Economics Bulletin

Volume 39, Issue 1

A Comparison of Tract-Level, Nationwide Indices of Economic Deprivation

Scott W Hegerty
Northeastern Illinois University

Abstract

Indices of socioeconomic deprivation, which combine a number of variables into a single measure, are often used in public health and other fields to examine geographic disparities in health outcomes and quality of life. Much of the research using these indices has been conducted outside the United States, and often focuses heavily on urban areas. This study uses Principal Component Analysis (PCA) to combine a set of socioeconomic variables for more than 72,000 Census tracts in all 50 U.S. states to construct a set of deprivation indices for the year 2015. These measures are highly correlated with one another and with measures that use a different weighting scheme. A comparison of our main index with a simpler measure—tract-level poverty rates—show the two to be highly correlated, but that the deprivation index value is higher than predicted by poverty alone. This is particularly true when spatial autocorrelation is incorporated into the model. An analysis of only the 14,000 tracts within the largest cities shows less of a discrepancy between these two measures, but that spatial autocorrelation is still an issue. Deprivation indices, therefore, are shown to capture more than just poverty, particularly when geography is taken into account, for both urban and rural areas.

The helpful comments of an anonymous referee are greatly appreciated. All remaining errors are, however, mine.
Citation: Scott W Hegerty, (2019) "A Comparison of Tract-Level, Nationwide Indices of Economic Deprivation", *Economics Bulletin*, Volume 39, Issue 1, pages 256-263
Contact: Scott W Hegerty - S-Hegerty@neiu.edu.
Submitted: December 03, 2018. **Published:** February 18, 2019.

1. Introduction

Individuals' physical and emotional well-being are often strongly influenced by their socioeconomic circumstances, which can be related to their physical location. Quantifying these relationships often requires an appropriate multidimensional measure that captures such factors as relative income, quality of life, and housing or family status across space. In fields such as public health, correlations between such measures and health outcomes can influence policy and help save lives. Indices of (multiple) deprivation combine a number of socioeconomic variables over a geographic space. But the variables, the method by which they are combined, and the study area can have an effect on the index itself. In particular, we note four main issues: The choice of variables for the index, the weighting scheme used to combine them, a relative lack of U.S.-based measures, and differences between urban and rural areas.

One "benchmark" deprivation measure was introduced by Townsend (1987), who measures material and social disadvantage in Britain using 77 indicators including diet, clothing, housing, education, and integration and social inclusion. Over time, these measures have become more streamlined, incorporating fewer variables while still capturing deficiencies in well-being. They are often used as covariates to assess various disparities. Morris and Carstairs (1991) find that deprivation is highly correlated with various health measures, but that some measures perform better than others depending on their choices of variables. Carstairs (1995) also finds high correlations between various deprivation indices and mortality. Aaberge and Brandolini (2014) rigorously compare approaches to calculating deprivation measures, noting the controversies involved with the selection of dimensions and weights.

As Bell *et al.* note, principal component analysis (PCA) is the most commonly used weighting scheme, although a variety of alternatives exist. Salmond *et al.* (1998), for example, use PCA to incorporate nine variables for New Zealand, including the receipt of benefits, the unemployment rate, schooling, occupancy, housing tenure, car access, single-parent families, and divorce status; they also find correlations between this measure and cancer and mortality rates. Langlois and Kitchen (2001) also use PCA to construct a deprivation index for Montreal in the 1990s. At the same time, Broadway and Jesty (1998) examine unemployment, rates, the lack of a ninth-grade education, and low income separately (and uweighted) in Canadian inner cities.

While the country of interest is the same, the choice of measures also differs between Pacione (2002) and Seaman *et al.* (2015), who use PCA and the Scottish Index of Multiple Deprivation, respectively. The former study also focuses on rural areas and notes that variable choice and measurements might differ substantially between rural and urban areas. Bertin *et al.* (2014) argue that rural areas cannot be reliably served by measures designed for cities. In addition, measures designed for other countries, which form a large share of the literature, might not capture conditions in the United States. Smith (2009), for example, uses PCA to construct a deprivation measure and examine environmental inequality in Detroit and Portland (Oregon), but such studies are relatively limited and there is a lack of countrywide analyses.

This study constructs a measure of deprivation for the 72,226 Census tracts for which there are sufficient data in all 50 United States for the year 2015. Of the variables typically included in the literature, we focus on the lack of necessities such as food, education, employment, and neighborhood stability rather than variables that infer reduced quality of life based solely on ethnicity or marital or family status. We find that, when comparing alternative selections of variables, the resulting deprivation indices are highly correlated with one another. We also compare two alterative weighting schemes (PCA and variance-smoothing weights) in constructing our indices. These are shown to be highly correlated as well. To test whether multivariate deprivation indices have any advantage over simpler choices of variables, we then compare one PCA measure with the poverty rate, for all U.S. tracts, as well as for the subset of 13,885 tracts that are located within large cities. While the two measures are indeed highly correlated, regression techniques show the deprivation index to be higher than what poverty rates might predict. This divergence is even stronger when spatial autocorrelation is taken into account. At the same time, a purely urban model shows deprivation and the poverty rate to be more closely linked, and that spatial correlation plays a smaller, yet still significant, role.

2. Methodology

Using tract-level data from the U.S. Census Bureau (2015 ACS 5-year estimates), we use between three and five component variables to include in our single deprivation index:

1) The percent of adults 25 to 64 without a high-school diploma
2) The percent of recipients of nutrition benefits (SNAP)
3) The unemployment rate
4) The vacancy rate
5) The percent of households earning less than $15,000 annually.

These are based on the approach of Salmond *et al.* (1998), and some compared by Morris and Carstairs (1991), and capture various types of economic and noneconomic deprivation. Here, we omit variables that might be considered benign—many families choose not own a car, for example, particularly in urban areas. In addition, we do not equate minority residents with "deprivation," as has sometimes been the case in the literature. While we briefly considered a measure of "crowding," which is the percentage of households with more than 1.5 persons per room, this not only was noted by Blake *et al.* (2007) not to be an effective measure in the United States, but was also shown in our preliminary estimates to have a median value of zero. We therefore use the tract-level vacancy rate to capture neighborhood blight and instability.

We begin our main analysis by generating the first principal component of all five series (which we name *PC5*) for the 72,226 tracts with complete data in 2015. We do the same for the first three and the first four variables listed above, generating *PC3* and *PC4*, respectively. We also use "variance-smoothing" weights for the selected components for k = [3,4,5], deflating each by its own standard deviation to ensure that the component with the highest variance does not dominate the index, to generate *SD3*, *SD4*, and *SD5*:

$$SD(I) = \sum_{i=1}^{k} \frac{1}{\sigma_i} X_i \tag{1}$$

All variables are normalized to cover the [0,100] range. We then compare all six indices, as well as the poverty rate, using nonparametric (Spearman) correlations. We also calculate the degree of spatial autocorrelation for each variable using Moran's *I*:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{2}$$

Here, as in our spatial lag regression below, we use Queen contiguity of order one for our weights, which meets the recommendation of Lesage (2014) that the simplest matrix be used, but is also empirically tractable given the sample size.

We next select a “best” measure of multivariate deprivation to compare against a univariate alternative. Although they are highly correlated with variance-weighted measures, we prefer the PCA-based measures because of their mathematical sophistication and their wide use in the literature. We choose the four-variable model because we do not wish to duplicate an income variable when comparing multivariate deprivation and the poverty rate, and because PC4 clearly has one valid principal component (*PCA5* has two eigenvalues above one). All variables in this index have sufficiently large factor loadings (including the vacancy rate, although it is shown below to have the smallest among the four components).We next compare this index to the simple tract-level poverty rate for both the United States and a subset of large-city tracts.

This comparison helps to address the degree of improvement provided by using such an index, which includes education and other factors, instead of purely *economic* deprivation. In addition to including Pearson and Spearman correlations, we also conduct bivariate Ordinary Least Squares (OLS) and spatially lagged regressions. This spatial method is explained further by Ward and Gledtisch (2008) and others. In this model, which incorporates correlations among neighboring tracts as ρ, the regression equation can be written as:

$$y = (I - \rho W)^{-1}(X\beta + \varepsilon) \qquad (3).$$

Overall, we find our measures to be nearly perfectly correlated. Correlations are somewhat lower, but still high overall, between deprivation and the poverty rate. There is a fraction of high-poverty tracts that do not have correspondingly high deprivation scores; these are disproportionately urban. In addition, PCA measures tend to exceed their predicted values, especially when spatial autocorrelation is taken into account. When we repeat the bivariate regressions for only the 13,931 large-city tracts (19.3 percent of the total), we find that these show a stronger relationship between deprivation and poverty rates, but that discrepancies between spatial and nonspatial models persist.

3. Results

Table 1 provides the PCA results and component weights. In all but one case, exactly one eigenvalue is greater than one, so the first principal component is valid for our analysis. The percentages of SNAP recipients load the highest on all three 50-state PCA-based measures, but unemployment and education have the highest loadings for the urban tracts. The unemployment rate has the largest inverse standard deviation for all tracts, while the vacancy rate has the highest value for urban tracts. Clearly urban deprivation differs from deprivation in the country overall.

Table 1: Principal Components Analysis and Inverse Standard-Deviation Weights.

		All Tracts (N = 72,226)							PC4, City Tracts (N = 13,931)		
		PC5		PC4		PC3					
PC	Variable	EV	Loadings	EV	Loadings	EV	Loadings	1/SD	EV	Loadings	1/SD
1	NOHS	1.223	0.516	1.222	0.518	1.211	0.547	0.092	1.302	0.537	0.067
2	PERCSNAP	1.002	0.600	0.983	0.601	0.867	0.618	0.081	0.892	0.437	0.073
3	UNEMP	0.980	0.552	0.859	0.553	0.732	0.564	0.175	0.742	0.543	0.063
4	PERCVAC	0.859	0.256	0.732	0.255			0.094	0.661	0.475	0.137
5	PERC15K	0.732	0.060					0.044			

Table 2 provides Spearman correlation coefficients; all correlations between deprivation measures are above 0.9. Figure 1 depicts scatterplots of three pairs of deprivation indices; while

Table 2: Spearman Correlations among Full-Sample Measures.

Measure	PC4	PC3	SD5	SD4	SD3	PERCPOV
PC5	1.000	0.977	0.986	0.986	0.977	0.819
PC4	1	0.977	0.985	0.986	0.977	0.818
PC3		1	0.929	0.930	1.000	0.814
SD5			1	0.999	0.929	0.802
SD4				1	0.930	0.798
SD3					1	0.810

PC4 is somewhat larger than *PC3* at low values, and somewhat higher than *SD4* in the middle range, there is a clear linear relationship among the variables.

Figure 1: PCA vs. Standard Deviation-Weighted Deprivation Measures.

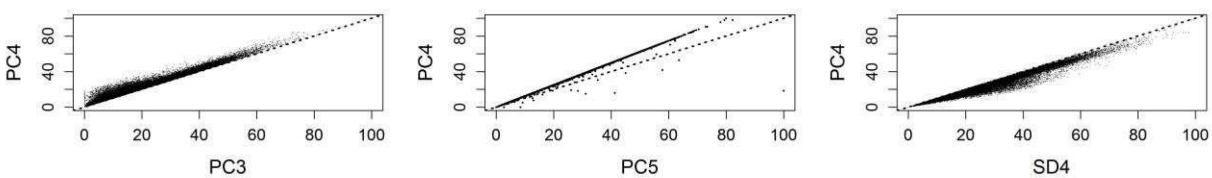


Table 3 provides a summary of our six country-wide deprivation measures, as well as poverty rates, and the *PCA4* measure and the poverty rate for tracts within cities larger than 250,000 inhabitants. One key, yet not necessarily unexpected, finding is that the *PCA4_250* and urban poverty values are generally higher than those for the full sample. Both deprivation and poverty rates are spatially autocorrelated, as shown by their Moran's *I* coefficients, but autocorrelation is higher among the urban tracts.

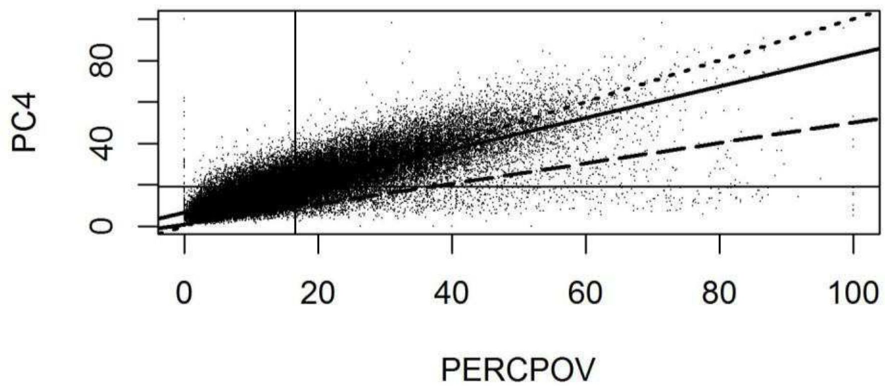
Table 3: Summary Statistics for Socioeconomic Deprivation Measures.

	1Q	Median	Mean	3Q	SD	Moran's I
PC5	8.22	13.19	15.43	20.37	9.54	0.655
PC4	10.26	16.47	19.26	25.43	11.90	0.656
PC3	8.49	14.03	16.96	22.63	11.29	0.645
SD5	1.19	1.89	2.18	2.89	1.36	0.593
SD4	11.92	18.96	21.81	29.03	12.87	0.665
SD3	8.46	13.95	16.83	22.42	11.16	0.645
PCA4_250	11.29	21.49	24.32	35.09	15.51	0.693
PERCPOV	7.10	13.20	16.57	22.70	12.65	0.577
PERCPOV_250	9.80	19.00	21.99	31.5	14.85	0.610

Figure 2 plots *PC4* against the poverty rate for all 72,226 census tracts. While the two measures appear highly correlated, there are a small number of "outlier" tracts with relatively high poverty rates and relatively low PCA scores. We examine those tracts with both above-average deprivation and below-average poverty rates and find that a total of 5,496 tracts (7.6% of the total) meet both criteria; 1,290 of these (a disproportionately high 23.5%) are located in large cities. Two regression lines—from OLS and a spatial lag model—are included in Figure 2 as well. The low slope of the latter suggests that geographic factors, particularly spatial autocorrelation, might be key when analyzing these variables. Both regressions show that the PCA index exceeds its predicted values, particularly for all census tracts.

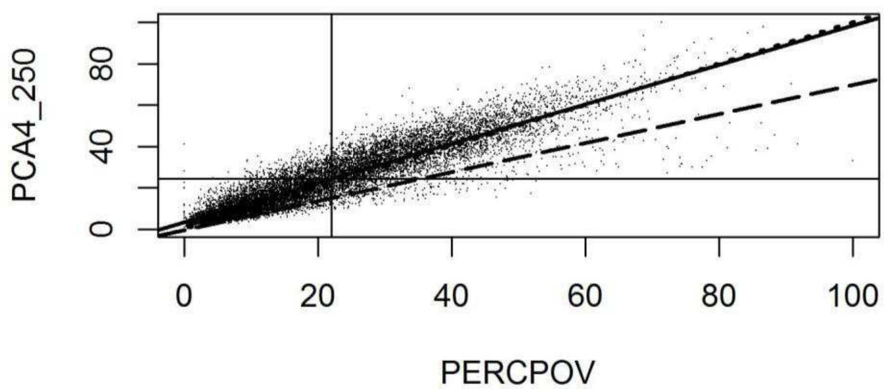
Figure 3 presents a similar graph for only the large-city tracts (*PCA4_250*). This deprivation index is more closely connected to poverty, with an OLS slope coefficient of nearly one. The spatial lag regression, which performs the best of all four 50-state and urban models in terms of R-squared, also diverges less in the urban specification than in the full sample. This suggests that in urban areas, indices of socioeconomic deprivation might be less informative than in rural ones and that urban poverty might better capture deprivation in large cities.

Figure 2: (Normalized) PCA Deprivation Measure vs. Percent Poverty for All Tracts.



Correlation = 0.808 (Pearson) 0.818 (Spearman)	45-degree line (short dashed)		
Regression line (solid):	*PC4* = 0.660+ 0.760**PERCPOV*		R^2 = 0.653
Spatial regression (dashed):	*PC4* = 1.058 + 0.491**PERCPOV*	ρ = 0.527	R^2 = 0.785
Horizontal and vertical lines: Mean values of variables.			

Figure 3: PCA Deprivation Measure vs. Percent Poverty for 13,931 Large-City Tracts.



Correlation = 0.911 (Pearson) 0.920 (Spearman)	45-degree line (short dashed)		
Regression line (solid):	*PCA4_250* = 3.397 + 0.951**PERCPOV*		R^2 = 0.830
Spatial regression (dashed):	*PCA4_250* = -0.547 + 0.702**PERCPOV*	ρ = 0.387	R^2 = 0.886
Horizontal and vertical lines: Mean values of variables.			

Urban deprivation might itself differ from rural deprivation: In Figure 4 at the end of this study, we see that many of the tracts with the highest index values appear to be rural: Appalachia and the Southeast, as well as the South and West, have large contiguous areas with values in the top quartile. Further research is necessary to assess how these tracts differ from large cities or suburban areas. Nonetheless, there is sufficient evidence that multivariate indices provide useful information, beyond that given by single measures, for all types of geographic area. The choice of component variables might matter most outside of urban areas.

4. Conclusion

Indices of socioeconomic deprivation are calculated and applied in a wide range of fields to assess relationships with health, quality of life, or other variables that vary geographically. While Principal Component Analysis (PCA) is often applied to reduce a number of economic and other variables into a single measure, the exact choice of variables and weighting scheme are subject to considerable debate. In addition, many previous studies have done so for cities in the United Kingdom and other countries; relatively little has been done to calculate such a measure for non-urban areas, or for the United States in general.

This paper does both, calculating deprivation indices for more than 72,000 Census tracts, as well as for the nearly 14,000 tracts that are located within large U.S. cities. PCA and a variance-smoothing method produce very similar results, even when various selections of socioeconomic variables are used in the index. Of five potential explanatory variables, which capture deficiencies in education, employment, food access, neighborhood quality, and income, we select the measure that includes the first four. Comparing this multivariate PCA measure with a univariate measure (the poverty rate), we find the deprivation index tends to be higher than what the poverty rate might predict, but not in certain high-poverty, urban tracts. Estimating the model for only urban tracts shows a much closer connection between deprivation and poverty. Taking spatial autocorrelation into account increases this disparity, but the divergence between deprivation and poverty is more distinct for the nationwide sample than for the urban subset of tracts.

Not only does this study generate a useful, nationwide index of socioeconomic deprivation for the entire United States, its findings suggest that methods developed for urban areas produce different results elsewhere in the country. In addition, controlling for spatial correlation is important when modeling deprivation, both in urban areas and elsewhere. Nonetheless, such indices provide more information about a geographic area than does a simpler statistic such as the poverty rate alone. Further research could examine this divergence further.

References:

Aaberge, R. and A. Brandolini (2014) "Multidimensional Poverty and Inequality" Bank of Italy Temi di Discussione (Working Paper) No. 976.

Bell, N., N. Schuurmanm, and M.V. Hayes (2007) "Using GIS-Based Methods of Multicriteria Analysis to Construct Socio-Economic Deprivation Indices" *International Journal of Health Geographics* **6**:17.

Bertin, M., C. Chevrier, P. Fabienne, T. Serrano-Chavez, S. Cordier, and J.-F. Viel (2014) "Can a Deprivation Index Be Used Legitimately over Both Urban and Rural Areas?" *International Journal of Health Geographics* **13**:22.

Blake, K.S. and R.L. Kellerson (2007) "Measuring Overcrowding in Housing" Econometrica, Inc.: Bethesda, Maryland, prepared for the U.S. Department of Housing and Urban Development Office of Policy Development and Research.

Broadway, M.J. and G. Jesty (1998) "Are Canadian Inner Cities Becoming More Dissimilar? An Analysis of Urban Deprivation Indicators" *Urban Studies* **35**, 1423-1438.

Carstairs, V. (1995) "Deprivation Indices: Their Interpretation and Use in Relation to Health" *Journal of Epidemiology and Community Health* **49**, S3-S8.

Langlois, A. and P. Kitchen (2001) “Identifying and Measuring Dimensions of Urban Deprivation in Montreal: An Analysis of the 1996 Census Data” *Urban Studies* **38**, 119-139.

LeSage, J.P. (2014) “What Regional Scientists Need to Know about Spatial Econometrics” *The Review of Regional Studies* **44**, 13-32.

Morris, R. and V. Carstairs (1991) “Which Deprivation? A Comparison of Selected Deprivation Indexes” *Journal of Public Health Medicine* **13**, 318-326.

Pacione, M. (2004) The Geography of Disadvantage in Rural Scotland” *Tijdschrift voor Economische en Sociale Geografie* **95**, 375-391.

Salmond, C., and P. Crampton (1998) “NZDep91: A New Zealand Index of Deprivation” *Australian and New Zealand Journal of Public Health* **22**, 833-837.

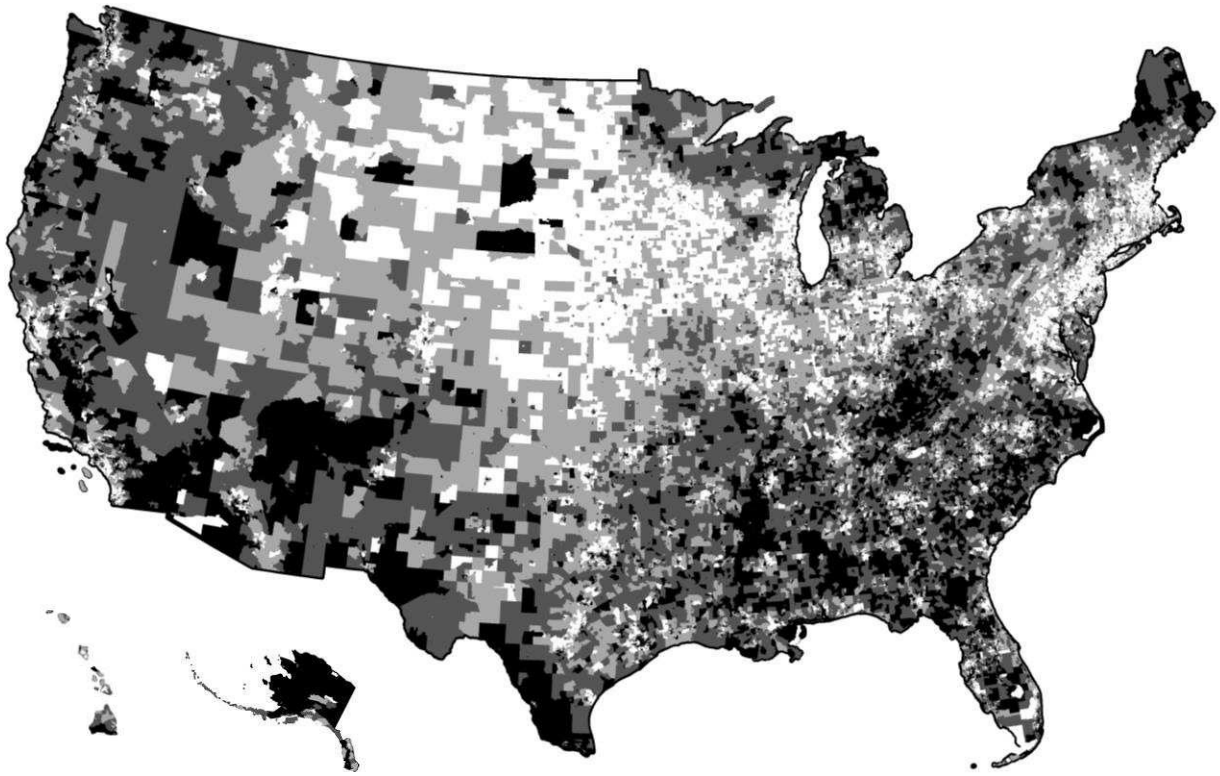
Seaman, R., R. Mitchell, R. Dundas, A.H. Leyland, and F. Popham (2015) “How Much of the Difference in Life Expectancy between Scottish Cities Does Deprivation Explain?” *BMC Public Health* **15**,1057.

Smith, C.L. (2009) “Economic Deprivation and Racial Segregation: Comparing Superfund Sites in Portland, Oregon and Detroit, Michigan” *Social Science Research* **38**, 681-692.

Townsend, P. (1987) “Deprivation” *Journal of Social Policy* **16**, 125-146.

Ward, M.D. and K.S. Gledtisch (2008) *Spatial Regression Models*, Sage Publications: Thousand Oaks, Calif.

Figure 4: PC4 Deprivation Index Measure for 72,226 U.S. Census Tracts.



Key: lowest quartile = white; highest quartile = black.

Table 4: List of Cities with Population above 250,000 (N = 78).

Albuquerque, NM
Anaheim, California
Anchorage, Alaska
Arlington, Texas
Atlanta, Georgia
Aurora, Colorado
Austin, Texas
Bakersfield, California
Baltimore, Maryland
Boston, Massachusetts
Buffalo, New York
Corpus Christi, Texas
Chicago, Illinois
Chula Vista, California
Cincinnati, Ohio
Cleveland, Ohio
Colorado Springs, Colorado
Charlotte, N.C.
Columbus, Ohio
Dallas, Texas
Denver, Colorado
Detroit, Michigan
El Paso, Texas
Fresno, California
Fort Worth, Texas
Fort Wayne, Indiana

Source: U.S. Census ACS
5-year estimates,2015.

Greensboro, NC
Henderson, Nevada
Honolulu, Hawaii
Houston, Texas
Indianapolis, Indiana
Jacksonville, Florida
Jersey City, New Jersey
Kansas City, Missouri
Long Beach, California
Los Angeles, California
Lexington, Kentucky
Lincoln, Nebraska
Louisville, Kentucky
Las Vegas, Nevada
Memphis, Tennessee
Mesa, Arizona
Miami, Florida
Minneapolis, Minnesota
Milwaukee, Wisconsin
Nashville, Tennessee
New Orleans, Louisiana
Newark, New Jersey
New York, New York
Oakland, California
Oklahoma City, Oklahoma
Omaha, Nebraska

Orlando, Florida
Portland, Oregon
Philadelphia, PA
Phoenix, Arizona
Pittsburgh, PA
Plano, Texas
Raleigh, North Carolina
Riverside, California
Sacramento, California
San Antonio, Texas
San Diego, California
Seattle, Washington
San Francisco, CA
San Jose, California
Santa Ana, California
St. Louis, Missouri
Stockton, California
St. Paul, Minnesota
St. Petersburg, Florida
Tampa, Florida
Toledo, Ohio
Tucson, Arizona
Tulsa, Oklahoma
Virginia Beach, Virginia
Washington, DC
Wichita, Kansas