What Can 70 Million or So Neighborhoods Tell Us about the Performance of Income Segregation Measures?

Sergio J. Rey
GeoDa Center
School of Geographical Sciences
Arizona State University

David C. Folch
GeoDa Center
School of Geographical Sciences
Arizona State University

June 7, 2009

Abstract

Segregation indices are at the core of many studies on the wellbeing of urban areas. However, their distributional properties have been largely ignored in the literature. It is often taken for granted that urban areas of all kinds can be run through the same index with truly comparable results. This paper explores the impact of varying forms of spatial effects on the sampling distributions of four income segregation indices to identify where researchers should be concerned when using these measures. This paper specifically addresses four indices: Dissimilarity Index (D), Adjusted Dissimilarity Index (D(adj)), Neighborhood Sorting Index (NSI) and Generalized Neighborhood Sorting Index (GNSI). However, some of the conclusions apply to the broader field of spatial measures of segregation. It is found that the indices generally hold up well to variation in the number or area with the region indicating their comparability across regions of different sizes. They failed to accommodate for the situation where "real world" neighborhoods are arbitrary grouped into larger administrative areas. The level of integration of the region, as measured by three spatial autocorrelation models, showed some impact on the index values, but not to a large degree. Finally, the magnitude and direction of spatial autocorrelation (ρ) was picked up by the spatialized segregation indices, but not in an intuitive manner.

The use of segregation indices is often done with little consideration of the underlying properties affecting the level of *measured* segregation. Large regions are compared to small ones, declining ones to those experiencing rapid growth, dense to sprawling, and the list goes on. Can the indices be used in a one-size-fits-all fashion? This paper explores some of the properties of residential income segregation indices in the presence of varying degrees of spatial effects.

There is an extensive literature on the measurement of segregation. Included in this are taxonomies of measures, rules defining what makes an appropriate measure and of course the proposal of new measures. However, there is little controlled simulation testing of the measures. Some authors offer small simulations based on a limited number of extreme scenarios to test their measures (e.g. Wong (1993)), while others use real world data from a diverse set of regions to explore the properties of the measures (e.g. Krupka (2007)). While these provide direction, the existing simulation work does not provide sufficient real world variation to explore the properties, and the real world data likely have too much unexplained variation to isolate the factors affecting the properties of the indices.

In this paper, 285,000 regions containing a total of 69,825,000 neighborhoods and billions of households are simulated to examine the properties of four income segregation indices. They are tested along four spatial dimensions: region size defined in terms of number of neighborhoods, variation in neighborhood interconnectivity through three spatial autocorrelation models, variation in spatial clustering of like neighborhoods via 19 values of an autocorrelation parameter and the effect of scaling small neighborhoods into larger administrative areas.

The remainder of the paper is laid out in five sections beginning with a review of the challenges faced in the measurement of segregation. This is followed by a description of the simulation procedure in Section 2, which leads into a summary of the results in Section 3. A brief discussion follows the results looking into ways to improve upon the current spatial measures of segregation. The paper concludes in Section 5 with a summary of results and future directions.

1 Motivation

The concept of segregation is relatively intuitive. However, the fact the it operates along two entirely separate dimensions makes it difficult to quantify. First, it is concerned with the classification of a person (or firm or thing) between two or more categories, e.g. rich-poor, black-white, etc. The second concern is where that person is (loosely defined), e.g. central city versus suburb, white collar versus blue collar job, etc. All this provides considerable opportunity to debate a method for encapsulating all the information into a single number. We will briefly review this debate here, but more thorough treatments can be found in Massey and Denton (1988) and Reardon (2006), among others.

Within the broad field of segregation measurement, this paper primarily concerns itself with *residential* segregation, or the concentration of persons (or households) across neighborhoods within an urban area. Examples of other applications include the racial distribution of students among various schools or gender distribution of workers among job types. One distinguishing characteristic of residential segregation is that the "where" of a person is necessarily an areal unit. In the counter-examples, schools and job types do not necessarily have a spatial dimension and thus avoid some of the complicating factors surrounding the arrangement of the people to one another. It might be better stated that a map provides for a clear view of how the people are arranged, whereas in the counter-examples an arrangement is not as transparent.

As a subset of residential segregation, this paper is further focused on *income* segregation. Typically, income segregation studies implement a discrete classification of households into categories such as rich-poor, low-middle-high income or an even greater number of income categories. This is a direct translation of the *classic* race-based segregation indices into the income domain. There are of course drawbacks to this method since income is not a discrete viable like race, and so certain decisions must be made by the researcher as to the definition of the categories. Jargowsky (1996) and Jargowsky and Kim (2005) have provided indices which allow for income to be treated more naturally as a continuous variable. As will be seen in the simulation section, we focus on two discrete indices and the two put forward by Jargowsky. Two points should be noted: 1) in general, the remainder of this section addresses the discrete categorization of people which is the typical way residential segregation of all types is measured, and 2) the terms discrete and continuous will also be applied to the segmentation of space, which is a separate concept from discrete and continuous categories.

This section is divided into two parts. The first looks at the measurement of residential segregation and those factors which confound its measurement when space is involved. The second probes issues surrounding how to use the segregation data generated from the chosen index. This split should not imply the separation of these two components in an empirical study of segregation. In fact, the second section will likely drive the choices made in the first, with some feedback along the way as the strengths and weaknesses of the various segregation measures are considered.

1.1 Measuring Segregation

While the intuition about residential segregation is relatively straightforward, the lack of a unifying definition confounds the ability to provide a single measure for it (Massey and Denton, 1988). From a methodological perspective, one must transform this intuition into available data, explicit mathematical functions and polygons on a map. From this departure point we will briefly explore the latter two issues: choosing among the wealth of indices and the explicit incorporation of space into the computation. The first issue will be delved into more deeply in the section on analyzing the segregation. However, it should be noted that the availability of data and computational resources has colored much of the debate in this section also (Cortese et al., 1978; Wong, 2003).

Indexing indices of segregation is no simple task. Duncan and Duncan (1955) discuss six, James and Taeuber (1985) five, Massey and Denton (1988) 20 and Reardon and Firebaugh (2002) six, which are only but a few of the higher profile reviews of the measures. While the various choices are correlated, they capture different dimensions of the intuition behind segregation.

1.1.1 Treatment of Space

By their very nature, any measure of residential segregation is spatial. This is the spatial regimes notion of "spatial" where basic observational units, in this case people or households, are aggregated into neighborhoods. The principles of transfers and exchanges (Reardon and Firebaugh, 2002; James and Taeuber, 1985) state that a segregation index value should change if a person moves from one neighborhood to another or if two people exchange places, assuming they are not the same type of person. Issues relating to spatial scale have been an important part of the literature, with recent contributions by Wong (2004) focusing on the size of the neighborhoods and Krupka (2007) on the size of the region encompassing the neighborhoods. While this is certainly one dimension highlighting the richness of spatial data, there are other spatial interactions which influence the conceptual idea of segregation.

As was stated above, the input to most segregation indices requires having all people grouped by their classification and area. While in the case of residential segregation the areas are areal units (i.e. neighborhoods), the classic residential measures ignore this by treating the different areas as independent. Parallel to developments in exploratory spatial data analysis (ESDA), segregation researchers have recognized that the spatial configuration of the areas should be central to the empirical analysis. This reflects the recognition that traditional segregation measures are examples of "whole map" statistics in the sense that a permutation of the neighborhoods across the map would leave the aggregate segregation value unchanged.

Conceptually, the objective of spatial indices is to account for the intuitive observation that a poor neighborhood surrounded by poor neighborhoods is more segregated than one surrounded by rich neighborhoods. Figure 1 illustrates the concept by asking the question: is the region on the right more or less segregated than the one on the left? Aspatial measures treat both the same, spatial ones should identify the region on the right as more segregated.

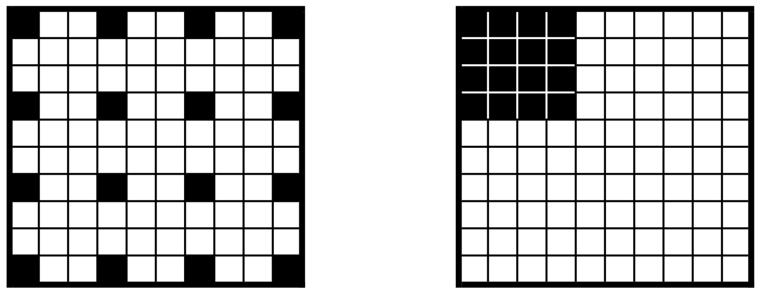


Figure 1. Illustration of Spatial Segregation Scenarios

Spatially explicit indices have been suggested throughout the social sciences for decades. However, the presence in the residential segregation literature was spotty until Massey and Denton (1988) took a broad view of the concept and brought the scattered ideas into a unified framework. In addition to identifying the existing segregation indices, they borrowed from other literatures to develop a few more.

The more recent literature on spatial measures of segregation has opened a new split in the discipline. To some extent is can be characterized as a person-centric versus neighborhood-centric view on the computation of segregation. Is it enough to just know the neighborhood in which a person lives, or is it necessary to know that person's exact location *within* the neighborhood? This difference is operationalized by treating the urban area as either a discrete surface made-up of known neighborhoods or a continuous surface that estimates the location of individuals. The indices organized by Massey and Denton (1988) generally treat space as discrete units and model personal interactions vis-a-vis their home neighborhood and that neighborhood's location relative to other neighborhoods within the urban area. Spatial relationships were defined in terms of adjacency and distance between neighborhood centroids. Recent spatial segregation indices use spatial interpolation techniques to interpolate polygon based data to any point within the urban area (for examples see Reardon and O'Sullivan, 2004; O'Sullivan and Wong, 2007; Lee et al., 2008). This offers the ability to break free of the spatial units provided by the data source, and allows one to estimate segregation measures for researcher-defined spatial scales. One significant hurdle these new spatial measures face is that data are rarely supplied for individual persons or households, and the estimation necessary to transform the areal data to points may introduce significant error into the final measured segregation.

1.2 Analyzing Segregation

In their 1955 paper Duncan and Duncan stated, "the problem of validating segregation indexes is viewed as one of some importance, not only in its own right, but also as an illustration of the difficulties in finding an adequate rationale for much sociological research using index numbers" (Duncan and Duncan, 1955, p. 210). While that paper dealt more with topics from the preceding section, this sentence gives the motivation for those debates. Specifically, are the inferences made using segregation indices appropriate, valid and relevant?

This section takes a high level view of the ways segregation indices are used in the empirical literature. To the largest extent possible, this section attempts to remain neutral on the index chosen, and focuses more on the interrelationships between index values. We start from a premiss that a single value from a single segregation index is of little interest. Value is derived from comparisons of many regions at a particular time point or a blend of the cross-sectional and temporal frames into a panel data configuration. Not only must the scale and temporal nature of the problem be chosen, but its treatment in a descriptive or confirmatory framework is based on the goals of a particular study.

1.2.1 Space or Space and Time

The two cases for residential segregation studies are the pure cross-sectional approach and the panel data perspective where variation in segregation indices is both cross-sectional (i.e., across urban areas) as well as temporal. In most such cases these panels are cross-sectionally dominated in that the number of cross-sectional observations (urban areas) often dwarfs the smaller number of temporal observations, typically census or decadal in nature. Since cross-sectional analysis is just a special case of panel analysis with a single time period, the more complicated case will be discussed here.

The literature on the spatio-temporal nature of segregation can be divided into two groups based on the spatial scale of the comparisons. The first looks at how urban areas change over time and the second looks at how neighborhoods change over time. The former is quite prevalent in the empirical literature due to the ease of acquiring the data at that scale. In contrast, the ease of acquiring data on individual neighborhoods at different points in time is not as straightforward.

Studies of urban area change from 1970 to 2000 have typically shown a pattern of declining racial segregation through the decades (a representative example is Timberlake and Iceland, 2007). Cutler et al. (1999) look back even further to map out black segregation from 1890 to 1990. These studies typically hold the definition of the urban area constant through time, but allow the subdivision of that area to change at each time point. This is a reasonable methodology considering that there is a single segregation value for each urban area for each year. Tracing a single neighborhood through time adds little value to these computations.

1.2.2 Descriptive or Confirmatory

Fischer et al. (2004) is an example of the main type of empirical residential segregation studies. Using the Theil Index, they compare five decades, five geographic scales, four race/ethnicity types, three measures of social class and four indicators of life cycle position. Comparisons of all these segregation values allow the authors to reach conclusions about the pattern of segregation in the United States. For example, they can distinguish between the decreasing segregation of blacks and the increasing segregation of foreign born residents through time, and identify the spatial scale at which the changes are detectable.

In contrast to the descriptive study mentioned above, confirmatory studies utilize the cross-sectional variation in global segregation patterns to consider the impact of covariates. The study by Pendall and Carruthers (2003) is representative of this branch of the literature where segregation indices for different cities are calculated and formal confirmatory models are estimated in an attempt to uncover the direction and magnitude of relationships between segregation and other aspects of urban communities.

1.2.3 Inference

The question of whether two segregation indices differ is an important one, yet to date most comparative studies focused on this question have adopted a largely descriptive approach. While there has been some recognition that residential segregation indices can be treated as random variables (Massey, 1978), in practice this has not been exploited to carry out formal inferential tests of comparative residential segregation patterns.[1] However, this subject has been broached in the occupational segregation literature where the observations, e.g. workers or firms, are clearly a sample from a larger population (see Boisso et al., 1994; Ransom, 2000). Where inference has been used is when residential segregation indices are related

[1]Ransom (2000) postulates that this is due to the computationally burdensome approaches which have been proposed previously in the literature. In response to the critics of an early attempt at assessing the distributional properties of the dissimilarity index, (Cortese et al., 1978, p. 591) state, "while this computation remains difficult for the moment, we strongly urge researchers in the area to be sensitive to the issues we raise and not attempt to dismiss our substantive points on an issue of practicality which can and will be resolved."

to covariates in regression models, but not in direct comparisons between the indices in a univariate context.

2 Simulation Procedure

This section details the simulation procedure to generate regions and areas according to the conditions necessary to test hypotheses on the relationship between urban structure and measurement of segregation. In the following discussion, the term "region" is analogous to "urban area" and "area" to "neighborhood." Studies using census data typically define an urban area as a Metropolitan Statistical Area (MSA) and a neighborhood as a Census Tract.

2.1 Segregation Indices

For this exercise, four segregation indices were tested: Dissimilarity Index (D), its spatial analogue D(adj), Neighborhood Sorting Index (NSI) and its spatial analogue Generalized Neighborhood Sorting Index (GNSI). These four indices provide a comparison of ways to measure income segregation. First the indices will be defined mathematically and then some of the intuition behind their selection will be provided. In all cases the indices range from 0 indicating no segregation to 1 indicating perfect segregation.

D (Duncan and Duncan, 1955) can be interpreted as the share of the poor population which would have to move to a new neighborhood to achieve an even distribution of population types (rich and poor) across all neighborhoods. Let $p_{i,g}$ denote the population (or count of households) from group g in tract i and $p_{i,\bar{g}} = p_i - p_{i,g}$ as the population in tract i that are not members of group g, where p_i is the total population of tract i and p_g the total population of group g in the region. The dissimilarity index for this group is given as:

$$D_g = \frac{1}{2} \sum_i \left| \frac{p_{i,g}}{p_g} - \frac{p_{i,\bar{g}}}{p_{\bar{g}}} \right| \tag{1}$$

D(adj) (Morrill, 1991) *adjusts* D to explicitly incorporate space into the index. It reduces D by an amount related to the exposure population $p_{i,g}$ has to population $p_{i,\bar{g}}$ in its neighboring areas, where the presence of any residents of population $p_{\bar{g}}$, no matter their ratio, is considered a contributor to greater integration and thus a reduction in segregation. D(adj) is defined as:

$$D_g(adj) = D_g - \frac{\sum_i \sum_j |w_{ij}(s_{i,g} - s_{j,g})|}{\sum_i \sum_j w_{ij}}, \tag{2}$$

where $s_{i,g}$ is the share of population g in area i (i.e. $s_{i,g} = p_{i,g}/p_g$) and $w_{i,j}$ is the value in the ith row and jth column of a spatial weights matrix where $w_{i,j} = 1$ if i and j are neighbors and 0 otherwise, and $w_{ii} = 0$.

NSI (Jargowsky, 1996) is driven by magnitudes of household income. It can be viewed as a ratio of the standard deviation of neighborhood mean household income to the standard deviation of individual household incomes. NSI is defined as:

$$NSI = \sqrt{\frac{\sum_{i=1}^{N} h_i(\bar{y}_i - Y)^2}{\sum_{k=1}^{H}(y_k - Y)^2}}, \tag{3}$$

where N is the number of neighborhoods in the region, H is the number of households in the region, h_i is the number of households in area i, y_k is the income of household k, $\bar{y}_i$ is the mean household income for area i and Y is the mean household income for the region. If individual neighborhoods have mean values ($\bar{y}_i$) far from the region mean, that implies a concentration of either high or low income households in those neighborhoods, i.e. segregation. This would raise the index value indicating higher segregation. Conversely, heterogeneous neighborhoods would have means near the region mean resulting in a low numerator, and hence low segregation.

GNSI (Jargowsky and Kim, 2005) *generalizes* NSI into a form which allows for the inclusion of information from the spatial neighbors of each area. To some extent, this is a smoothing technique which averages area i's average household income $\bar{y}_i$ with that of its neighbors to arrive at a new income value for i, namely $\bar{y}_i^*$. GNSI is given as:

$$GNSI = \sqrt{\frac{\sum_{i=1}^{N} h_i(\bar{y}_i^* - Y)^2}{\sum_{k=1}^{H}(y_k - Y)^2}}, \tag{4}$$

$$\text{where } \bar{y}_i^* = \frac{h_i\bar{y}_i + \sum_{j=1}^{N} w_{ij}h_j\bar{y}_j}{h_i + \sum_{j=1}^{N} w_{ij}h_j}.$$

The only difference between Equations 3 and 4 is the replacement of $\bar{y}_i$ with $\bar{y}_i^*$ which is the average of i with its neighbors.

As can be seen, the dissimilarity indices are structured entirely differently from the NSI indices. The primary difference between the pairs is the root data on which they operate. D is a measure of the binary "rich" and "poor," while NSI uses the continuous variable income as its base. NSI is targeted directly at income segregation, while D is a more flexible measure which could be equally adapted to racial binaries such as "white" and "non-white." D was chosen for this exercise because it is the most commonly encountered segregation index.

D and NSI are also ideal candidates because they have direct spatial alternatives in D(adj) and GNSI. Equations 1 and 3 do not use any spatial information and so the spatial patterning of the areas should have no bearing on measured segregation. In contrast, the other two indices explicitly include the spatial structure defined through the spatial weights matrix W. D(adj) is the most straightforward of a family of similar spatializations of D presented in Wong (1993). Wong (1993) extends Morrill (1991) by developing two variations on D(adj) which consider different ways of defining neighborhood relationships.

Another reason to use these measures is that they are both feasible to implement using publicly available census data. This is obvious in the cases of D and D(adj) which only use data at the area level. NSI and GNSI are designed for use with individual household data for the computation of the denominator. However, for empirical implementation, Jargowsky (1996) uses census data on the number of households within various income ranges to get at this denominator. For the simulations presented here, actual individual households are simulated.

2.2 Region Size and Shape

A number of characteristics are consistent across every simulation. Each region is a square lattice made up of N areas, implying the shape of the region is $\sqrt{N} \times \sqrt{N}$. Seven square regions are simulated, where the region size N is set to 25, 81, 169, 289, 441, 576 and 625. To provide some perspective, the Phoenix-Mesa MSA contains 661 Census Tracts. Each area within a region is assumed to have the same number of residents or households, set at 5,000. The empirical segregation literature is filled with cross-sectional studies comparing segregation between urban areas. This spatial dimension is studied to help identify any systematic variation in the indices when regions of different sizes are compared.

2.3 Sampling from Spatially Correlated Distribution

The simulation of spatially correlated variables falls within the general area of simulation of correlated variables. The first step is to construct the correlation structure, with the second being the generation of values which fit both this structure and a known distribution function. The following describes these two steps.

2.3.1 Correlation Structure

In order to model a spatial data generating process (DGP), a specification of the correlation structure is required. In this case three specifications are examined: the spatial autoregressive (SAR), spatial moving average (SMA) and conditional spatial autoregressive (CAR). Each of these models implies something different about the extent and structure of the spatial interconnectivity of the region.

The SAR model can be characterized as a measure of global autocorrelation for the region (Anselin, 2003). The estimation of any particular observation is dependent on its “neighbors’” values, which in turn are dependent on their neighbors’ values. This structure links the entire region together. The variance-covariance matrix of y for the SAR model is:

$$\text{VC}[y] = \sigma^2 \left[(I - \rho W)'(I - \rho W)\right]^{-1}, \tag{5}$$

where y is an $N \times 1$ vector of observations on a variable, σ^2 is the scalar variance, ρ is a measure of spatial autocorrelation ranging from -1 to 1, W is an $N \times N$ matrix of spatial weights describing the neighbor structure of the region and I is the $N \times N$ identity matrix.

The second model under consideration is the spatial moving average (SMA) model, the second most popular spatial model after the SAR model. In this model, the spatial structure is a function of the error term alone, and as such the spatial impacts dissipate quickly. Anselin (2003) labels this a local model of autocorrelation since the correlation structure dies out after the second order neighbors. The specification of the spatial structure is a follows:

$$\text{VC}[y] = \sigma^2[(I + \rho W)'(I + \rho W)] \tag{6}$$

The final model studied is the conditional autoregressive model (CAR). This model is distinct from the previous two in that it treats the neighboring values of a particular observation as exogenous. Therefore the explicit interlinking nature of the other models is gone. This specification could be considered the most “local” of those used in this exercise. The variance-covariance matrix is:

$$\text{VC}[y] = \sigma^2 \left[I - \rho W\right]^{-1} \tag{7}$$

In cross-sectional studies, one might be mixing regions which follow different spatial models. One might expect that a neighborhood is influenced by a higher order of contiguous neighborhoods in a denser city. In contrast, interactions may dissipate quickly in a sprawling western U.S. city where people move around by car and are able to easily leapfrog neighborhoods for shopping, services and employment.

As can be seen in the equations above, the term ρW is a key driver of the correlation structure, where the sign and magnitude of ρ dictate the strength and direction of the relationships. The simulations are generated with the following sequence of values for ρ: $-0.9, -0.8, \ldots, -0.1, 0, 0.1, \ldots, 0.8, 0.9$. Positive values of ρ imply spatial clustering of like values, either high or low, while a negative ρ indicates spatial dispersion where high and low values are interlaced. This spatial dimension is the key target of spatialized segregation indices. The idea is that a poor neighborhood surrounded by poor neighborhoods (i.e.

$\rho > 0$) should be considered more "segregated" than one surrounded by rich neighborhoods (i.e. $\rho < 0$). In the latter scenario, the poor neighborhood is less isolated from the rich population and therefore has greater opportunity for interaction. For all simulations, W is a row-standardized queen based contiguity weights matrix describing the neighbor structure of the lattice. See Anselin (2003) for a thorough treatment of these models and Rey and Dev (2006) for an application in the context of economic convergence.

2.3.2 Sampling Distribution

For the segregation indices chosen for this paper, it was necessary to simulate two values for each region: 1) the share of area population fitting a certain category, in this case "low income," and 2) the average income for an area. Both of these require a distribution with a minimum value of zero, and in the former case, one with a distribution capped at 1.0. This precludes the normal distribution which produces values over the interval $(-\infty, \infty)$ even though there is a well known algorithm for generating correlated values via the Cholesky Decomposition. The key challenge facing other candidate distributions is the ability to generate correlated (both negative and positive) random variables in a reliable fashion for large problems.[2] The uniform distribution is the only one known to the author which meets the minimum criteria necessary for this exercise.

The process, in pseudo-code, for simulating correlated values from a uniform distribution is as follows: [3]

1. Generate an $N \times N$ variance-covariance matrix $\sigma^2\Omega$ via one of the spatial models defined above
2. Transform $\sigma^2\Omega$ to a correlation matrix Σ
3. Transform each value in Σ via $\hat{\rho}_{ij}^{adj} = 2\sin((\pi/6)\hat{\rho}_{ij})$ to create Σ^{adj} [4]
4. Compute the Cholesky Decomposition C of Σ^{adj}
5. Generate Z, where $Z \sim N(0, 1)$ and is of length N
6. Transform Z into Z^{adj} via C so that $Z^{adj} \sim N(0, \Sigma^{adj})$
7. Transform Z^{adj} to uniform variates

[2]Lunn and Davies (1998) provides a method for generating Poisson random variables, but it does not apply to negative correlation and Dias et al. (2008) has a method for the flexible beta distribution, but the method is not reliable for all cases.

[3]See Schumann (2009) for further description of this method along with code for implementation in MatLab and R.

[4]$\hat{\rho}$ is used here simply to differentiate it from the autocorrelation coefficient ρ.

One consequence of the above procedure, specifically the conversion of Σ to Σ^{adj}, is that Σ^{adj} is not always positive definite. This is a necessary condition for the Cholesky Decomposition, and therefore the procedure cannot proceed without it. For the simulations performed for this exercise, this only occurs in the case of extreme negative autocorrelation for the SMA case, specifically for $\rho = -0.8$ and $\rho = -0.9$.

As a result of sampling from $y \sim U(0,1)$, the expected value of y is 0.5. For the Dissimilarity index (D) and D(adj), y is the percent of the neighborhood population designated "rich" and $1-y$ represents the "poor" population. While the expectation is a neighborhood containing equal shares of rich and poor residents, the uniform distribution generates a diverse set of neighborhoods with any ratio of rich to poor equally likely to be drawn. For the cases of the NSI and GNSI, y is the average household income of the tract measured in \$100,000's. The implication of $y \sim U(0,1)$ is that average incomes are confined to the interval [\$0, \$100,000] which is not ideal, but clearly reasonable for most neighborhoods. Again, the uniform distribution generates a wide range of neighborhoods with and overall expected value of \$50,000.

2.4 Sampling within Areas

The NSI and GNSI indices require a distribution of household income within the individual areas. Since this distribution does not necessitate a spatial correlation structure between the simulated values, more flexibility is available in the selection of a distribution. In this case the log-normal distribution is used. This distribution falls on the interval $[0, \infty)$ [5] and often matches real world income distributions due to its positive skew and long right side tail (Cowell, 1995).[6] Given a known mean (E[y]) and variance (Var[y]), it is possible to build a specific log-normal distribution for each simulated area according to the following rule (Greene, 2003):

$$\begin{aligned} \mu &= \ln(\mathrm{E}[y])^2 - \frac{1}{2}\ln\left((\mathrm{E}[y])^2 + \mathrm{Var}[y]\right) \\ \sigma^2 &= \ln\left(1 + \frac{\mathrm{Var}[y]}{(\mathrm{E}[y])^2}\right), \end{aligned} \tag{8}$$

where $y \sim LN(\mu, \sigma^2)$.

The implementation of the within area sampling is a four step process. The first step is the generation of a spatially correlated simulation of average household income for each area,

[5] It is acknowledged that real world household incomes can go negative, but for this simulation exercise it was deemed reasonable to assume $y \geq 0, \forall y$.

[6] Cowell (1995) finds that incomes typically follow the log-normal or Pareto distributions, with the Pareto occurring more at the higher income levels. For this exercise, the log-normal is applied to all areas.

as described in the previous section. Each area income, E[y], is then plugged into Equation 8 along with Var[y] = 1 to define its households' reference distribution. Next, 5,000 household incomes are drawn from the distribution for each area. Finally, in order to ensure the final segregation index is balanced, the original average income for each area is replaced by the average of its 5,000 simulated household incomes. In total, these four steps generate an entire region of household incomes while maintaining the spatial correlation structure at the area (i.e. neighborhood) level.

2.5 Scaling

The motivation for these simulations is to test the idea that the administrative areas (e.g. Census Tracts) may not follow the actual neighborhood areas within the region. Krupka (2007) provides evidence that this is the case by testing data for US regions at different US Census data scales (Block Groups, Census Tracts and counties). He hypothesizes that larger urban areas have larger neighborhoods and thus one complete neighborhood can fit into an administrative boundary such as a Census Tract, but that neighborhoods are smaller in smaller urban areas and thus there is mixture at the administrative boundary level.

A related issue in segregation measurement is downscaling. If a region continues to be subdivided into smaller and smaller areas, in the limit, each area contains only one household. Since that household can only be rich or poor, the index must have a value of unity, or perfect segregation. Therefore, upscaling is expected to reduce segregation and downscaling to increase it.

Simulations incorporating a conversion of scale are run only on the 24×24 region consisting of 576 areas. To test the *upscaling* hypothesis, a simulated region is built using 576 areas and one of the spatial correlation structures described above. This run is intended to model the actual spatial patterning of neighborhoods. These areas are then scaled up by factors of 4, 9 and 16 producing regions of 144 (12×12), 64 (8×8) and 36 (6×6). The upscaling is performed by averaging the values of the original areas into the larger areas. These new areas are intended to mimic the actual administrative boundaries which might group mismatched neighborhoods together. It should be noted that all other simulations run for this exercise could be viewed has having a scale factor of unity.

2.6 Iterations

Combining all the pieces together results in 570 simulated regions. There are seven region sizes, three spatial correlation models and 19 values for ρ (i.e. strength and direction of correlation) resulting in 399 regions where scale was held constant. The scale case results in an additional 171 regions for the three degrees of scale intensity studied.

For each region, 500 simulations are run and a value for each of the four segregation indices is computed. The mean and standard deviation of the 500 values is then computed with the results presented in the following section.

3 Simulation Results

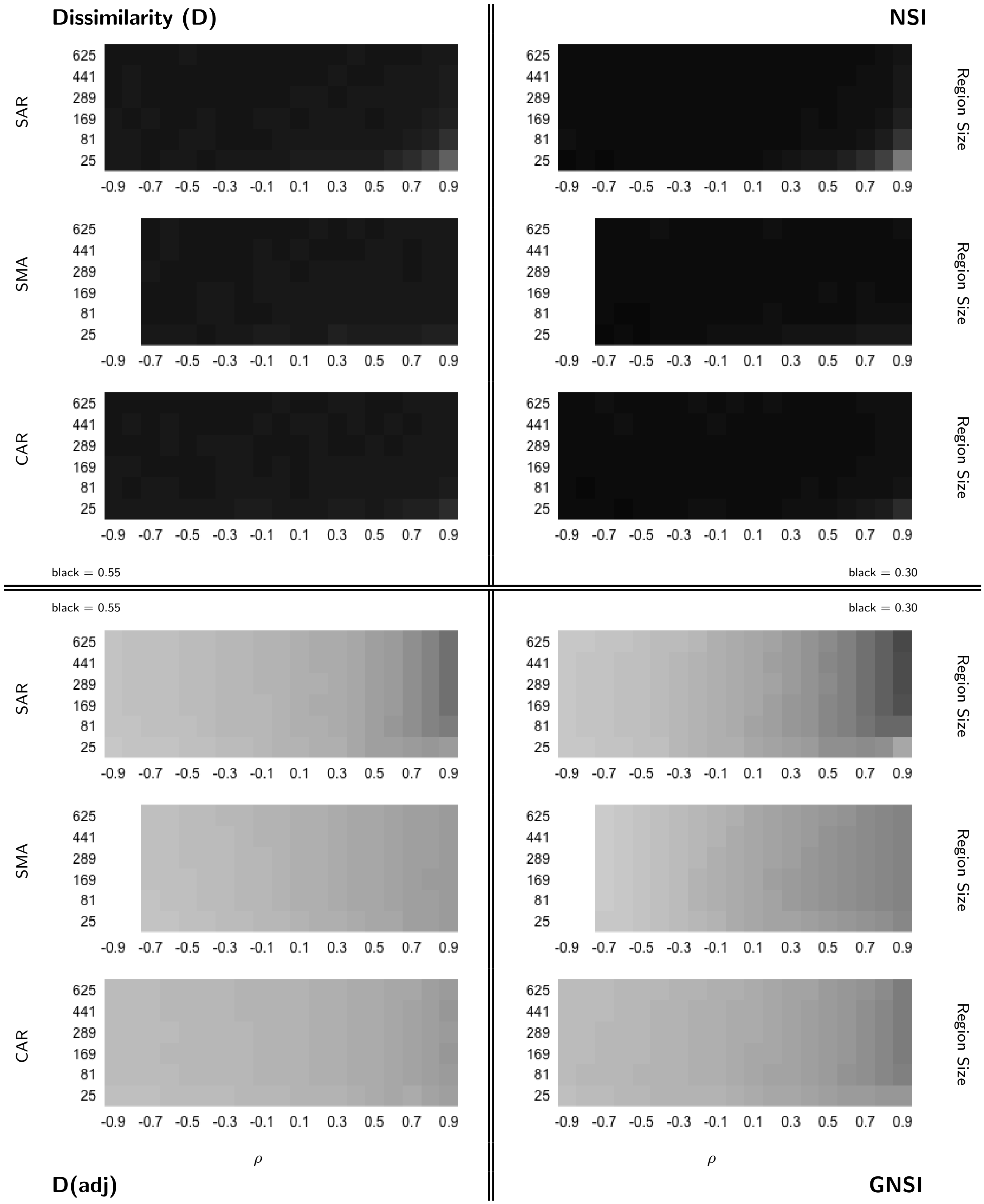
This section will begin with an introduction to the figures used to present the simulation results, followed by a summary of key findings from this exercise. The results are presented in four parts based on the four spatial dimensions investigated, with an attempt to hold the other four constant in the discussion.

3.1 Presentation of Results

Figure 2 presents the mean value for each of the four segregation indices for 342 regions. This represents 1,368 "pixels" in the figure, where each pixel is the mean value of a segregation index computed for each of 500 simulations of the region. The color intensity of each pixel is the magnitude of that mean value. It should be noted that the variation discussed in this section is visible on the screen, but the printed resolution may not represent the shades of gray appropriately.

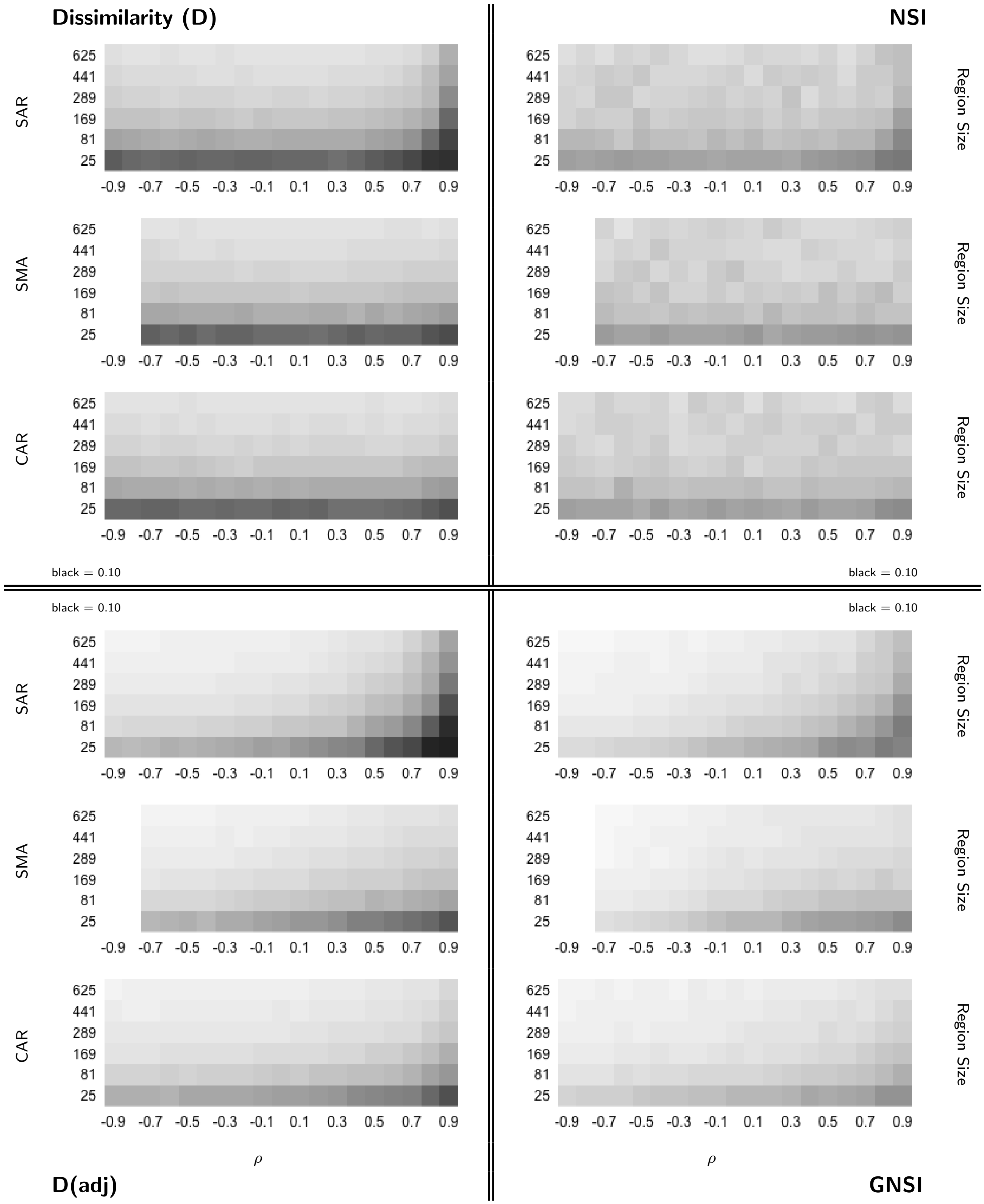
The figure is laid-out in four quadrants with each quadrant representing a particular segregation index. The aspatial indices are in the upper quadrants, with the spatial measures below. The two forms of the dissimilarity index are on the left with the NSI and GNSI on the right. Within each quadrant are three graphs representing the spatial data generating process, SAR, SMA and CAR. The graphs themselves are laid out with rows presenting the various region sizes, 25, 81, 169, 289, 441 and 625, and the columns presenting the value of ρ which takes 19 values between -0.9 and 0.9. For the two dissimilarity indices, a pure black pixel would represent a mean value of 0.55, while pure black for NSI and GNSI would represent 0.30. In all cases pure white would represent an index value of zero.

Figure 3 is identical in structure to Figure 2 except that standard deviations replace the means, and the meaning of pure black now indicates a value of 0.05 or 0.10 depending on the index. Figures 4 and 5 have a similar layout to the previous two, except that rows now represent scaling factors of 1:1, 4:1, 9:1 and 16:1, which are labeled 1, 4, 9 and 16 respectively.



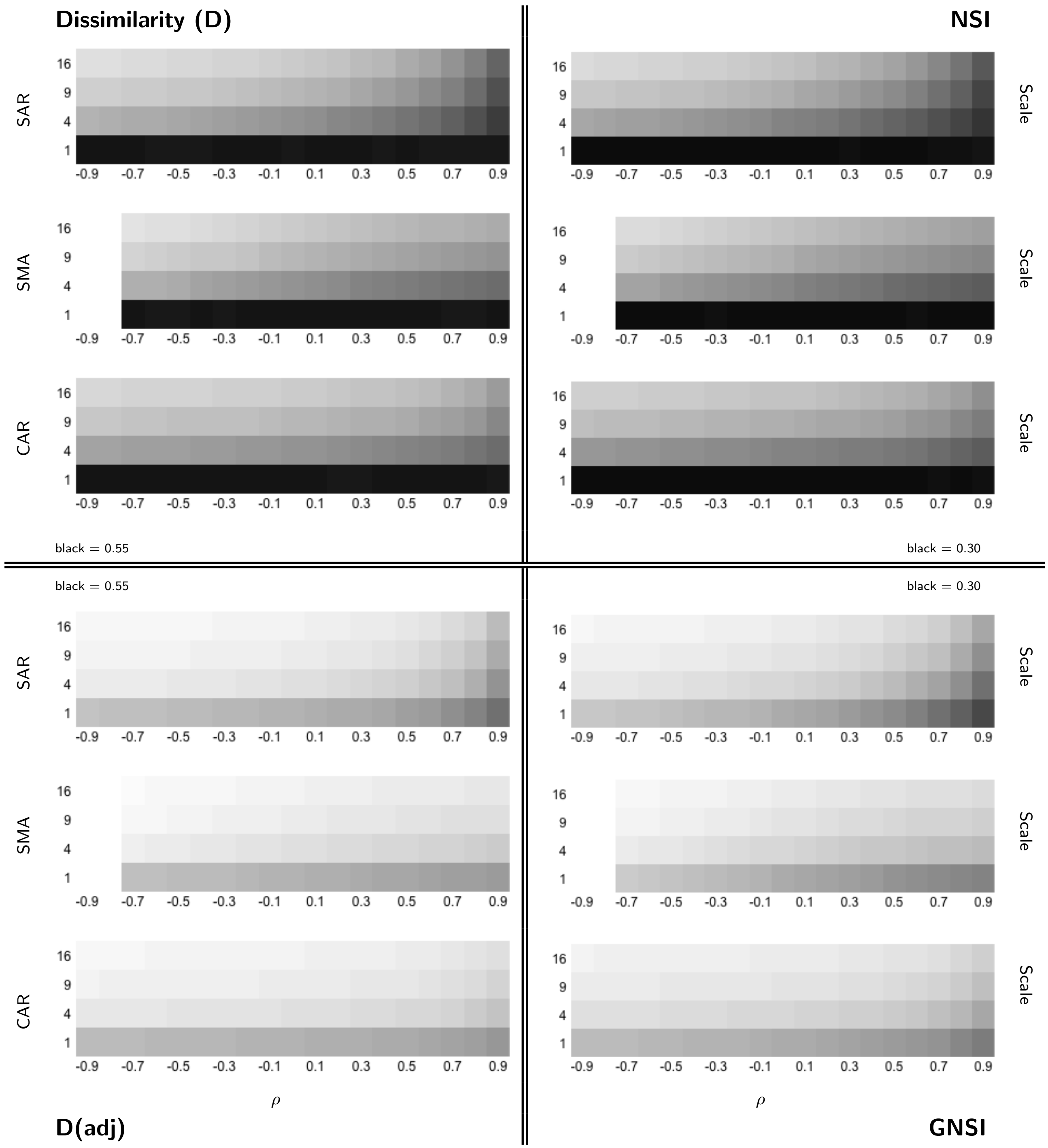
Notes: Each quadrant contains the results for one index. The three graphs within a quadrant present the results for the SAR, SMA and CAR cases. For each graph, rows represent the region size and columns the value of ρ, with color intensity being the magnitude of the mean value from 500 simulations.

Figure 2. Simulation Results: Mean Values for Region Size Case



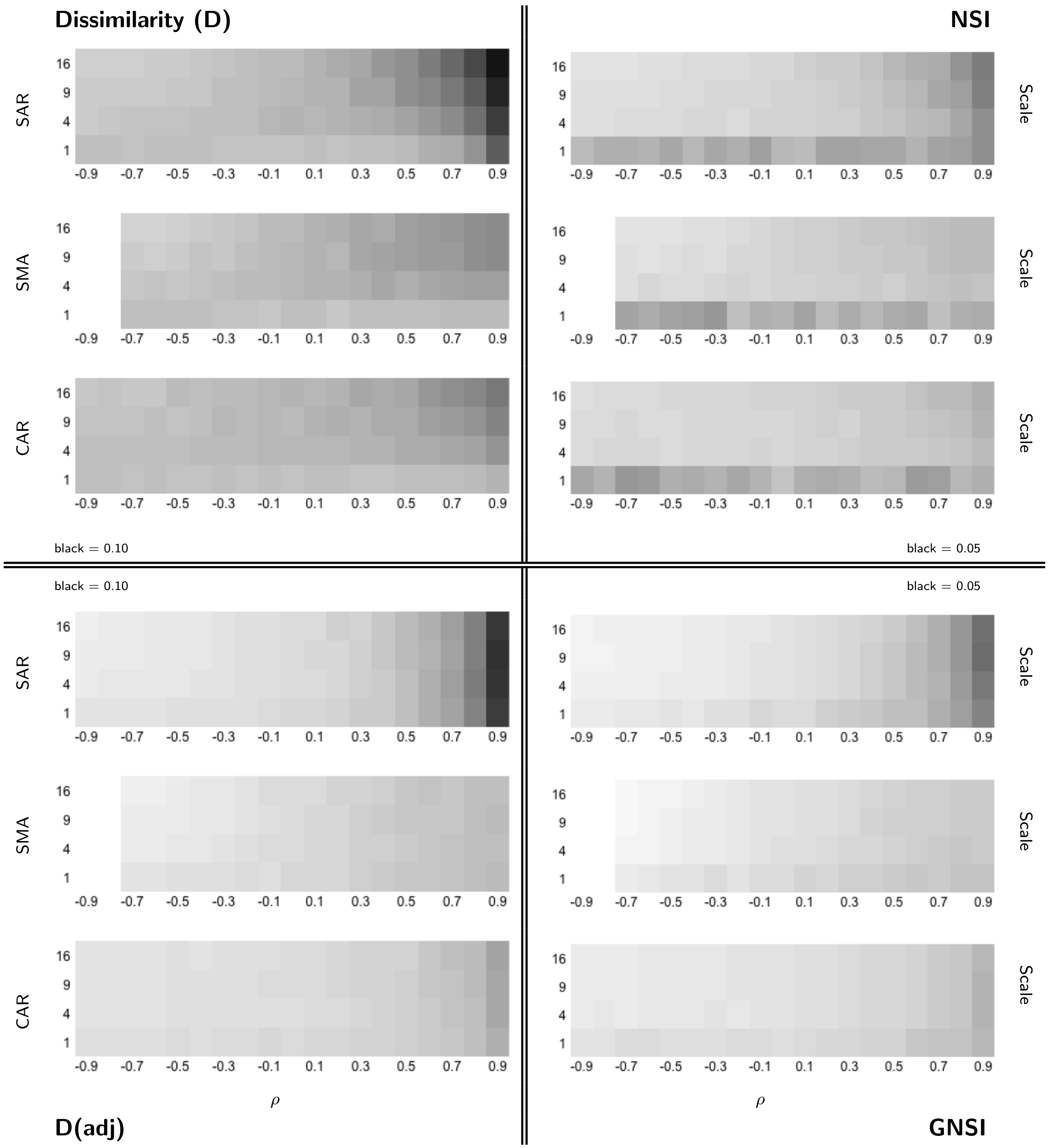
Notes: Each quadrant contains the results for one index. The three graphs within a quadrant present the results for the SAR, SMA and CAR cases. For each graph, rows represent the region size and columns the value of ρ, with color intensity being the magnitude of the standard deviation value from 500 simulations.

Figure 3. Simulation Results: Standard Deviation Values for Region Size Case



Notes: Each quadrant contains the results for one index. The three graphs within a quadrant present the results for the SAR, SMA and CAR cases. For each graph, rows represent the scaling factor and columns the value of ρ, with color intensity being the magnitude of the mean value from 500 simulations. The orginal region size is 578 areas.

Figure 4. Simulation Results: Mean Values for Scaled Case



Notes: Each quadrant contains the results for one index. The three graphs within a quadrant present the results for the SAR, SMA and CAR cases. For each graph, rows represent the scaling factor and columns the value of ρ, with color intensity being the magnitude of the standard deviation value from 500 simulations. The orginal region size is 578 areas.

Figure 5. Simulation Results: Standard Deviation Values for Scaled Case

3.2 Discussion of Results

Variation in Region Size. Varying the region size appears to have little affect on the mean values of the indices, except in the extreme case of the 5×5 region with 25 areas, and for some instances of high values of ρ (see Figure 2). This is an ideal result indicating that the measures chosen are generally unaffected by region size, which facilities comparisons across regions. That being said, the standard deviation results in Figure 3 do show clear horizontal banding, except in the case of NSI. As region size decreases, standard deviation increases. This is an expected result considering that smaller regions are made-up of less areas allowing for greater impact of outliers in the computations. Combined, this indicates the mean values are largely unaffected by region size, but their reliability decreases with region size.

Variation in Spatial Scale. The impacts of varying spatial scale are presented in Figures 4 and 5 and they appear to be strong as horizontal banding is visible in all 12 graphs in Figure 4. Each row of a particular graph presents the same baseline raw data (i.e. before applying the segregation index) which has been aggregated from a baseline of region size 576 (i.e. the 1:1 case). Recall that this is an investigation of the mismatch between the actual neighborhood structure, and the one revealed to the researcher via administrative data, where the implication is that actual neighborhoods are smaller than administrative areas.

The general trend of decreasing segregation as aggregation level increases is the expected result, and this is found. The idea is that as more neighborhoods are mixed together, ramping up from 4 to 9 to 16 neighborhoods blended into one administrative area, the level of "measured" segregation should decrease.

For the aspatial measures (upper half of Figure 5), there is a slight trend toward increased standard deviation as a upscaling increases. Again, this is an expected result as the number of areas used to compute the segregation indices decreases as scale increases. For the spatial measures of segregation (lower half of Figure 5) there is little impact from the change of scale above the 1:1 case. This is a welcome result, implying that while segregation declines with scale, its measurement becomings more precise. This is likely driven by the double-smoothing happening in this case: one type of smoothing from the scale change and a second from the inclusion of neighbor values in the computation of segregation.

A final note, while the "expected" results are found, this does not imply that the concept of segregation has been measured correctly for the region. The large decline from the 1:1 to 4:1 case shows that segregation could be grossly under-measured even in the smallest aggregation scenarios. This issue is compounded in cross-region comparisons if one region is measured at 1:1 while another is at 4:1.

Variation in the Spatial DGP. The impact of the spatial DGP, i.e. SAR, SMA and CAR, is shown in all four results figures. A broad statement covering this type of variation is that SMA and CAR, the two local autocorrelation models, impact the indices of segregation similarly, and are distinct from the impact of the global measure modeled via SAR. Another broad commonality, is that the local and global models are difficult to differentiate below $\rho = 0.7$. The deviations above $\rho = 0.7$ in the D and NSI cases for SAR in smaller regions are a bit troubling since the expectation is a constant value for any DGP or ρ. That being said, the problem appears to go away as region size increases.

The previous withstanding, the minimal lack of impact from the varying spatial DGP models could be explained by the simulation structure and the segregation indices themselves. In all cases calling for a spatial weights matrix, the first order queen contiguity was used. This implies that the segregation indices do not extend beyond the first order neighbors where more dramatic impacts of the DGPs are expected to be found.

Variation in the Magnitude and Direction of Spatial Autocorrelation. Variation in the magnitude and direction of spatial autocorrelation should have no impact on aspatial measures of segregation since the neighborhood structure is not involved in the computation. This result is largely borne out as can be seen in the upper halves of Figures 2 and 4. Figure 2 shows a nearly homogeneous result for any value of ρ, except on the high end where index values decline in the SAR case.

Similarly, the 1:1 rows in the upper half of Figure 4 are solid all the way across. In contrast, the increasing index values in Figure 4 as ρ increases from -0.9 to 0.9 are functions of the increase in clustering of like areas together at the smaller scale as discussed in the previous section. When the population values (in the case of D) and income values (in the case of NSI) are averaged to create the upscaled regions, low values of ρ imply that dissimilar neighborhoods are combined giving a result of greater integration. The opposite is the case for high values of ρ where similar areas are grouped together implying that the upscaled regions are nearly the same as the parts from which they were made.

The standard deviation in Figures 3 and 5 shows a seemingly random pattern in the horizontal direction for NSI for the SMA and CAR cases which implies no effect from changes in ρ. The SAR case shows a slight upward trend in Figure 3. The standard deviation of D in Figure 3 appears to not be affected by ρ except in the SAR case where it increases for high values of ρ. This is not the expected result which is that mean and standard deviation should be independent of ρ in the aspatial cases. This could imply a slight bias in the DGP for the SAR case.

Looking at the six graphs in the lower half of Figure 2, it can be seen that varying the value of ρ has almost no impact on the measure of segregation. There is a slight increase for most graphs, but it is only pronounced in the SAR case. The intention of these measures is to capture this effect, but they seem to have a difficult time differentiating the cases, except in the case of global autocorrelation. A similar pattern is noticeable in Figure 4 with slight increases in the measure. The standard deviation results show a slightly more pronounced banding effect, implying that uncertainty is increasing as ρ increases. In a few cases the standard deviations for the spatial indices increase greatly for high values of ρ

4 Discussion

As the previous section implied, the results from the simulation exercise were less than convincing about the chosen segregation indices' ability to appropriately measure the impacts of spatial variation for which they were designed. While the segregation indices performed as designed, one might argue that they miss the mark in terms of the intention of spatialzing a segregation index. The two key challenges they face are 1) the limited impact of ρ on the index values and 2) that segregation is always reduced when going from an aspatial index to its spatial counterpart. While variation in ρ is manifest in the spatial indices, it is always less than the aspatial value. An ideal spatial index would return a value lower than its aspatial counterpart for $\rho < 0$, higher for $\rho > 0$ and the same for $\rho = 0$. This section highlights issues related to the spatialized indices which prevent or hinder them from arriving at more intuitive results.

Revisiting D(adj) from Equation 2 reveals that spatial segregation must always be lower than D except in the case of D=0, i.e. $s_{i,g} = x, \forall i$. The primary driver of this is the absolute value in the second term on the right hand side of Equation 2 combined with the summation over all pairs of i and j where $w_{ij} = 1$. Dropping the absolute value alone would result in the i and j values canceling out. As a result, any difference in the shares of two areas causes measured segregation to decrease. A more intuitive measure would allow segregation to either decrease or increase depending on the relative intensity of its neighbors' population shares.

GNSI falls into a similar trap, although its impact is a little more subtle. Looking back at Equations 3 and 4 it is easily seen that the only difference is a change in the term $\bar{y}_i$ to $\bar{y}_i^*$ where $\bar{y}_i$ is the average income for region i and $\bar{y}_i^*$ is the weighted average of neighboring regions' average income with itself. The indices subtract this local mean value from the regional mean and square the result. However, a subtle issue arises in that:

$$\mathrm{E}[\bar{y}_i^*] < \bar{y}_i \text{ for } \bar{y}_i > Y$$
$$\mathrm{E}[\bar{y}_i^*] > \bar{y}_i \text{ for } \bar{y}_i < Y$$

due to the smoothing of neighboring values. Since $\bar{y}_i^*$ values tend to be closer to the regional mean than $\bar{y}_i$ values, GNSI will also tend to be lower than NSI. Unlike D(adj), it is possible for GNSI to be greater than its aspatial counterpart, however it is not the expected result over many trials as shown in the simulation results.

Reardon and O'Sullivan (2004) present a list of conditions which an ideal spatial measure should meet. While most of their list is reasonable[7], some of the issues identified here indicate that another condition (or set of conditions) should be added. The following list of conditions is proposed, where ASI is defined as any aspatial measure of segregation and SSI is its spatial counterpart:

- $\mathrm{E}[SSI] = \mathrm{E}[ASI]$ for $\rho = 0$
- $\mathrm{E}[SSI] > \mathrm{E}[ASI]$ for $\rho > 0$
- $\mathrm{E}[SSI] < \mathrm{E}[ASI]$ for $\rho < 0$
- $ASI = SSI = 0$ for the no segregation case (when all areas are the same)
- $SSI = 1$ for the perfect segregation and perfect clustering case (when every area is homogeneous and like areas form a single contiguous subregion)

The last criteria above conflicts with the standard segregation criteria which state that $ASI = 1$ for perfect segregation. The issue here is that perfect segregation can happen for any value of ρ as long as each area is homogeneous. Of course, this last criteria simply adds in the convenience of knowing the index is bounded on both ends, but it is not related to the intuition of segregation like the previous four.

5 Conclusion

In this paper, the performance of four segregation indices, Dissimilarity Index (D), Adjusted Dissimilarity Index (D(adj)), Neighborhood Sorting Index (NSI) and Generalized Neighborhood Sorting Index (GNSI), are tested against varying degrees of spatial effects. The indices

[7] Their argument for decomposability of an index into subregions appears to only be on the list to give a boost to entropy-like measures based on the Theil Index.

generally held up well to variation in region size indicating their comparability across regions of different sizes. They failed to accommodate for the situation where "real world" neighborhoods are arbitrary grouped into larger administrative regions. In fact, the biggest decline in measured segregation happened for the jump from a 1:1 scaling factor to 4:1 indicating that this could be a significant problem in empirical studies. The level of integration of the region, as measured by the spatial data generating process, showed some impact on the index values, but not to large degree. This could be influenced by the local nature of the measures of segregation which do not include higher order neighbors in the segregation measurement. Finally, the magnitude and direction of spatial autocorrelation (ρ) was picked up by the spatialized segregation indices, but not in an intuitive manner.

One disappointment in the design of the two spatial indices is that the explicit addition of spatial effects causes measured segregation to be reduced. Intuitively, one would expect that bringing space into the formulation should allow segregation to either rise or decline, but this is not the case. To address this, criteria are proposed which should be considered for any measure of spatial segregation.

The current work can be expanded in a number of directions. First, additional measures of income segregation can be incorporated into the existing framework. Entropy measures (i.e. those based on the Theil Index), Gini index, coefficient of variation and many others are available. Second, multi-group measures of segregation are a relatively recent development in the literature, and these would lend themselves well to the study of income segregation as they allow for multiple income ranges. Third, variation of the *overall* population is not addressed in this work. Specifically, all regions have the same expected value for average income or share of poor residents (even though the areas within could vary greatly). Real world regions vary along this dimension and it would be interesting to explore how these indices preform at different locations along the income or poverty distribution. Finally, the continuous versus discrete segmentation of space is an important and emerging direction in the literature. It is not at all clear if new continuous space methods are an improvement over the discrete space alternatives when one considers that publicly available data is only available in the discrete space form (e.g Census Tracts). Does the application of an arguably better statistic applied to estimated data perform better than an arguably worse statistic applied to true data?

References

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26(2):153.

Boisso, D., Hayes, K., Hirschberg, J., and Silber, J. (1994). Occupational segregation in the multidimensional case: Decomposition and tests of significance. *Journal of Econometrics*, 61(1):161–171.

Cortese, C., Falk, R., and Cohen, J. (1978). Understanding the standardized index of dissimilarity: Reply to massey. *American Sociological Review*, 43(4):590–592.

Cowell, F. A. (1995). *Measuring inequality*. Prentice Hall/Harvester Wheatsheaf, London, 2nd ed edition.

Cutler, D., Glaeser, E., and Vigdor, J. (1999). The rise and decline of the american ghetto. *Journal of Political Economy*, 107(3):455–506.

Dias, C., Samaranayaka, A., and Manly, B. (2008). On the use of correlated beta random variables with animal population modelling. *Ecological Modelling*, 215(4):293–300.

Duncan, O. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Fischer, C., Stockmayer, G., Stiles, J., and Hout, M. (2004). Distinguishing the geographic levels and social dimensions of us metropolitan segregation, 1960-2000. *Demography*, 41(1):37–59.

Greene, W. (2003). *Econometric analysis*. Prentice Hall Upper Saddle River, NJ.

James, D. and Taeuber, K. (1985). Measures of segregation. *Sociological Methodology*, 15(1):1–32.

Jargowsky, P. and Kim, J. (2005). A measure of spatial segregation: The generalized neighborhood sorting index. *Ann Arbor, MI: University of Michigan, National Poverty Center, Working Paper Series*.

Jargowsky, P. A. (1996). Take the money and run: Economic segregation in u.s. metropolitan areas. *American Sociological Review*, 61(6):984–998.

Krupka, D. (2007). Are big cities more segregated? Neighbourhood scale and the measurement of segregation. *Urban Studies*, 44(1):187.

Lee, B., Reardon, S., Firebaugh, G., Farrell, C., Matthews, S., and O'Sullivan, D. (2008). Beyond the census tract: Patterns and determinants of racial segregation at multiple geographic scales. *American Sociological Review*, 73(5):766–791.

Lunn, A. and Davies, S. (1998). A note on generating correlated binary variables. *Biometrika*, pages 487–490.

Massey, D. (1978). On the measurement of segregation as a random variable. *American Sociological Review*, 43(4):587–590.

Massey, D. and Denton, N. (1988). Dimensions of residential segregation, the. *Social Forces*, 67:281.

Morrill, R. (1991). On the measure of spatial segregation. *Geography Research Forum*, 11:25–36.

O'Sullivan, D. and Wong, D. W. S. (2007). A surface-based approach to measuring spatial segregation. *Geographical Analysis*, 39(2):147–168.

Pendall, R. and Carruthers, J. (2003). Does density exacerbate income segregation? Evidence from US metropolitan areas, 1980 to 2000. *Housing Policy Debate*, 14(4):541–590.

Ransom, M. (2000). Sampling distributions of segregation indexes. *Sociological Methods & Research*, 28(4):454.

Reardon, S. F. (2006). A conceptual framework for measuring segregation and its association with population outcomes. In Oakes, J. M. and Kaufman, J. S., editors, *Methods in social epidemiology*. Jossey-Bass, San Francisco, CA, 1st ed edition.

Reardon, S. F. and Firebaugh, G. (2002). Measures of multigroup segregation. In Stolzenberg, R. M., editor, *Sociological Methodology*, pages 33–67. Blackwell Publishing, Boston.

Reardon, S. F. and O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological Methodology*, 34:121–162.

Rey, S. and Dev, B. (2006). σ-convergence in the presence of spatial effects. *Papers in Regional Science*, 85(2):217–234.

Schumann, E. (2009). *Generating Correlated Uniform Variates*. Last checked: March 12, 2009. http://comisef.wikidot.com/tutorial:correlateduniformvariates.

Timberlake, J. and Iceland, J. (2007). Change in racial and ethnic residential inequality in american cities, 1970–2000. *City & Community*, 6(4):335–365.

Wong, D. (2004). Comparing traditional and spatial segregation measures: A spatial scale perspective. *Urban Geography*, 25(1):66–82.

Wong, D. S. (1993). Spatial indices of segregation. *Urban Studies*, 30:55–572.

Wong, D. S. (2003). Implementing spatial segregation measures in GIS. *Computers, Environment and Urban Systems*, 27:53–70.