Economics* 26, 3 (November): 361-385.
Mills, Edwin S. 1972. *Studies in the Structure of the Urban Economy.* Johns Hopkins Press for Resources for the Future.
Muth, Richard F. 1969. *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*. University of Chicago Press.
Ottensmann, John R. 2022a. the decline of the density gradient is not a good measure of decentralization. Available at *https://urbanpatternsblog.files.wordpress.com/2022/06/density-gradient-centralization.pages_.pdf*.
Ottensmann, John R. 2022b. Nonlinear estimation of the negative exponential model of population density decline. Available at *https://urbanpatternsblog.files.wordpress.com/2022/05/nonlinear-negative-exponential.pdf*.
Ottensmann, John R. 2023a. Urban patterns of the largest urban areas in the United States, 1950-2020: a new dataset. Available at *https://urbanpatternsblog.files.wordpress.com/2023/02/new-urban-patterns-dataset.pdf*.
Ottensmann, John R, 2023b. Density of large urban areas in the U.S. and barriers to urban expansion, 1950-2020. Available at *https://urbanpatternsblog.files.wordpress.com/2023/05/density-and-barriers.pdf*.
Smith, Betty E. 1997. A review of monocentric urban density analysis. *Journal of Planning Literature* 12, 2: 115-135.
Thrall, Grant Ian. 1988. Statistical and theoretical issues in verifying the population density function. *Urban Geography* 9, 5: 518-537.
U.S. Census Bureau. 2022. Urban Area criteria for the 2020 Census—final criteria. *Federal Register* 97, 57 (March 24): 16706-16715.
Urban Institute and Geolytics. 2003. *Census CD Neighborhood Change Database (NCDB): 1970-2000 Tract Data*. Somerville, NJ: Geolytics.
White, Lawrence J. 1977. How good are two-point estimates of urban density gradients and central densities? *Journal of Urban Economics* 4, 3: 292-309.

Appendix: Nonlinear versus log linear estimation

Log linear estimation of the negative exponential model gives diminished weight to errors in the prediction of higher densities near the center, resulting in the potential underestimate of the central density. For this reason, nonlinear regression is used to estimate the values for the negative exponential model for the large urban areas.

Results for the New York urban area described above provide the specific (and extreme) example of this problem. In this appendix, the results for nonlinear and log linear estimation of the negative exponential model for all forty urban areas in 2020 are compared. Table A-1 gives the basic statistics for the nonlinear and log linear estimates, the difference, and the percent difference for the central density, the density gradient, and the R^2 values. The log linear estimates are most often lower than the nonlinear estimates, with some of the differences substantial.

The mean difference between the nonlinear and log linear estimates of the central density is over 4,000 units per square mile. The mean percent difference has the log linear estimates 41 percent lower. The log linear estimates are lower for every urban area, though the differences ranged from and eight percent decline to an 81 percent decline. The maximum decline is for the New York area, with Philadelphia in second place. These two also have the greatest percent decline but swapping positions, with Philadelphia highest.

The mean log linear estimate of the density gradient of 0.062 is just over half the nonlinear gradient of 0.112. The mean percent decline from the nonlinear to log linear gradients was -31 percent, a drop of nearly a third. However, the maximum difference and percent difference of 0.019 and 44.7 percent show that not all urban areas have

Table A-1. Basic statistics for nonlinear and log linear estimates of central density, density gradient, prediction R^2 in 2020 and the differences and percent differences.

Year	Mean	Minimum	Median	Maximum
Central density				
Nonlinear estimate	7,464	1,938	4,713	58,533
Log linear estimate	3,224	1,212	2,451	13,584
Difference	-4,241	-44,949	-1,763	-205
Percent difference	-40.6	-81.0	-39.4	-8.2
Density gradient				
Nonlinear estimate	0.112	0.029	0.090	0.284
Log linear estimate	0.062	0.019	0.053	0.212
Difference	-0.049	-0.220	-0.034	0.019
Percent difference	-31.6	-77.6	-39.2	44.7
R^2				
Nonlinear estimate	0.257	0.050	0.232	0.593
Log linear estimate	0.178	0.016	0.160	0.491
Difference	-0.079	-0.345	-0.083	0.177
Percent difference	-26.6	-82.4	-35.4	81.8

lower log linear estimates of the density gradient. For seven areas, the log linear estimate of the gradient is larger than the nonlinear estimate.

Mean R^2 values are 0.26 for the nonlinear estimates and 0.18 for the log linear estimates, with the mean percent difference of -27 percent. As with the density gradient, the log linear estimates the R^2 values are greater for some urban areas. Seven areas also have increases, though only two urban areas, Albuquerque and Jacksonville, increase for both the density gradient and R^2. (These two areas have shown up as having extreme values in some of the earlier results.)

The similarities in the percent changes between the nonlinear and log linear estimates for some of the basic statistics is striking. The mean percent change for the central density is -41 percent, for the density gradient -32 percent, and the R^2 values -27 percent, differences of generally comparable magnitude. But the comparisons are more striking for the minima and medians. For the minima, the mean percent changes are -81, -78, and -82 percent for central density, density gradient, and R^2. These results for the median were -39, -35, and -35 percent respectively. Only for the maxima was there a divergence. All of the central density differences were negative. The density gradient and R^2 had some positive changes, the significant maxima of 45 and 82 percent for the density gradient and R^2. Those positive values account for greater differences in mean percent change among the means compared to the medians.

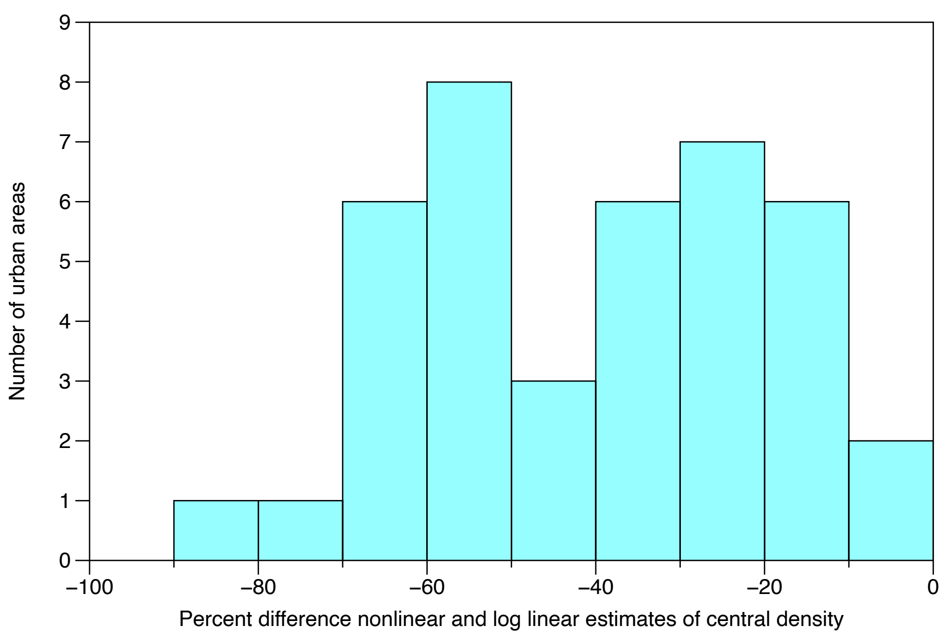

Figure A-1. Histogram of distribution of urban areas by percent difference between nonlinear and log linear estimates of central density in 2020.

The estimate of the central density can influence the density gradient which in turn can influence R^2. This makes the variation in the differences in the nonlinear and log linear estimates of the central density especially important. Figure A-1 is a histogram of the distribution of the urban areas by their percent difference in the central density estimates. From Table A-1, the values range from -8 percent to -81 percent. Only two of the urban areas have declines of less than ten percent and only two areas have declines greater than seventy percent. The remaining urban areas are relatively uniformly distributed across the interval from -10 to -70. The differences between the nonlinear and log linear estimates vary widely but are generally more than minimal.