National Poverty Center Working Paper Series

#05-3

March 2005

A Measure of Spatial Segregation: The Generalized Neighborhood Sorting Index

Paul A. Jargowsky, University of Texas at Dallas

Jeongdai Kim, University of Texas at Dallas

A Measure of Spatial Segregation: The Generalized Neighborhood Sorting Index*

Paul A. Jargowsky

Associate Professor of Political Economy

jargo@utdallas.edu

Jeongdai Kim

Research Associate

jdkim@utdallas.edu

University of Texas at Dallas
School of Social Sciences GR31
2601 N. Floyd Road
Richardson, TX 75080

March 21, 2005

* The authors would like to acknowledge financial support from the Century Foundation. We also with to thank Nathan Berg, Kurt Beron, Brian Berry, Ron Briggs, Marie Chevrier, and James Murdoch for helpful comments and suggestions. Comments welcome to the email addresses listed above.

A Measure of Spatial Segregation: The Generalized Neighborhood Sorting Index

ABSTRACT

The measurement of spatial segregation has often used measures such as the Neighborhood Sorting Index (NSI) or the Index of Dissimilarity (D) that do not take full account of space. We propose a Generalized Neighborhood Sorting Index (GNSI), an extension of NSI, in which each neighborhood is part of a broader community, the extent of which can be easily changed by increasing the order of contiguity or the allowable distance between neighborhood centroids. The GNSI has desirable characteristics as a valid spatial measure of segregation and significantly alleviates the checkerboard problem while NSI and similar measures are insensitive to the spatial arrangement of the geographic units in the analysis, such as census tracts. An application of GNSI to the ten largest metropolitan areas reveals that the GNSI gives a more complete account of the extent of economic segregation and the changes over time, and ranks the metropolitan areas differently in term of the level and changes in economic segregation.

A Measure of Spatial Segregation:
The Generalized Neighborhood Sorting Index

INTRODUCTION

For any given individual, success or failure in life is determined by a complex array of factors, chance not least among them. Economists tend to focus on human capital, the skills and aptitudes that a person brings to the labor market, and take the individual's tastes and preferences as a given. Sociologists wonder about how the distribution of human capital is shaped by social processes and institutions, and they also worry more than economists about how an individual's tastes and preferences are formed. Both groups, along with political scientists, have recently focused more attention on the neighborhood as a factor. Neighborhoods are the setting of peer interactions which can influence a child's attitudes and aspirations. Neighborhoods are also the setting for schools, which clearly play a role in the accumulation of skills. The increased attention to neighborhoods follows in part from the finding that neighborhoods are becoming increasingly differentiated along economic lines (Jargowsky 1996, 1997; Wilson 1987).

There is a vast and growing literature on whether neighborhood conditions matter for specific social and economic outcomes, after controlling for individual and family characteristics (Jencks and Mayer 1990; Brooks-Gunn et al. 1997a, 1997b). Less attention has been focused on how economic segregation ought to be measured. Yet if poorer neighborhoods do exert an independent effect on individuals, then greater variation in neighborhood conditions will produce more unequal outcomes than would otherwise be expected. Moreover, the children of the poor are disproportionately exposed to these conditions, setting up a potentially inequality-enhancing feedback mechanism. It is, therefore, important to study the levels and trends in economic segregation. This paper addresses a major deficiency in common measures of economic segregation. Specifically, we argue that current techniques for measuring economic segregation have a blind spot when it comes to spatial data.

Studies of racial and economic segregation have historically used measures such as the index of dissimilarity (e.g. Duncan and Duncan 1955; Lichter 1955; Massey and Denton 1987; Morrill 1995), the exposure index (Erbe 1975), the entropy index (Miller and Quigley 1990), and the correlation ratio (Farley 1977). All of these measures are based on subdividing of the unit of analysis – typically cities or metropolitan areas – into smaller parcels representing neighborhoods. Statistics are computed for the parcels, and then aggregated to compute the global segregation measure for the larger unit (James and Taueber 1985). In the typical application, none of these measures of segregation – ostensibly a spatial concept – takes into account the topological relationship of the parcels to one another. In this paper, we propose a Generalized Neighborhood Sorting Index (GNSI), a modification of an existing measure of economic segregation, the Neighborhood Sorting Index (NSI). The NSI is a unitless measure that varies from 0 to 1, with 1 indicating perfect economic segregation and 0 indicating perfect economic integration (Jargowsky 1996). GNSI extends NSI by incorporating the geographic relationships among the parcels. We provide a methodological analysis of our measure and empirical results that demonstrate the importance of more fully incorporating the spatial arrangement of the parcels into segregation measures.

NSI is based on the observation that the distribution of mean neighborhood incomes and the distribution of individual household incomes necessarily have the same mean, but differ in their degree dispersion. Because people of different incomes are inevitably combined within one neighborhood, the standard deviation of the neighborhood-level distribution is less than that of the household-level distribution. The ratio of these standard deviations represents the degree of economic segregation. The NSI is based on deviations from the mean income, and the denominator controls for the total amount of income inequality, so it does not confound changes in the income distribution with economic segregation *per se*. The measure is also appropriate for a continuous variable such as income, and does not require the division of income into arbitrary income classes.

Despite these advantages, NSI suffers from a number of disadvantages that are common to most segregation measures. First, NSI is sensitive to the population size of the parcels, as are all segregation measures. This is known as the Modifiable Areal Unit Problem (MAUP), which has been discussed extensively in the literature (King, 1997:249-255; Openshaw 1984a, 1984b; Wong 1997; Yule 1950). More importantly, the calculated value of NSI is insensitive to the physical location of census tracts vis-à-vis one another.

White (1983: 1010) characterizes this deficiency as the “checkerboard problem.” A checkerboard with all black squares on one half and all white on the other should show a greater level of segregation than a normal checkerboard, in which the colors alternate. However, NSI would show both patterns as completely segregated, since no individual square on the board contains a mixture of colors. The same is true for the Index of Dissimilarity (D). A similar problem arises when the Chi Square statistic is applied to spatial data (Rogerson 1999). To overcome the aspatial characteristics of D index, “distance-based” (Jakubs 1981, Morgan 1983) and “boundary-modified” versions of D (Morrill 1991; Wong 1993, 1998, 2002) have been introduced. However, these measures are based on dichotomous groups and are not well suited for studying income segregation.

The GNSI is a measure of economic segregation that addresses both deficiencies, based on the concept of a moving window centered on each parcel of the metropolitan space. Because the windows overlap, the spatial relationship of the parcels enters explicitly into the calculation. Therefore, rearranging the parcels could dramatically alter the calculated value of segregation even if the contents of the individual units were unchanged. The GNSI is a flexible measure, allowing empirical researchers to adjust the size of the moving window and, consequently, the degree of spatial dependency reflected in the measure. To illustrate the importance of incorporating the topology of parcels into segregation measures, we calculate the NSI and GNSI for the 10 largest metropolitan areas.

The remainder of the paper is organized as follows. First, we develop the GNSI. Second, we explore the characteristics and spatial interpretations of GNSI with reference to proposed requirements for an appropriate spatial segregation measure. Third, we provide a few simple illustrations. Finally, we provide empirical results using 1990 and 2000 Census data for the ten largest metropolitan areas, showing that taking space into account does make a difference in our understanding of income segregation.

THE GENERALIZED NEIGHBORHOOD SORTING INDEX (GNSI)

Jargowsky (1996) argued that segregation measures that had been developed for discrete groups were problematic when applied to segregation along a continuous dimension, such as income. A common approach to measuring segregation by income is to divide the income distribution into two or more discrete categories and then to compute the Index of Dissimilarity between pair-wise combinations of the categories (Massey and Eggers 1991). However, a shift in the parameters of the income distribution results in the reclassification of many households from one arbitrarily-defined income category to another, and results in a changed measure of economic segregation even if no one has moved.

An alterative approach is the Neighborhood Sorting Index (NSI), defined as follows:

$$NSI \equiv \frac{\sigma_N}{\sigma_H} \equiv \left(\frac{\sum_{n=1}^{N} h_n (m_n - M)^2 \Big/ H}{\sum_{i=1}^{H} (y_i - M)^2 \Big/ H} \right)^{0.5} = \left(\frac{\sum_{n=1}^{N} h_n (m_n - M)^2}{\sum_{i=1}^{H} (y_i - M)^2} \right)^{0.5} \quad (1)$$

where

H = total households in the entire area;
h_n = number of households in neighborhood n, where $n = \{1,2,...,N\}$;
m_n = mean household income of neighborhood n;
y_i = income of household i, $i = \{1,2,...,H\}$;
M = mean household income over entire area;
N = number of neighborhoods into which area is divided.

The NSI is ratio of the standard deviation of the parcel means to the standard deviation of the individual incomes.[1] Thus, if each individual lives in a neighborhood in which the mean income is identical to his own, the index is equal to one, reflecting total economic homogeneity within neighborhoods and 100% of all variation in income between neighborhoods. If all neighborhoods have identical mean incomes, the NSI is 0, reflecting no economic segregation.

The NSI has advantages and disadvantages. By measuring neighborhood and individual incomes relative to the mean and by controlling for the total variation among households, NSI is not sensitive to changes in the parameters of the income distribution that do not alter spatial relations. NSI, however, treats neighborhoods as individual atoms that do not interact. The

---

[1] We assume at least some income heterogeneity, i.e. $\sigma^2_H \neq 0$

neighborhoods could be randomly shuffled and there would be no effect on the measured level of economic segregation.

The square of NSI is closely related to a measure which has been called a variety of different names: the variance ratio index (James and Taeuber 1985); eta squared (Duncan and Duncan 1955); S (Zoloth 1976); and the correlation ratio (Farley 1977). Applied to a dichotomous variable, the equivalent measure is:

$$V = \frac{\sum_{n=1}^{N} h_n \left(p_i - P\right)^2 / H}{P\left(1-P\right)} = \frac{\sigma_{p_i}^2}{\sigma_P^2} \qquad (2)$$

in which p_i is the neighborhood proportion, P is the population proportion (the mean of the binomial variable), and $P(1\text{-}P)$ is the variance. Virtually all of the literature relevant to this class of measures is based on the binomial form.

In a sense, the NSI is merely a measure of the heterogeneity of the neighborhood means, normalized by the income variance.[2] It is a measure of spatial segregation only to the extent that it tells you how much information about variation in household income is lost by aggregating data to a non-overlapping spatial lattice of neighborhoods, such as census tracts. However, it fails to capture larger scale features of the spatial arrangement of neighborhoods. The NSI is not affected if all high-income neighborhoods are clustered in one part of the metropolitan area or if they are scattered randomly around the map.

To address this problem, we propose the GNSI as a geographically sensitive measure of spatial segregation. The key difference between GNSI and NSI is that the numerator of GNSI incorporates a flexible moving window for the calculation of a neighborhood's economic level which is larger than the neighborhood itself. The larger entity we refer to as a community. For

---

[2] Morrill (1991) and Wong (1997) make a similar point with respect to the Index of Dissimilarity.

example, the community could include all contiguous neighborhoods, or all neighborhoods whose centroids are within a certain distance r of the given neighborhood's centroid. The community can be expanded by going to higher levels of contiguity or higher multiples of the radius *r*.[3]

Conceptually, the GNSI of order *k* is then defined as:

$$GNSI_k \equiv \left(\frac{\sum_{i=1}^{H} (m_{ki} - M)^2}{\sum_{i=1}^{H} (y_i - M)^2} \right)^{0.5} \qquad (3)$$

where m_{ki} is the mean household income in the kth order-expanded community from a household *i*. The order defines the spatial extent of the community, defined either in terms of distance from each household or in terms of the order of contiguity. For example, in the first order contiguity expansion, the moving window for each household consists of directly contiguous neighbors including itself. The first order distance expansion includes all households within a circle of one unit radius from each household. (The radius itself is arbitrary.) Second and higher order expansions are defined in an analogous manner.

In practice, however, individual household income data with latitude and longitude information is often unavailable, and, thus, equation (3) can not be implemented as it stands. Usually, income data are available only as summaries for geographical neighborhood boundaries, e.g. census tracts. Thus, we need a working definition of GNSI as follows:

$$GNSI_k \equiv \left(\frac{\sum_{n=1}^{N} h_n (m_{kn} - M)^2}{\sum_{i=1}^{H} (y_i - M)^2} \right)^{0.5} \qquad (4)$$

[3] More complex conceptions of geography can be incorporated, recognizing natural and man-made boundaries, through a spatial weight matrix (Reardon and O'Sullivan 2004). However, this does not affect the conceptual issues discussed here.

where m_{kn} is the mean household income in the k^{th} order expansion from a neighborhood n. The first order distance expansion, for instance, includes all neighborhoods whose centroids are within a unit radius from the centroid of the given neighborhood. Note that GNSI differs from NSI only in that m_{kn} is substituted for m_n in Equation (1). In fact, NSI is a special case of GNSI in which the order of expansion defined as zero. In that case, the community is simply the neighborhood and the two measures are identical. For any order of expansion greater than 0, however, the communities are interrelated by the spatial structure of the neighborhoods. Each neighborhood is considered part of a larger community, and the communities form an overlapping set.

COMPARISON OF GNSI AND NSI

GNSI overcomes a number of the shortcomings of NSI. First, GNSI is sensitive to the spatial relationships of the tracts, as will be illustrated below. Second, by increasing the order of expansion, the degree of segregation can be measured at various spatial levels from small scale to large scale, reducing the dependence on arbitrarily defined administrative units such as census tracts. In effect, GNSI incorporates two types of information about the spatial segregation of household income. First, like NSI, it reflects the heterogeneity of the parcels representing neighborhoods. Second, it reflects the spatial patterning of the neighborhoods themselves. The latter point can be illustrated by expanding the formula for GNSI. We multiply by one, inserting the sum of the squared deviations of the neighborhood means into both the numerator and denominator, and rearrange terms as follows:

$$GNSI_k \equiv \left(\frac{\sum_{n=1}^{N} h_n (m_n - M)^2}{\sum_{i=1}^{H} (y_i - M)^2} \right)^{0.5} \left(\frac{\sum_{n=1}^{N} h_n (m_{kn} - M)^2}{\sum_{i=1}^{H} h_n (m_n - M)^2} \right)^{0.5}$$

$$= (NSI)(C_0^k) \qquad (5)$$

The first term is the familiar NSI, measuring the deviation of neighborhood means from the grand mean. The second term is the ratio of the sum of the squared deviations of the community means, given expansion of order *k*, relative to the sum of the squared deviations of the neighborhood means, effectively expansion of order 0.

The heterogeneity of the neighborhoods captured by NSI and the spatial clustering of tracts represented by C_o^k may seem like dissimilar concepts that ought not to be combined in one measure.[4] On the contrary, they get at the same underlying phenomenon (Reardon and O'Sullivan 2004). Essentially we have a continuous space and households of different incomes are scattered in two dimensions. If there is spatial clustering of individual households along the income dimension, this could be manifested in two ways. First, there will be differences between the district means. Second, there could also be a spatial clustering of the mean incomes of the tracts themselves.[5] It depends on whether the household clustering process operates on scales larger than the boundaries of the districts. Since the districts are often arbitrary administrative units of which residents take little note, if they are aware of them at all, there is no

---

[4] Massey and Denton (1988), for example, surveyed existing measures of segregation and classified them into five dimensions - evenness, exposure, concentration, centralization, and clustering.

[5] Measures of clustering have been developed by Geary (1954), White (1986), Wong (1993, 1999), and recently O'Sullivan and Wong (2004). However, the measures are for two group segregation, not for a continuous variable such as income.

reason to think that the spatial clustering process would respect these boundaries. Thus, the capacity of GNSI to capture segregation at both the sub- and super-district levels is an advantage of GNSI over other measures, because both are manifestations of the same underlying phenomenon: the clustering of individual households.[6]

GNSI is calculated by means of a spatial weight matrix that incorporates the spatial structure of the neighborhoods. We restate Equation (3) in matrix notation in which the individual, rather than the neighborhood, is the basic unit of observation:

$$GNSI_k \equiv \left(\frac{\sum_{n=1}^{N} h_n (m_{kn} - M)^2}{\sum_{i=1}^{H} (y_i - M)^2} \right)^{0.5} = \left(\frac{y' W_k' W_k y}{y' y} \right)^{0.5} \qquad (6)$$

where y is an H by 1 vector representing the deviations of individual household incomes from the metropolitan mean and Wk is an H by H spatial weight matrix for the k^{th} order expansion. Recall that H is the total number of households. The $(i,j)^{th}$ element of the weight matrix indicates whether household i and household j are members of the same community. If they are not, the element is zero. If they are members of the same community, the element is $1/h_c$, where h_c is the total number of households in the k^{th} order-expanded community of individual i. In other words, the matrix is row-standardized, and the numerator in GNSI is the household-weighted sum of squared deviations of the community means from the grand mean.

When the order of expansion is zero (no expansion beyond the individual neighborhood), the GNSI is identical to NSI. In that case, assuming that the households are sorted in the data by neighborhood, the weight matrix has a block diagonal structure:

---

[6] Rogerson (1999) makes a similar point: "It would be desirable to have a statistic with which we would conclude that...the combination of the aspatial deviations *and* the spatial pattern of the deviations could not have occurred by chance (131, emphasis in the original).

$$W_o = \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{NN} \end{bmatrix}, \quad w_{nn} = \frac{1}{h_n} i_{h_n} i_{h_n}' \tag{7}$$

As before, h_n is the number of households in neighborhood n and i_{h_n} is a vector of ones of dimension h_n. Thus, W_0 is a symmetric idempotent matrix and $y'W_0'W_0 y$ equals $y'W_0 y$ and is the sum of the squared deviations of the neighborhood means from the metropolitan mean, as in the numerator of NSI in Equation (1). All individuals within a given individual's neighborhood are given equal weight in the individual's contribution to the summary measure, while all individuals not in the given individual's immediate neighborhood are given zero weight, as shown by the zero elements off the main diagonal. This implies, as we have already argued above, that NSI (=$GSNI_0$) is insensitive to the spatial arrangement of the parcels. Here this property is seen to be the result of a zero weight in the relevant cells of the weight matrix.

In any order of expansion beyond zero, all households in neighborhoods included in a given neighborhood's community (defined by the expansion), will have a non-zero weight. All households in the community contribute to each household's component of the segregation score. In expansions based on contiguity, each neighborhood's community overlaps, but is different from, the community of each of its neighbors. In expansions based on distances between neighborhood centroids, a circular window with a radius of $k*r$ moves over the region from neighborhood to neighborhood, and again the communities overlap. Because the communities are interwoven, the spatial relationships of the tracts are taken explicitly into account. Rearranging the parcels now changes the measured level of segregation. From a sociological point of view, the overlapping window reflects the fact that in the sequence {A, B, C}, B can be a neighbor to both A and C, even though A and C are not considered neighbors. Communities are matters of individual perception, and need not conform to the administrative dictum that spatial divisions should be mutually exclusive.

CHARACTERISTICS OF GNSI

Previous literature has defined sets of criteria for the evaluation of segregation measures (Frankel and Volij 2004; James and Taeuber 1985; Reardon and O'Sullivan 2004; Schwartz and Winship 1980). The criteria are designed to assess the conceptual and operational characteristics of the numerous alternative measures, with a view towards helping researchers choose among them. Since most of criteria have been developed without explicit attention to spatial considerations, some need modification and others may not be useful in the context of spatial segregation measures.

Scale Interpretability (Reardon and O'Sullivan 2004)

GNSI being equal to zero indicates all the community mean incomes (m_n) are the same as the total average (M), and a value of one indicates that all the households reside in strictly homogenous communities, with the each household's income exactly equal to the community's mean income. Thus, GNSI is bounded between zero and one and the scale is easily interpretable.

Independence of Arbitrary Boundaries

This criterion is related to modifiable aerial unit problem (MAUP).[7] Although King (1997) argued that MAUP can be solved based on aggregated data, it is largely agreed that MAUP cannot be solved unless all the individual data become available or boundaries are exactly matched to the boundaries of interest (Anselin 2000). Any measures based data from arbitrary

[7] MAUP refers to problems not only from size but also from shape modification.

spatial boundaries will suffer from MAUP. The conceptual form of GNSI is based on the exact locations of individuals, and would not be dependent on arbitrary boundaries. The application of the working definition of GNSI, however, assumes data based on geographic boundaries, such as census tracts, and therefore will not be entirely free from MAUP. However, GNSI provides a partial solution, by allowing the size of the moving window to be expanded, so that the computed statistics can reflect larger units composed in various ways, but of course not smaller ones.

Decomposability

Reardon and O'Sullivan (2004) argue that segregation measures should be decomposable into a sum of within- and between-area components.[8] GNSI is decomposable both by spatial scale and by spatial units. First, allowing $\tilde{m}_{kn}$ and $\tilde{y}_i$ to represent deviations from the overall metropolitan mean income for notational convenience, GNSI of order k can be rewritten as:

$$\begin{aligned}\left(\frac{\sum_{i=1}^{N} h_n \tilde{m}_{kn}^{\;2}}{\sum_{i=1}^{H} \tilde{y}_i^{\;2}}\right)^{0.5} &= \left(\frac{\sum_{n=1}^{N} h_n \tilde{m}_{0n}^{\;2}}{\sum_{i=1}^{H} \tilde{y}_i^{\;2}}\right)^{0.5} \left(\frac{\sum_{n=1}^{N} h_n \tilde{m}_{1n}^{\;2}}{\sum_{n=1}^{N} h_n \tilde{m}_{0n}^{\;2}}\right)^{0.5} \left(\frac{\sum_{n=1}^{N} h_n \tilde{m}_{2n}^{\;2}}{\sum_{n=1}^{N} h_n \tilde{m}_{1n}^{\;2}}\right)^{0.5} \cdots \left(\frac{\sum_{n=1}^{N} h_n \tilde{m}_{kn}^{\;2}}{\sum_{n=1}^{N} h_n \tilde{m}_{(k-1)n}^{\;2}}\right)^{0.5} \\ &= (NSI)\left(C_0^1\right)\left(C_1^2\right)\cdots\left(C_{k-1}^k\right) \qquad \text{[From (5)]} \\ &= (\sigma_N/\sigma_H)\left(C_0^1\right)\left(C_1^2\right)\cdots\left(C_{k-1}^k\right) \qquad \text{[From (1)]}\end{aligned} \tag{8}$$

Each term on the right hand side provides the change in segregation as one degree of expansion order increases, as the scale of community increases: the first term, from individual incomes to an areal unit (e.g., census tract), the second term, from the areal unit to its first order neighbors, etc. The first term, i.e., NSI is decomposed again into the heterogeneity of individual income (σ_H) and the heterogeneity of areal units (σ_N). The scale decomposability of GNSI

---

[8] Although it is desirable to have the decomposability, valid measures without it may still provide good measures of spatial segregation.

allows the investigation on the changes in spatial structure of segregation in a region or on the different levels of segregation among regions, as illustrated in empirical application section.

On left hand side, $GNSI_k^2$ can be decomposed so that the value for each tract, $h_n \tilde{m}_{kn}^{\,2} / \sum \tilde{y}_i^2$ is a localized measure of segregation, reflecting the contribution of each neighborhood to the global segregation score.[9] Because of the unit decomposability of GNSI, it is possible to test the statistical significance of segregation in each unit by comparing the values with the spatial random distribution (Anselin 1995).

Organization Equivalence and Size Invariance[10]

James and Taeuber (1985) argued that since a segregation measure should permit comparison of districts that differ in the number of schools and the number of students, the measured segregation level should be unchanged if two organizational subunits with identical population composition and size are combined into a single unit or a single unit is split into two identical units. This is known as organization equivalence. In our application, it would require that the measured level of economic segregation for a metropolitan area should be unchanged if two identical census tracts are combined. In a similar vein, the combination of two identical districts into one should yield the same degree of segregation for the combined populations. For example, if the population of each school or census tract was doubled, without changing the relative distribution of students or persons on the relevant characteristics, segregation should not change. This is known as size invariance.

---

[9] For certain applications, particularly maps showing the degree of segregation of the parcels, the local value should be normalized by the number of households.

[10] Organization equivalence is also called location equivalence in Reardon and O'Sullivan (2004). Size invariance is also named population symmetry (Schwartz and Winship 1980), or population density invariance (Reardon and O'Sullivan 2004).

In spatial context, organization equivalence is a problematic concept. First, if combining or splitting spatial units changes the structure of the spatial distribution of population, then a valid segregation measure will change the level of measured segregation accordingly. Two different census tracts, by definition, can not be identical in terms of their spatial location. Thus, the organization equivalence condition cannot be a valid criterion for desirable spatial measures, because a split or join necessarily changes the spatial relationships of the units – the location of the combined centroid will differ from the original tracts, and the orders of contiguity will likely be affected as well.

GNSI, in general, does not possess organization invariance, except in certain unusual circumstances where the split or join does not affect the spatial weight matrix. Even then, when two census tracts are joined, the new tract has a larger population and will change the community average of all tracts that include the joined tract in their communities, unless: a) both tracts were part of the original communities to start with, or b) the joined unit has the same mean income as all other members of the communities of the affected tracts. These are rather restrictive conditions.

GNSI does possess size invariance, which can be shown as follows. Suppose every household is replicated R times. Then, the community mean, m_{kn}, is not affected because the neighborhood means are not affected; the weight assigned to each neighborhood in the calculation of the community mean is scaled equivalently. Finally, the R term cancels out:

$$GNSI_k^R = \frac{\sum_{n=1}^{N} Rh_n \left(m_{kn} - M\right)^2}{\sum_{i=1}^{H} R\left(y_i - M\right)^2} = GNSI_k \qquad (9)$$

Transfers / Exchanges

If poor households move from poor neighborhoods to affluent neighborhoods or if the rich move in the opposite direction, a valid measure of economic segregation should decline. Reardon and O'Sullivan (2004) translate this rule in the spatial context into type 1 and 2 exchange rules depending on the environment of the exchange. GNSI meets both types of exchange rules, since the exchange reduces the magnitude of the numerator in both the conceptual and working definitions, Equations (3) and (4) respectively, as neighborhoods become more heterogeneous. Note that, unlike D, the exchange between the rich in affluent units and the poor in low income units always reduces the measured segregation irrespective of both units being below or above the average income.

Composition Invariance

In considering racial segregation, James and Taeuber (1985) argued that a desirable property of a segregation index is that it should be independent of the underlying population proportion of black residents. For example, in examining the segregation of blacks from whites, the index should not change if the number of blacks in each unit is increased proportionately while holding the number of whites constant. D and the Gini coefficient, among others, have this quality. According to an analysis by James and Taeuber (1985), the Variance Ratio Index – a measure closely related to NSI – and the Entropy Index do not. However, their analysis is based on the dichotomous case. By changing P, the proportion black, they simultaneously change both the mean (P) and the variance (P(1-P)) of the population composition.

In the case of income, a continuous variable, we can consider the effects of the mean and variance separately. Clearly, GNSI satisfies the condition of composition invariance with respect to the mean. If a constant value, a, is added to every household's income, the neighborhood means, the community means, and the overall mean are all raised by the same constant amount. Therefore:

$$GNSI_k\left[M+a\right]=\frac{\sum_{n=1}^{N}t_n\left(\left(m_{kn}+a\right)-\left(M+a\right)\right)^2}{\sum_{i=1}^{H}\left(\left(y_i+a\right)-\left(M+a\right)\right)^2}=GNSI_k\left[M\right] \tag{10}$$

Assessing the effect of a change in the variance is more complex. Both NSI and GNSI are partitions of the total variance in household income. In the case of NSI, the variance is partitioned into the between- and within-neighborhood parts, and the NSI is defined as the square root of the variance between neighborhoods divided by the total variance. For GNSI, the variance is divided into the between and within community parts. To increase the total variance, you either have to increase the between unit variance, the within unit variance, or a combination of both. If you hold the ratio of the two constant, then by definition the NSI or GNSI is held constant. If you change the ratio, then by definition the NSI or GNSI changes.

Logically, the criterion for composition invariance should be a test for a change in the variance of an income distribution holding other factors constant. For example, we can convert each income to a z score. We can increase the variance of the income distribution by changing the incomes of each household to reflect the higher variance while holding the relative positions of each household constant, relative to the remaining households, and by keeping those households grouped in the same units (neighborhoods) as before.

For a z-score preserving transformation, both NSI and GNSI are composition invariant with respect to the household income variance. The z scores of all households are unaffected by the variance change. Thus, the average z score for each neighborhood is also unaffected. Regardless of the order of expansion, the community average z score of each household will include the same neighborhoods if the spatial relationships are unaltered; therefore, the community averages will be the same as well. In terms of z scores, the GNSI is expressed as:

$$GNSI_k = \frac{\sum_{n=1}^{N} h_n z_{kn}^2}{\sum_{i=1}^{H} z_i^2} = \frac{\sum_{n=1}^{N} h_n z_{kn}^2 \Big/ H}{\sum_{i=1}^{H} z_i^2 \Big/ H} = \frac{\sum_{n=1}^{N} h_n z_{kn}^2}{H} \qquad (11)$$

where z_{kn} is the community average z score for the k^{th} order expansion for neighborhood n. Therefore, neither GNSI nor NSI will be affected by a change in the variance of the income distribution unless that increase in the variance differentially affects households in one area of the city, thereby changing the relative positions of at least some households. But such a change would represent an actual change in the degree of economic segregation, which a valid measure should reflect.

Thus, James and Taeuber(1985) were incorrect in claiming that the Variance Ratio Index was not composition invariant. The requirement is that "segregation indexes should not be affected by the value of the measure of [population] composition about which the variation is calculated" (1985:15). As shown above, the Variance Index Ratio and measures like it, such as the NSI and GNSI, are invariant with respect to a change in the mean of the underlying variable. There results were a function of the manner in which they modeled the change in the mean, which not only changed the variance – which is inevitable in the case of a binomial variable – but also changed the spatial relations between high and low value tracts. A change in the mean carried out while maintaining the z scores of the schools in terms of percent black at their initial values has no effect on the Variance Ratio.

ILLUSTRATIVE EXAMPLES

This section employs a hypothetical and highly simplified example to demonstrate the advantages of GNSI over NSI. Figure 1 shows a hypothetical MSA composed of 100 square neighborhoods, arranged in a 10 by 10 grid, each of which is occupied by 10 persons. For the sake of the example, all 10 persons within a given neighborhood are either all white or all nonwhite.[11] If they are white, the square is shaded and valued one, otherwise the square is not shaded and is valued at zero. (The assignment of values is arbitrary and does not affect the analysis.) Figure 1 illustrates three possible arrangements of an equal number of white and non-white residents. The first case is a traditional checkerboard pattern of alternating white and non-white neighborhoods. In the second example, the metropolitan area is divided into four quadrants consisting of five by five grids of neighborhoods. The northwest and southeast quadrants are all white, while the other two quadrants are non-white. Finally, in the last example, the entire east half of the city is all white and the west side non-white.

The Neighborhood Sorting Index, as well as other traditional measures of segregation such as the Index of Dissimilarity, would rate all three metropolitan areas as equally segregated. Indeed, each metropolitan area in the example would be judged to be totally segregated; that is, NSI=1 or Index of Dissimilarity=1. In such measures, all that matters is the composition of the parcels and not their spatial arrangements. There is a sense in which these measures correctly report the degree of segregation in that no person lives in a neighborhood with a person of a different color. However, these measures fail to capture the higher-order segregation of the

[11]Given that the focus of this paper is economic segregation, we could equally well specify the condition as rich vs. poor, or assign specific dollar income amounts, but the arguments would be exactly the same and the black/white example is familiar and consistent with the discussion of other segregation measures in the literature.

parcels themselves. Clearly, rating these three scenarios as equally segregated is conceptually incorrect in any conceivable sociological application. Neighborhood boundaries are permeable; in most empirical applications, the neighborhood boundaries are arbitrary administrative demarcations with limited relevance to neighborhood residents. To the extent that interactions across neighborhood boundaries are mediated by spatial distance, then the failure of traditional measures to account for the clustering of parcels is a failure to properly measure segregation.

Applying GNSI to the three scenarios yields quite different results. Here we employ first and second order expansions based on contiguity. Two spatial weight matrices are employed, using "rook" and "queen" criteria for defining whether one neighborhood is contiguous to another. In the former, any square which shares a side with the given square is considered a neighbor; in the latter, any square which either shares a side or touches at a corner is considered a neighbor.

Table 1 shows the measured levels of segregation. Compared with NSI, which shows a segregation level of 1 for all three scenarios, GNSI with both first and second order expansion clearly ranks the metropolitan areas in a way that reflects the clustering of the neighborhoods. With rook criteria, the measured segregation level of the checkerboard pattern drops to 0.56 in the first expansion and 0.34 in the second order expansion. With queen criteria, both expansions yield segregation levels less than 0.10 in the checkerboard pattern. In the east-west scenario, the measured level of segregation remains high, even with a second-order expansion using queen criteria. Even taking larger communities into account, the east-west scenario is still highly segregated. The results for the Quadrants scenario fall between these two extremes.

Rook and queen criteria yield qualitatively similar results except in the case of the checkerboard pattern, which results from the alternating white and non-white neighborhoods. In the first order expansion, with rook criteria every white neighborhood is part of a community that is four fifths non-white and vice versa. With queen criteria, each neighborhood is about half white because of the inclusion of the diagonal squares. With irregularly shaped units, the choice between rook and queen is not likely to make any difference. However, if the units are formed by street intersections, as census tracts often are, the decision may be consequential.

The GNSI reduces the influence of arbitrary boundaries, i.e. the MAUP problem. Figure 2 examines what happens when the same households are allocated to larger neighborhood units. This would be analogous to looking at the same metropolitan area with census tracts rather than blocks, or counties rather than tracts. The shadings represent the original distribution of households, and are the same as in Figure 1. Only the boundary lines have changed, so that each area is now divided into a 5 by 5 grid of neighborhoods. As shown in Table 2, the NSI for the checkerboard case is suddenly measured as zero, rather than one, a rather extreme example of the modifiable aerial unit problem. GNSI also measures zero segregation in the checkerboard case, but the measure was already low, especially using queen criteria. In general, the change in the measure when changing the grid size is smaller for the spatially sensitive measures than for $GNSI_0$, i.e. NSI. This is a key advantage of an extensible, spatially-aware measure.

Figure 3 shows a section of the Dallas Metropolitan Area census tract map for illustrative purposes. In the first-order contiguity expansion, census tract 11's community includes census tracts 6, 8, 10, 12, 16, and 18. Census tract 8 has neighbors 4, 5, 6, 10, and 11. Thus, the communities of neighborhoods 8 and 11 overlap by sharing 6 and 10, as well as by being neighbors to each other. But each neighborhood also has a few neighbors not shared by the other. Table 3 shows the number of households and the mean household income for each census tract.

The table also shows the community mean for the first-order contiguity expansion.[12] Because of the effect of averaging neighborhoods together, the community means are less variable than the neighborhood means.

Figure 4 illustrates the construction of W_k, the spatial weight matrix for GNSI. In this necessarily small example, we have 5 neighborhoods denoted A through E. Neighborhood A contains one person, C contains 3 persons, and the rest contain 2 persons each. The top panel of Table 4 shows the weight matrix for the zero-order expansion, which would have the effect of summing up the squared deviations of the tract means from the grand mean. The bottom panel shows the first-order contiguity expansion assuming queen criteria. Note that the weights sum to one across any row, a property known as "row standardization."

EMPIRICAL APPLICATION

As we argue above, the GNSI is conceptually superior to the NSI because it takes into account the spatial interrelationships of the parcels. Whether or not this conceptual superiority has any practical significance is an empirical question. For example, it could be the case that virtually all metropolitan areas have their poor neighborhoods in one or two main clusters. If there is relatively little variation among metropolitan areas in the super-clustering of neighborhoods, then application of the GNSI may not affect on the rank-ordering of cities in terms of economic segregation. Accordingly, in this section we examine 10 large metropolitan areas to show how the GNSI changes and improves our understanding of economic segregation.

[12]Note that there are two intersections where the rook and queen criteria would yield different results; in these calculations, we use the queen criterion. Also, the communities of census tracts on the edge of this illustration include neighborhoods not shown in the diagram. These are included in the calculations.

We estimate the GNSI for the ten largest metropolitan statistical areas for 1990 and 2000. We also replicate Jargowsky's calculation of NSI in 1990 for these metropolitan areas. Census tracts are used as proxies for neighborhoods (White 1987). Both measures require the variance of individual household income, which is not directly available in the census tract data. Instead, we estimate the variance from the grouped data (Jargowsky 1995, Appendix A).[13] Prior to estimating the variance, we adjust MSA boundaries to make them comparable over time.

ArcGIS software is used to identify orders of contiguity among the census tracts, using "Queen" criteria.[14] Once the spatial structure of the neighborhoods has been captured in a spatial weight matrix using GIS software, the data can be exported to any statistical software package for the remainder of the analysis.

Order Expansion

Tables 5 through 8 report the GNSI by race/ethnicity and order of expansion for the 10 largest metropolitan areas. The first column in each time period shows GNSI with a 0-order expansion, which is identical to NSI. Only the census tract itself contributes information to the value of mean income recorded at that location. The second column represents a first order expansion in which the mean income is a function of the tract itself and immediately contiguous tracts. The third column is GNSI with a second order expansion. The tract itself, its immediately contiguous tracts, and any additional tracts immediately contiguous to the first order tracts are all included in the calculation of mean income at a point in space.

---

[13] Alternatively, one could use the Public Use Microdata Sample (PUMS), but the PUMs sample areas do not correspond exactly to metro areas.

[14] A few tracts are islands in lakes or off the coast and technically have no contiguous neighbors. For the purpose of this analysis, we treat them as isolated.

For all race/ethnic groups and in both years, the mean level of measured segregation declines as the order of expansion is increased. The decline is a natural consequence of expanding the area over which community mean incomes are calculated, and is in effect another manifestation of the MAUP discussed earlier. As the area expands, more diverse households are captured in the moving window resulting in the appearance of lower segregation. At the extreme, as the community is expanded to include the whole metropolitan area, the GNSI would approach zero, which is conceptually correct. Thus, we do not attach any weight to the declines in the measured levels of economic segregation as the order of expansion increases. More important than the mean level of measured segregation are the trends over time and the changes in the ranks of the metropolitan areas as the moving window is expanded.

Trends over time

On average, overall economic segregation in the 10 metropolitan areas decreased significantly, regardless of the measure employed, as shown in Table 5. $GNSI_0$, the aspatial measure, declined by nearly one third, falling from 0.45 in 1990 to 0.31 in 2000. This is a significant departure from the trend in economic segregation between 1970 and 1990 (Jargowsky 1996). However, the 1990s were a decade of sustained economic growth, low unemployment, and rising wages for less skilled workers. Many urban areas experienced a Renaissance, resulting in redevelopment of central city neighborhoods, i.e. "gentrification." Further, concentration of the poor in isolated high-poverty neighborhoods also declined markedly (Jargowsky 2003; Kingsley and Pettit 2003). All of these factors are consistent with a reversal of the earlier trend, resulting in a decline in economic segregation in the 1990s.

As the order of expansion increases, the relative changes in economic segregation for the overall population are about the same, even though the absolute changes are smaller. As noted, the overall $GNSI_0$, declined 31 percent from 0.45 to 0.31. In comparison, $GNSI_1$ declined 31% from 0.36 to 0.25, and $GNSI_2$ declined 29% from 0.31 to 0.22. However, as explained below, the overall figure represents two different and offsetting trends among whites and minority groups.

Some of the decline in economic segregation could have been driven by changes in racial and ethnic segregation, because members of minority groups tend to have lower household income. The general trend is toward lower levels of segregation by race since 1970 (Farley and Frey 1994; Iceland, Weinberg, and Steinmetz 2002; Massey and Denton, 1987). Thus, it is probably better to examine economic segregation within racial and ethnic groups, to isolate the effect of economic segregation from changing patterns of racial segregation. Tables 6, 7, and 8 show that there were also declines in $GNSI_0$ for whites, blacks, and Hispanics, respectively. The declines in $GNSI_0$ are somewhat smaller than the overall figure for whites (28%) and blacks (25%) on a percentage basis, but larger for Hispanics (37%).

Interestingly, Whites and minority groups show different patterns on higher-level expansions. Economic segregation of whites declines 28 percent using GNSI0, but only 25 percent with $GNSI_2$. Thus, the decline in economic segregation is not quite as great when larger spatial structures are taken into account. Some of the apparent decline in white economic segregation was therefore offset by segregation at a higher level, implying that the white settlement pattern was spreading out.

In contrast, among Blacks and Hispanics, the decline in segregation is even larger when the order of expansion is raised. For Blacks, the decline is 25 percent for $GNSI_0$, 28 percent for $GNSI_1$, and 29 percent for $GNSI_2$. For Hispanics, the decline is 37 percent for $GNSI_0$, 38 percent for $GNSI_1$ and 42 percent for $GNSI_2$. Minorities were less segregated in their immediate neighborhoods, and their immediate neighborhoods were parts of larger communities that were becoming less economically segregated. To some extent, this may reflect the movement of lower-income blacks out of traditional inner-city areas to inner-ring suburbs, where they may be closer to more middle-class black neighborhoods.

Following the method described in Equation (8), $GNSI_2$ can be decomposed as follows:

$$GNSI_2 \equiv \left(\frac{\sum_{n=1}^{N} h_n (m_n - M)^2}{\sum_{i=1}^{H} (y_i - M)^2} \right)^{0.5} \left(\frac{\sum_{n=1}^{N} h_n (m_{n1} - M)^2}{\sum_{i=1}^{H} h_n (m_n - M)^2} \right)^{0.5} \left(\frac{\sum_{n=1}^{N} h_n (m_{n2} - M)^2}{\sum_{i=1}^{H} h_n (m_{n1} - M)^2} \right)^{0.5}$$

$$= (\sigma_N / \sigma_H)(C_0^1)(C_1^2) \qquad (12)$$

Equation (12) shows the spatial structure of the changes in economic segregation. The second term on the right hand side of the equation indicates the segregation level in the first order expansion compared to the average of census tract level while the last term shows the segregation level in the second order expansion compared to the first order expansion.

Tables 9 - 11 show that the primary cause of the declines in economic segregation was not the changes in spatial clustering of census tracts, but rather increases in the heterogeneity of individual incomes (σ_H) that were not reflected in corresponding changes in heterogeneity of neighborhood mean incomes (σ_N). Across the 10 cities, the increase in the standard deviation of household income was more about 60 percent overall as well as for whites and blacks, and nearly 50 percent for Hispanics. At the same time, the increase in the standard deviation of the neighborhood means was 10 percent overall, 15 percent for whites, 18 percent for blacks, and -2 percent for Hispanics. Since the overall variance is the sum of the within-neighborhood and between-neighborhood components, the obvious implication is the variance within neighborhoods must have increased during this period for all groups.

In contrast, the degree of spatial clustering among relatively poor and rich census tracts in the first order expansion (Table 10) and in the second order expansion (Table 11) hardly changed from 1990 to 2000. On average, the changes in the super-clustering of tracts are less than 3% for all race-ethnic groups. However, while small, the changes are positive for whites and negative for minorities, consistent with the differential trends in effect of expansion noted above.

The decomposition suggests that, during the very strong economy of the 1990s, there was a general increase in inequality driven by increases in income that were broadly distributed across neighborhoods. Apparently, better-off residents in many locations benefited from the new economy, increasing inequality with changing the disparity of neighborhoods. In addition, gentrification, when higher income persons move closer to the central-city core and have neighbors with lower incomes, may have played a role in constraining the growth of neighborhood-level inequality.

In conclusion, the main drive for the changes in segregation between 1990 and 2000 is the significant increase in income variation among individual households not reflected in the variation of neighborhood mean incomes. The gap between the rich and the poor has increased faster than the gap between high average income tracts and low average income tracts.

Changes in metropolitan areas

The analysis thus far has looked only at the averages of the 10 metropolitan areas included in our analysis. The differences among the metropolitan areas also reveal how taking space directly into account alters our understanding of economic segregation.

As noted above, the absolute levels of economic segregation are lower when a first- or second order expansion of the neighborhood is conducted, a natural consequence of employing a broader spatial ranges. Metropolitan areas, however, differ a great deal in how much the level of segregation declines as the degree of overlap increases. Figure 5 shows the level of economic segregation for the first and second order expansions relative to $GNSI_0$. The figures for $GNSI_1$ are always lower, and the second order is always lower than the first order, which is to be expected. However, taking the spatial arrangement of census tracts into account makes a much larger difference in Dallas and Houston than it does in New York and Chicago, as indicated by the much steeper decline in the GNSI as the order of the statistic is increased. The western and southern metropolitan areas decline far more rapidly than the northern and eastern metropolitan areas. The implication is that the tracts of similar incomes are more likely to be clustered in Midwest and Northeast, whereas tracts with similar mean incomes are less clustered in the South and West.

Thus, the answers to questions about which places and regions are the most segregated are dependent upon the scale employed in the analysis. Because segregation by income occurs at different scales, and because the relative importance of these scales differs among metropolitan areas, the rank ordering of metropolitan areas changes if space is taken into account. For example, in 1990 Chicago was ranked 7th in terms of economic segregation of the total population based on $GNSI_0$, but moved up to 3rd place based on $GNSI_2$. Chicago's segregation extends over larger scales than most other cities, and $GNSI_2$ picks this up better than the other measures. Likewise, Houston drops from 4th to 7th place in 1990 after space is taken explicitly into account. Comparative work on segregation needs to explicitly consider the issue of scales raised here, either by using a spatially expandable statistic or other means, or they risk giving a limited and potentially misleading view of the relative degree of segregation across metropolitan areas.

CONCLUSION

The measurement of spatial segregation has often used measures such as the Neighborhood Sorting Index (NSI) or the Index of Dissimilarity (D) that do not take full account of space. Those traditional measures of economic segregation are flawed because they incorporate space in only a very limited way. They measure the heterogeneity of the parcels used in the analysis, which does capture some of the geographic clustering of households. Clustering, rather than being a separate dimension from evenness or homogeneity of parcels, is just segregation at a scale larger than the unit used in the analysis.

We propose GNSI, an extension of NSI, in which each neighborhood is part of other neighborhoods, the extent of which can be easily changed by increasing the order of contiguity or the allowable distance between points (or centroids) in space. The central innovation is that neighborhoods are overlapped with other neighborhoods, so that they are woven together in a way that reflects the sociological concept of neighborhood.

As demonstrated in Section 3, the GNSI easily solves the checkerboard problem in which NSI and similar measures are insensitive to the spatial arrangement of the geographic units in the analysis, such as census tracts. NSI is a special case of GNSI in which the order of expansion is zero. For orders of expansion greater than 0, GNSI incorporates information about the heterogeneity of the neighborhoods, the spatial relationship between the neighborhoods, and the interaction between those two levels.

From a calculation standpoint, a GIS is needed to construct the spatial weight matrix. After that, any statistical software may be used to compute the measure. As discussed in Section 4, GNSI has scale interpretability, decomposability, organization equivalence, size invariance, and satisfies the transfer/exchange condition. Further we argue that the measure is composition invariant as well, and that prior literature has been in error in arguing that measures in the same class failed the composition invariance test due to a misspecification of the composition change.

An application of this technique to the ten largest metropolitan areas reveals that previous work on economic segregation has painted a somewhat incomplete picture. Both spatial and non-spatial measures show a dramatic reversal of the increases in minority economic segregation reported by Jargowsky (1996). In all metropolitan areas the segregation level reduced significantly. The decomposition of GNSI reveals that the reversal was driven by the increased income diversity within census tracts rather than changes in tract clusters by relative income, although the decreases were greater for minorities, and less for whites, when space was taken more into account. Moreover, the relative rankings of cities differ depending on the degree to which space is taken into account. This general approach can be extended to other measures of segregation as well, a topic we hope to address in future research.

REFERENCES

Anselin, L. 2000. "The Alchemy of Statistics, or Creating Data Where No Data Exist." *Annals, Association of American Geographers* 90:586-592.

Anselin, L. 1995. "Local indicators of spatial association-LISA." *Geographical Analysis*, 27:93-115.

Brooks-Gunn, J., G. J. Duncan, and J. L. Aber, eds. 1997a. *Neighborhood Poverty*. Vol. 1, Context and Consequences for Children. New York: Russell Sage Foundation.

-------, eds. 1997b. *Neighborhood Poverty*. Vol. 2, Policy Implications in Studying Neighborhoods. New York: Russell Sage Foundation.

Duncan, O. D. and B. Duncan. 1955. "A Methodological Analysis of Segregation Measures." *American Sociological Review* 20:210-217.

Erbe, B. M. 1975. "Race and Socioeconomic Segregation." *American Sociological Review* 40:801-812.

Farley, R. 1977. "Residential Segregation in Urbanized Areas of the United States in 1970: An Analysis of Social Class and Racial Differences." *Demography* 14:497-517.

Farley, R. and W. H. Frey 1994. "Changes in the Segregation of Whites and Blacks During the 1980s: Small Steps Toward a More Integrated Society." *American Sociological Review* 59:23-45.

Frankel, D. M. and O. Volij 2004. "Measuring Segregation." Economics Dept in Iowa State University Working paper.

Geary, R.C. 1954. "The Contiguity Ratio and Statistical Mapping." *Incorporated Statistician* 5:114-141.

Iceland, J., D. H. Weinberg, and E. Steinmetz. 2002. *Racial and Ethnic Residential Segregation in the United States, 1980-2000.* Washington, DC: U.S. Bureau of the Census.

Jakubs, J.F. 1981. "A distance based segregation index." *Journal of Socio-Economic Planning Sciences* 15:129-136.

James, D. R. and K. E. Taeuber. 1985. "Measures of Segregation." *Sociological Methodology* 15:1-32.

Jargowsky, P. A. 1995. "Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas." Discussion Paper No. 1056-95 (January, 1995). Madison, Wisconsin: Institute for Research on Poverty. [http://www.irp.wisc.edu.]

-------. A.1996. "Take the Money and Run: Economic Segregation in U.S. Metropolitan Areas." *American Sociological Review* 61:984-998.

-------. 1997. *Poverty and Place: Ghettos, Barrios, and the American City*. New York: Russell Sage Foundation.

-------. 2003. "Stunning Progress, Hidden Problems: The Dramatic Decline of Concentrated Poverty in the 1990s." Living Cities Census Series, Metropolitan Policy Program. Washington, DC: The Brookings Institution. May, 2003.

Jencks, C. and S. E. Mayer 1990. "The Social Consequences of Growing up in a Poor Neighborhood." Pp. 111-186 in *Inner-City Poverty in America*, edited by J. Lynn, Laurence E. and M. McGeary, G. H. Washington, DC: National Academy Press.

King, G. 1997. *A Solution to the Ecological Inference Problem*. Princeton, New Jersey: Princeton University Press.

C. T. Kingsley and K. L. S. Pettit. "Concentrated Poverty: A Change in Course."

Lichter, D. T. 1985. "Racial Concentration and Segregation Across U.S. Countries, 1950-1980." *Demography* 22: 603-609.

Massey, D. S. and N. A. Denton 1987. "Trends in the Residential Segregation of Blacks, Hispanics, and Asians: 1970-1980." *American Sociological Review* 52:802-825.

Massey, D. S. and M. L. Eggers 1991. "The Ecology of Inequality: Minorities and the Concentration of Poverty, 1970-1980." *American Journal of Sociology* 95:1153-1188.

Miller, V. P. and J. M. Quigley 1990. "Segregation by Racial and Demographic Group: Evidence from the San Francisco Bay Area." *Urban Studies* 27:3-21.

Morgan, B.S. 1983. "An Alternate Approach to the Development of a Distance-Based Measure of Racial Segregation." *American Journal of Sociology*, 88:1237-1249.

Morrill R. 1991. "On the Measure of Geographic Segregation." *Geography Research Forum* 11:25-36.

Morrill R. 1995. "Racial Segregation and Class in a Liberal Metropolis." *Geographical Analysis* 27:22-41.

O'Sullivan, D. and D. S. Wong 2004. “A Surface-based Approach to Measuring Spatial Segregation.” Presented at the 2004 conference, Association of American Geographers.

Openshaw, S. 1984a. “ The Modifiable Areal Unit Problem.” *Concepts and Techniques in Modern Geography*, No. 38. Norwich, England: Geo Books.

Openshaw, S. 1984b. “Ecological Fallacies and the Analysis of Areal Census Data.” *Environment and Planning* A 6:17-31.

Reardon, S. F. and D. O’Sullivan 2004. “Measures of Spatial Segregation.” Population Research Institute Working Paper 04-01, Pennsylvania State University.

Rogerson, P. A. 1999. "The Detection of Clusters Using a Spatial Version of the Chi-Square Goodness-of-Fit Statistic." *Geographical Analysis* 31:130-147.

Schwartz, J. and C. Winship 1980. “The Welfare Approach to Measuring Inequality.” *Sociological Methodology* 11:1-36.

White, M.. J. 1983. "The Measurement of Spatial Segregation." *American Journal of Sociology* 88:1008-1018.

White, M.. J. 1986. "Segregation and Diversity Measures in Population Distribution." *Population Index* 52:198-221.

White, M.. J. 1987. American Neighborhoods and Residential Differentiation. in *The Population of the United States in the 1980s*. Edited by National Committee for Research on the 1980 Census. New York: Russell Sage Foundation.

Wilson, W. J. 1987. *The Truly Disadvantaged: The Inner-City, the Underclass and Public Policy*. Chicago: University of Chicago Press.

Wong, D. W. S. 1993. "Spatial Indices of Segregation." *Urban Studies* 30:559-572.

-------. 1997. "Spatial Dependency of Segregation Indices." *The Canadian Geographer* 41:128-136.

-------. 1998. "Measuring Multiethnic Spatial Segregation." *Urban Geography* 19:77-87.

-------. 1999. "Geostatistics as Measures of Spatial Segregation." *Urban Geography* 20:635-647.

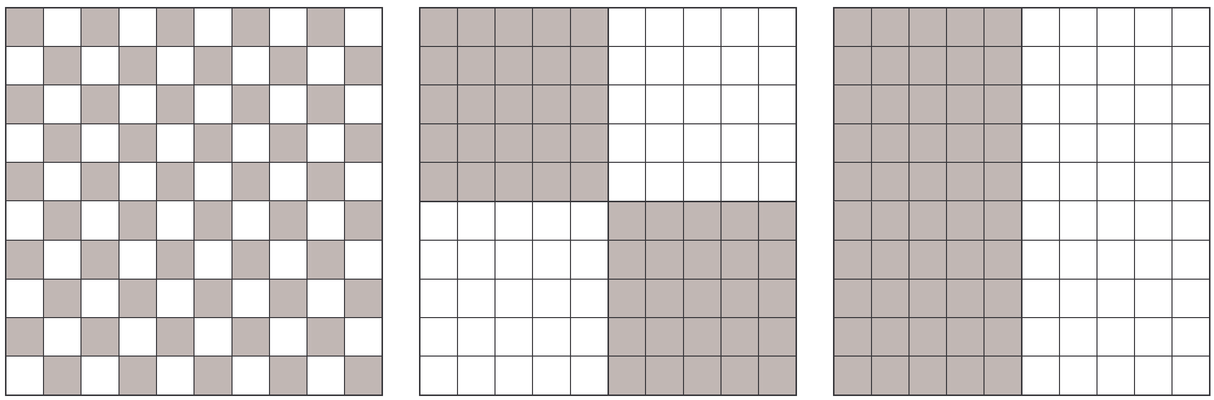
-------. 2002. "Modeling Local Segregation: A Spatial Interaction Approach." *Geographical and Environmental Modeling*, 6:81-97.

Yule, G. U. and M. G. Kendall 1950. *An Introduction to the Theory of Statistics*. London: Griffin.

Zoloth, B. S. "Alternative Measures of School Segregation." *Land Economics* 52:278-298.

Figure 1: Checkerboard Experiments: 10 by 10

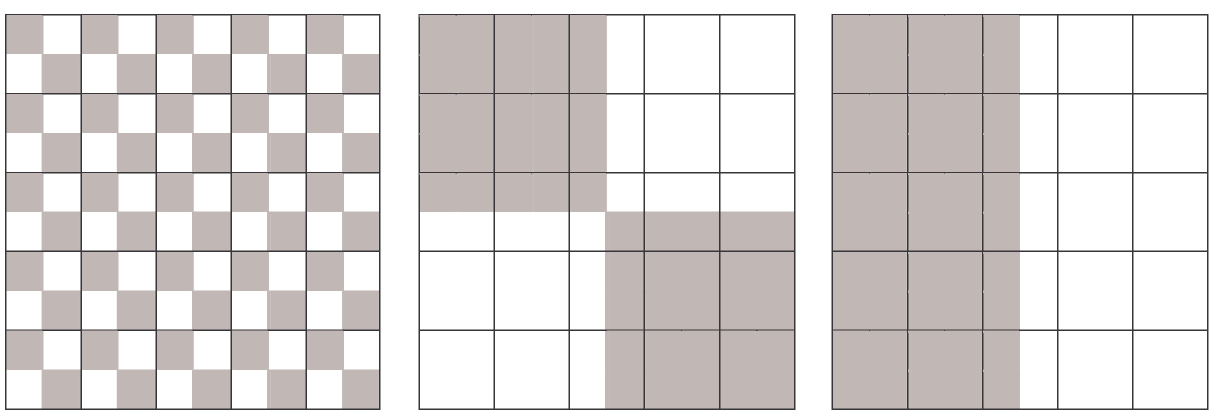
[Checkerboard] [Quadrant] [East - West]



NSI = 1 NSI = 1 NSI = 1

Figure 2: Checkerboard Experiments: 5 by 5

[Checkerboard] [Quadrant] [East - West]



NSI = 0 NSI = 0.64 NSI = 0.80

Figure 3: Census Tracts from the City of Dallas, TX

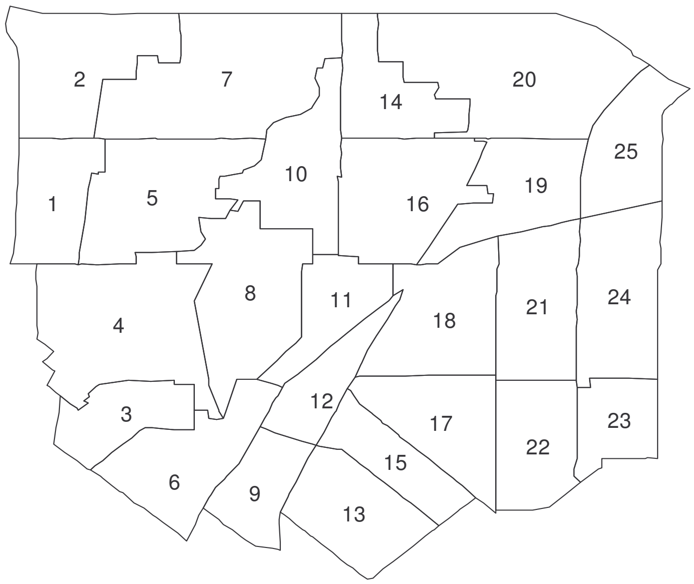


Figure 4: Spatial Arrangement of Census Tracts

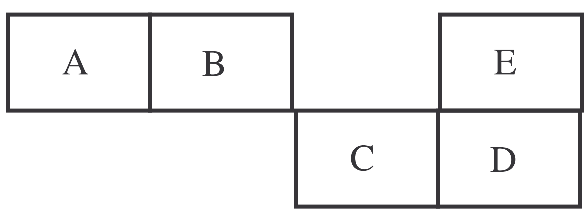


Figure 5: Expansions of GNSI Relative to $GNSI_0$, by Metropolitan Area, for Total Population, 2000

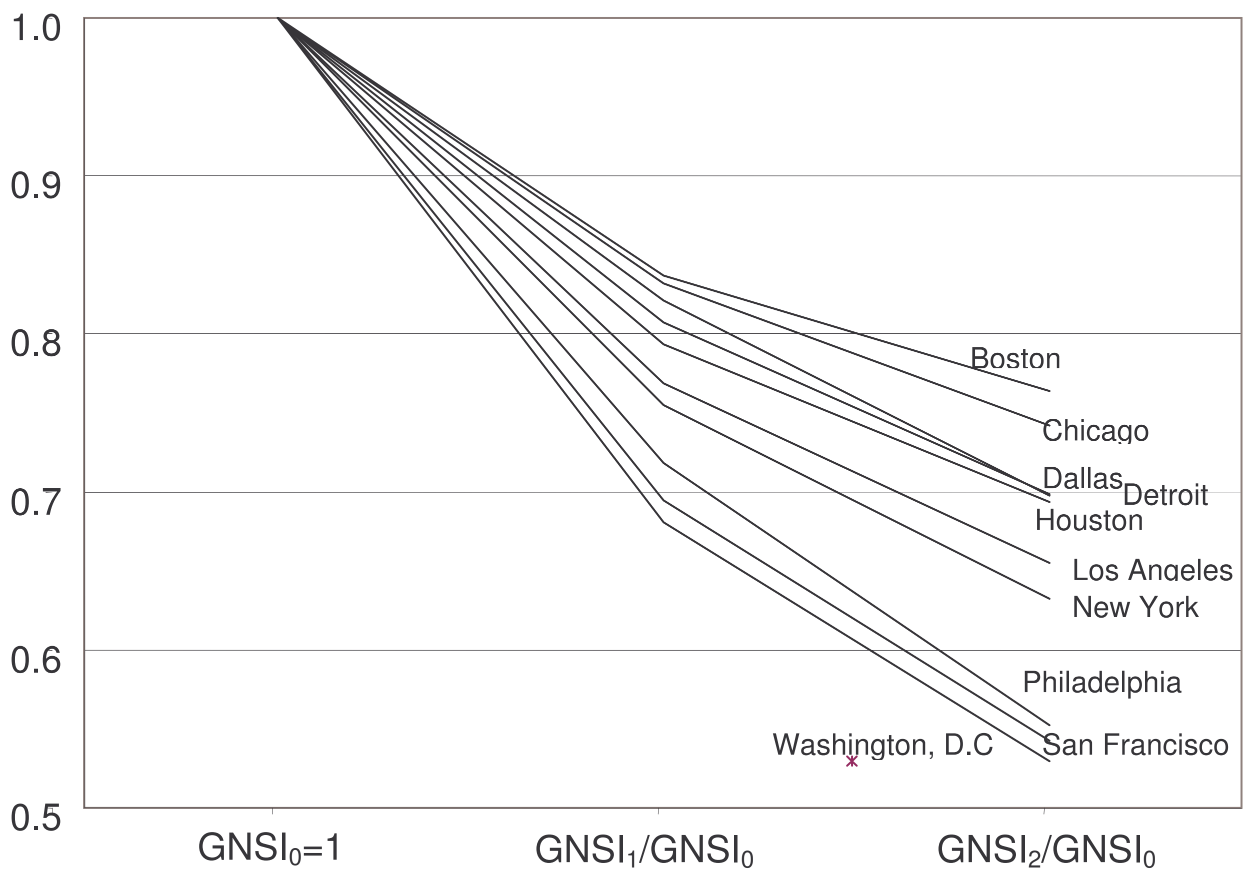


Table 1: The GNSI and the Spatial Contiguity Structure: 10 X 10

Criteria	Cell Structure	$GNSI_0$	$GNSI_1$	$GNSI_2$
Rook	Checkerboard	1.00	0.56	0.34
	Quadrants	1.00	0.86	0.76
	East - West	1.00	0.93	0.87
Queen	Checkerboard	1.00	0.09	0.05
	Quadrants	1.00	0.82	0.68
	East - West	1.00	0.91	0.82

Table 2: The GNSI and the Spatial Contiguity Structure: 5 X 5

Criteria	Cell Structure	$GNSI_0$	$GNSI_1$	$GNSI_2$
Rook	Checkerboard	0.00	0.00	0.00
	Quadrants	0.64	0.63	0.35
	East - West	0.80	0.80	0.64
Queen	Checkerboard	0.00	0.00	0.00
	Quadrants	0.64	0.58	0.20
	East - West	0.80	0.76	0.45

Table 3: Dallas Example: Neighborhood and 1st Order Community Expansion

Census Tract ID	Households	Neighborhood Mean Income (thousands)	Community Mean Income (thousands)
1	1128	32.977	43.986
2	972	28.074	38.743
3	2533	19.798	39.876
4	1971	54.864	33.994
5	1983	43.38	44.932
6	2208	26.496	33.503
7	2151	41.767	34.939
8	818	76.837	39.195
9	1587	18.835	17.875
10	1624	29.305	40.384
11	962	44.306	26.812
12	1673	15.602	20.247
13	1885	13.647	16.055
14	944	36.316	24.603
15	1982	14.212	15.783
16	1879	19.807	23.326
17	1911	15.515	16.762
18	1691	19.504	20.603
19	2675	20.382	20.240
20	6053	18.731	24.689
21	1916	19.235	19.743
22	1788	15.429	17.270
23	1173	14.217	17.235
24	1621	19.226	19.957
25	1133	33.266	19.235

Source: U.S. Census, 1990 Summary File 3. Tabulations by the authors. See text for explanation of neighborhood and community.

Table 4: Spatial Weight Matrices for Calculating GNSI

a) Weight Matrix of Zero Order Expansions (W_0)

Tract	A	B		C			D		E	
A	1	0	0	0	0	0	0	0	0	0
B	0	1/2	1/2	0	0	0	0	0	0	0
	0	1/2	1/2	0	0	0	0	0	0	0
C	0	0	0	1/3	1/3	1/3	0	0	0	0
	0	0	0	1/3	1/3	1/3	0	0	0	0
	0	0	0	1/3	1/3	1/3	0	0	0	0
D	0	0	0	0	0	0	1/2	1/2	0	0
	0	0	0	0	0	0	1/2	1/2	0	0
E	0	0	0	0	0	0	0	0	1/2	½
	0	0	0	0	0	0	0	0	1/2	½

b) Weight Matrix of the First order Queen Expansion *(W_1)*

Tract	A	B		C			D		E	
A	1/3	1/3	1/3	0	0	0	0	0	0	0
B	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0
	1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0
C	0	1/7	1/7	1/7	1/7	1/7	1/7	1/7	0	0
	0	1/7	1/7	1/7	1/7	1/7	1/7	1/7	0	0
	0	1/7	1/7	1/7	1/7	1/7	1/7	1/7	0	0
D	0	0	0	1/7	1/7	1/7	1/7	1/7	1/7	1/7
	0	0	0	1/7	1/7	1/7	1/7	1/7	1/7	1/7
E	0	0	0	0	0	0	1/4	1/4	1/4	1/4
	0	0	0	0	0	0	1/4	1/4	1/4	1/4

Table 5: GNSI for Total Population, 1990 - 2000

		Total					
		$GNSI_0$		$GNSI_1$		$GNSI_2$	
MSA/PMSA	ID	1990	2000	1990	2000	1990	2000
Boston	1120	0.35	0.23	0.26	0.18	0.22	0.15
Chicago	1600	0.44	0.27	0.36	0.22	0.32	0.20
Dallas	1920	0.43	0.28	0.30	0.21	0.23	0.16
Detroit	2160	0.53	0.37	0.44	0.30	0.37	0.26
Houston	3360	0.45	0.35	0.30	0.25	0.24	0.20
Los Angeles	4480	0.48	0.21	0.37	0.16	0.32	0.14
New York	5600	0.48	0.39	0.40	0.33	0.37	0.31
Philadelphia	6160	0.44	0.33	0.35	0.27	0.30	0.24
San Francisco	7360	0.45	0.37	0.32	0.28	0.25	0.21
Washington, D.C	8840	0.43	0.40	0.34	0.32	0.30	0.28
Average		0.45	0.32	0.35	0.25	0.29	0.21
HHS weighted Avg		0.45	0.31	0.36	0.25	0.31	0.22

Table 6: GNSI for White Population, 1990 - 2000

		White					
		$GNSI_0$		$GNSI_1$		$GNSI_2$	
MSA/PMSA	ID	1990	2000	1990	2000	1990	2000
Boston	1120	0.32	0.22	0.24	0.16	0.20	0.13
Chicago	1600	0.40	0.23	0.32	0.18	0.28	0.16
Dallas	1920	0.40	0.24	0.27	0.17	0.21	0.13
Detroit	2160	0.50	0.35	0.39	0.28	0.32	0.23
Houston	3360	0.41	0.31	0.27	0.21	0.21	0.16
Los Angeles	4480	0.48	0.38	0.36	0.28	0.30	0.23
New York	5600	0.49	0.40	0.40	0.34	0.37	0.32
Philadelphia	6160	0.40	0.30	0.31	0.24	0.26	0.20
San Francisco	7360	0.44	0.38	0.31	0.28	0.23	0.22
Washington, D.C	8840	0.42	0.36	0.32	0.28	0.27	0.25
Average		0.43	0.32	0.32	0.24	0.27	0.20
HHS weighted Avg		0.43	0.31	0.33	0.24	0.28	0.21

Table 7: GNSI for Black Population, 1990 - 2000

		Black					
		$GNSI_0$		$GNSI_1$		$GNSI_2$	
MSA/PMSA	ID	1990	2000	1990	2000	1990	2000
Boston	1120	0.50	0.31	0.29	0.19	0.24	0.15
Chicago	1600	0.48	0.37	0.36	0.25	0.32	0.21
Dallas	1920	0.50	0.32	0.36	0.24	0.30	0.21
Detroit	2160	0.55	0.35	0.42	0.25	0.36	0.21
Houston	3360	0.51	0.40	0.35	0.27	0.28	0.22
Los Angeles	4480	0.52	0.37	0.35	0.24	0.28	0.19
New York	5600	0.44	0.37	0.34	0.26	0.30	0.23
Philadelphia	6160	0.48	0.30	0.37	0.23	0.33	0.21
San Francisco	7360	0.52	0.40	0.35	0.21	0.29	0.17
Washington, D.C	8840	0.47	0.46	0.35	0.35	0.30	0.30
Average		0.50	0.36	0.35	0.25	0.30	0.21
HHS weighted Avg		0.49	0.37	0.36	0.26	0.31	0.22

Table 8: GNSI for Hispanic Population, 1990 - 2000

		Hispanic					
		$GNSI_0$		$GNSI_1$		$GNSI_2$	
MSA/PMSA	ID	1990	2000	1990	2000	1990	2000
Boston	1120	0.53	0.46	0.33	0.30	0.27	0.24
Chicago	1600	0.52	0.34	0.35	0.21	0.30	0.17
Dallas	1920	0.56	0.38	0.30	0.22	0.21	0.17
Detroit	2160	0.82	0.68	0.51	0.39	0.39	0.30
Houston	3360	0.51	0.20	0.29	0.12	0.22	0.09
Los Angeles	4480	0.46	0.27	0.33	0.19	0.27	0.15
New York	5600	0.51	0.34	0.38	0.24	0.34	0.20
Philadelphia	6160	0.81	0.48	0.56	0.34	0.49	0.30
San Francisco	7360	0.44	0.34	0.27	0.19	0.22	0.14
Washington, D.C	8840	0.55	0.44	0.32	0.24	0.26	0.19
Average		0.57	0.39	0.36	0.24	0.30	0.20
HHS weighted Avg		0.51	0.32	0.34	0.21	0.29	0.17

Table 9: Percentage Changes of Components in $GNSI_0$

	Total		White		Black		Hispanic	
	σ_N	σ_H	σ_N	σ_H	σ_N	σ_H	σ_N	σ_H
Boston	21.5	82.4	23.6	85.4	19.2	91.7	0.0	15.0
Chicago	4.8	70.6	10.0	95.6	18.3	56.0	-1.3	51.5
Dallas	15.2	75.7	20.4	102.5	27.4	100.1	4.2	52.3
Detroit	-0.1	45.3	3.5	47.7	0.7	58.4	-10.9	8.2
Houston	4.9	34.2	9.4	44.8	14.5	47.4	-5.9	143.1
Los Angeles	-2.6	128.0	9.5	40.5	3.9	47.1	-1.7	66.4
New York	9.4	35.7	18.7	43.4	5.1	26.8	5.7	58.0
Philadelphia	6.8	43.2	7.3	43.7	6.5	67.5	-16.5	41.4
San Francisco	24.8	52.4	29.2	50.5	69.0	119.1	9.1	39.1
Washington, D.C	12.3	20.5	18.0	38.1	12.4	16.0	-4.2	18.1
Average	9.7	58.8	15.0	59.2	17.7	63.0	-2.1	49.3

Table 10: Decomposition of GNSI: C_0^1

	Total		White		Black		Hispanic	
	1990	2000	1990	2000	1990	2000	1990	2000
Boston	0.76	0.75	0.74	0.74	0.57	0.61	0.62	0.65
Chicago	0.83	0.82	0.81	0.80	0.75	0.67	0.66	0.61
Dallas	0.69	0.73	0.68	0.71	0.72	0.74	0.54	0.57
Detroit	0.82	0.82	0.79	0.79	0.77	0.72	0.62	0.57
Houston	0.68	0.72	0.65	0.66	0.69	0.69	0.56	0.61
Los Angeles	0.77	0.79	0.75	0.75	0.66	0.66	0.71	0.69
New York	0.84	0.86	0.81	0.83	0.77	0.71	0.74	0.69
Philadelphia	0.81	0.83	0.77	0.79	0.78	0.77	0.69	0.71
San Francisco	0.72	0.75	0.70	0.74	0.66	0.52	0.62	0.57
Washington, D.C	0.79	0.80	0.77	0.78	0.74	0.77	0.59	0.55
Average	0.77	0.79	0.75	0.76	0.71	0.69	0.63	0.62
HHS weighted avg	0.79	0.80	0.76	0.77	0.74	0.71	0.68	0.65

Table 11: Decomposition of GNSI: C_1^2

	Total		White		Black		Hispanic	
	1990	2000	1990	2000	1990	2000	1990	2000
Boston	0.84	0.84	0.82	0.83	0.84	0.81	0.83	0.81
Chicago	0.89	0.89	0.87	0.87	0.87	0.85	0.86	0.84
Dallas	0.78	0.78	0.79	0.80	0.83	0.87	0.70	0.78
Detroit	0.85	0.86	0.82	0.85	0.85	0.84	0.76	0.78
Houston	0.78	0.79	0.78	0.79	0.79	0.79	0.78	0.77
Los Angeles	0.85	0.85	0.83	0.81	0.81	0.78	0.84	0.82
New York	0.91	0.92	0.93	0.94	0.89	0.88	0.89	0.85
Philadelphia	0.87	0.87	0.84	0.84	0.89	0.90	0.89	0.88
San Francisco	0.77	0.76	0.76	0.76	0.83	0.83	0.79	0.69
Washington, D.C	0.87	0.88	0.86	0.88	0.85	0.87	0.80	0.79
Average	0.84	0.85	0.83	0.84	0.85	0.84	0.81	0.80
HHS weighted avg	0.86	0.86	0.85	0.85	0.86	0.85	0.84	0.82